

Near-Optimal Low-Complexity Hybrid Precoding for THz Massive MIMO Systems

Yuke Sun, Aihua Zhang*, Hao Yang, Di Tian, and Haowen Xia

School of Electronic and Information Engineering, Zhongyuan University of Technology
Zhengzhou, Henan 450007

[e-mail: zhah@zut.edu.cn, syk@zut.edu.cn]

*Corresponding author: Aihua Zhang

*Received August 28, 2023; revised November 28, 2023; accepted April 1, 2024;
published April 30, 2024*

Abstract

Terahertz (THz) communication is becoming a key technology for future 6G wireless networks because of its ultra-wide band. However, the implementation of THz communication systems confronts formidable challenges, notably beam splitting effects and high computational complexity associated with them. Our primary objective is to design a hybrid precoder that minimizes the Euclidean distance from the fully digital precoder. The analog precoding part adopts the delay-phase alternating minimization (DP-AltMin) algorithm, which divides the analog precoder into phase shifters and time delayers. This effectively addresses the beam splitting effects within THz communication by incorporating time delays. The traditional digital precoding solution, however, needs matrix inversion in THz massive multiple-input multiple-output (MIMO) communication systems, resulting in significant computational complexity and complicating the design of the analog precoder. To address this issue, we exploit the characteristics of THz massive MIMO communication systems and construct the digital precoder as a product of scale factors and semi-unitary matrices. We utilize Schatten norm and Hölder's inequality to create semi-unitary matrices after initializing the scale factors depending on the power allocation. Finally, the analog precoder and digital precoder are alternately optimized to obtain the ultimate hybrid precoding scheme. Extensive numerical simulations have demonstrated that our proposed algorithm outperforms existing methods in mitigating the beam splitting issue, improving system performance, and exhibiting lower complexity. Furthermore, our approach exhibits a more favorable alignment with practical application requirements, underlying its practicality and efficiency.

Keywords: Massive MIMO, terahertz, hybrid precoding, low complexity

1. Introduction

Terahertz (THz) technology has emerged as a potential technique for next-generation wireless communication systems, with the goal of supporting Tbps data rates and catering to emerging ultra-high-speed applications. THz technology is well-suited for achieving high data rates due to its utilization of abundant spectrum resources [1]. THz communication, on the other hand, has obstacles such as path loss and beam splitting, making it difficult to propagate high gain THz signals over long periods of time. To address the path loss issue, researchers have leveraged massive multiple-input multiple-output (MIMO) technology to extend propagation distance in THz communication and efficiently minimize the impact of severe signal attenuation [2]. Subsequently, directional beams with high array gain are generated by precoding technology to achieve full array gain.

Hybrid precoding technology is acknowledged as a pivotal component for THz wideband mobile communication [3]. Hybrid precoding technology consists of analog RF precoding and digital baseband precoding. Traditional analog precoding, implemented with a phase shifter array, is only suitable for narrowband systems with small differences in subcarrier frequencies, resulting in beam splitting and affecting array gain in THz communication systems with significant differences in subcarrier frequencies from the center carrier frequency [4]. Furthermore, using traditional digital precoding schemes in wider antenna array systems also results in extremely high system computational complexity. Therefore, it is critical to investigate effective hybrid precoding schemes that can mitigate beam splitting while also reducing complexity [5].

In previous studies, various solutions were proposed to tackle the beam splitting problem. An antenna array structure operating in an extremely wide instantaneous bandwidth was proposed in [6], the study introduced the principles of a space-time array processor based on true-time delay (TTD) and an ultra-wideband beamformer. This structure facilitates independent control of the phase and amplitude of signals on each path. In [7], a fully connected and subarray structure was presented to enable simultaneous adjustment of amplification and phase in analog precoding. An algorithm was devised to jointly optimize phase shifters (PS) and TTD values under the constraints of each TTD device, effectively handling practical delay constraints [8]. A Delay-Phase Alternating Minimization (DP-AltMin) algorithm was proposed to address the beam splitting issue caused by the angle expansion of multipath under a channel cluster model [9]. The authors in [10] suggested a technique in which all subcarriers were projected onto a central subcarrier, aiming to simplify hardware requirements and minimize power consumption. They also constructed a generalized analog precoder based on the channel covariance matrix. However, compared to beamforming techniques based on instantaneous channel state information (CSI), this approach may result in performance degradation. A sparse RF chain antenna structure based on serial equidistant delay elements was introduced to reduce the number of RF chains and delay elements, consequently lowering hardware costs [11]. In [12], a hardware structure called true-time-delayers based delay-phase precoding (TTD-DPP) was introduced, which effectively reduces hardware costs by using some delay elements for frequency-dependent phase shifting. A joint hybrid precoding scheme based on an equivalent channel was presented in [13], which improved spectral efficiency and reduced complexity in scenarios with fewer RF chains by optimizing the hybrid precoding and composite tasks, however, the influence of beam splitting was not considered. Therefore, existing solutions either have limited performance or excessive complexity.

To address the aforementioned issues, we propose a near-optimal hybrid precoding structure with low complexity. It first performs singular value decomposition on the CSI of all subcarriers to obtain the fully digital precoder. The analog precoder is then determined using the DP-AltMin algorithm to address the impact of beam splitting. Finally, the digital precoder is designed as the product of a scaling factor and a semi-unitary matrix with a semi-orthogonal structure. By alternately optimizing the scaling factor and the semi-unitary matrix, a near-optimal digital precoding matrix is obtained. Simulation results indicate that the proposed approach offers notable benefits, including reduced computational complexity and processing time, as well as improved system performance.

This paper is structured as follows. Section II introduces the THz communication system model and channel model. In Section III, we introduce the near-optimal low-complexity hybrid precoding scheme employed in this paper. Section IV provides a simulation analysis that evaluates the system's performance and computational complexity. Finally, Section V summarizes the paper.

Notation: $|\cdot|$ denotes the magnitude of a vector, $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $\|\cdot\|_F$ denotes the conjugate, transpose, conjugate transpose and Frobenius norm of a matrix, respectively. $\mathbf{a} \otimes \mathbf{b}$ and $\mathbf{a} \odot \mathbf{b}$ denote the Kronecker product and the Hadamard product of \mathbf{a} and \mathbf{b} . $\angle \mathbf{a}$ denotes the phase of \mathbf{a} .

Table 1. Definition of parameters in the channel

Parameters	Definition	Parameters	Definition
N_t	Number of transmit antennas	\mathbf{F}_{opt}	Fully digital precoder
N_r	Number of receive antennas	\mathbf{V}_{RF}	Analog precoder
N_{RF}	Number of RF chains	\mathbf{V}_A	Phase shifting matrix
N_s	Number of data streams	\mathbf{T}_k	Delay matrix
\mathbf{y}_k	Received data at user k	\mathbf{V}_{BB}	Digital precoder
\mathbf{H}_k	Frequency-domain channel for user k	γ	Scaling factor
$\mathbf{a}(N, \theta)$	Antenna array response	\mathbf{V}_{DD}	Semi-unitary matrix

2. System Model and Channel Model

In this section, we describe the system model and channel model for the THz communication system. **Table 1** presents the definitions of the channel parameters.

