

BER-Minimizing Precoded Wideband Generalized BeamSpace Modulation for Hybrid mmWave Massive MIMO

Shijian Gao¹, Graduate Student Member, IEEE, Jinpeng Li², Xiang Cheng², Senior Member, IEEE, and Liuqing Yang¹, Fellow, IEEE

Abstract—Generalized beamspace modulation (GBM) unleashes a higher multiplexing gain via fewer radio-frequency chains, making it an attractive uplink transmission solution to hybrid mmWave massive multiple-input multiple-output systems. Leaping from its wideband form, namely, wideband GBM (WGBM), a further enhanced version termed precoded WGBM (P-WGBM) will be studied, striving for performance improvement by actively involving the digital precoder. Unfortunately, the seemingly “straightforward” extension proves to be challenging because of the precoding constraint and computation complexity. To overcome these obstacles, this letter has carefully formulated an optimization problem, aiming to find an optimal precoder that retains GBM’s multiplexing merit. By exploiting the beamspace properties, we arrive at an efficient technique to achieve near-optimal precoding. Whist keeping all major advantages of WGBM, simulations show that P-WGBM can achieve a remarkable coding gain over WGBM at a small complexity cost.

Index Terms—Index modulation, mmWave, hybrid massive MIMO, generalized beamspace modulation, precoding.

I. INTRODUCTION

AS NUMEROUS emerging applications put forward more stringent requirements for data rate and latency, communication towards the mmWave band is becoming the new mainstream [1]–[3]. Relative to the sub-6GHz counterpart, mmWave has considerably shorter wavelength and higher sensitivity to propagation environments [4]. These two prominent features render the combination of mmWave and massive multiple-input multiple-out (MIMO) not only necessary but also feasible. To save in power consumption and hardware expenditure, mmWave massive MIMO (m-MIMO) typically adopts a hybrid (digital/analog) structure [5].

Manuscript received August 15, 2021; revised October 29, 2021; accepted October 30, 2021. Date of publication November 8, 2021; date of current version February 17, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62125101, and in part by the National Science Foundation under Grant CPS-2103256 and Grant ECCS-2102312. The associate editor coordinating the review of this article and approving it for publication was G. Zheng. (Corresponding author: Xiang Cheng.)

Shijian Gao and Liuqing Yang are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: gao00379@umn.edu; qingqing@umn.edu).

Jinpeng Li and Xiang Cheng are with the State Key Laboratory of Advanced Optical Communication Systems and Networks, Department of Electronics, School of Electronics Engineering and Computing Sciences, Peking University, Beijing 100080, China (e-mail: jinpengli_pku@163.com; xiangcheng@pku.edu.cn).

Digital Object Identifier 10.1109/LWC.2021.3125778

Compared with digital m-MIMO, the hybrid structure falls short in multiplexing capability. This soft spot will be aggravated in uplink transmission, where the terminal typically has only a single radio-frequency (RF) chain. Over the past years, this issue has not been well addressed until the introduction of GBM [6], resorting to which the multiplexing gain will be dictated by the channel itself. Breaking away from the classic framework, GBM draws the idea of spatial modulation (SM) [7] by applying index modulation in beamspace. This domain is in fact a natural outcome through exploiting the unique system and channel properties under hybrid mmWave m-MIMO. Specifically, the constellation symbols of GBM will be up-converted by RF chains and used for activating the selected beams. As for the zero elements, whose location convey the index bits [8], will be transmitted via RF switches. As such, more data streams can be sent by fewer RF chains as SM does [9], [10]. Recently, GBM has been successfully tailored for frequency-selective channels thanks to a novel orthogonal frequency-division multiplexing (OFDM) symbol-based mechanism [11]. The resulting WGBM inherits GBM’s superiority in both multiplexing and error performance.

Note that WGBM fixes the digital precoder as an identity matrix. Motivated by the advantage of applying precoding in index modulation [12], [13], one would expect a positive yield to WGBM as well by activating the digital precoder. This one-step extension, however, is easier said than done for several reasons. First, the digital precoder is found to be a diagonal matrix to retain the multiplexing capability, and such a restriction significantly lowers the design flexibility. Secondly, WGBM adopts a symbol-based modulating mode, therefore digital precoding needs to jointly evaluate all sub-carriers. In light of these restrictions, we carefully formulate an optimization problem, seeking for a precoder minimizing the asymptotic pair-wise error probability (APEP). Although a closed-form APEP can be readily derived, the analytical intractability arising from the special modulation mechanism hinders any further process. To bypass the formidable exponential complexity in APEP computation, we carry out a two-step simplification by utilizing the beamspace correlation and sparsity in the frequency domain. As a result, we arrive at an efficient way of quantifying APEP. By demonstrating the asymptotic convexity of the alternative APEP, we end up with a near-optimal precoder via the projected gradient-descent algorithm. Owing to careful treatment, P-WGBM retains all significant advantages possessed by WGBM. Simulations show that PWGBM can achieve a notable coding gain over WGBM at slightly increased cost.

