SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

February 2025, Vol. 68, Iss. 2, 122301:1–122301:14
https://doi.org/10.1007/s11432-024-4229-1

Cross-subcarrier precoder design for massive MIMO-OFDM downlink with symplectic optimization

Yuxuan ZHANG[1,2], An-An LU[1,2], Xiang-Gen XIA[3] & Xiqi GAO[1,2*]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China*
[2]*Purple Mountain Laboratories, Nanjing 211100, China*
[3]*Department of Electrical and Computer Engineering, University of Delaware, Newark DE 19716, USA*

Received 13 July 2024/Revised 21 September 2024/Accepted 21 November 2024/Published online 9 January 2025

Abstract We propose a cross-subcarrier precoder design (CSPD) for massive multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) downlink. This work aims to significantly improve the channel estimation and signal detection performance by enhancing the smoothness of the frequency domain effective channel. This is accomplished by designing a few vectors known as the transform domain precoding vectors (TDPVs), which are then transformed into the frequency domain to generate the precoders for a set of subcarriers. To combat the effect of channel aging, the TDPVs are optimized under imperfect channel state information (CSI). The optimal precoder structure is derived by maximizing an upper bound of the ergodic weighted sum-rate (WSR) and utilizing the a posteriori beam-based statistical channel model (BSCM). To avoid the large-dimensional matrix inversion, we propose an algorithm with symplectic optimization. Simulation results indicate that the proposed cross-subcarrier precoder design significantly outperforms conventional methods.

Keywords massive MIMO-OFDM, cross subcarrier precoder design, transform domain precoding vectors, effective channel, symplectic optimization

Citation Zhang Y X, Lu A-A, Xia X-G, et al. Cross-subcarrier precoder design for massive MIMO-OFDM downlink with symplectic optimization. Sci China Inf Sci, 2025, 68(2): 122301, https://doi.org/10.1007/s11432-024-4229-1

1 Introduction

Massive multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) is crucial for the fifth-generation (5G) cellular communication systems and will be pivotal for future sixth-generation (6G) systems [1]. By deploying numerous antennas at a base station (BS) to serve many users simultaneously, it significantly enhances spectral and energy efficiencies. There have been many studies for massive MIMO-OFDM systems on subjects such as channel estimation, precoding, and positionings [2–5]. As interference among users can impair the performance of a communication system, precoder design at the BS is essential [6–12].

Conventional massive MIMO precoder designs, such as the regularized zero-forcing (RZF) precoder [13], the signal-to-leakage and-noise ratio (SLNR) precoder [14], and the weighted minimum mean-squared-error (WMMSE) precoder [15] relying on perfect channel state information (CSI), have been intensively studied. The precoders in these studies are per-subcarrier precoder designs, where each precoder is independently designed for each subcarrier. As a result, the smoothness of effective channels between adjacent subcarriers is disrupted. This complicates channel estimation and degrades system performance at the receiver side [16].

To preserve the smoothness of the effective channel, a common remedy is to design a frequency-flat precoder for all subcarriers [17]. In [18], by exploiting the high correlation between adjacent subcarriers, a singular value decomposition algorithm is proposed, based on which a frequency smoothed beamformer is designed to ensure smooth effective channels across all subcarriers. However, a rate degradation appears due to the frequency-selectivity of the channels. In [19], the authors proposed to use the same precoder for a small group of adjacent subcarriers. Compared with a single precoder design for all subcarriers, the

* Corresponding author (email: xqgao@seu.edu.cn)

subcarrier grouping method achieves a better performance. However, subcarriers at the edge of the group experience channel discontinuities, which hinder the channel estimation. To address these discontinuities, a phase-rotated singular value decomposition (SVD) is proposed. This approach leverages the phase non-uniqueness of SVD to smooth the effective channel, thereby enhancing channel estimation in single-user scenarios [20]. In [21], a smooth beamforming technique is proposed by incrementally computing the beamforming weights across subcarriers to ensure the smoothness of the frequency domain effective channel.

Due to the channel aging, it is challenging to obtain instantaneous CSI at the transmitter. The robust RZF precoder and the robust SLNR precoder were proposed to improve the weighted sum-rate (WSR) performance under imperfect CSI [22, 23]. The stochastic WMMSE precoder was proposed to further enhance the WSR performance in [24]. However, this precoder design involves iterations with multiple channel realizations, which might be complicated. In [25], the robust precoder design under imperfect CSI that maximizes an upper bound of the ergodic WSR is proposed. These precoders over subcarriers are designed individually.

In this paper, we investigate the precoder design for massive MIMO-OFDM under imperfect CSI. As we can see that the precoder design in [26] for each individual subcarrier extends the delay spread in the inverse discrete Fourier transform (IDFT) domain, we introduce a cross-subcarrier precoder design (CSPD) by designing the transform domain precoding vectors (TDPVs). The delay spread of the effective channel in the IDFT domain may be reduced with the proposed precoders. We want to emphasize that although each subcarrier is precoded individually for a low complexity, the precoders over different subcarriers are designed jointly. The optimal precoder structure is obtained by maximizing an upper bound of the ergodic WSR, considering the beam-based statistical channel model (BSCM) [27]. To avoid the large-dimensional matrix inversion in the optimal precoder structure, an algorithm based on symplectic optimization [28] is proposed. This algorithm draws from concepts in Hamiltonian dynamics and symplectic integration. Simulation results demonstrate that the proposed precoder facilitates the effective channel estimation and improve the signal detection performance significantly. Meanwhile, the transform-based CSPD decreases computational complexity while incurring a slight performance loss compared with conventional precoder designs.

The remainder of the paper is structured as follows. Section 2 presents the system model and problem formulation. Section 3 proposes the transform-based CSPD and the symplectic optimization algorithm for precoder design with low complexity. Section 4 presents simulation results illustrating the benefits of the proposed precoder. The paper concludes in Section 5.

Notations. The conjugate, transpose, and hermitian of matrix $\boldsymbol{A}$ are respectively defined as $\boldsymbol{A}^*$, $\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{H}}$. The trace of $\boldsymbol{A}$ is represented as the operator $\mathrm{tr}(\boldsymbol{A})$. The operators $\boldsymbol{A} \otimes \boldsymbol{B}$ and $\boldsymbol{A} \odot \boldsymbol{B}$ denote the Kronecker and the Hadamard product of two matrices, respectively. The vectorization of matrix $\boldsymbol{A}$ represented by $\mathrm{vec}\{\boldsymbol{A}\}$ is obtained by stacking its columns into a single column vector. The identity matrix of size M is defined as $\boldsymbol{I}_M$. The expectation operation is defined as $\mathbb{E}\{\cdot\}$. The operation diag(a) generates a diagonal matrix with its main diagonal filled by the elements of $\boldsymbol{a}$ accordingly. The matrix $\boldsymbol{A} = \mathrm{Bdiag}\{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N\}$ generates a block diagonal matrix with its main block diagonal filled by $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N$. The operators $\mathrm{Re}\{\boldsymbol{a}\}$ and $\mathrm{Im}\{\boldsymbol{a}\}$ represent the real and imaginary parts of complex vector $\boldsymbol{a}$, respectively. The inner product of $\boldsymbol{a} \in \mathbb{R}^{M\times 1}$ and $\boldsymbol{b} \in \mathbb{C}^{M\times 1}$ are defined as $\langle \boldsymbol{a}, \boldsymbol{a} \rangle_R = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ and $\langle \boldsymbol{b}, \boldsymbol{b} \rangle = \boldsymbol{b}^{\mathrm{H}}\boldsymbol{b}$, respectively.

2 System model and problem formulation

In this section, we present the system configuration and the BSCM. Further, we formulate the cross-subcarrier precoder design.

2.1 System configuration

The considered massive MIMO system includes one BS with a uniform planar array (UPA) of $M = M_z M_x$ antennas, serving K single-antenna users. Here, M_x and M_z are the numbers of horizontal and vertical antennas, respectively. The system employs OFDM modulation with N_c subcarriers where N_v subcarriers are used for transmission. Each slot consists of N_b OFDM symbols. Channel parameters remain constant within each OFDM symbol but vary across symbols.

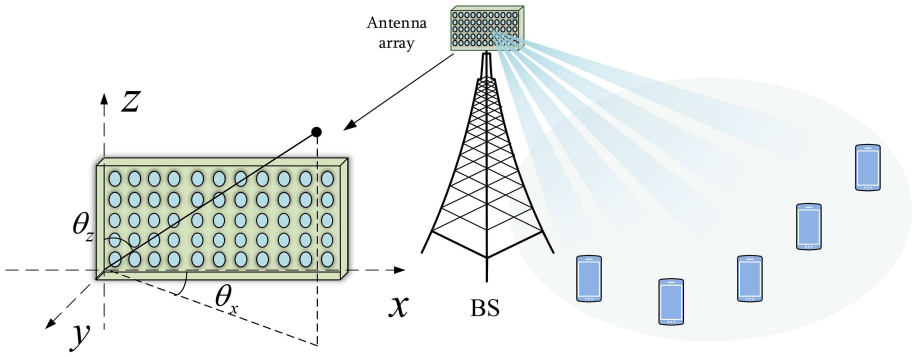

Figure 1 (Color online) Massive MIMO system.