2.1 System Model

We consider the delay-phase precoding (DPP) architecture [12], as shown in **Fig. 1**. The base station (BS) is equipped with N_t transmitting antennas and N_{RF} RF chains to serve users, while each user is equipped with N_r receiving antennas. The system adopts MIMO-assisted orthogonal frequency-division multiplexing (OFDM) modulation with K subcarriers [14]. For each subcarrier, the BS transmits N_s data streams to the receiver, typically with

$N_s \leq N_{RF} \ll N_t$. We assume that the set of transmitted data is \mathbf{s}_k for the k -th subcarrier. The received data vector \mathbf{y}_k can be represented as follows

$$\mathbf{y}_k = \sqrt{P_t} \mathbf{H}_k \mathbf{V}_{RF} \mathbf{V}_{BB} \mathbf{s}_k + \mathbf{n}_k, \quad (1)$$

where $\sqrt{P_t}$ denotes the transmit power, $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ represents the frequency-domain channel. $\mathbf{V}_{BB} \in \mathbb{C}^{N_{RF} \times N_s}$ is the digital precoder, while $\|\mathbf{V}_{RF} \mathbf{V}_{BB}\|_F^2 = N_s$ represents the power constraint relationship for the digital precoder. $\mathbf{s}_k \in \mathbb{C}^{N_s \times 1}$, is the transmitted data, and $\mathbb{E}(\mathbf{s}\mathbf{s}^H) = \frac{1}{N_s} \mathbf{I}_{N_s}$. $\mathbf{n}_k \in \mathbb{C}^{N_s \times 1}$ represents the additive white Gaussian noise.

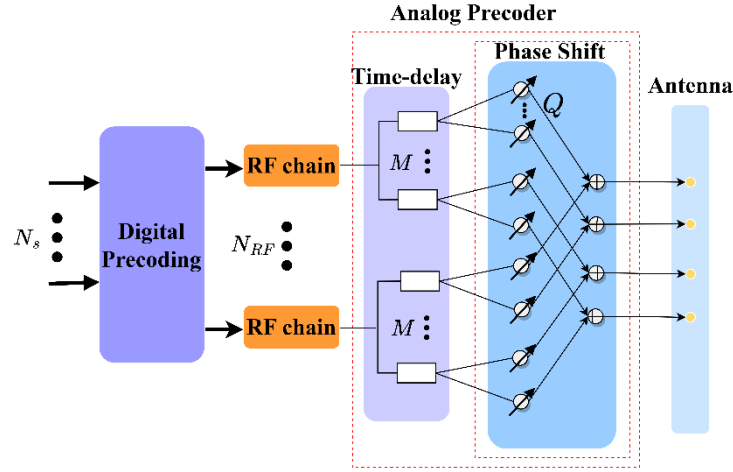


Fig. 1. Delay-phase precoding architecture for THz massive MIMO.

The analog precoding matrix \mathbf{V}_{RF} is composed of two parts: a phase-shifting matrix that is independent of frequency and a delay matrix that is dependent on the carrier frequency. \mathbf{V}_{RF} can be represented as

$$\mathbf{V}_{RF} = \mathbf{V}_A \odot (\mathbf{T}_k \otimes \mathbf{e}_p), \quad (2)$$

where $\mathbf{V}_A \in \mathbb{C}^{N_t \times N_{RF}}$ is a phase-shifting matrix with a constraint $|\mathbf{V}_A(i, j)| = 1$. The phase-shifting matrix only adjusts the phase related to the center carrier frequency f_c . $\mathbf{e}_p \in \mathbb{C}^{P \times 1}$ is a vector with all elements equal to 1. $\mathbf{T}_k \in \mathbb{C}^{M \times N_{RF}}$ is a frequency-dependent phase-shifting matrix generated by the delay array at the frequency point f_k . To save hardware costs, a partial connectivity approach is employed; each RF chain is connected to M delay elements. The phase shifters are divided into M groups, and each delay element is connected to $Q = N_t/M$ phase shifters. The delay array satisfies the delay constraint $\mathbf{T}_k(m, n) = e^{-j2\pi f_k T(m, n)}$, where

$T(m, n)$ represents the actual delay of the m -th delay element on the n -th RF chain. Each delay element $T(m, n)$ is directly connected to Q phase shifters through the Kronecker product of \mathbf{T}_k and \mathbf{e}_p . The phase-shifting matrix and delay matrix work together to achieve frequency-dependent phase shift, compensating for the impact of the beam splitting effect in the wideband.

2.2 Channel Model

The design of hybrid precoders requires accurate channel information, particularly in THz communication where the channel exhibits unique characteristics, including significant path loss, severe atmospheric absorption, and limited diffraction capabilities. To accurately simulate the channel properties of THz communication systems, we employ the classical ray-tracing model. This model, which is based on geometric optics theory, is a commonly used wireless signal propagation model [15]. It simulates signal propagation in the environment by considering the paths of rays and phenomena such as reflection and refraction. In the THz frequency range, this model enables accurate simulation of signal propagation paths, attenuation, and interference. By considering crucial factors such as signal path gain, arrival angle, and delay, we can construct an accurate representation of the THz communication system channel. Building upon the ray tracing model, on the k -th subcarrier, the channel matrix $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ can be represented as

$$\mathbf{H}_k = \mathbf{H}_k^L + \mathbf{H}_k^N, \quad (3)$$

where the channel consists of both line-of-sight (LoS) and non-line-of-sight (NLoS) components. \mathbf{H}_k^L represents the LoS component, corresponding to single-path transmission. On the other hand, \mathbf{H}_k^N represents the NLoS components, modeled as a cluster and involving multipath transmission [16]. Their representations are

$$\mathbf{H}_k^L = \alpha_0 e^{-j2\pi f_k \tau_0} \mathbf{a}_r(N_r, \theta_0^r) \mathbf{a}_t^H(N_t, \theta_0^t) \quad (4)$$

$$\mathbf{H}_k^N = \sum_{l=1}^{L_c} \sum_{q=1}^{L_q} \alpha_{l,q} e^{-j2\pi f_k \tau_{l,q}} \mathbf{a}_r(N_r, \theta_{l,q}^r) \mathbf{a}_t^H(N_t, \theta_{l,q}^t). \quad (5)$$