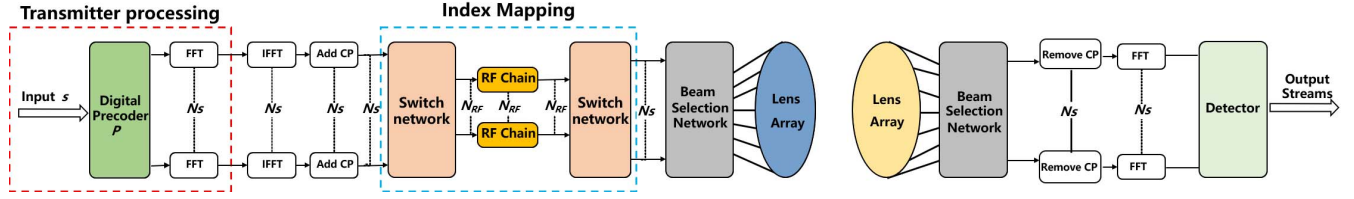


Fig. 1. The schematic of P-WGBM for wideband hybrid mmWave m-MIMO.

Notations: a , $\mathbf{a}$, $\mathbf{A}$ represent a scalar, a vector and a matrix, respectively. $\mathbf{a}_{a:b} = [\mathbf{a}[a], \dots, \mathbf{a}[b]]^T$. $\mathbf{A}_{a:b,c:d}$ is a sub-matrix of $\mathbf{A}$ by slicing from the a -th to the b -th row and from the c -th to the d -th column. $\{\mathbf{A}_k\}_{k=1}^K = [\mathbf{A}_1^T, \mathbf{A}_2^T, \dots, \mathbf{A}_K^T]^T$. $\mathbf{A}^T$ and $\mathbf{A}^H$ represent the transpose and Hermitian transpose of $\mathbf{A}$, respectively. $\mathbf{F}_K$ is the K -point fast Fourier Transform (FFT) matrix. $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{\mu^2}{2}} d\mu$ is the Gaussian Q-function. $\|\cdot\|_F$ is the Frobenius norm. $\otimes$ and $\odot$ denote Kronecker product and Hadamard product, respectively. diag is the operation of generating a diagonal matrix. $\mathcal{U}$ and $\mathcal{CN}$ represent the uniform distribution and circular complex Gaussian distribution, respectively.

II. SYSTEM DESCRIPTION AND PROBLEM FORMULATION

In this section, we first introduce the system model for P-WGBM. After a quick review of the newly proposed WGBM, we then start formulating the precoding optimization problem for P-WGBM that promises a better error performance.

A. A Recap of WGBM

We consider an uplink hybrid mmWave massive MIMO system with the schematic shown in Fig. 1. The base station (BS) and the mobile station (MS) deploy N_b - and N_m -dimensional lens-antenna arrays playing a role of spatial FFT per the array dimension.

At each of K subcarriers, the MS transmits $\mathbf{s}_k$ to the BS. In the context of WGBM, the mapping rule follows other representative index modulations, including SM and GBM. Specifically, $\mathbf{s}_k$ is an N_s -dimensional vector containing N_{RF} ($N_{RF} \leq N_s \ll \min\{N_b, N_m\}$) symbols chosen from a normalized M -ary phase shift keying/quadrature amplitude modulation (PSK/QAM) constellation [11]. The remaining $N_s - N_{RF}$ elements are filled with zeros with their indices randomly selected. As a result, $\forall k$, $\mathbf{s}_k$ conveys

$$\eta = N_{RF} \log_2 M + \lfloor \log_2 C_{N_s}^{N_{RF}} \rfloor \text{ bits.} \quad (1)$$

Before going to RF domain, $\mathbf{s}_k$ is initially precoded by $\mathbf{P}_k \in \mathbb{C}^{N_s \times N_s}$. Stacking all precoded $\mathbf{s}_k$'s gives rise to a KN_s -dimensional vector

$$\mathbf{x} = \mathbf{P} \mathbf{s}, \quad (2)$$

where $\mathbf{P} = \text{diag}(\mathbf{P}_1, \dots, \mathbf{P}_K)$ and $\mathbf{s} = \{\mathbf{s}_k\}_{k=1}^K$. With N_{RF} RF chains available, the K -point inverse IFFT output must maintain the same sparse structure as $\mathbf{s}_k$. Thanks to pre-FFT on $\mathbf{x}$, the N_s -dimensional symbol sampled at the k -th

time instant is exactly N_{RF} -sparse. The index mapping can be interpreted as activating N_{RF} out of N_s beams.