2.2 Beam-based statistical channel model

We briefly introduce the BSCM proposed in [27] for massive MIMO system. The UPA is placed on an xz plane. Define the spacings of antennas in x and z directions as d_x and d_z, respectively. Let θ_z and θ_x be the elevation and azimuth angles. The directional cosines of the UPA are defined as $\alpha = \cos\theta_z$ and $\beta = \sin\theta_z \cos\theta_x$, respectively. Figure 1 depicts the system diagram of the massive MIMO system. The carrier frequency is f_c, and the wavelength is $\lambda_c = \frac{v_c}{f_c}$, where v_c denotes the speed of light. Let $\Delta_x = \frac{d_x}{\lambda_c}$ and $\Delta_z = \frac{d_z}{\lambda_c}$. Let

$$\boldsymbol{v}_x(\alpha) = \frac{1}{M_x}[1, \mathrm{e}^{-\mathrm{j}2\pi\Delta_x\alpha}, \ldots, \mathrm{e}^{-(M_x-1)\mathrm{j}2\pi\Delta_x\alpha}]^{\mathrm{T}}, \quad \boldsymbol{v}_z(\beta) = \frac{1}{M_z}[1, \mathrm{e}^{-\mathrm{j}2\pi\Delta_z\beta}, \ldots, \mathrm{e}^{-(M_z-1)\mathrm{j}2\pi\Delta_z\beta}]^{\mathrm{T}}.$$

Then, the steering vector $\boldsymbol{v}(\alpha, \beta)$ of the UPA is decomposed as

$$\boldsymbol{v}(\alpha, \beta) = \boldsymbol{v}_x(\alpha) \otimes \boldsymbol{v}_z(\beta) \in \mathbb{C}^{M\times 1}. \tag{1}$$

Let $\tilde{M}_x \geqslant M_x$ and $\tilde{M}_z \geqslant M_z$ be the sampling numbers in α and β, respectively. Let $\alpha_i = (2i - 1 - \tilde{M}_x)/\tilde{M}_x$ and $\beta_j = (2j-1-\tilde{M}_z)/\tilde{M}_z$ be the sampled directional cosines. Based on (1), we define the matrix $\boldsymbol{V}$ as $\boldsymbol{V} = \boldsymbol{V}_x \otimes \boldsymbol{V}_z \in \mathbb{C}^{M\times \tilde{M}_x\tilde{M}_z}$ where $\boldsymbol{V}_x = [\boldsymbol{v}_x(\alpha_1), \ldots, \boldsymbol{v}_x(\alpha_{\tilde{M}_x})]$ and $\boldsymbol{V}_z = [\boldsymbol{v}_z(\beta_1), \ldots, \boldsymbol{v}_z(\beta_{\tilde{M}_z})]$.

The channel vector from the BS to user k at the n-th OFDM symbol of the m-th time slot on the c-th subcarrier is given as

$$\boldsymbol{h}_{k,m,n,c} = \boldsymbol{g}_{k,m,n,c}\boldsymbol{V}^{\mathrm{H}} = (\boldsymbol{m}_k \odot \boldsymbol{w}_{k,m,n,c})\boldsymbol{V}^{\mathrm{H}}, \tag{2}$$

where $\boldsymbol{g}_{k,m,n,c} = \boldsymbol{m}_k \odot \boldsymbol{w}_{k,m,n,c}$ denotes the beam domain channel, $\boldsymbol{m}_k$ is a deterministic vector of size $\tilde{M}_z\tilde{M}_x$, and $\boldsymbol{w}_{k,m,n,c}$ consists of independent and identically distributed elements following $\mathcal{CN}(0, 1)$ distribution. Let

$$\boldsymbol{\omega}_{k,m,n,c} = \mathbb{E}\{\boldsymbol{g}_{k,m,n,c} \odot \boldsymbol{g}^*_{k,m,n,c}\} = \boldsymbol{m}_k \odot \boldsymbol{m}_k \in \mathbb{R}^{1\times \tilde{M}_x\tilde{M}_z}, \tag{3}$$

which is independent of the OFDM symbols and subcarriers, and remains constant over multiple time slots. Thus, let $\boldsymbol{\omega}_k = \boldsymbol{\omega}_{k,m,n,c}$ for simplicity by dropping indices m, n, and c, which is referred to as the statistical CSI.

2.3 A posteriori beam based statistical channel model

To consider the impact of imperfect CSI, we follow the a posteriori BSCM in [25]. This model is derived from the prior channel model described in (2) and incorporates a Gauss-Markov process.

Let $\tilde{\boldsymbol{g}}_{k,m,1,c}$ be the ideal beam domain channel at the first OFDM symbol. The a posteriori channel model is given as

$$\boldsymbol{h}_{k,m,n,c} = \rho_k\tilde{\boldsymbol{g}}_{k,m,1,c}\boldsymbol{V}^{\mathrm{H}} + \sqrt{1-\rho_k^2}(\boldsymbol{m}_k \odot \boldsymbol{w}_{k,m,n,c})\boldsymbol{V}^{\mathrm{H}}, \tag{4}$$

where the temporal correlation coefficient is defined as ρ_k, which is related to Jakes' autocorrelation model [29]. Specifically, $\rho_k = J_0(2\pi v_k f_c T/v_c)$, where $J_0(\cdot)$ represents the zero-th order Bessel function

of the first kind, v_k is the moving speed of the k-th user. The coefficient ρ_k approaches 1 in quasi-static scenarios but decreases to a smaller value as users move at higher speeds. Adjusting ρ_k allows modeling channels with different mobility speeds. Importantly, the a posteriori BSCM is a general model applicable to both perfect and imperfect CSI scenarios.

Let $\bar{\boldsymbol{h}}_{k,m,n,c} = \rho_k \tilde{\boldsymbol{g}}_{k,m,1,c} \boldsymbol{V}^{\mathrm{H}}$ and $\tilde{\boldsymbol{h}}_{k,m,n,c} = \sqrt{1-\rho_k^2}(\boldsymbol{m}_k \odot \boldsymbol{w}_{k,m,n,c})\boldsymbol{V}^{\mathrm{H}}$ be the deterministic and the stochastic components of the channel, respectively. The posterior channel model in (4) can be decomposed as

$$\boldsymbol{h}_{k,m,n,c} = \bar{\boldsymbol{h}}_{k,m,n,c} + \tilde{\boldsymbol{h}}_{k,m,n,c}. \tag{5}$$

For convenience, let

$$\eta_k(\boldsymbol{T}) = \mathbb{E}\{\tilde{\boldsymbol{h}}_{k,m,n,c} \boldsymbol{T} \tilde{\boldsymbol{h}}_{k,m,n,c}^{\mathrm{H}}\} = (1-\rho_k^2) \sum_{j=1}^{\tilde{M}_z \tilde{M}_x} [\boldsymbol{\omega}_k^{\mathrm{T}}]_j [\boldsymbol{V}^{\mathrm{H}} \boldsymbol{T} \boldsymbol{V}]_{jj}, \tag{6}$$

where $\boldsymbol{T} \in \mathbb{C}^{M \times M}$ is an argument matrix. Additionally, we define $\boldsymbol{\Xi}_k \in \mathbb{C}^{M \times M}$ as

$$\boldsymbol{\Xi}_k = \mathbb{E}\{\tilde{\boldsymbol{h}}_{k,m,n,c}^{\mathrm{H}} \tilde{\boldsymbol{h}}_{k,m,n,c}\} = (1-\rho_k^2) \boldsymbol{V} \mathrm{diag}(\boldsymbol{\omega}_k^{\mathrm{T}}) \boldsymbol{V}^{\mathrm{H}}. \tag{7}$$

The parameterized channel covariance matrix and the channel covariance matrix definitions in (6) and (7) are used for the stochastic channel-related calculation in Subsection 3.2.

2.4 Problem formulation

We define $\boldsymbol{h}_{k,c} = \boldsymbol{h}_{k,m,n,c}$ for simplicity by dropping indices m, n. Define $\boldsymbol{p}_k^c \in \mathbb{C}^{M \times 1}$ and $x_{k,c}$ as the precoding vector and the transmitted signal for the k-th user on the c-th subcarrier, respectively. The received signal $y_{k,c}$ is expressed as

$$y_{k,c} = \boldsymbol{h}_{k,c} \boldsymbol{p}_k^c x_{k,c} + \boldsymbol{h}_{k,c} \sum_{l \neq k}^{K} \boldsymbol{p}_l^c x_{l,c} + z_{k,c}, \tag{8}$$

where $z_{k,c}$ is modeled as a complex Gaussian random variable distributed as $\mathcal{CN}(0, \sigma_z^2)$. Let $\tilde{z}_{k,c} = \boldsymbol{h}_{k,c} \sum_{l \neq k}^{K} \boldsymbol{p}_l^c x_{l,c} + z_{k,c}$ be the Gaussian noise [25]. The variance of $\tilde{z}_{k,c}$ is given as

$$r_{k,c} = \sigma_z^2 + \sum_{l \neq k}^{K} \mathbb{E}\{\boldsymbol{h}_{k,c} \boldsymbol{p}_l^c (\boldsymbol{p}_l^c)^{\mathrm{H}} \boldsymbol{h}_{k,c}^{\mathrm{H}}\}. \tag{9}$$