We adopted the cluster channel model in this work, primarily focusing on the analysis of (5), where α_0 and $\alpha_{l,q}$ represent the complex gains for the path model and the cluster model, respectively. f_c is the center carrier frequency, and the frequency of the k -th subcarrier is denoted by $f_k = f_c - \frac{B}{2} + \frac{k}{K}$ ($k = 0, 1, \dots, K-1$). τ_0 and $\tau_{l,q}$ represents the path delays for the single-path model and the delay for the cluster model, respectively. θ_0^r, θ_0^t refer to the receiving and transmitting angles for the single-path model, while $\theta_{l,q}^r, \theta_{l,q}^t$ representing the receiving and transmitting angles for the cluster model. In practical channel cluster models, a clustered distribution of scattering objects causes the delays and path angles of the actual channel to have a clustered distribution; i.e., angle spread occurs [17]. An angle within the l -th cluster is modeled as $\theta_{l,q}^t = \theta_l^t + \Delta\theta_q^t$, where θ_l^t is the emission angle of the center cluster

and $\Delta\theta'_q$ follows a Laplace distribution [18]. The vectors $\mathbf{a}_r(N_r, \theta^r)$ and $\mathbf{a}_t(N_t, \theta^t)$ represent the antenna array response. Using uniform linear arrays (ULAs), they can be represented as

$$\mathbf{a}(N, \theta) = \frac{1}{\sqrt{N}} \begin{bmatrix} 1, e^{j\pi \frac{f_k}{f_c} \sin \theta}, e^{j2\pi \frac{f_k}{f_c} \sin \theta}, \dots, e^{j(N-1)\pi \frac{f_k}{f_c} \sin \theta} \end{bmatrix}^T. \quad (6)$$

It can be observed that $\mathbf{a}(N, \theta)$ is dependent on the subcarrier frequency f_k . Let us define the equivalent angle on the subcarrier as

$$\sin \theta_k = \frac{f_k}{f_c} \sin \theta, \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. \quad (7)$$

2.3 Problem Modeling

Assuming a time-division duplexing (TDD) mode, the BS estimates the downlink channel through the uplink channel [19]. For a THz massive MIMO system with K subcarriers, the sum-rate R is expressed as

$$R(\mathbf{V}_{RF}, \mathbf{V}_{BB}) = \sum_{k=0}^{K-1} \log_2 \left| I_{N_r} + \frac{P_t}{N_s \sigma^2} \mathbf{H}_k \mathbf{V}_{RF} \mathbf{V}_{BB} \mathbf{V}_{BB}^H \mathbf{V}_{RF}^H \mathbf{H}_k^H \right|. \quad (8)$$

The goal is to optimize the system's sum-rate. To approximate the optimal fully digital precoder, this study discusses the search for a hybrid precoder. This involves the joint optimization of matrix variables \mathbf{V}_{RF} and \mathbf{V}_{BB} to address the challenge of maximizing the sum-rate. The optimization problem of precoding can be represented by

$$\begin{aligned} (\mathbf{V}_{RF}^{opt}, \mathbf{V}_{BB}^{opt}) &= \arg \max_{\mathbf{V}_{RF}, \mathbf{V}_{BB}} R(\mathbf{V}_{RF}, \mathbf{V}_{BB}) \\ s.t. \quad &\|\mathbf{V}_{RF} \mathbf{V}_{BB}\|_F^2 = N_s. \end{aligned} \quad (9)$$

In general, solving optimization problems with such constraints directly can be challenging. To solve the optimization problem, we can transform it into a minimum Euclidean distance problem for designing a hybrid precoder and a fully digital precoder [20]. This approximate transformation converts the original nonconvex optimization problem into a matrix factorization problem. This approach can effectively reduce the computational complexity of precoding algorithms while achieving high-gain THz wideband precoding. We convert (9) to

$$\min_{\mathbf{V}_{RF}, \mathbf{V}_{BB}} \sum_{k=0}^{K-1} \|\mathbf{F}_{opt} - \mathbf{V}_{RF} \mathbf{V}_{BB}\|_F^2, \quad (10)$$

where \mathbf{F}_{opt} is the optimal fully digital precoder for each subcarrier, and it can be obtained through the singular value decomposition (SVD) of the channel matrix. The singular value decomposition of \mathbf{H}_k is given by

$$\mathbf{H}_k = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H, \quad (11)$$

where \mathbf{V} and \mathbf{U} are unitary matrices with dimensions N_t and N_r , respectively. According to the properties of SVD, the optimal fully digital precoder is formed by selecting the first N_s columns of the right unitary matrix of the channel matrix \mathbf{H}_k , i.e.,

$$\mathbf{F}_{opt} = [\mathbf{V}]_{:, 1:N_s}. \quad (12)$$

To address the beam splitting effect, the analog precoding matrix adopts DP-AltMin [9]. The digital precoding matrix using traditional least square (LS) methods is extremely computationally intensive. In this paper, a semi-orthogonal structure is applied to the digital precoding matrix, representing it as a product of a scaling factor γ and a semi-unitary matrix \mathbf{V}_{DD} , which are then optimized alternately. The digital precoding matrix is defined as

$$\mathbf{V}_{BB} \triangleq \gamma \mathbf{V}_{DD}, \quad (13)$$

where γ is a non-zero scaling factor, and \mathbf{V}_{DD} is a semi-unitary matrix that satisfies the constraint $\mathbf{V}_{DD}^H \mathbf{V}_{DD} = \mathbf{I}_{N_s}$. Thus, (13) can be represented as

$$\mathbf{V}_{BB}^H \mathbf{V}_{BB} = \gamma^2 \mathbf{I}_{N_s}, \quad \gamma \neq 0, \quad (14)$$

the optimization problem in (10) can ultimately be formulated as the following constrained optimization problem

$$\begin{aligned} \min_{\gamma, \mathbf{V}_{RF}, \mathbf{V}_{DD}} \sum_{k=0}^{K-1} \left\| \mathbf{F}_{opt} - \mathbf{V}_{RF} \gamma \mathbf{V}_{DD} \right\|_F^2 \\ \text{s.t.} \left\| \mathbf{V}_{RF} \mathbf{V}_{BB} \right\|_F^2 = N_s. \end{aligned} \quad (15)$$

3. The Proposed Hybrid Precoding Structure for THz Communication System

3.1 Analog Precoder Design

The analog precoding matrix is composed of a delay matrix and a phase shift matrix, with the following basic principle:

Given \mathbf{V}_{BB} and \mathbf{T}_k , the design of \mathbf{V}_A according to (10) can be equivalently expressed as

$$\min_{\mathbf{V}_A} \sum_{k=0}^{K-1} \left\| \mathbf{F}_{opt} - \left(\mathbf{V}_A \odot \left(\mathbf{T}_k \otimes \mathbf{e}_p \right) \right) \mathbf{V}_{BB} \right\|_F^2. \quad (16)$$

