Spatially Sparse Precoding in Wideband Hybrid Terahertz Massive MIMO Systems

Jiaobao Gao, Caijun Zhong, Senior Member, IEEE, Geoffrey Ye Li, Fellow, IEEE,
Joseph B. Soriaga, and Arash Behboodi, Member, IEEE

Abstract—In terahertz (THz) massive multiple-input multiple-output (MIMO) systems, the combination of huge bandwidth and massive antennas results in severe beam split, thus making the conventional phase-shifter based hybrid precoding architecture ineffective. With the incorporation of true-time-delay (TTD) lines in the hardware implementation of the analog precoders, delay-phase precoding (DPP) emerges as a promising architecture to effectively overcome beam split. However, existing DPP approaches suffer from poor performance, high complexity, and weak robustness in practical THz channels. In this paper, we propose a novel DPP approach in wideband THz massive MIMO systems. First, the matrix decomposition optimization problem is converted into a compressive sensing (CS) form, which can be solved by the proposed extended spatially sparse precoding (SSP) algorithm. To compensate for beam split, frequency-dependent measurement matrices are designed, which can be approximately realized by feasible phase and delay codebooks. Furthermore, several efficient atom selection techniques are developed to further reduce the complexity of the extended SSP algorithm. In simulation, the proposed DPP approach achieves superior performance, complexity, and robustness by using it alone or in combination with existing DPP approaches under various settings.

Index Terms—THz, hybrid massive MIMO, delay phase precoding, beam split, compressive sensing.

I. INTRODUCTION

In future 6G wireless communication systems, it is foreseeable that ultra-high data rates will be required to support many cutting-edge applications, such as virtual reality, augmented reality, and digital twins [1]. With abundant bandwidth resources, terahertz (THz) is a promising frequency band to meet the sharply increased data rate requirement [2], [3].

Manuscript received 2 March 2023; revised 10 May 2023; accepted 29 June 2023. Date of publication 19 July 2023; date of current version 12 March 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1803400. The associate editor coordinating the review of this article and approving it for publication was S.-H. Park. (Corresponding author: Caijun Zhong.)

Jiaobao Gao and Caijun Zhong are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310007, China (e-mail: gao.jiaobao@zju.edu.cn; caijuzhong@zju.edu.cn).

Geoffrey Ye Li is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BU London, U.K. (e-mail: Geoffrey.Li@imperial.ac.uk).

Joseph B. Soriaga is with Qualcomm Technologies Inc., San Diego, CA 92122 USA (e-mail: jsoriaga@qti.qualcomm.com).

Arash Behboodi is with Qualcomm Technologies Netherlands B.V., 1098 XH Amsterdam, The Netherlands (e-mail: behboodi@qti.qualcomm.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2023.3292834.

Digital Object Identifier 10.1109/TWC.2023.3292834

Specifically, the bandwidth provided by the THz band could reach up to several tens of GHz, which is much larger than that in the millimeter-wave (mmWave) band used in 5G [2]. However, higher carrier frequencies make the THz signals suffer from more severe attenuation than signals in lower frequency bands [4]. To alleviate this issue, THz is usually combined with massive multiple-input multiple-output (MIMO), where directional beams with high array gains can be generated by massive antennas at transceivers [4], [5], [6]. Furthermore, to relieve the manufacturing cost and power consumption, it is a common practice to adopt the hybrid precoding architecture in THz massive MIMO systems, where precoding is decomposed into a high-dimensional analog precoder and a low-dimensional digital precoder [7], [8]. Thanks to the sparsity of THz channels in the angular domain, satisfactory performance can still be achieved with only a few radio-frequency (RF) chains using hybrid precoding. Next, we will briefly review prior works about hybrid precoding in THz massive MIMO systems. Since mmWave channels and THz channels have many similar properties, some works based on the mmWave model are included as well.