Let $\mathbf{U}_t$ and $\mathbf{U}_r$ respectively represent the analog precoder and combiner, whose column stands for an FFT basis. Define $\mathbf{H}_k$ to be the channel at subcarrier- k , then the sub-beamspace channel can be written as

$$\overline{\mathbf{H}}_k = \mathbf{U}_r^H \mathbf{H}_k \mathbf{U}_t. \quad (3)$$

After channel propagation and OFDM demodulation, the received signal can be represented as

$$\mathbf{y} = \underbrace{\mathbf{A}(\mathbf{F}_K \otimes \mathbf{I}_{N_s})}_{\mathbf{A}} \mathbf{x} + \mathbf{n}, \quad (4)$$

where $\mathbf{A} = \text{diag}(\overline{\mathbf{H}}_1, \dots, \overline{\mathbf{H}}_K)$, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_{KN_s})$ is the white Gaussian noise vector. With the established I-O relationship, the analog parts, i.e., $\mathbf{U}_t$ and $\mathbf{U}_r$, shared by all subcarriers are determined as ones minimizing the following mean square error [14]:

$$\mathbb{E} \left\{ \left\| \overline{\mathbf{A}}^H (\overline{\mathbf{A}} \overline{\mathbf{A}}^H + N_0 N_s \mathbf{I}_{KN_s})^{-1} \mathbf{y} - \mathbf{x} \right\|_F^2 \right\}. \quad (5)$$

B. From WGBM to P-WGBM

Recall that the digital precoder $\mathbf{P}_k$ is a fixed identity matrix in WGBM. If relaxing this configuration, one could anticipate a performance gain. To enable the transmission of N_s streams via N_{RF} chains and RF switches, P-WGBM confines $\mathbf{P}_k$ to be a diagonal matrix such that the N_s -dimensional IFFT output remains N_{RF} -sparse. Without causing ambiguity, the rest manuscript will not highlight “digital” in precoding design.

As $\mathbf{P}_k$ is a diagonal matrix, the precoding design boils down to finding an optimal KN_s -dimension vector. Let $P_e(\mathbf{p})$ stand for the corresponding error rate, which is commonly quantified as APEP. Then we can formulate the following optimization problem to guide the precoding design.

[P. 1: Original APEP-minimized precoding optimization]

$$\arg \min_{\mathbf{p}} P_e(\mathbf{p})$$

$$\text{s.t.} \quad \left\| \mathbf{p}_{(k-1)N_s+1:kN_s} \right\|_F^2 \leq N_s, \quad \forall k \in [1, K].$$

The constraint ensures that the transmit power in P-WGBM remains equal to WGBM.

III. OPTIMAL PRECODED WGBM (P-WGBM)

Based on P. 1, we start with revealing its analytical intractability. We carry out a two-step simplification by tactfully exploiting the beamspace properties. Hence, a near-optimal precoder can be secured efficiently.

A. Problem Transformation

The top priority in dealing with P. 1 is of course to find a tractable alternative of $P_e(\mathbf{p})$. Denote S as the set containing all possible $\mathbf{s}$'s defined in Eq. (2), and assume that the maximum-likelihood (ML) detection is adopted by the BS. Then the APEP for P-WGBM can be written as follows:

$$\begin{aligned} P_e(\mathbf{p}) &= \frac{1}{K\eta 2^{K\eta}} \sum_{\mathbf{s} \in S} \sum_{\hat{\mathbf{s}} \neq \mathbf{s}} P_r(\mathbf{s} \rightarrow \hat{\mathbf{s}}) e(\mathbf{s}, \hat{\mathbf{s}}) \\ &= \frac{1}{K\eta 2^{K\eta}} \sum_{\mathbf{s} \in S} \sum_{\hat{\mathbf{s}} \neq \mathbf{s}} Q\left(\sqrt{\frac{d(\mathbf{s}, \hat{\mathbf{s}})}{2N_0}}\right) e(\mathbf{s}, \hat{\mathbf{s}}), \quad (6) \end{aligned}$$

where $d(\mathbf{s}, \hat{\mathbf{s}}) = \|\bar{\mathbf{A}}\mathbf{P}(\mathbf{s} - \hat{\mathbf{s}})\|^2$, K is the subcarrier number, and η is the information bits per subcarrier. $e(\mathbf{s}, \hat{\mathbf{s}}) = \sum_{k=1}^K e(\mathbf{s}_k, \hat{\mathbf{s}}_k)$ denotes the sum of bit-type Hamming distance, with $e(\mathbf{s}_k, \hat{\mathbf{s}}_k)$ calculated similar to [6] and [7].