For the objective function of (16) and utilizing the Cauchy-Schwarz inequality, it can be expressed as

$$\sum_{k=0}^{K-1} \left\| \left(\left(\mathbf{F}_{opt} \mathbf{V}_{BB}^\dagger \right) \odot \left(\mathbf{T}_k \otimes \mathbf{e}_p \right)^* - \mathbf{V}_A \right) \mathbf{V}_{BB} \right\|_F^2$$

$$\leq \sum_{k=0}^{K-1} \left\| \left(\mathbf{F}_{opt} \mathbf{V}_{BB}^\dagger \right) \odot \left(\mathbf{T}_k \otimes \mathbf{e}_p \right)^* - \mathbf{V}_A \right\|_F^2 \cdot \left\| \mathbf{V}_{BB} \right\|_F^2. \quad (17)$$

Then, the solution of problem (16) can be determined as

$$\mathbf{V}_A = \exp \left\{ j \angle \left(\sum_{k=0}^{K-1} \left\| \mathbf{V}_{BB} \right\|_F^2 \left(\mathbf{F}_{opt} \mathbf{V}_{BB}^\dagger \right) \odot \left(\mathbf{T}_k \otimes \mathbf{e}_p \right)^* \right) \right\}. \quad (18)$$

Similar to the optimization process of \mathbf{V}_A , neglecting the constant term $\left\| \mathbf{V}_{BB} \right\|_F^2$, we can convert (16) into

$$\min_{\mathbf{T}_k} \sum_{k=0}^{K-1} \left\| \left(\mathbf{F}_{opt} \mathbf{V}_{BB}^\dagger \right) \odot \mathbf{V}_A^* - \left(\mathbf{T}_k \otimes \mathbf{e}_p \right) \right\|_F^2 = \min_{\mathbf{T}_k} \sum_{k=0}^{K-1} \left(\left\| \left(\mathbf{F}_{opt} \mathbf{V}_{BB}^\dagger \right) \odot \mathbf{V}_A^* \right\|_F^2 - 2 \operatorname{Re} \left(\operatorname{tr} \left(\left(\mathbf{F}_{opt} \mathbf{V}_{BB}^\dagger \odot \mathbf{V}_A^* \right)^H \left(\mathbf{T}_k \otimes \mathbf{e}_p \right) \right) \right) + \left\| \mathbf{T}_k \otimes \mathbf{e}_p \right\|_F^2 \right), \quad (19)$$

the design of \mathbf{T}_k according to (19) can be expressed as

$$\max_{\mathbf{T}_k} \sum_{k=0}^{K-1} \operatorname{Re} \left(\operatorname{tr} \left(\left(\mathbf{F}_{opt} \mathbf{V}_{BB}^\dagger \odot \mathbf{V}_A^* \right)^H \left(\mathbf{T}_k \otimes \mathbf{e}_p \right) \right) \right), \quad (20)$$

from (20), we have $\mathbf{D}_k^H \left(\mathbf{T}_k \otimes \mathbf{e}_p \right) = \Theta_k^H \mathbf{T}_k$, where $\mathbf{D}_k = \left(\mathbf{F}_{opt} \mathbf{V}_{BB}^\dagger \right) \odot \mathbf{V}_A^*$,

$\Theta_k(m, n) = \sum_{q=1}^Q D_k((M-1)Q + q, n)$. Substituting $\mathbf{T}_k(m, n) = e^{-j2\pi f_k T(m, n)}$ into (20), the optimization problem in (20) is found to be equivalent to the maximization problem in (21)

$$\max_{\mathbf{T}_k} \sum_{k=0}^{K-1} \operatorname{Re} \left(\operatorname{tr} \left(\Theta_k^H \mathbf{T}_k \right) \right) = \max_{T(m, n)} \sum_{m=1}^M \sum_{n=1}^{N_{RF}} \sum_{k=0}^{K-1} \operatorname{Re} \left(\Theta_k^*(m, n) e^{-j2\pi f_k T(m, n)} \right). \quad (21)$$

Assuming the delay of the delay element $T(m, n)$ is denoted as t , where t takes values within the range $[0, T_{\max}]$, and there are a total of S delay values. When S is sufficiently large, evenly traversing the S delay values allows us to obtain the optimal solution for t . Thus, the optimal values of $T(m, n)$ and \mathbf{T}_k can be obtained.

3.2 Design of the Initial Value of the Scaling Factor

Given \mathbf{V}_{RF} , the goal of maximizing system performance and sum-rate, the design of \mathbf{V}_{BB} can be determined as follows.

$$\sum_{k=0}^{K-1} \log_2 \left| I_{N_r} + \frac{P_t}{N_s \sigma^2} \mathbf{H}_k \mathbf{V}_{RF} \mathbf{V}_{BB} \mathbf{V}_{BB}^H \mathbf{V}_{RF}^H \mathbf{H}_k^H \right|. \quad (22)$$

According to the sum-rate formula and the water-filling power allocation principle, the optimal baseband precoding matrix is given by

$$\mathbf{V}_{BB} = \left(\mathbf{V}_{RF}^H \mathbf{V}_{RF} \right)^{-1/2} \mathbf{U}_e \mathbf{\Gamma}_e, \quad (23)$$

where \mathbf{U}_e represents the set of right singular vectors corresponding to the first N_s largest singular values of $\mathbf{H}_k \left(\mathbf{V}_{RF}^H \mathbf{V}_{RF} \right)^{-1/2}$, and $\mathbf{\Gamma}_e$ is the diagonal matrix that represents the power allocation to each data stream using the water-filling solution. Generally, we have $\mathbf{\Gamma}_e \approx \mathbf{I}_{N_s}$. The number of diagonal elements of $\mathbf{V}_{RF}^H \mathbf{V}_{RF}$ is exactly N_t , and the off-diagonal elements can be approximated as the sum of N_t independent terms. Because of the sparsity of massive MIMO systems, the probability of having fewer independent terms than N_t is high, so it can be approximated as $\mathbf{V}_{RF}^H \mathbf{V}_{RF} \approx N_t \mathbf{I}$. This characteristic also proves that the optimal digital precoder for $N_{RF} = N_s$ usually satisfies $\mathbf{V}_{BB}^H \mathbf{V}_{BB} \propto \mathbf{I}$, and the scaling factor can be further assumed to be obtained under the condition of equal power allocation for all data streams, i.e., $\mathbf{\Gamma}_e \approx \sqrt{P/N_{RF}} \mathbf{I}$. Therefore, the $\mathbf{V}_{BB} \approx \gamma \mathbf{U}_e$, and we have

$$\gamma = \sqrt{P / (N_t N_{RF})}. \quad (24)$$

3.3 Semi-Unitary Matrix and Scaling Factor Optimization for Digital Precoder Design

Given the analog precoder matrix and initial values of the scaling factors, the (15) can be further expressed as