Assume that $r_{k,c}$ is obtained for the k-th user. Let $\mathcal{R}_{k,c} = \mathbb{E}\{\log(1 + r_{k,c}^{-1} \boldsymbol{h}_{k,c} \boldsymbol{p}_k^c (\boldsymbol{p}_k^c)^{\mathrm{H}} \boldsymbol{h}_{k,c}^{\mathrm{H}})\}$. Then, the ergodic rate for user k is given as

$$\mathcal{R}_k = \sum_{c=1}^{N_v} \mathcal{R}_{k,c}. \tag{10}$$

Define w_k as the weight factor for the k-th user. We formulate the precoder design as an ergodic WSR maximization problem, expressed as follows:

$$\max_{\{\boldsymbol{p}_k^c | 1 \leqslant c \leqslant N_v, 1 \leqslant k \leqslant K\}} \sum_{k=1}^{K} w_k \mathcal{R}_k \qquad \text{s.t.} \ \sum_{k=1}^{K} \sum_{c=1}^{N_v} (\boldsymbol{p}_k^c)^{\mathrm{H}} \boldsymbol{p}_k^c \leqslant P, \tag{11}$$

where P is the power budget. The conventional per-subcarrier precoder design in terms of the ergodic WSR maximization at each subcarrier with total power constraint is

$$\max_{\{\boldsymbol{p}_k^c | 1 \leqslant k \leqslant K\}} \sum_{k=1}^{K} w_k \mathcal{R}_{k,c} \qquad \text{s.t.} \ \sum_{k=1}^{K} (\boldsymbol{p}_k^c)^{\mathrm{H}} \boldsymbol{p}_k^c \leqslant P_c, \tag{12}$$

where P_c is the power budget at the c-th subcarrier. The subcarrier grouping precoder design is based on the per-subcarrier precoder design, using the same precoder for a small group of adjacent subcarriers. In (11), the precoders over all used subcarriers are jointly optimized. In this way, the smoothness of effective channels between adjacent subcarriers can be maintained by carefully designing precoders. Although we design precoder vectors for all used subcarriers, the following method applies to any set of subcarriers. Let $\boldsymbol{p}_k = [(\boldsymbol{p}_k^1)^{\mathrm{T}}, (\boldsymbol{p}_k^2)^{\mathrm{T}}, \ldots, (\boldsymbol{p}_k^{N_v})^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{N_v M \times 1}$ be the vector consisting of precoders for the k-th user across subcarriers. To make the effective channel smooth in the frequency domain, let

$$\boldsymbol{p}_k = (\boldsymbol{F} \otimes \boldsymbol{I}_M)\boldsymbol{q}_k, \tag{13}$$

where $\boldsymbol{q}_k = [\boldsymbol{q}_{k,1}^{\mathrm{T}}, \boldsymbol{q}_{k,2}^{\mathrm{T}}, \ldots, \boldsymbol{q}_{k,N}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{NM \times 1}$, $\boldsymbol{q}_{k,i}$ are defined as TDPVs, and $\boldsymbol{F} \in \mathbb{C}^{N_v \times N}$ is a partial discrete Fourier transform (DFT) matrix. For simplicity, $\boldsymbol{F}$ is obtained by selecting the first N columns of the DFT matrix $\boldsymbol{U}$ of size N_v. Here, N represents the degrees of freedom for designing the precoding vectors $\boldsymbol{p}_k$ for each k and is chosen to be much smaller than the number of used subcarriers N_v, which improves the smoothness of the effective channels, as detailed later.

According to (13), $\boldsymbol{p}_k^c$ can be expressed as

$$\boldsymbol{p}_k^c = \boldsymbol{E}_c(\boldsymbol{F} \otimes \boldsymbol{I}_M)\boldsymbol{q}_k, \tag{14}$$

where $\boldsymbol{E}_c \in \mathbb{C}^{M \times N_v M}$ is an extracting matrix defined as $[\boldsymbol{0}, \ldots, \boldsymbol{I}_M, \ldots, \boldsymbol{0}]$, and $\boldsymbol{I}_M$ is the c-th submatrix. The smoothness of the effective channels across subcarriers can be adjusted by setting different numbers N of TDPVs, which will be clarified at the end of this subsection. Let $\boldsymbol{F}_c = \boldsymbol{E}_c(\boldsymbol{F} \otimes \boldsymbol{I}_M)$ for simplicity. We rewrite $r_{k,c}$ by substituting (14) into (9) as

$$r_{k,c} = \sigma_z^2 + \sum_{l \neq k}^{K} \mathbb{E}\{\boldsymbol{h}_{k,c}\boldsymbol{F}_c\boldsymbol{q}_l\boldsymbol{q}_l^{\mathrm{H}}\boldsymbol{F}_c^{\mathrm{H}}\boldsymbol{h}_{k,c}^{\mathrm{H}}\}. \tag{15}$$

Let $\tilde{r}_{k,c} = r_{k,c} + \mathbb{E}\{\boldsymbol{h}_{k,c}\boldsymbol{F}_c\boldsymbol{q}_k\boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{F}_c^{\mathrm{H}}\boldsymbol{h}_{k,c}^{\mathrm{H}}\}$. Then, $\mathcal{R}_k$ is rewritten as

$$\mathcal{R}_k = \mathbb{E}\left\{\sum_{c=1}^{N_v} \log(1 + r_{k,c}^{-1}\boldsymbol{h}_{k,c}\boldsymbol{F}_c\boldsymbol{q}_k\boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{F}_c^{\mathrm{H}}\boldsymbol{h}_{k,c}^{\mathrm{H}})\right\}. \tag{16}$$

Since $\log(\cdot)$ is a concave function, we establish an upper bound of $\mathcal{R}_k$ to circumvent computationally expensive expectations using Jensen's inequality as

$$\mathcal{R}_k^{\mathrm{ub}} = \sum_{c=1}^{N_v} \log(1 + r_{k,c}^{-1}\mathbb{E}\{\boldsymbol{h}_{k,c}\boldsymbol{F}_c\boldsymbol{q}_k\boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{F}_c^{\mathrm{H}}\boldsymbol{h}_{k,c}^{\mathrm{H}}\}). \tag{17}$$

Compared with $\mathcal{R}_k$, the upper bound $\mathcal{R}_k^{\mathrm{ub}}$ is rather tight in the low and middle signal-to-noise ratio (SNR) regimes and has a slight degradation in the high SNR regime. Let $f(\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_K) = \sum_{k=1}^{K} w_k\mathcal{R}_k^{\mathrm{ub}}$ denote the upper bound of the ergodic WSR. Then, the problem for designing precoders across subcarriers is reformulated to obtain the stacked TDPVs that maximize the upper bound of the ergodic WSR,

$$\arg\max_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_K} f(\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_K) \qquad \text{s.t.} \ \sum_{k=1}^{K} \boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{q}_k \leqslant P. \tag{18}$$

In the following, we reveal that the smoothness of the effective channels across subcarriers can be adjusted by setting different numbers N of TDPVs. Let $\boldsymbol{H}_k = \mathrm{Bdiag}\{\boldsymbol{h}_{k,1}, \boldsymbol{h}_{k,2}, \ldots, \boldsymbol{h}_{k,N_v}\} \in \mathbb{C}^{N_v \times N_v M}$. Let $\boldsymbol{H}_k^m = \mathrm{Bdiag}\{\boldsymbol{h}_{k,1}[m], \boldsymbol{h}_{k,2}[m], \ldots, \boldsymbol{h}_{k,N_v}[m]\} \in \mathbb{C}^{N_v \times N_v}$ and $\boldsymbol{q}_k^m = [\boldsymbol{q}_{k,1}[m], \boldsymbol{q}_{k,2}[m], \ldots, \boldsymbol{q}_{k,N}[m]]^{\mathrm{T}} \in \mathbb{C}^{N \times 1}$ denote the frequency domain channel and the TDPVs at the m-th antenna, respectively. $\boldsymbol{h}[m]$ is the m-th element of $\boldsymbol{h}$. Let $\boldsymbol{U}$ be the DFT matrix of size N_v. The frequency domain effective channel $\boldsymbol{h}_k^f$ of user k is given as

$$\boldsymbol{h}_k^f = \boldsymbol{H}_k(\boldsymbol{F} \otimes \boldsymbol{I}_M)\boldsymbol{q}_k = \sum_{m=1}^{M} \boldsymbol{H}_k^m\boldsymbol{U}\bar{\boldsymbol{q}}_k^m, \tag{19}$$

where $\bar{\boldsymbol{q}}_k^m = [(\boldsymbol{q}_k^m)^{\mathrm{T}}, 0, \ldots, 0]^{\mathrm{T}} \in \mathbb{C}^{N_v \times 1}$. Let $\boldsymbol{h}_k^d = \boldsymbol{U}^{\mathrm{H}} \boldsymbol{H}_k (\boldsymbol{F} \otimes \boldsymbol{I}_M) \boldsymbol{q}_k$ denote the delay domain effective channel. The relationship between $\boldsymbol{h}_k^f$ and $\boldsymbol{h}_k^d$ is given as

$$\boldsymbol{h}_k^f = \boldsymbol{U} \boldsymbol{h}_k^d. \tag{20}$$

Further, the frequency domain channel at the m-th antenna $\boldsymbol{H}_k^m$ is decompose as

$$\boldsymbol{H}_k^m = \boldsymbol{U} \bar{\boldsymbol{H}}_k^m \boldsymbol{U}^{\mathrm{H}}, \tag{21}$$

where $\bar{\boldsymbol{H}}_k^m$ is a circulant matrix of the delay domain channel at the m-th antenna. By substituting (21) into (19), $\boldsymbol{h}_k^f$ is rewritten as

$$\boldsymbol{h}_k^f = \sum_{m=1}^{M} \boldsymbol{U} \bar{\boldsymbol{H}}_k^m \bar{\boldsymbol{q}}_k^m. \tag{22}$$

Since $\bar{\boldsymbol{H}}_k^m$ is a circulant matrix, we have $\bar{\boldsymbol{H}}_k^m \bar{\boldsymbol{q}}_k^m = \bar{\boldsymbol{H}}_k^m[:, 1] * \bar{\boldsymbol{q}}_k^m$, where $*$ denotes the circular convolution operator and $\bar{\boldsymbol{H}}[:, i]$ is the i-th column vector of $\bar{\boldsymbol{H}}$. Further, we have

$$\boldsymbol{h}_k^d = \sum_{m=1}^{M} \bar{\boldsymbol{H}}_k^m[:, 1] * \bar{\boldsymbol{q}}_k^m. \tag{23}$$

Let N_{de} be the delay spread of the delay domain channel. According to (23) and the property of circular convolution, the delay spread of $\boldsymbol{h}_k^d$ is no larger that $N_{\mathrm{de}} + N - 1$. For the per-subcarrier precoder design that corresponds to the case when $N = N_v$ above, the delay spread of the delay domain effective channel may be as large as N_v. Since N_{de} is much smaller than N_v in a practical massive MIMO system, the delay spread of $\boldsymbol{h}_k^d$ in the IDFT domain can be reduced by adjusting N with our proposed joint design of the precoding vectors across subcarriers. Thus, the smoothness of the effective channels across subcarriers can be adjusted by setting different numbers N of TDPVs.