Taking a closer look at Eq. (6), we discover its calculation involves $2^{2K\eta-1}$ items, while only $K2^{2\eta-1}$ items are required in precoded SM [15]. Such a vast disparity arises from their different modulation mechanisms. That is, *P-WGBM is associated with the entire OFDM symbol while precoded SM is associated with one subcarrier*. As the subcarrier number K would be in the hundreds in mmWave systems, the primitive APEP is barely useful, forcing us to find a tractable alternative. Fortunately, the following two conclusions provide us with such an opportunity.

Proposition 1: $\bar{\mathbf{A}}^2 = \bar{\mathbf{A}}^H \bar{\mathbf{A}}$ is a sparse matrix whose dominant energy is captured by the sub-matrices along its diagonal line.

Proof: Let $\mathbf{f}_i$ be the i -th row of $\mathbf{F}_K$, then

$$\begin{aligned} \bar{\mathbf{A}}^2 &= (\mathbf{F}_K^H \otimes \mathbf{I}_{N_s}) \mathbf{A}^H \mathbf{A} (\mathbf{F}_K \otimes \mathbf{I}_{N_s}) \\ &= \begin{bmatrix} \mathbf{f}_1 \otimes \mathbf{I}_{N_s} \\ \mathbf{f}_2 \otimes \mathbf{I}_{N_s} \\ \vdots \\ \mathbf{f}_K \otimes \mathbf{I}_{N_s} \end{bmatrix}^H \begin{bmatrix} \bar{\mathbf{H}}_1^H \bar{\mathbf{H}}_1 & & \\ & \ddots & \\ & & \bar{\mathbf{H}}_K^H \bar{\mathbf{H}}_K \end{bmatrix} \begin{bmatrix} \mathbf{f}_1 \otimes \mathbf{I}_{N_s} \\ \mathbf{f}_2 \otimes \mathbf{I}_{N_s} \\ \vdots \\ \mathbf{f}_K \otimes \mathbf{I}_{N_s} \end{bmatrix} \\ &= \sum_{k=1}^K (\mathbf{f}_k^H \mathbf{f}_k) \otimes \bar{\mathbf{H}}_k^H \bar{\mathbf{H}}_k \quad (7) \end{aligned}$$

Defining $\omega = e^{j\frac{2\pi}{K}}$, we have

$$\begin{aligned} \bar{\mathbf{A}}_{m,n}^2 &= [\bar{\mathbf{A}}^2]_{(m-1)N_s+1:mN_s, (n-1)N_s+1:nN_s} \\ &= \sum_{k=1}^K \omega^{(n-m) \cdot (k-1)} \bar{\mathbf{H}}_k^H \bar{\mathbf{H}}_k, \quad m \neq n. \quad (8) \end{aligned}$$

$\forall k$, $\bar{\mathbf{H}}_k^H \bar{\mathbf{H}}_k$ is demonstrated to be asymptotically equal as antennas approach infinity [16]. Since $\frac{1}{K} \sum_{k=1}^K \omega^{n \cdot (k-1)} = 0$ if $k \neq 0$, $\bar{\mathbf{A}}_{m,n}^2$ is proved to be close to a zero matrix. ■

Following the derivation above, we can instantly get another useful result to simplify APEP.

Lemma 1: The sub-matrices $\bar{\mathbf{A}}_{k,k}^2$'s on the diagonal line of $\bar{\mathbf{A}}^2$ are exactly the same.

Proof: We have already arrived at $\bar{\mathbf{A}}^2 = \sum_{i=1}^K (\mathbf{f}_i^H \mathbf{f}_i) \otimes (\mathbf{A}_i^H \mathbf{A}_i)$. Since matrix $\mathbf{f}_i^H \mathbf{f}_i$ has an all-one diagonal line, the sub-matrices on the diagonal line of $\bar{\mathbf{A}}^2$ can be represented by

$$\bar{\mathbf{A}}_{k,k}^2 = [\bar{\mathbf{A}}^2]_{(k-1)N_s+1:kN_s, (k-1)N_s+1:kN_s} = \sum_{i=1}^K \bar{\mathbf{H}}_i^H \bar{\mathbf{H}}_i. \quad (9)$$