$$\begin{aligned} \min_{\gamma, \mathbf{V}_{RF}, \mathbf{V}_{DD}} \left\| \mathbf{F}_{opt} - \mathbf{V}_{RF} \gamma \mathbf{V}_{DD} \right\|_F^2 &= \min_{\mathbf{V}_{DD}} \text{tr} \left(\left(\mathbf{F}_{opt} - \gamma \mathbf{V}_{RF} \mathbf{V}_{DD} \right)^H \left(\mathbf{F}_{opt} - \gamma \mathbf{V}_{RF} \mathbf{V}_{DD} \right) \right) \\ &= \min_{\mathbf{V}_{DD}} \left\| \mathbf{F}_{opt} \right\|_F^2 - 2\gamma \text{Re} \left(\text{tr} \left(\mathbf{V}_{DD} \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right) \right) + \left\| \mathbf{V}_{RF} \mathbf{V}_{DD} \right\|_F^2, \end{aligned} \quad (25)$$

where the constant terms $\left\| \mathbf{F}_{opt} \right\|_F^2 = N_s$ and $\left\| \mathbf{V}_{RF} \mathbf{V}_{DD} \right\|_F^2 = N_s$ can be obtained from the power constraint. Therefore, the above minimization problem can be rewritten as the following maximization problem

$$\mathbf{V}_{DD}^{opt} = \arg \max_{\mathbf{V}_{DD}} \gamma \text{Re} \left(\text{tr} \left(\mathbf{V}_{DD} \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right) \right), \quad (26)$$

for (26), we have

$$\begin{aligned} \gamma \text{Re} \left(\text{tr} \left(\mathbf{V}_{DD} \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right) \right) &\leq \left| \text{tr} \left(\gamma \mathbf{V}_{DD} \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right) \right| \\ &\stackrel{(a)}{\leq} \left(\text{tr} \left| \mathbf{V}_{DD} \right|^p \right)^{1/p} \left(\text{tr} \left| \gamma \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right|^q \right)^{1/q} \\ &\stackrel{(b)}{=} \left\| \mathbf{V}_{DD} \right\|_\infty \times \left\| \gamma \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right\|_1 \\ &= \left\| \gamma \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right\|_1 = \sum_{i=1}^{N_s} \sigma_i, \end{aligned} \quad (27)$$

where (a) follows the Hölder inequality, and $p > 0$, $1/p + 1/q = 1$, while $\|\cdot\|_\infty$ and $\|\cdot\|_1$ in (b) represent the L_∞ and L_1 of the Schatten norm [21], respectively. σ_i is the i -th singular value of $\gamma \mathbf{F}_{opt}^H \mathbf{V}_{RF}$, where $i = 1, 2, \dots, N_s$.

The truncated SVD of $\gamma \mathbf{F}_{opt}^H \mathbf{V}_{RF}$ is given by

$$\gamma \mathbf{F}_{opt}^H \mathbf{V}_{RF} = \mathbf{U}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^H, \quad (28)$$

and when (a) holds, we obtain

$$\mathbf{V}_{DD} = \mathbf{V}_1 \mathbf{U}_1^H. \quad (29)$$

From (29), we can obtain the digital precoding matrix \mathbf{V}_{BB}

$$\mathbf{V}_{BB} = \gamma \mathbf{V}_{DD} = \gamma \mathbf{V}_1 \mathbf{U}_1^H. \quad (30)$$

This shows that the solution to the digital precoding matrix avoids the problem of high-dimensional matrix inversion and only involves singular value decomposition operations, resulting in a significant reduction in system computational complexity. However, the scaling factor is still not optimal and needs further optimization.

The initial value of the scaling factor can be determined according to the water-filling power allocation rule. However, due to the certain error between the optimized semi-unitary matrix \mathbf{V}_{DD} and the actual γ , further optimization of γ is needed. The optimization problem in (15) is simplified to

$$\begin{aligned} \min_{\gamma} \|\mathbf{F}_{opt} - \mathbf{V}_{RF} \gamma \mathbf{V}_{DD}\|_F^2 &= \min_{\gamma} \text{tr} \left(\left(\mathbf{F}_{opt} - \gamma \mathbf{V}_{RF} \mathbf{V}_{DD} \right)^H \left(\mathbf{F}_{opt} - \gamma \mathbf{V}_{RF} \mathbf{V}_{DD} \right) \right) \\ &= \min_{\gamma} \|\mathbf{F}_{opt}\|_F^2 - 2\gamma \text{Re} \left(\text{tr} \left(\mathbf{V}_{DD} \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right) \right) + \gamma^2 \|\mathbf{V}_{RF} \mathbf{V}_{DD}\|_F^2, \end{aligned} \quad (31)$$

and the optimal value of γ is γ_{opt} from (32)

$$\gamma_{opt} = \frac{\text{Re} \left(\text{tr} \left(\mathbf{V}_{DD} \mathbf{F}_{opt}^H \mathbf{V}_{RF} \right) \right)}{\|\mathbf{V}_{RF} \mathbf{V}_{DD}\|_F^2}. \quad (32)$$

At this point, the minimum value of (31) is achieved.

The overall framework of the proposed algorithm is as follows:

Algorithm 1: Low-Complexity Optimization Algorithm

- 1: **Input:** Fully digital precoder \mathbf{F}_{opt} , channel matrix \mathbf{H}_k of the k -th subcarrier
 - 2: **Output:** Analog precoding matrix \mathbf{V}_{RF} , semi-unitary digital precoding matrix \mathbf{V}_{DD} , scaling factor γ
 - 3: Initialization: $\mathbf{V}_{RF} = \mathbf{I}$, $\mathbf{V}_{DD} = \mathbf{I}$
 - 4: Calculate the initial value of γ according to (24)
 - 5: **For** $n = 1, 2, 3, \dots$, **do**
 - 6: Update the analog precoding matrix \mathbf{V}_{RF} based on (18) and (21)
 - 7: Update the semi-unitary matrix $\mathbf{V}_{BB} = \gamma \mathbf{V}_1 \mathbf{U}_1^H$ based on (30)
 - 8: Update the scaling factor $\gamma_{opt} = \frac{\text{Re}\left(\text{tr}\left(\mathbf{V}_{DD} \mathbf{F}_{opt}^H \mathbf{V}_{RF}\right)\right)}{\|\mathbf{V}_{RF} \mathbf{V}_{DD}\|_F^2}$ based on (32)
 - 9: **End for**
 - 10: **Return** \mathbf{V}_{RF} , \mathbf{V}_{DD} , γ
-

4. Simulation results

In this section, we evaluate the performance of the proposed method in THz massive MIMO systems through numerical simulation. We compare it with several other schemes, including fully digital precoding, traditional spatially sparse precoding [20], DP-AltMin [9], TTD-DPP[12], and an improved OMP algorithm [22]. The parameter settings in this paper are shown in Table 2.