3 Cross subcarrier precoder design

In this section, we solve the optimization problem for the transform-based CSPD, and a precoder structure is obtained. To circumvent the large-dimensional matrix inversion in the precoder structure, a low-complexity algorithm using symplectic optimization is proposed.

3.1 Precoder structure

The minorize-maximization (MM) is an iterative optimization technique to find the local maximum of a nonconvex function. Let $\boldsymbol{q} \triangleq \{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_K\}$ denote the sequence of the stacked TDPVs. Further, the sequence of the stacked TDPVs at the d-th iteration is defined as $\boldsymbol{q}[d]$. A function $g(\boldsymbol{q}|\boldsymbol{q}[d])$ is deemed as a minorizing function of $f(\boldsymbol{q})$ at the d-th iteration of the stacked TDPVs if $g(\boldsymbol{q}|\boldsymbol{q}[d])$ satisfies the following conditions:

$$\begin{aligned} &g(\boldsymbol{q}|\boldsymbol{q}[d]) \leqslant f(\boldsymbol{q}), \\ &g(\boldsymbol{q}[d]|\boldsymbol{q}[d]) = f(\boldsymbol{q}[d]), \\ &\frac{\partial g(\boldsymbol{q}|\boldsymbol{q}[d])}{\partial \boldsymbol{q}_k^*}\bigg|_{\boldsymbol{q}_k = \boldsymbol{q}_k[d]} = \frac{\partial f(\boldsymbol{q})}{\partial \boldsymbol{q}_k^*}\bigg|_{\boldsymbol{q}_k = \boldsymbol{q}_k[d]}. \end{aligned} \tag{24}$$

Then, the stacked TDPVs in the next iteration are updated as

$$\boldsymbol{q}[d+1] = \arg\max_{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_K} g(\boldsymbol{q}|\boldsymbol{q}[d]). \tag{25}$$

Note that the surrogate function in each iteration is convex. Ultimately, the sequence in (25) has at least one feasible limit point if g is concave as the number of iterations increases. This indicates that a converging subsequence of the sequence converges to a point in the set of all stationary points [30].

Define matrices $\boldsymbol{A}_{k,c}[d]$ and $\boldsymbol{D}_c[d]$ as

$$\boldsymbol{A}_{k,c}[d] = r_{k,c}^{-1}[d]\boldsymbol{R}_{k,c}, \tag{26}$$

$$\boldsymbol{C}_{k,c}[d] = (r_{k,c}^{-1}[d] - \tilde{r}_{k,c}^{-1}[d])\boldsymbol{R}_{k,c}, \tag{27}$$

$$\boldsymbol{D}_c[d] = \sum_{k=1}^{K} w_k \boldsymbol{C}_{k,c}[d], \tag{28}$$

where $\boldsymbol{R}_{k,c} = \mathbb{E}\{\boldsymbol{h}_{k,c}^{\mathrm{H}}\boldsymbol{h}_{k,c}\}$ denotes the covariance matrix. From [25], we provide a minorizing function g satisfying conditions (24) as

$$\begin{aligned} g =& \gamma[d] + \sum_{k=1}^{K}\sum_{c=1}^{N_v} w_k \mathrm{tr}(\boldsymbol{A}_{k,c}[d]\boldsymbol{F}_c\boldsymbol{q}_k(\boldsymbol{q}_k[d])^{\mathrm{H}}\boldsymbol{F}_c^{\mathrm{H}}) + \sum_{k=1}^{K}\sum_{c=1}^{N_v} w_k \mathrm{tr}(\boldsymbol{A}_{k,c}[d]\boldsymbol{F}_c\boldsymbol{q}_k[d]\boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{F}_c^{\mathrm{H}}) \\ &- \sum_{k=1}^{K}\sum_{c=1}^{N_v} \mathrm{tr}(\boldsymbol{D}_c[d]\boldsymbol{F}_c\boldsymbol{q}_k\boldsymbol{q}_k^{\mathrm{H}}\boldsymbol{F}_c^{\mathrm{H}}), \end{aligned} \tag{29}$$

where $\gamma[d]$ is a constant, and $\boldsymbol{A}_{k,c}[d]$ and $\boldsymbol{D}_c[d]$ are provided in (26) and (28), respectively. The derivation process is given in Appendix A. Based on the minorizing function and the MM methodology, we reformulate the optimization problem (18) and solve the problem using the Lagrangian method. The precoder structure is provided below.

Theorem 1. The solution of the optimization problem (18) has the following form:

$$\boldsymbol{q}_k[d+1] = w_k(\hat{\boldsymbol{D}}[d] + \mu^*\boldsymbol{I}_{NM})^{-1}\hat{\boldsymbol{A}}_k[d], \boldsymbol{q}_k[d], \tag{30}$$

where $\hat{\boldsymbol{D}}[d]$ and $\hat{\boldsymbol{A}}_k[d]$ are calculated respectively as

$$\hat{\boldsymbol{D}}[d] = \sum_{c=1}^{N_v} \boldsymbol{F}_c^{\mathrm{H}}\boldsymbol{D}_c[d]\boldsymbol{F}_c, \tag{31}$$

$$\hat{\boldsymbol{A}}_k[d] = \sum_{c=1}^{N_v} \boldsymbol{F}_c^{\mathrm{H}}\boldsymbol{A}_{k,c}[d]\boldsymbol{F}_c. \tag{32}$$

Proof. See Appendix A.

The solution given in (30) considers the available CSI for all subcarriers. Hence, the potential performance gain can be exploited. The dimension of the matrix inversion in (30) is NM, resulting in high complexity with larger values of N.

3.2 Low-complexity precoder design with symplectic optimization

We propose a low-complexity transform-based CSPD using symplectic optimization, inspired by concepts from Hamiltonian dynamical systems and symplectic integration. Symplectic optimization preserves the continuous symmetries of the original dynamical system, enabling larger step sizes. Consequently, algorithms based on symplectic optimization can move through the configuration space more quickly. This method achieves a provably faster rate than gradient descent and meets the oracle lower bound. Additionally, symplectic optimization performs efficiently when applied to large-scale data analysis [31]. To efficiently calculate the solution provided in (30), we define a quadratic function as

$$s(\boldsymbol{q}_k) = \mathrm{tr}((\hat{\boldsymbol{D}}[d] + \mu^*\boldsymbol{I}_{NM})\boldsymbol{q}_k\boldsymbol{q}_k^{\mathrm{H}}) - \mathrm{tr}(\hat{\boldsymbol{A}}_k[d]\boldsymbol{q}_k[d]\boldsymbol{q}_k^{\mathrm{H}}). \tag{33}$$

The precoder $\boldsymbol{q}_k[d+1]$ in (30) can be viewed as the unique minimizer of the optimization problem,

$$\boldsymbol{q}_k^o = \arg\min_{\boldsymbol{q}_k} s(\boldsymbol{q}_k). \tag{34}$$

For simplicity, the subscript k of $\boldsymbol{q}_k$ is omitted in the following derivations. To obtain the minimizer, a Hamiltonian dynamics system can be constructed where $\boldsymbol{q}$ is the location and $s(\boldsymbol{q})$ serves as the potential energy.

Let $\boldsymbol{r}$ be the momenta of the system. The Bregman Hamiltonian is evolved from the Bregman Lagrangian as [28]

$$H(\boldsymbol{q},\boldsymbol{r},t)=\mathrm{e}^{\alpha(t)+\gamma(t)}(k(\boldsymbol{q},\boldsymbol{r})+\mathrm{e}^{\beta(t)}s(\boldsymbol{q})), \tag{35}$$

where $\alpha(t)$, $\gamma(t)$, and $\beta(t)$ are scaling functions, $k(\boldsymbol{q},\boldsymbol{r})$ is the kinetic energy. Since $\mathrm{e}^{\beta(t)}s(\boldsymbol{q})$ serves as a time-dependent potential energy. As the time t goes on, the time dependence of the Bregman Hamiltonian allows the dynamics to rapidly converge to a minimum [32].

Let $\boldsymbol{v}$ denote the system velocity. The kinetic energy is defined as

$$k(\boldsymbol{q},\boldsymbol{v})=D_h(\mathrm{e}^{-\alpha(t)}\boldsymbol{v}_R+\boldsymbol{q}_R,\boldsymbol{q}_R), \tag{36}$$

where $\boldsymbol{q}_R,\boldsymbol{v}_R$, and the function $D_h(\cdot)$ are defined in Appendix B. Using the relation between $\boldsymbol{v}$ and $\boldsymbol{r}$, the kinetic energy is reformulated as

$$k(\boldsymbol{q},\boldsymbol{r})=\frac{\mathrm{e}^{-2\gamma(t)}\langle\boldsymbol{r},\boldsymbol{r}\rangle}{2}. \tag{37}$$

The concise derivation is given in Appendix B. Define the scaling factors as $\alpha(t)=\log p-\log t$, $\beta(t)=p\log t+\log C$, and $\gamma(t)=p\log t$, where $p,C\in\mathbb{R}^+$ [33].