As revealed by Proposition 1 that $\bar{\mathbf{A}}^2$ is near block-diagonal, the computation of Euclidean distance between two OFDM symbols can be transformed into per subcarrier. Therefore, the original APEP given in Eq. (6) can be approximated as

$$\frac{1}{K^2 \eta 2^\eta} \sum_{k=1}^K \sum_{\mathbf{s}_k \in \mathcal{G}} \sum_{\hat{\mathbf{s}}_k \neq \mathbf{s}_k} Q\left(\sqrt{\frac{d(\mathbf{s}_k, \hat{\mathbf{s}}_k)}{2N_0}}\right) e(\mathbf{s}_k, \hat{\mathbf{s}}_k), \quad (10)$$

where $\mathcal{G}$ represents the ensemble of $\mathbf{s}_k$, being common to all subcarriers; $d(\mathbf{s}_k, \hat{\mathbf{s}}_k) = \mathbf{p}_k^H \bar{\mathbf{A}}_{k,k}^H \bar{\mathbf{A}}_{k,k} \odot [(\mathbf{s}_k - \hat{\mathbf{s}}_k)(\mathbf{s}_k - \hat{\mathbf{s}}_k)^H]^T \mathbf{p}_k$. This step of approximation reduces the number of summations from $2^{\eta K}(2^\eta - 1)$ to $K2^\eta(2^\eta - 1)$.

Recall that lemma 1 points out all subcarriers share the same $\bar{\mathbf{A}}_{k,k}^2$. Also, they have a shared error pattern set. Hence, their behavior in terms of the error performance is the same, leading to $\mathbf{p}_1 = \dots = \mathbf{p}_K = \bar{\mathbf{p}}$. Using this property, we are able to absorb the summation over k on the outermost layer of Eq. (10), giving rise to a more compact form as follows:

$$\frac{1}{K\eta 2^\eta} \sum_{\mathbf{s}_i \in \mathcal{G}} \sum_{\mathbf{s}_j \neq \mathbf{s}_i} Q\left(\sqrt{\frac{\bar{\mathbf{p}}^H \bar{\mathbf{D}}_{i,j} \bar{\mathbf{p}}}{2N_0}}\right) e(\mathbf{s}_i, \mathbf{s}_j), \quad (11)$$

where $\bar{\mathbf{D}}_{i,j} = \bar{\mathbf{A}}_{i,i}^H \bar{\mathbf{A}}_{i,i} \odot [(\mathbf{s}_i - \mathbf{s}_j)(\mathbf{s}_i - \mathbf{s}_j)^H]^T$. As a result, APEP can be computed by summing up only $2^\eta(2^\eta - 1)$ items, even fewer than that required in precoded SM.

We replace $P_e(\mathbf{p})$ with Eq. (11) and further define

$$\bar{\mathbf{q}} = [\text{Re}(\bar{\mathbf{p}})^T, \text{Im}(\bar{\mathbf{p}})^T]^T \in \mathbb{R}^{2N_s \times 1}, \quad (12a)$$

$$\bar{\mathbf{R}}_{i,j} = \begin{bmatrix} \text{Re}(\bar{\mathbf{D}}_{i,j}) & -\text{Im}(\bar{\mathbf{D}}_{i,j}) \\ \text{Im}(\bar{\mathbf{D}}_{i,j}) & \text{Re}(\bar{\mathbf{D}}_{i,j}) \end{bmatrix}, \quad (12b)$$

then similar to [10], we rewrite the optimization problem from complex domain into its equivalent real-valued form as

[P. 2: Simplified APEP-minimized precoding optimization]

$$\begin{aligned} \arg \min_{\bar{\mathbf{q}}} \tilde{P}_e(\bar{\mathbf{q}}) &= \frac{1}{K\eta 2^\eta} \sum_{\mathbf{s}_i \in \mathcal{G}} \sum_{\mathbf{s}_j \neq \mathbf{s}_i} Q\left(\sqrt{\frac{\bar{\mathbf{q}}^T \bar{\mathbf{R}}_{i,j} \bar{\mathbf{q}}}{2N_0}}\right) e(\mathbf{s}_i, \mathbf{s}_j) \\ \text{s.t.} \quad &\|\bar{\mathbf{q}}\|^2 \leq N_s. \end{aligned}$$

B. Problem Solving

Our ultimate goal is to solve P. 2, whose global optimum is generally not achievable. However, we will show that the obtained solution can be sufficiently close to the optimum if the signal-to-noise ratio (SNR) is high enough. In fact, this

SNR range is also within our interest because index modulation is well-known to show an advantage over conventional modulations in that SNR range.