Table 2. Simulation parameters

Simulation Parameters	Parameter Settings
The number of transmit antennas N_t	256
The number of receiving antennas N_r	4
The number of carriers K	128
The central frequency f_c	0.1 THz
AoD, AoA	$\mathcal{U}(-\pi/2, \pi/2)$
The bandwidth B	$B = 10$ GHz
The path gain $\alpha_{l,p}$	$\alpha_{l,p} \sim \mathcal{CN}(0,1)$
The number of sub-paths per cluster	10

4.1 Achievable Sum-Rate Comparison

In Fig. 2 and Fig. 3, we present the relationship between system throughput and signal-to-noise ratio (SNR) for cases where the $N_{RF} = 4$ and $N_{RF} = 8$, respectively. The configuration assumes that the N_s per subcarrier is $N_s = 4$. Each RF chain has $M = 16$ delay elements

connected to it, with each delay element connected to N_t/M phase shifters. The SNR is defined as $\text{SNR} = 10 \log_{10} P_t / \sigma^2$, where P_t is the system transmission power and σ^2 is the noise power. It can be seen from Fig. 2, when $N_{RF} = 4$, as SNR increases, the achievable system sum-rate of all algorithms exhibits a linear upward trend. The hybrid precoding scheme proposed in this paper achieves higher system rates at different SNRs, particularly at high SNRs, where the sum-rate achieved by the proposed scheme is closest to that of the fully digital precoder. Fig. 3 illustrates the relationship between system sum-rate and SNR when $N_{RF} = 8$. With an increase in the number of RF chains, the overall system sum-rate improves. The algorithm proposed in this paper enables a closer approximation to fully digital precoders. This is because the proposed scheme combines a delayed-phase analog precoding approach with semi-unitary matrix digital precoding algorithm, which effectively solves the beam splitting problem and improves the system sum-rate while reducing the computational complexity and enhancing the system performance.

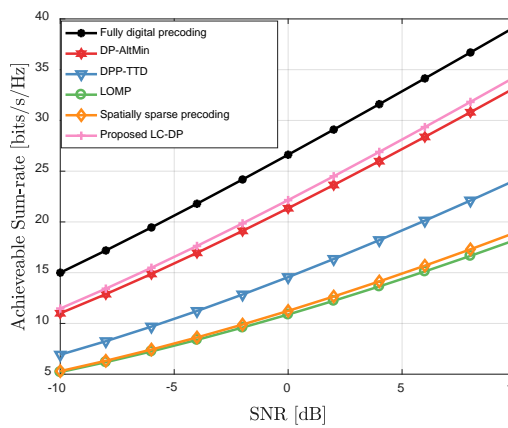


Fig. 2. Comparisons of achievable sum-rates for different architectures with $N_{RF} = 4$.

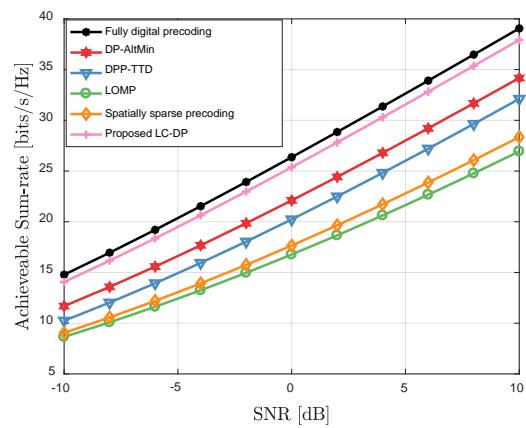


Fig. 3. Comparisons of achievable sum-rates for different architectures with $N_{RF} = 8$.

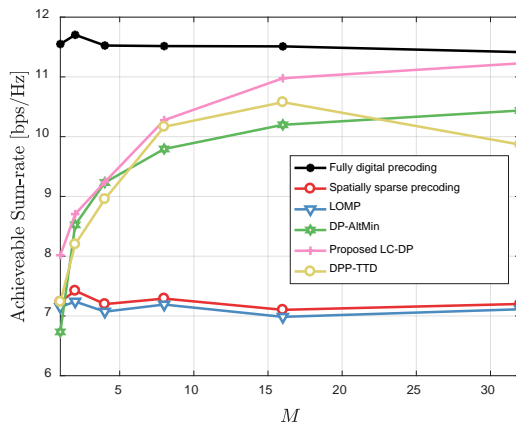


Fig. 4. Achievable sum-rate versus the number of M with $\text{SNR} = -10$ dB.

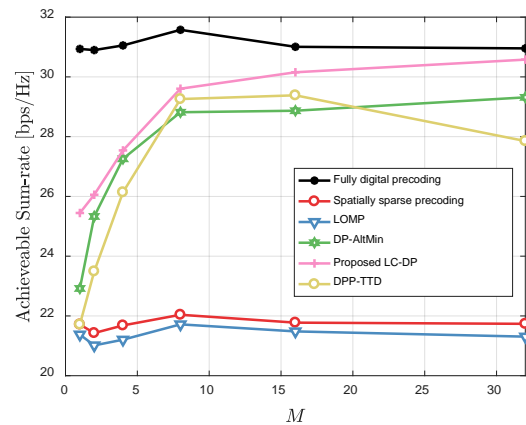


Fig. 5. Achievable sum-rate versus the number of M with $\text{SNR} = 10$ dB.

Fig. 4 and **Fig. 5** illustrate the correlation between system sum-rate and the quantity of delay elements per RF chain at low and high SNR, respectively. We gradually increased the number of delay elements per RF chain from 1 to 32. At SNR = -10 dB, **Fig. 4** illustrates that the system sum-rate shows an increasing trend as the number of delay elements per RF chain increases, the proposed approach in this paper is closest to a fully digital precoder. **Fig. 5** illustrates that when the SNR reaches 10 dB, the proposed algorithm achieves the greatest system sum-rate. Considering that a higher number of delay elements can lead to increased system power consumption, a balance between achievable rates and hardware cost is crucial. Therefore, we generally set the number of delay elements per RF chain to 16. Through simulation comparisons, it is evident that the proposed scheme achieves a system sum-rate that roughly approximates that of a fully digital precoder, whereas the performance of existing hybrid precoding schemes is significantly lower

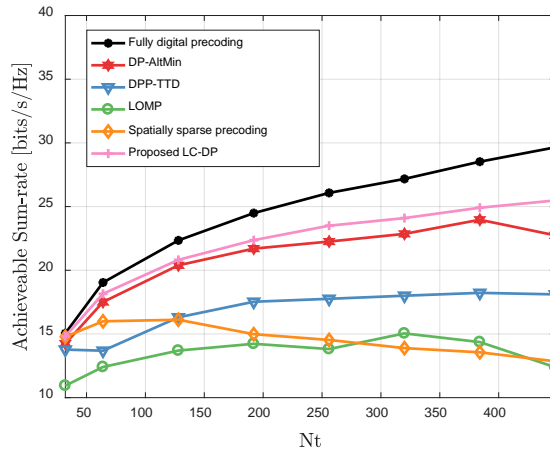


Fig. 6. Achievable sum-rate versus the number of N_t .