The Bregman dynamics should satisfy the equations given as [28]

$$\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t}=+\frac{\partial H}{\partial\boldsymbol{r}}(\boldsymbol{q},\boldsymbol{r},t),\qquad \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}=-\frac{\partial H}{\partial\boldsymbol{q}}(\boldsymbol{q},\boldsymbol{r},t). \tag{38}$$

Because the above Hamiltonian system is dependent on time, the standard methods that preserve the critical symmetries cannot be directly used. To overcome this problem, it is transformed into an autonomous one by introducing auxiliary variables. Let t be a new position variable, i.e., $(\boldsymbol{q},t)$ is the new location variable. The autonomous Hamiltonian is defined as

$$H_{\boldsymbol{\Xi}}=\varepsilon-H(\boldsymbol{q},t,\boldsymbol{r}). \tag{39}$$

The auxiliary variable ε is introduced to offset the time dependence of the original Hamiltonian, ensuring that the Hamiltonian $H_{\boldsymbol{\Xi}}$ is autonomous. Let τ be the new time variable to parameterize the motion. The autonomous Hamiltonian should satisfy the following equations [34]:

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}\tau}&=+\frac{\partial H_{\boldsymbol{\Xi}}}{\partial\boldsymbol{r}}(\boldsymbol{q},t,\boldsymbol{r},\varepsilon),\qquad \frac{\mathrm{d}t}{\mathrm{d}\tau}=+\frac{\partial H_{\boldsymbol{\Xi}}}{\partial\varepsilon}(\boldsymbol{q},t,\boldsymbol{r},\varepsilon),\\ \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}\tau}&=-\frac{\partial H_{\boldsymbol{\Xi}}}{\partial\boldsymbol{q}}(\boldsymbol{q},t,\boldsymbol{r},\varepsilon),\qquad \frac{\mathrm{d}\varepsilon}{\mathrm{d}\tau}=-\frac{\partial H_{\boldsymbol{\Xi}}}{\partial t}(\boldsymbol{q},t,\boldsymbol{r},\varepsilon).\end{aligned}$$

In the following, a symplectic integrator is constructed to model the Bregman dynamics. Such integrators are created by splitting the Hamiltonian into component parts [35]. The dynamics of each part can be solved precisely, or close to the exact solution. Let $H_{\boldsymbol{\Xi}}$ be rewritten as

$$H_{\boldsymbol{\Xi}}=H_A+H_B+H_C, \tag{40}$$

where

$$H_A(\varepsilon)=\varepsilon,\quad H_B(\boldsymbol{q},\boldsymbol{r},t)=\mathrm{e}^{\alpha(t)-\gamma(t)}\langle\boldsymbol{r},\boldsymbol{r}\rangle,\quad H_C(\boldsymbol{q},t)=\mathrm{e}^{\alpha(t)+\beta(t)+\gamma(t)}s(\boldsymbol{q}).$$

These three component Hamiltonians vary in six vector fields given as

$$\begin{aligned}&\boldsymbol{H}_A=\frac{\mathrm{d}}{\mathrm{d}t},\quad \boldsymbol{H}_{B1}=\boldsymbol{0},\quad \boldsymbol{H}_{B2}=\frac{p(p+1)}{t^{p+2}}\langle\boldsymbol{r},\boldsymbol{r}\rangle\frac{\mathrm{d}}{\mathrm{d}\varepsilon},\\ &\boldsymbol{H}_{B3}=\frac{p}{t^{p+1}}\boldsymbol{r}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{q}},\quad \boldsymbol{H}_{C1}=-Cpt^{2p-1}\nabla s(\boldsymbol{q})\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{r}},\quad \boldsymbol{H}_{C2}=-Cp(2p-1)t^{2p-2}s(\boldsymbol{q})\frac{\mathrm{d}}{\mathrm{d}\varepsilon}.\end{aligned}$$

The vector fields $\boldsymbol{H}_A$, $\boldsymbol{H}_{B2}$, $\boldsymbol{H}_{C1}$, and $\boldsymbol{H}_{C2}$ are straightforward, allowing for exact solutions to their evolution. In contrast, the dynamics of $\boldsymbol{H}_{B1}$ and $\boldsymbol{H}_{B3}$ may be more complex, but can still be addressed

Algorithm 1 Symplectic optimization method.

1: Set iteration $n = 0$; initialize t_0, $\boldsymbol{r}_0$, and $\boldsymbol{q}_0$;
2: Update $t_{n+1/2} = t_n + \epsilon$;
3: Update $\boldsymbol{r}_{n+1/2} = \boldsymbol{r}_n - \epsilon Cpt^{2p-1}\nabla s(\boldsymbol{q}_n)$;
4: Update $\boldsymbol{q}_{n+1} = \boldsymbol{q}_n + \epsilon\frac{p}{t^{p+1}}\boldsymbol{r}_{n+1/2}$;
5: Update $\boldsymbol{r}_{n+1} = \boldsymbol{r}_{n+1/2} - \epsilon Cpt^{2p-1}\nabla s(\boldsymbol{q}_{n+1})$;
6: Update $t_{n+1} = t_{n+1/2} + \epsilon$ and update the iteration as $n = n + 1$;
7: Repeat Steps 2–6 until the preset goal is achieved or convergence is reached.

implicitly through fixed-point iteration techniques. Combining these component Hamiltonians and the second-order leapfrog integrator [36] then gives Algorithm 1. Note that the update of auxiliary variable ε is omitted in Algorithm 1.

In this way, the large-dimensional matrix inversion in (30) can be avoided with symplectic optimization. However, the complexity is not satisfied yet. fast Fourier transform (FFT) can be introduced to simplify the matrix multiplications and the computational complexity of transform-based CSPD with symplectic optimization can be further reduced. In the following, we explain this in more detail.

Let matrices $\boldsymbol{\Lambda}[d] = \mathrm{Bdiag}\{\boldsymbol{D}_1[d], \ldots, \boldsymbol{D}_{N_v}[d]\}$ and $\boldsymbol{\Gamma}_k[d] = \mathrm{Bdiag}\{\boldsymbol{A}_{k,1}[d], \ldots, \boldsymbol{A}_{k,N_v}[d]\}$ be the block diagonal matrices. The matrices $\bar{\boldsymbol{\Lambda}}[d]$ and $\bar{\boldsymbol{\Gamma}}_k[d]$ related to the deterministic channel and $\tilde{\boldsymbol{\Lambda}}[d]$ and $\tilde{\boldsymbol{\Gamma}}_k[d]$ related to the stochastic channel are defined similarly from $\bar{\boldsymbol{D}}_c[d]$, $\tilde{\boldsymbol{D}}_c[d]$, $\bar{\boldsymbol{A}}_{k,c}[d]$, and $\tilde{\boldsymbol{A}}_{k,c}[d]$, which are given as

$$\bar{\boldsymbol{D}}_c[d] = \sum_{k=1}^{K} w_k(r_{k,c}^{-1}[d] - \tilde{r}_{k,c}^{-1}[d])\bar{\boldsymbol{h}}_{k,c}^{\mathrm{H}}\bar{\boldsymbol{h}}_{k,c}, \quad \tilde{\boldsymbol{D}}_c[d] = \sum_{k=1}^{K} w_k(r_{k,c}^{-1}[d] - \tilde{r}_{k,c}^{-1}[d])\boldsymbol{\Xi}_k,$$
$$\bar{\boldsymbol{A}}_{k,c}[d] = r_{k,c}^{-1}[d]\bar{\boldsymbol{h}}_{k,c}^{\mathrm{H}}\bar{\boldsymbol{h}}_{k,c}, \quad \tilde{\boldsymbol{A}}_{k,c}[d] = r_{k,c}^{-1}[d]\boldsymbol{\Xi}_k.$$

Let

$$\boldsymbol{\Lambda}[d] = \bar{\boldsymbol{\Lambda}}[d] + \tilde{\boldsymbol{\Lambda}}[d], \tag{41}$$
$$\boldsymbol{\Gamma}_k[d] = \bar{\boldsymbol{\Gamma}}_k[d] + \tilde{\boldsymbol{\Gamma}}_k[d]. \tag{42}$$

Based on the above decompositions, the calculations of $\hat{\boldsymbol{D}}[d]$ and $\hat{\boldsymbol{A}}_k[d]$ in the solution (30) are divided into two parts as shown below.

Theorem 2. The solution of optimization problem (18) can be further expressed as

$$\boldsymbol{q}_k[d+1] = w_k(\hat{\boldsymbol{D}}^{\mathrm{det}}[d] + \hat{\boldsymbol{D}}^{\mathrm{sto}}[d] + \mu^*\boldsymbol{I}_{NM})^{-1}(\hat{\boldsymbol{A}}_k^{\mathrm{det}}[d] + \hat{\boldsymbol{A}}_k^{\mathrm{sto}}[d])\boldsymbol{q}_k[d], \tag{43}$$

where

$$\hat{\boldsymbol{D}}^{\mathrm{det}}[d] = (\boldsymbol{F}^{\mathrm{H}} \otimes \boldsymbol{I}_M)\bar{\boldsymbol{\Lambda}}[d](\boldsymbol{F} \otimes \boldsymbol{I}_M), \tag{44}$$
$$\hat{\boldsymbol{D}}^{\mathrm{sto}}[d] = \sum_{k'=1}^{K}(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Phi}_{k'}[d]\boldsymbol{F}) \otimes \boldsymbol{\Xi}_{k'}, \tag{45}$$
$$\hat{\boldsymbol{A}}_k^{\mathrm{det}}[d] = (\boldsymbol{F}^{\mathrm{H}} \otimes \boldsymbol{I}_M)\bar{\boldsymbol{\Gamma}}_k[d](\boldsymbol{F} \otimes \boldsymbol{I}_M), \tag{46}$$
$$\hat{\boldsymbol{A}}_k^{\mathrm{sto}}[d] = (\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Theta}_k[d]\boldsymbol{F}) \otimes \boldsymbol{\Xi}_k \tag{47}$$

with two diagonal matrices

$$\boldsymbol{\Phi}_{k'}[d] = \mathrm{diag}\{w_k(r_{k',1}^{-1}[d] - \tilde{r}_{k',1}^{-1}[d]), \ldots, w_k(r_{k',N_v}^{-1}[d] - \tilde{r}_{k',N_v}^{-1}[d])\}, \tag{48}$$
$$\boldsymbol{\Theta}_k[d] = \mathrm{Diag}\{r_{k,1}^{-1}[d], \ldots, r_{k,N_v}^{-1}[d]\}. \tag{49}$$

Proof. See Appendix C.