Under the high-SNR assumption, the objective function of P.2 can be tightly approximated as

$$\begin{aligned} \tilde{P}_e(\bar{\mathbf{q}}) &\approx \tilde{P}_{e,h}(\bar{\mathbf{q}}) \\ &= \frac{1}{K\eta 2^\eta} \sum_{s_i \in \mathcal{G}} \sum_{s_j \neq s_i} Q\left(\sqrt{\frac{\bar{\mathbf{q}}^T \bar{\mathbf{R}}_{i,j} \bar{\mathbf{q}}}{2N_0}}\right) e(s_i, s_j) \mathbb{I}(s_i, s_j) \end{aligned} \quad (13)$$

where $\mathbb{I}(s_i, s_j)$ is a binary indicator function taking on 1 if s_j and s_i differ in only one element. Once $\mathbb{I}(s_i, s_j) = 1$, $\bar{\mathbf{D}}_{i,j} = \bar{\mathbf{A}}_{i,i}^H \bar{\mathbf{A}}_{i,i} \odot [(s_i - s_j)(s_i - s_j)^H]^T$ contains a single non-zero element on its diagonal line. Then one can verify that altering the sign of an arbitrary element of $\bar{\mathbf{q}}$ does not change $\tilde{P}_{e,h}$. In consequence, minimizing $\tilde{P}_{e,h}$ over $\|\bar{\mathbf{q}}\|^2 \leq N_s$ equates to minimizing over $\mathcal{S} = \{\mathbf{q} | \|\bar{\mathbf{q}}\|^2 \leq N_s, \mathbf{q} \succ 0\}$.

Proposition 2: $\tilde{P}_{e,h}(\bar{\mathbf{q}})$ is asymptotically convex w.r.t. $\bar{\mathbf{q}}$ over $\mathcal{S}$ at high SNR.

Proof: The heart of demonstration lies in proving the convexity of Q -term. Let $\mathbf{R}_{i,j} = \bar{\mathbf{R}}_{i,j} + \bar{\mathbf{R}}_{i,j}^T$. Taking the second derivative of the Q -term w.r.t. $\bar{\mathbf{q}}$ results in

$$\begin{aligned} \frac{\partial^2 Q}{\partial^2 \bar{\mathbf{q}}} &\propto e^{-\frac{\bar{\mathbf{q}}^T \bar{\mathbf{R}}_{i,j} \bar{\mathbf{q}}}{4N_0}} \left(\frac{\bar{\mathbf{q}}^T \bar{\mathbf{R}}_{i,j} \bar{\mathbf{q}}}{2N_0}\right)^{-\frac{1}{2}} \left(\frac{\mathbf{R}_{i,j} \bar{\mathbf{q}} \bar{\mathbf{q}}^T \mathbf{R}_{i,j}}{4N_0} - \mathbf{R}_{i,j}\right) \\ &+ \left(\frac{\bar{\mathbf{q}}^T \bar{\mathbf{R}}_{i,j} \bar{\mathbf{q}}}{2N_0}\right)^{-1} \frac{\mathbf{R}_{i,j} \bar{\mathbf{q}} \bar{\mathbf{q}}^T \mathbf{R}_{i,j}}{4N_0} \propto \mathbf{R}_{i,j} \bar{\mathbf{q}} \bar{\mathbf{q}}^T \mathbf{R}_{i,j} - o(\text{SNR}^{-1}) \mathbf{R}_{i,j} \end{aligned}$$

Note that if $\mathbb{I}(s_i, s_j) = 1$, $\mathbf{R}_{i,j}$ is rank-deficient, so $\mathbf{R}_{i,j} \bar{\mathbf{q}} \bar{\mathbf{q}}^T \mathbf{R}_{i,j} \succeq \mathbf{0}$. The Hessian matrix is positive semi-definite, leading to the convexity (not strict convexity) of $\tilde{P}_{e,h}(\bar{\mathbf{q}})$. ■

Thanks to the asymptotic convexity of $P_{e,h}(\bar{\mathbf{q}})$, we can get a near-optimal solution to P. 2, denoted as $\bar{\mathbf{q}}$, at high SNR via a simple projected gradient algorithm. Accordingly, the precoder can be determined as

$$\mathbf{P}_k = \text{diag}(\bar{\mathbf{q}}_{1:N_s}) + j \text{diag}(\bar{\mathbf{q}}_{N_s+1:2N_s}), \quad \forall k. \quad (14)$$