In **Fig. 6** we present the sum-rate against N_t . Our observations indicate that the proposed methodology exhibits superior performance compared to other hybrid precoding schemes across various values of N_t . With the increasing of N_t , all schemes show an ascending trend in achievable system sum-rate. Notably, as the number of antennas increases, the performance advantages of the proposed method become increasingly apparent due to the consideration of eliminating the impact of beam splitting. Consequently, for the prevalent deployment of massive MIMO systems, the methodology in this paper presents heightened practicality and efficacy.

4.2 Energy Efficiency Comparison

Fig. 7 and **Fig. 8** illustrate the relationship between system energy efficiency (EE) and the number of delay elements per RF chain at low and high SNR. The SNR values are set to -10 dB and 10 dB, respectively. EE is defined as $\eta = R/P_{total}$, where P_{total} is the power consumption. $P_{total} = P_t + P_c + N_{RF}P_{RF} + N_{RF}N_tP_{PS} + N_{RF}MP_{TD} + P_{BB}$, and P_t is the system calculation cost, comprising two parts: $p_c = 14.1$ mW/MOps, the power consumption of the digital signal processor (DSP) for every 10 million operations (MOps), and the computational complexity N_c [23]. The power consumptions of each RF chain, phase shifter, delay element,

and baseband processor are $P_{RF} = 230$ mW, $P_{PS} = 10$ mW, $P_{TD} = 100$ mW, and $P_{BB} = 300$ mW, respectively. By gradually increasing the number of delay elements per RF chain from 1 to 32, we computed the system EE under various precoding schemes. The results indicate a gradual increase in system EE with the growing number of delay elements. The system EE reaches its peak when the number of delay elements is 8 and then experiences a decline. The comparison conducted indicates that the proposed approach achieves superior EE both in high and low SNR scenarios. Combining this observation with the analysis presented in Fig. 4 and Fig. 5, it becomes evident that a higher number of delay elements per RF chain leads to a higher achievable system sum-rate, a peak in system EE occurs when the number of delay elements is approximately 8.

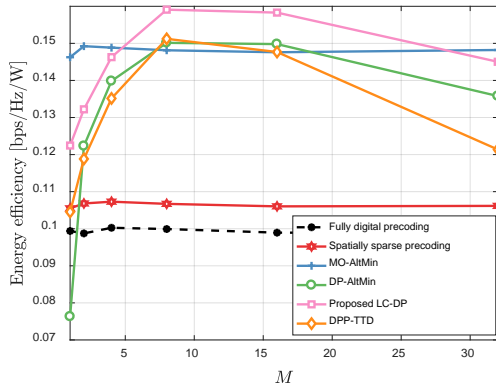


Fig. 7. Comparisons of EE with different M at SNR = -10 dB .

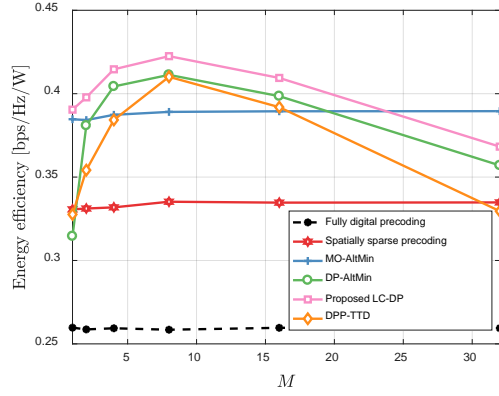


Fig. 8. Comparisons of EE with different M as SNR = 10 dB .

4.3 Computational Complexity Comparison

In the low-complexity hybrid precoding scheme, assuming that the number of iterations is I_{niter} , the computational complexity can be expressed as follows

$$\mathcal{O}\left(KN_t N_r^2 + I_{niter} KN_s N_{RF} N_t + I_{niter} N_{RF} KMS + KN_s N_t N_{RF}\right), \quad (33)$$

where $\mathcal{O}\left(KN_t N_r^2\right)$ is the computational complexity of calculating \mathbf{F}_{opt} for K subcarriers, $\mathcal{O}\left(I_{niter} N_{RF} K\left(N_s N_t + MS\right)\right)$ is the computational complexity of updating \mathbf{V}_{RF} . $\mathcal{O}\left(KN_s N_t N_{RF}\right)$ is the computational complexity of the digital precoder, which mainly involves singular value decomposition of $\mathbf{F}_{opt}^H \mathbf{V}_{RF}$.

To ensure an accurate comparison of the computational complexity among these schemes, we conducted simulations using identical computer hardware and software setups and measured the running time of each simulation.

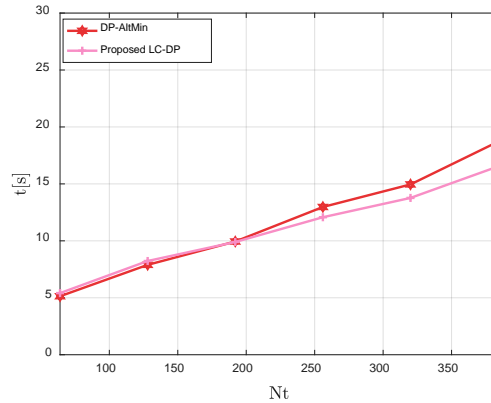


Fig. 9. Runtime versus N_t .

Fig. 9 illustrates the system runtime, where different numbers of BS transmitting antennas are considered. To enhance the reliability of the simulation results, we conducted 50 Monte Carlo simulations to compare the runtime of the proposed scheme with that of DP-AltMin. As shown in **Fig. 9**, the runtimes of both algorithms increase as the number of antennas grows. However, the proposed algorithm demonstrates a lower increase rate than the other precoding scheme. This finding is particularly relevant in the context of THz massive MIMO systems that employ a large number of antennas, as the proposed algorithm effectively reduces system complexity and shortens runtime.

5. Conclusion

This paper introduces a new and efficient hybrid precoding scheme for terahertz massive MIMO communication systems. This method utilizes delay-phase alternating minimization hybrid precoding to address the issue of beam splitting in THz systems. Instead of using the original LS algorithm, we employ matrix singular value decomposition to obtain the fully digital precoder. Subsequently, the digital precoder is optimized to minimize the Euclidean distance between the hybrid precoder and the fully digital precoder. The Simulation results illustrate that the proposed scheme achieves superior performance and runtime efficiency compared to other THz hybrid precoding strategies. Future research can explore the extension of this method to multi-user scenarios.