The calculations in Algorithm 1 with symplectic optimization can be divided into the deterministic and the stochastic channel-related calculations based on Theorem 2. Let $\boldsymbol{x}_k \in \mathbb{C}^{NM\times 1}$ be an arbitrary vector. The deterministic channel related calculation of $\hat{\boldsymbol{D}}\boldsymbol{x}_k$ is given as

$$(\boldsymbol{F}^{\mathrm{H}} \otimes \boldsymbol{I}_M)\bar{\boldsymbol{\Lambda}}(\boldsymbol{F} \otimes \boldsymbol{I}_M)\boldsymbol{x}_k. \tag{50}$$

Let $\boldsymbol{X}_k \in \mathbb{C}^{M\times N}$ be the matrix reconstructed from $\boldsymbol{x}_k$. We introduce a property

$$(\boldsymbol{F}\otimes\boldsymbol{I}_M)\mathrm{vec}(\boldsymbol{X}_k)=\mathrm{vec}(\boldsymbol{X}_k\boldsymbol{F}^{\mathrm{T}})=\mathrm{vec}([\boldsymbol{X}_k,\boldsymbol{0}_{N\times(N_v-N)}]\boldsymbol{E}_{N_v}^{\mathrm{T}}), \tag{51}$$

where $\boldsymbol{E}_{N_v}$ is a DFT matrix and the relationship between $\boldsymbol{E}_{N_v}$ and $\boldsymbol{F}$ is $\boldsymbol{F}=\boldsymbol{E}_{N_v}[\boldsymbol{I}_N,\boldsymbol{0}_{N\times(N_v-N)}]^{\mathrm{T}}$. Using this property, the calculation of $(\boldsymbol{F}\otimes\boldsymbol{I}_M)\boldsymbol{x}_k$ can be simplified by several FFT operations. Its complexity is $O(N_IKN_vM\log N_v)$ where N_I is the number of iterations for symplectic optimization. The complexity of the matrix multiplication of $\bar{\boldsymbol{\Lambda}}(\boldsymbol{F}\otimes\boldsymbol{I}_M)\boldsymbol{x}_k$ is $O(N_IK^2N_vM)$. The complexity of calculating the deterministic channel related part of $\hat{\boldsymbol{D}}\boldsymbol{x}_k$ is $O(N_I(KN_vM\log N_v+K^2N_vM))$.

The stochastic channel related calculation of $\hat{\boldsymbol{D}}\boldsymbol{x}_k$ is given as

$$(\boldsymbol{F}^{\mathrm{H}}\otimes\boldsymbol{I}_M)\tilde{\boldsymbol{\Lambda}}(\boldsymbol{F}\otimes\boldsymbol{I}_M)\boldsymbol{x}_k=\left(\sum_{k'=1}^{K}(\boldsymbol{F}^{\mathrm{H}}\tilde{\boldsymbol{\Lambda}}_{k'}\boldsymbol{F})\otimes\boldsymbol{\Xi}_{k'}\right)\mathrm{vec}(\boldsymbol{X}_k)=\mathrm{vec}\left\{\sum_{k'=1}^{K}\boldsymbol{\Xi}_{k'}\boldsymbol{X}_k(\boldsymbol{F}^{\mathrm{H}}\tilde{\boldsymbol{\Lambda}}_{k'}\boldsymbol{F})^{\mathrm{T}}\right\}, \tag{52}$$

where the matrix

$$\tilde{\boldsymbol{\Lambda}}_{k'}=\mathrm{diag}\{w_k(r_{k',1}^{-1}-\tilde{r}_{k',1}^{-1}),\ldots,w_k(r_{k',N_v}^{-1}-\tilde{r}_{k',N_v}^{-1})\}\in\mathbb{C}^{N_v\times N_v} \tag{53}$$

is diagonal. The complexity of stochastic channel-related calculation for $\hat{\boldsymbol{D}}\boldsymbol{x}_k$ is concentrated on the matrix multiplication which is $O(N_I(KN^2\log N_v+K^2N^2M))$.

The complexity of the calculation in (52) might be unaffordable as the number of TDPVs increases. Hence, we provide another way to calculate the stochastic channel-related calculation of $\hat{\boldsymbol{D}}\boldsymbol{x}_k$. According to (7), the block elements of matrix $\tilde{\boldsymbol{\Lambda}}=\mathrm{Bdiag}\{\tilde{\boldsymbol{D}}_1,\ldots,\tilde{\boldsymbol{D}}_{N_v}\}$ can be written as

$$\tilde{\boldsymbol{D}}_c=\boldsymbol{V}\boldsymbol{\Delta}_c\boldsymbol{V}^{\mathrm{H}}, \tag{54}$$

where $\boldsymbol{\Delta}_c=\sum_{k=1}^{K}w_k(1-\rho_k^2)(r_{k,c}^{-1}-\tilde{r}_{k,c}^{-1})\mathrm{diag}(\boldsymbol{\omega}_k^{\mathrm{T}})$ denotes a diagonal matrix. Let $\mathrm{Bdiag}\{\boldsymbol{V}\boldsymbol{\Delta}_1\boldsymbol{V}^{\mathrm{H}},\ldots,\boldsymbol{V}\boldsymbol{\Delta}_{N_v}\boldsymbol{V}^{\mathrm{H}}\}$ denote the block diagonal matrix $\boldsymbol{\Delta}$. The stochastic channel related calculation of $\hat{\boldsymbol{D}}\boldsymbol{x}_k$ is given as

$$(\boldsymbol{F}^{\mathrm{H}}\otimes\boldsymbol{I}_M)\tilde{\boldsymbol{\Lambda}}(\boldsymbol{F}\otimes\boldsymbol{I}_M)\boldsymbol{x}_k=(\boldsymbol{F}^{\mathrm{H}}\otimes\boldsymbol{I}_M)\boldsymbol{\Delta}(\boldsymbol{F}\otimes\boldsymbol{I}_M)\boldsymbol{x}_k. \tag{55}$$

Note that multiplying $\boldsymbol{V}$ can be implemented using FFT, and the statistical CSI is sparse. Its complexity is $O(N_I(KN_vM\log M+KN_vM\log N_v))$.

Algorithm 2 summarizes the transform-based CSPD. We consider a more general case when the number of TDPVs is not sufficiently small instead of a special case as mentioned in (52). The complexity of CSPD is primarily determined by calculating $(\hat{\boldsymbol{D}}[d]+\mu^*\boldsymbol{I}_{NM})\boldsymbol{u}_k[d]$, which can be given as $O(N_I(KN_vM\log N_v+KN_vM\log M+K^2N_vM)+KNN_vM)$. We choose the precoder design in [25] for comparison. It is a robust precoder design for each subcarrier and the complexity is $O(\frac{N_v}{N_G}(M^3+KM^2+K^2M))$, where N_G is the number of subcarriers in each group.

4 Simulation results

In this section, simulation results are presented to evaluate the performance of the transform-based CSPD.

4.1 Simulation setup

We utilize the QuaDRiGa [37], which operates based on a geometry-based stochastic channel model (GSCM). The channels generated by QuaDriGa are capable of closely matching the statistical parameters of the beam-based statistical channel model. The antenna array configuration is "3gpp-3d" with a UPA setup where $M=128$, $M_x=16$, and $M_z=8$. We set the carrier frequency and the subcarrier spacing as 2.53 GHz and 30 kHz, respectively, and use $N_v=256$ subcarriers for the transmission.

Algorithm 2 Symplectic optimization method with FFT.

1: Set iteration $d=0$; initialize $\boldsymbol{q}_{k,i}[d], 1 \leqslant k \leqslant K, 1 \leqslant i \leqslant N$;
2: Calculate $r_{k,c}[d]$ and $\tilde{r}_{kc}[d]$ as

$$r_{k,c}[d] = \sigma_z^2 + \sum\nolimits_{l \neq k}^{K} \{\bar{\boldsymbol{h}}_{k,c} \boldsymbol{F}_c \boldsymbol{q}_l[d] \boldsymbol{q}_l^{\mathrm{H}}[d] \boldsymbol{F}_c^{\mathrm{H}} \bar{\boldsymbol{h}}_{k,c}^{\mathrm{H}} + \eta_k (\boldsymbol{F}_c \boldsymbol{q}_l[d] \boldsymbol{q}_l^{\mathrm{H}}[d] \boldsymbol{F}_c^{\mathrm{H}})\},$$
$$\tilde{r}_{k,c}[d] = r_{k,c} + \bar{\boldsymbol{h}}_{k,c} \boldsymbol{F}_c \boldsymbol{q}_k[d] \boldsymbol{q}_k^{\mathrm{H}}[d] \boldsymbol{F}_c^{\mathrm{H}} \bar{\boldsymbol{h}}_{k,c}^{\mathrm{H}} + \eta_k (\boldsymbol{F}_c \boldsymbol{q}_k[d] \boldsymbol{q}_k^{\mathrm{H}}[d] \boldsymbol{F}_c^{\mathrm{H}});$$

3: Calculate $\boldsymbol{R}_{k,c}$ as

$$\boldsymbol{R}_{k,c} = \bar{\boldsymbol{h}}_{k,c}^{\mathrm{H}} \bar{\boldsymbol{h}}_{k,c} + \mathbb{E}\{\tilde{\boldsymbol{h}}_{k,c}^{\mathrm{H}} \tilde{\boldsymbol{h}}_{k,c}\} = \bar{\boldsymbol{h}}_{k,c}^{\mathrm{H}} \bar{\boldsymbol{h}}_{k,c} + \boldsymbol{\Xi}_k;$$

4: Calculate $\boldsymbol{A}_{k,c}[d]$ and $\boldsymbol{D}_c[d]$ as

$$\boldsymbol{A}_{k,c}[d] = r_{k,c}^{-1}[d] \boldsymbol{R}_{k,c}, \quad \boldsymbol{C}_{k,c}[d] = (r_{k,c}^{-1}[d] - \tilde{r}_{k,c}^{-1}[d]) \boldsymbol{R}_{k,c}, \quad \boldsymbol{D}_c[d] = \sum\nolimits_{k=1}^{K} w_k \boldsymbol{C}_{k,c}[d];$$

5: Update $\boldsymbol{q}_k[d+1]$ according to Algorithm 1 and update the iteration as $d = d+1$;
6: Repeat Steps 2–5 until the preset goal is achieved or convergence is reached.