IV. SIMULATIONS

In this section, numerical simulations will be carried out to test the bit error ratio (BER) performance of P-WGBM. The well-known geometrical channel model [17] is adopted, with the tap- d channel ($d < N_c$) and the corresponding k -th subcarrier channel represented as

$$\mathbf{H}[d] = \sqrt{\frac{N_b N_m}{N_p}} \sum_{l=1}^{N_p} \alpha_l p(dT_s - \tau_l) \mathbf{a}_r(\phi_l) \mathbf{a}_t^H(\theta_l) \quad (15a)$$

$$\mathbf{H}_k = \sum_{d=0}^{N_c-1} \mathbf{H}[d] e^{-j \frac{2\pi k}{K} d}. \quad (15b)$$

Specifically, $p(\cdot)$ is the raised-cosine filter with roll-off factor $\beta = 0.8$ and N_c is the maximum tap. The uniform linear arrays with half-wavelength spacing are employed at the transceivers, so $\mathbf{a}_r(\phi) = \frac{1}{\sqrt{N_r}} [1, e^{j\pi \sin \phi}, \dots, e^{j(N_r-1)\pi \sin \phi}]^T$. Other system and channel related parameters are listed in Table I. The SNR is

TABLE I
SYSTEM AND CHANNEL PARAMETERS FOR SIMULATION

Parameter	Value
BS-end antenna number N_b	64
MS-end antenna number N_m	32
Subcarrier number K	32
Maximum delay tap N_c	8
Angle of arrival (departure) ϕ_l (θ_l)	$U[0, 2\pi)$
Tap delay τ_l	$U[0, (N_c-1)T_s)$
Path amplitude α_l	$CN(0, 1)$
Sampling time T_s	10^{-9} s
Path number N_p	8

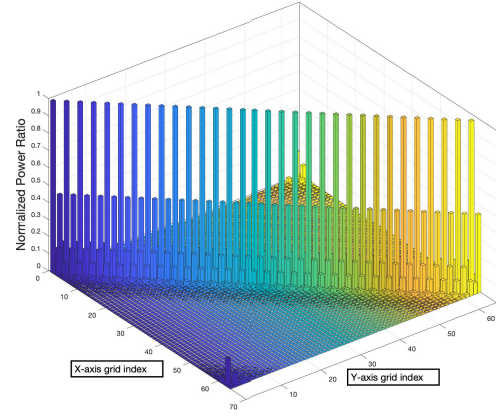


Fig. 2. Normalized power map of $\bar{\mathbf{A}}^2$.

defined as $E_b/n_0 = N_{RF}/(\eta N_0)$. All BER curves are the average of 10000 independent channel realizations, each with a block of 2000 symbols.

Recall that the success of APEP simplification counts on the unique structure of $\bar{\mathbf{A}}^2$. Actually, one can interpret $\bar{\mathbf{A}}^2$ as the autocorrelation of sub-beamspace channels. To gain more intuition, we visualize this matrix in a power map by averaging 10000 realizations. As seen from Fig. 2, all sub-matrices along the diagonal line are exactly the same, amassing the majority of channel power. In contrast, the off-diagonal sub-matrices are suppressed to a very low level. The result implies that all subcarriers have a similar behavior regarding the error performance, so they share a common digital precoder.

In Fig. 3, we present BER performance for P-WGBM and WGBM under $\{N_{RF} = 1, N_s = 2, \text{BPSK}\}$. Since ML detection is infeasible due to prohibitive complexity, we henceforth adopt near-ML detector proposed in [18] for both schemes. In brief, by leaving more effective candidates, whose number is denoted as m , the achieved performance is closer to the ML detector at the expense of higher complexity. Fig. 3 illustrates that the BER curve does not decline anymore as m comes to 4, so we can treat the corresponding BER as the near-ML one. As modest-to-high SNR, P-WGBM gains over 0.5dB advantage over WGBM when $\eta = 2$ bits/s/Hz. If lifting η up to 6 bits/s/Hz by setting $\{N_{RF} = 1, N_s = 4, 16\text{-PSK}\}$, the coding gain is up to 2dB at high SNR as reflected in Fig. 4.

More configurations have been tested but are not presented here for space limitation. In general, we find that a large N_s and a high modulation order typically lead to a high coding gain. This is because in this case the beam quality exhibits a

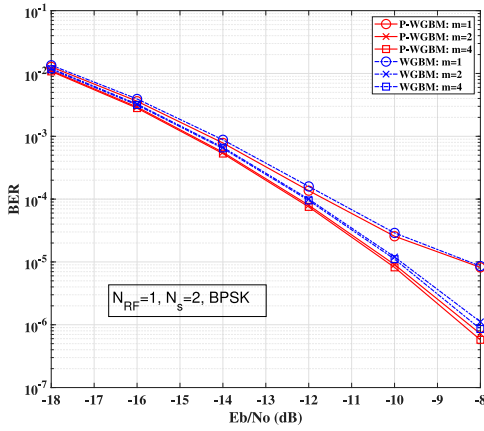


Fig. 3. BER comparison using the near-ML detector with $N_{RF} = 1$, $N_s = 2$, and BPSK: $\eta = 2$ bits/Hz/s.