References

- [1] Z. Zhang et al., "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28-41, 2019. [Article \(CrossRef Link\)](#).
- [2] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical communication*, vol. 12, pp. 16-32, 2014. [Article \(CrossRef Link\)](#).
- [3] T. S. Rappaport et al., "Wireless communications and applications above 100 GHz: opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729-78757, 2019. [Article \(CrossRef Link\)](#).
- [4] D. Headland, Y. Monnai, D. Abbott, C. Fumeaux, and W. Withayachumnankul, "Tutorial: Terahertz beamforming, from concepts to realizations," *APL Photonics*, vol. 3, no. 5, 2018. [Article \(CrossRef Link\)](#).

- [5] T. Cheng, Y. He, Y. Wu, S. Ning, Y. Sui and Y. Huang, "Low Complexity Hybrid Precoding in Millimeter Wave Massive MIMO Systems," *KSII Transactions on Internet and Information Systems*, vol. 16, no. 4, pp. 1330-1350, 2022. [Article \(CrossRef Link\)](#).
- [6] H. Hashemi, T. -s. Chu and J. Roderick, "Integrated true-time-delay-based ultra-wideband array processing," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 162-172, 2008. [Article \(CrossRef Link\)](#).
- [7] W. Hao, X. You, G. Sun and Z. Zheng, "Design of antenna structure and analysis of beam split effect in ultra-bandWidth terahertz communications," *Journal of Electronics & Information Technology*, vol. 45, no. 1, pp. 200-207, 2023. [Article \(CrossRef Link\)](#).
- [8] D. Q. Nguyen and T. Kim, "Joint delay and phase precoding under true-time delay constraints for THz massive MIMO," in *Proc. of ICC 2022 - IEEE International Conference on Communications*, Seoul, Korea, Republic of, pp. 3496-3501, 2022. [Article \(CrossRef Link\)](#).
- [9] M. Cui, J. Tan, L. Dai. "Wideband hybrid precoding for THz massive MIMO with angular spread (in Chinese)," *Scientia Sinica Informationis*, vol. 53, no. 4, pp. 772-786, 2023. [Article \(CrossRef Link\)](#).
- [10] Y. Chen, Y. Xiong, D. Chen, T. Jiang, S. X. Ng, and L. Hanzo, "Hybrid precoding for wideband millimeter wave MIMO systems in the face of beam squint," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1847-1860, 2021. [Article \(CrossRef Link\)](#).
- [11] R. Zhang, W. Hao, G. Sun and S. Yang, "Hybrid precoding design for wideband THz massive MIMO-OFDM systems with beam squint," *IEEE Systems Journal*, vol. 15, no. 3, pp. 3925-3928, 2021. [Article \(CrossRef Link\)](#).
- [12] J. Tan and L. Dai, "Delay-phase precoding for THz massive MIMO with beam split," in *Proc. of 2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, pp. 1-6, 2019. [Article \(CrossRef Link\)](#).
- [13] S. Wang, M. He, Y. Zhang, and R. Ruby, "Equivalent channel-based joint hybrid precoding/combining for large-scale MIMO systems," *Physical Communication*, vol. 47, 2021. [Article \(CrossRef Link\)](#).
- [14] H. Wang, S. Lim and K. Ko, "Improved Maximum Access Delay Time, Noise Variance, and Power Delay Profile Estimations for OFDM Systems," *KSII Transactions on Internet and Information Systems*, vol. 16, no. 12, pp. 4099-4113, 2022. [Article \(CrossRef Link\)](#).
- [15] B. Peng, K. Guan, and T. Kurner, "Cooperative dynamic angle of arrival estimation considering space-time correlations for terahertz communications," *IEEE Transactions on Wireless Communications*, vol. 17, no. 9, pp. 6029-6041, 2018. [Article \(CrossRef Link\)](#).
- [16] C. Lin and G. Y. L. Li, "Terahertz communications: An array-of-subarrays solution," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 124-131, 2016. [Article \(CrossRef Link\)](#).
- [17] Y. Xing and T. S. Rappaport, "Propagation measurement system and approach at 140 GHz-moving to 6G and above 100 GHz," in *Proc. of 2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, pp. 1-6, 2018. [Article \(CrossRef Link\)](#).
- [18] M. R. Akdeniz et al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164-1179, 2014. [Article \(CrossRef Link\)](#).
- [19] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831-846, 2014. [Article \(CrossRef Link\)](#).
- [20] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499-1513, 2014. [Article \(CrossRef Link\)](#).
- [21] R.A. Horn, C.R. Johnson, *Matrix Analysis*, New York: Cambridge University Press, 1985.
- [22] Q. Xiao, F. Tan, "Hybrid precoding in millimeter wave massive MIMO based on improved orthogonal matching pursuit algorithm," *Application Research of Computers*, vol. 40, no. 1, pp. 239-243, 2022. [Article \(CrossRef Link\)](#).

- [23] Y.-Y. Lee, C.-H. Wang, and Y.-H. Huang, "A hybrid RF/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 305-317, 2015. [Article \(CrossRef Link\)](#).



Yuke Sun received the B.S. degree in electronic information engineering from Zhongyuan University of Technology, Henan, China in 2021, and the M.S. degree in electronic information from Zhongyuan University of Technology, Henan, China in 2024. Her research interests include communication signal processing, channel estimation and hybrid precoding.



Aihua Zhang received the B.S. and Ph.D. degrees from Zhengzhou University of china in 1998 and 2014, respectively. She received the master degree from China University of Mining & Technology-Beijing in 2003. From 2003 to 2014, she was a Lecturer. Since 2014, she has been a Professor with Zhongyuan University of Technology. She was a visiting researcher with the Department of Electrical Engineering and Computer Science, University of Louisville, Kentucky, USA. Her current interests are in the areas of signal processing in communication, channel estimation, sparse signal processing.



Hao Yang received the B.S. degree in electronic information engineering from Zhongyuan University of Technology, Henan, China in 2020, and the M.S. degree in communication and information systems from Zhongyuan University of Technology, Henan, China in 2024. His research interests include communication signal processing, channel estimation, integrated sensing and communication.



Di Tian received the B.S. degree in electronic information engineering from Zhongyuan University of Technology, Zhengzhou, China, in 2022, and She is now studying for a master 's degree from Zhongyuan University of Technology. Her research interests include channel information acquisition and communication signal processing



Haowen Xia received the B.S. degree in communication engineering from Heilongjiang University of Science and Technology, Harbin, China, in 2023, and he is now studying for a master 's degree from Zhongyuan University of Technology. His research interests include integrated sensing and communication, the combination of AFDM and massive MIMO.