Table 1 Complexities for various precoding schemes.

Precoder	Parameter	Complexity
WMMSE	$N_G = 12$	5.03E7
WMMSE	$N_G = 8$	7.54E7
WMMSE	$N_G = 4$	1.51E8
WMMSE	$N_G = 1$	6.03E8
Transform-based CSPD	$N = 24$	7.08E7

4.2 WSR performance

The WSR performance of the proposed precoder design and the WMMSE precoder design in [15] is compared in Figure 2(a). The channels are quasi-static. The WMMSE precoder design is combined with the subcarrier grouping method. Let N_G denote the number of subcarriers in each group. The per-subcarrier precoder design is a special case with $N_G = 1$. The complexities of different precoding schemes are presented in Table 1. From Figure 2(a), the transform-based CSPD outperforms WMMSE design at a comparable complexity level. Specifically, with $N = 24$, the transform-based CSPD achieves approximately 22% higher performance than the WMMSE precoder design with $N_G = 8$ at SNR = 20 dB. Further, compared with the per-subcarrier precoder design, the proposed precoder design achieves acceptable WSR performance while reducing complexity.

In Figure 2(b), the WSR performance of the transform-based CSPD, the SLNR precoder, and the RZF precoder [13, 14] is presented. The RZF and SLNR precoders are per-subcarrier precoder designs based on instantaneous CSI. The effect of imperfect CSI is influenced by the speed. The moving speeds are chosen as 30, 120, and 250 km/h in the simulation. From Figure 2(b), we observe that the proposed precoder design outperforms the precoder designs relying on instantaneous CSI significantly. Specifically, in the 30 kmph scenario, the WSR of transform-based CSPD is approximately 1.28 times of that of the SLNR precoder at SNR = 20 dB. This superiority further rises to 1.63 times when the speed increases to 250 kmph. We find that the proposed precoder design outperforms precoder designs based on outdated instantaneous CSI, even with fewer design parameters. This performance gain is particularly evident in high mobility scenarios.

Figure 2(c) illustrates the WSR performance of transform-based CSPD with varying numbers of TD-PVs. It shows that the performance gap between $N = 24$ and $N = 48$ is relatively small. The WSR performance suffers a severe loss when N becomes less than 24.

In Figure 2(d), the WSR performance of the transform-based CSPD is depicted for different SNRs against the iteration count. We conclude from Figure 2(d) that the objective function values of the proposed algorithm quickly converge at low SNRs. More iterations are required for convergence as SNR increases. Specifically, 10 iterations are sufficient for the convergence of the proposed transform-based CSPD at SNR = 0 dB, whereas 30 iterations are needed as the SNR increases to 20 dB.

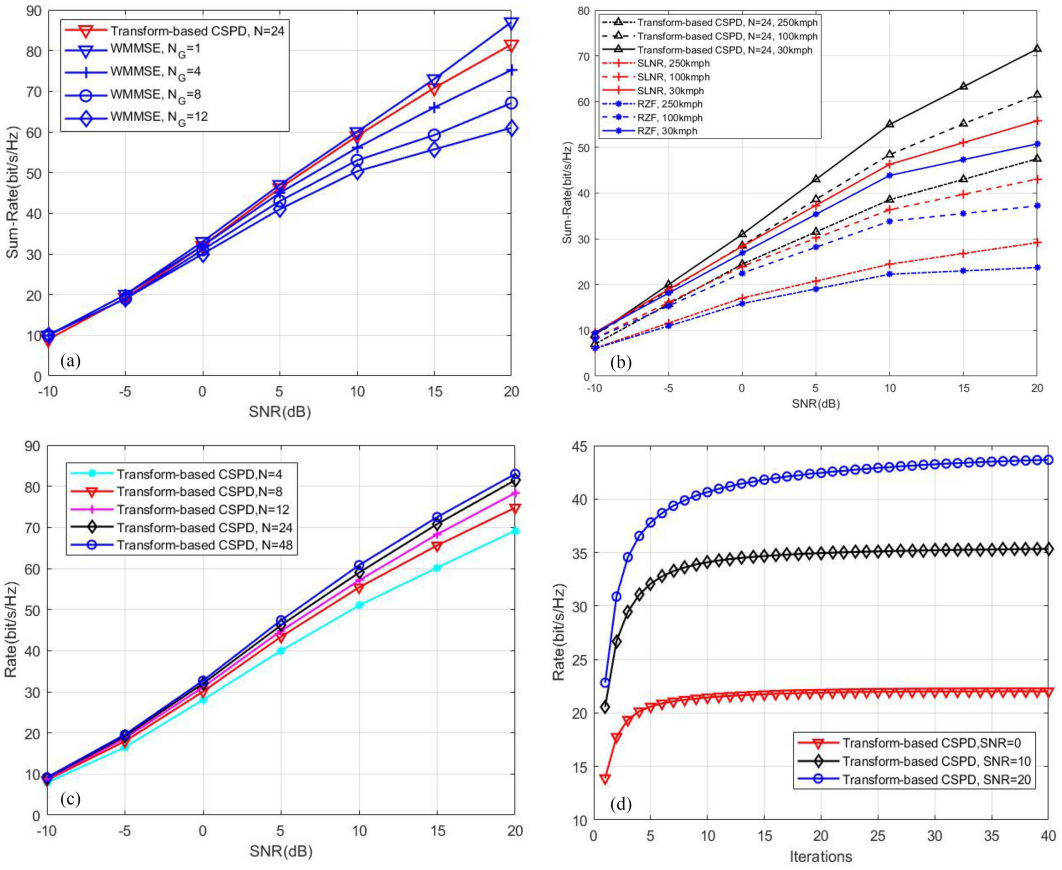

Figure 2 (Color online) (a) WSR performance of the transform-based CSPD compared with WMMSE, where $M = 128$ and $K = 10$; (b) WSR performance under different moving speeds, where $M = 128$ and $K = 10$; (c) WSR performance of the transform-based CSPD with different numbers of TDPVs; (d) convergency of the transform-base CSPD with $N = 24$ and $K = 10$.

4.3 Effective channel estimation performance

We discuss the effective channel estimation and signal detection when the proposed transform-based CSPD is used. Consider the MMSE channel estimation at the user ends. For simplicity, the subscript k dropped in the following derivation. Let y_c and $\bar{h}_c = \boldsymbol{h}_c \boldsymbol{p}_c$ denote the received signal and downlink effective channel at the user side for c-th subcarrier. Let $\boldsymbol{y} = [y_1, y_2, \ldots, y_{N_v}]^{\mathrm{T}}$, $\boldsymbol{h}_f = [\bar{h}_1, \bar{h}_2, \ldots, \bar{h}_{N_v}]^{\mathrm{T}}$, and $\boldsymbol{z} = [z_1, z_2, \ldots, z_{N_v}]^{\mathrm{T}}$. The transmission model for downlink effective channel estimation is given as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h}_f + \boldsymbol{z}, \tag{56}$$

where $\boldsymbol{X} = \mathrm{Diag}\{x_1, x_2, \ldots, x_{N_v}\}$ is the pilot signal whose elements can be provided based on the Zadoff-Chu (ZC) sequence. Let N_d denote the delay sampling number. The delay domain channel can be transformed into the frequency domain channel by

$$\boldsymbol{h}_f = \boldsymbol{U}_f \boldsymbol{h}_d, \tag{57}$$

where $\boldsymbol{U}_f \in \mathbb{C}^{N_v \times N_d}$ is a partial DFT and $\boldsymbol{h}_d \in \mathbb{C}^{N_d \times 1}$ is the delay domain channel. Further, the delay domain channel can be decomposed as

$$\boldsymbol{h}_d = \tilde{\boldsymbol{m}} \odot \tilde{\boldsymbol{w}}, \tag{58}$$

where $\tilde{\boldsymbol{m}}$ is a deterministic vector and $\tilde{\boldsymbol{w}}$ consists of i.i.d. entries following $\mathcal{CN}(0, 1)$ distribution. Let $\tilde{\boldsymbol{\omega}} = \tilde{\boldsymbol{m}} \odot \tilde{\boldsymbol{m}}$. We have $\mathbb{E}\{\boldsymbol{h}_d \boldsymbol{h}_d^{\mathrm{H}}\} = \boldsymbol{\Lambda}$ where the diagonal matrix $\boldsymbol{\Lambda}_{ii} = \tilde{\boldsymbol{\omega}}_i$.