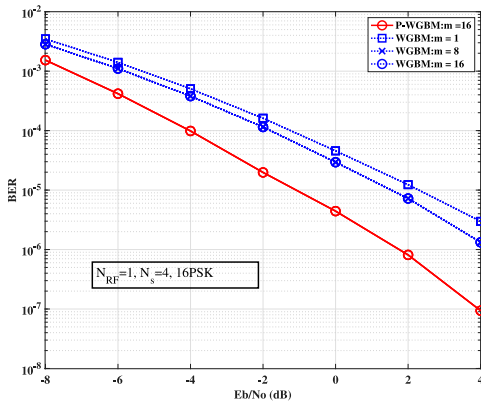


Fig. 4. BER comparison between P-WGBM and WGBM under $N_{RF} = 1$, $N_s = 4$, and 16-PSK: $\eta = 6$ bits/Hz/s.

higher degree of discrimination, hence digital precoding can help compensate for the weak beams and reduce their adversary effects on the error performance. From the hardware perspective, WGBM and P-WGBM have equal complexity because no additional component is needed. From the computation perspective, P-WGBM requires digital precoding whose complexity primarily relies on computing $\bar{D}_{i,j}$ and updating gradient descent, each needing the number of flops in the order of $o(KN_s^2)$ and $o(N_{ite}N_s^2)$. The beam selection and detector complexity are $o(K^3N_s^3)$ for both WGBM and P-WGBM. Although the latter is a bit more complicated, it shares the same cubic-order complexity concerning the subcarrier and stream number. Plus, the resulting nearly 1dB coding gain consolidates that such complexity cost is rather worthwhile.

V. CONCLUSION

This letter proposed a new index modulation termed P-WGBM for wideband hybrid mmWave m-MIMO. P-WGBM evolves from WGBM by actively involving the digital precoder. Starting with determining a feasible precoding set, an optimization problem has been formulated, seeking the optimal precoder that retains multiplexing merit.

To circumvent analytical intractability, we then devised a two-step simplification by exploiting the beamspace properties. Instead of resorting to complex processing, the resulting near-optimal precoder was efficiently obtained via projected gradient-descent algorithm. Simulations verified that a slightly increased complexity could yield more than 1dB coding gain over WGBM.

REFERENCES

- [1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [2] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.
- [3] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Apr. 2015.
- [4] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [5] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. 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.
- [6] S. Gao, X. Cheng, and L. Yang, "Spatial multiplexing with limited RF Chains: Generalized beamspace modulation (GBM) for mmWave massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2029–2039, Sep. 2019.
- [7] R. Y. 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.
- [8] E. Basar, "Index modulation techniques for 5G wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 168–175, Jul. 2016.
- [9] 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.
- [10] S. Guo, H. Zhang, P. Zhang, P. Zhao, L. Wang, and M.-S. Alouini, "Generalized beamspace modulation using multiplexing: A breakthrough in mmWave MIMO," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2014–2028, Sep. 2019.
- [11] Y. Fan, S. Gao, X. Cheng, L. Yang, and N. Wang, "Wideband generalized beamspace modulation (wGBM) for mmWave massive MIMO over doubly-selective channels," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6869–6880, Jul. 2021.
- [12] R. Zhang, L.-L. Yang, and L. Hanzo, "Generalised pre-coding aided spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5434–5443, Nov. 2013.
- [13] S. Guo, H. Zhang, P. Zhang, S. Dang, C. Xu, and M. Alouini, "Signal shaping for non-uniform beamspace modulated mmWave hybrid MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6660–6674, Oct. 2020.
- [14] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [15] P. Yang, Y. L. Guan, Y. Xiao, M. Di Renzo, S. Li, and L. Hanzo, "Transmit precoded spatial modulation: Maximizing the minimum Euclidean distance versus minimizing the bit error ratio," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2054–2068, Mar. 2016.
- [16] S. Gao, X. Cheng, and L. Yang, "Mutual information maximizing wideband multi-user (wMU) mmWave massive MIMO," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3067–3078, May 2021.
- [17] K. Venugopal, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
- [18] L. Xiao, L. Dan, Y. Zhang, Y. Xiao, P. Yang, and S. Li, "A low-complexity detection scheme for generalized spatial modulation aided single carrier systems," *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 1069–1072, Jun. 2015.