Let $\boldsymbol{R}_{\boldsymbol{h}_f} = \mathbb{E}\{\boldsymbol{h}_f \boldsymbol{h}_f^{\mathrm{H}}\}$. We have $\boldsymbol{R}_{\boldsymbol{h}_f} = \boldsymbol{U}_f \boldsymbol{\Lambda} \boldsymbol{U}_f^{\mathrm{H}}$. Using the MMSE channel estimation, the estimated frequency domain channel is given as

$$\hat{\boldsymbol{h}}_{f,\mathrm{mmse}} = \boldsymbol{R}_{\boldsymbol{h}_f}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{R}_{\boldsymbol{h}_f} + \delta_z^2 \boldsymbol{I})^{-1}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{U}_f \boldsymbol{\Lambda} \boldsymbol{U}_f^{\mathrm{H}}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{U}_f \boldsymbol{\Lambda} \boldsymbol{U}_f^{\mathrm{H}} + \delta_z^2 \boldsymbol{I})^{-1}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{y}. \tag{59}$$

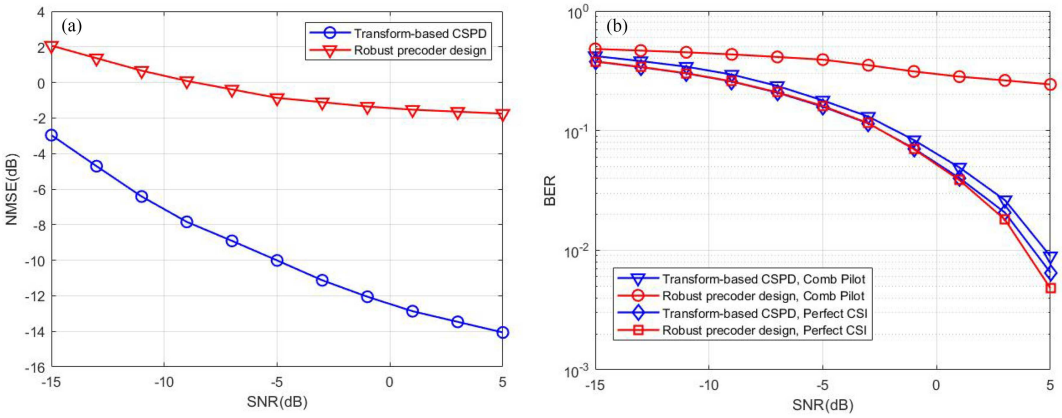

Figure 3 (Color online) (a) NMSE and (b) BER performance of transform-based CSPD compared with the robust precoder design in [25].

The normalized mean squared error (NMSE) performance of the channel estimation in the transform-based CSPD case and the robust per-subcarrier precoder design case is given in Figure 3(a). When the perfect CSI is known at the BS, the robust precoder design reduces to the WMMSE precoder. We utilize comb-type pilots for channel estimation with a pilot interval set to 2. The result shows that the effective channels can be estimated more precisely when the transform-based CSPD is used. The significant channel estimation performance improvement comes from the enhanced smoothness of the frequency domain effective channel using the proposed precoders.

Then, we perform the signal detection. With the definition of $\tilde{z}_{k,c}$ and the covariance $r_{k,c}$ provided in Section 2, the received signal is rewritten as

$$y_{k,c} = \hat{h}_{k,c} x_{k,c} + \tilde{z}_{k,c}, \tag{60}$$

where $\hat{h}_{k,c}$ denotes the estimated effective channel.

Considering the MMSE detection, the estimated signal is given as

$$\hat{x}_{k,c} = (\hat{h}^*_{k,c} \hat{h}_{k,c} + r_{k,c})^{-1} \hat{h}^*_{k,c} y_{k,c}. \tag{61}$$

The modulation scheme is the quadrature phase shift keying. The bit error rate (BER) performance of transform-based CSPD and the robust per-subcarrier precoder design is given in Figure 3(b). The BER performance of transform-based CSPD outperforms that of the robust per-subcarrier precoder design, primarily due to the improved effective channel estimation performance.

5 Conclusion

In this paper, we propose a transform-based CSPD for massive MIMO-OFDM downlink. The precoding vectors for a set of subcarriers are derived using a reduced number of the TDPVs, enhancing the frequency smoothness of the effective channel. Furthermore, the proposed precoders facilitate the channel estimation and the signal detection. The precoder structure is derived based on the MM methodology. To further avoid the large-dimensional matrix inversion, an algorithm by utilizing symplectic optimization is proposed.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62394294, 62394290, 62371125), Jiangsu Province Basic Research Project (Grant No. BK20192002), Jiangsu Province Major Science and Technology Project (Grant No. SBG2024000080), Fundamental Research Funds for the Central Universities (Grant No. 2242022k60007), Key R&D Plan of Jiangsu Province (Grant No. BE2022067), and Huawei Cooperation Project.

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Albreem M A, Juntti M, Shahabuddin S. Massive MIMO detection techniques: a survey. IEEE Commun Surv Tut, 2019, 21: 3109–3132

2 Liu X, Wang W, Gong X, et al. Structured hybrid message passing based channel estimation for massive MIMO-OFDM systems. IEEE Trans Veh Technol, 2023, 72: 7491–7507
3 Wu C, Yi X, Wang W, et al. Learning to localize: a 3D CNN approach to user positioning in massive MIMO-OFDM systems. IEEE Trans Wireless Commun, 2021, 20: 4556–4570
4 Ge Y, Zhang W, Gao F, et al. Beamforming network optimization for reducing channel time variation in high-mobility massive MIMO. IEEE Trans Commun, 2019, 67: 6781–6795
5 Xiao X, You L, Wang K Z, et al. Distortion-aware beamforming design for multi-beam satellite communications with nonlinear power amplifiers. Sci China Inf Sci, 2024, 67: 162302
6 Zhao X, Lu S, Shi Q, et al. Rethinking WMMSE: can its complexity scale linearly with the number of BS antennas? IEEE Trans Signal Process, 2023, 71: 433–446
7 Chen K, Yang J, Li Q, et al. Sub-array hybrid precoding for massive MIMO systems: a CNN-based approach. IEEE Commun Lett, 2021, 25: 191–195
8 Zhang Y, Lu A A, Liu B, et al. Cross-subcarrier precoder design for massive MIMO-OFDM downlink. In: Proceedings of IEEE 98th Vehicular Technology Conference, 2023. 1–6
9 Zhang Y X, Lu A A, Gao X. Sum-rate-optimal statistical precoding for FDD massive MIMO downlink with deterministic equivalents. IEEE Trans Veh Technol, 2022, 71: 7359–7370
10 Shi Q, Razaviyayn M, Luo Z Q, et al. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. IEEE Trans Signal Process, 2011, 59: 4331–4340
11 Jiang H, Dai L L. Transformer-based downlink precoding design in massive MIMO systems for 5G-advanced and 6G. Sci China Inf Sci, 2023, 66: 199301
12 Luo J, Fan J C, Wang Y W, et al. Manifold optimization assisted centralized hybrid precoding for cell-free massive MIMO systems. Sci China Inf Sci, 2023, 66: 219301
13 Nguyen L D, Tuan H D, Duong T Q, et al. Multi-user regularized zero-forcing beamforming. IEEE Trans Signal Process, 2019, 67: 2839–2853
14 Zhang C, Huang Y, Jing Y, et al. Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling. IEEE Trans Wireless Commun, 2017, 16: 2181–2194
15 Christensen S S, Agarwal R, de Carvalho E, et al. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design. IEEE Trans Wireless Commun, 2008, 7: 4792–4799
16 Li Y. Simplified channel estimation for OFDM systems with multiple transmit antennas. IEEE Trans Wireless Commun, 2002, 1: 67–75
17 Hu X, Dai X. Wideband precoding for MIMO-OFDM systems with per-antenna power constraints. IEEE Commun Lett, 2021, 25: 3423–3426
18 Shen C, Fitz M. MIMO-OFDM beamforming for improved channel estimation. IEEE J Sel Areas Commun, 2008, 26: 948–959
19 Venugopal K, Gonzalez-Prelcic N, Heath R W. Optimal frequency-flat precoding for frequency-selective millimeter wave channels. IEEE Trans Wireless Commun, 2019, 18: 5098–5112
20 Hu W, Li F, Jiang Y. Phase rotations of SVD-based precoders in MIMO-OFDM for improved channel estimation. IEEE Wireless Commun Lett, 2021, 10: 1805–1809
21 Sandell M, Ponnampalam V. Smooth beamforming for OFDM. IEEE Trans Wireless Commun, 2009, 8: 1133–1138
22 Wang Z, Chen W. Regularized zero-forcing for multiantenna broadcast channels with user selection. IEEE Wireless Commun Lett, 2012, 1: 129–132
23 Shu F, Tong J J, You X H, et al. Adaptive robust Max-SLNR precoder for MU-MIMO-OFDM systems with imperfect CSI. Sci China Inf Sci, 2016, 59: 062302
24 Razaviyayn M, Boroujeni M S, Luo Z Q. A stochastic weighted MMSE approach to sum rate maximization for a MIMO interference channel. In: Proceedings of IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2013. 325–329
25 Lu A A, Gao X, Xiao C. Robust linear precoder design for 3D massive MIMO downlink with a posteriori channel model. IEEE Trans Veh Technol, 2022, 71: 7274–7286
26 Jeon C, Li Z, Studer C. Approximate gram-matrix interpolation for wideband massive MU-MIMO systems. IEEE Trans Veh Technol, 2020, 69: 4677–4688
27 Lu A A, Gao X, Zhong W, et al. Robust transmission for massive MIMO downlink with imperfect CSI. IEEE Trans Commun, 2019, 67: 5362–5376
28 Duruisseaux V, Leok M. Practical perspectives on symplectic accelerated optimization. Optim Methods Software, 2023, 38: 1230–1268
29 Kobayashi M, Caire G. Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge. IEEE J Sel Areas Commun, 2007, 25: 1468–1477
30 Luo Z Q, Pang J S, Ralph D, et al. Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints. Math Programming, 1996, 75: 19–76
31 Jordan M I. Dynamical, symplectic and stochastic perspectives on gradient-based optimization. In: Proceedings of the International Congress of Mathematicians, 2018. 523–549
32 Guilherme F, Alessandro B, Mark G, et al. Optimization on manifolds: a symplectic approach. 2023. ArXiv:2107.11231v2
33 Wibisono A, Wilson A C, Jordan M I. A variational perspective on accelerated methods in optimization. Proc Natl Acad Sci USA, 2016, 113: E7351–E7358
34 Leon M D, Rodrigues P R. Methods of Differential Geometry in Analytical Mechanics. Amsterdam: Elsevier, 1989. 158
35 Betancourt M, Jordan M I, Wilson A C. On symplectic optimization. 2018. ArXiv:1802.03653v2
36 Hairer E, Wanner G, Lubich C. Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations. New York: Springer, 2006
37 Jaeckel S, Raschkowski L, Borner K, et al. QuaDRiGa: a 3-D multi-cell channel model with time evolution for enabling virtual field trials. IEEE Trans Antennas Propagat, 2014, 62: 3242–3256