A. Prior Works

In most prior works, the analog precoder is implemented by phase shifters. Since the optimal solutions are hard to obtain in the presence of the non-convex unit modulus constraint of phase shifts, the approximate orthogonal properties of the optimal analog and digital precoders are exploited in [7] to simplify the optimization process. In the well-known spatially sparse precoding (SSP) algorithm [9], the sum rate maximization problem is first converted to an approximately equivalent while much simpler matrix decomposition problem, where the distance between the fully-digital precoder and the hybrid precoder is minimized. Then, by exploiting the structures of optimal precoders, the optimization problem is further converted to a typical sparse recovery problem that can be readily solved by various compressive sensing (CS) algorithms. In [10], the complexity of SSP is reduced by using orthogonal codebooks and beamspace singular value decomposition (SVD). To achieve better performance without constraining the form of the analog precoder, an alternative minimization (AM) algorithm is proposed in [11] to optimize the analog and digital precoders alternately. To maximize the energy efficiency, the number of active RF chains and the digital-to-analog converter (DAC)/analog-to-digital converter (ADC) bit resolutions are
further dynamically optimized in [12] and [13], respectively. In [14] and [15], deep neural networks are used to predict hybrid precoders based on channel state information (CSI). Without the need of estimating the high-dimensional full CSI, codebook based approaches are appealing in practice, where efficient codebooks and beam sweeping procedures are developed to reduce the overhead [16], [17].

Although the above phase-shifter-based hybrid precoding architecture works well in most narrowband systems [8], it will suffer from severe performance degradation in wideband systems, especially when the number of antennas and the fractional bandwidth (ratio between bandwidth and central frequency) are large, e.g., THz massive MIMO. The reason is the existence of the beam split effect, which is proportional to both the number of antennas and the fractional bandwidth [18], [19], [20]. With beam split, there is a natural contradiction, such that the same physical channel path will have different equivalent angles at different subcarriers and make the optimal fully-digital precoders frequency-dependent while the analog precoder implemented by phase shifters is shared by all subcarriers and frequency-independent [20].

To minimize the impact of beam split, several methods have been proposed from the perspective of algorithm redesign [20], [21], [22], [23], [24]. In [20], a beam broadening approach has been proposed where wider beams are constructed through subarray coordination to cover the split directions at all subcarriers. Since the AM algorithm works with general channels including those with beam split, it is used in [21] to solve the wideband hybrid precoding problem. More generally, AM is used in [22] to develop mmWave and THz hybrid precoding algorithms for various analog architectures, such as the quantized phase shifters and the use of switches for dynamic connections. In [23] and [24], codebooks are revised to maximize the minimum array gain and the total array gain achieved across all subcarriers, respectively. Nevertheless, these methods still use a frequency-independent analog precoder, therefore will lose most of the effectiveness when beam split is relatively severe.

Since beam split essentially comes from the delay of electromagnetic wave traveling through the antenna array [18], a more direct and thorough solution is to compensate for the delay from the perspective of hardware redesign. In the emerging delay-phase precoding (DPP) architecture, true-time-delay (TTD) lines are incorporated apart from phase shifters to implement analog precoders. Since the phase shift of a TTD line is proportional to both its delay parameter and the subcarrier frequency, frequency-dependent analog precoders can be realized. Assuming the path channel model, heuristic DPP algorithms that exploit the angles of channel paths sorted by their path gains are proposed in [20] and [25], which can achieve near-optimal performance with low complexity. To support cluster channels, the AM algorithm is used in [26] to optimize the time delays, phase shifts, and digital precoders alternately. To realize adjustable TTD lines, a large number of cascaded time delay units and single-pole two-throw switches, as well as a movable parallel-coupled high-resistivity silicon slab are required, which causes high power consumption, insertion loss, and hardware complexity [27]. To alleviate these issues, low-cost fixed TTD lines that can be simply realized by a waveguide with fixed length or a microstrip line are used along with a switch network in [27]. In [28], multi-resolution time-delay codebooks are designed through subarray coordination for efficient beam sweeping. In [29], the level of beam split is controlled by delay-phase configuration to accelerate beam tracking, which is then extended to the near-field scenario in [30] by modifying the codebooks. In [31], the effectiveness of the DPP architecture is validated by practical 6G THz massive MIMO channels.

Among the existing DPP algorithms, it is hard to find one that satisfies both the strict requirements for performance and complexity in 6G communication systems with practical THz channels. Specifically, the performance of the heuristic DPP algorithms will degrade in cluster channels while iterative optimization based DPP algorithms usually suffer from high complexity. The SSP algorithm originally designed for narrowband systems has many attractive characteristics thanks to the exploitation of precoder structures [9], such as low complexity and satisfactory performance in cluster channels. Nevertheless, its extension to the wideband DPP scenario with beam split considered is not investigated yet in the literature.

B. Contributions

In this paper, the matrix decomposition formulation is selected for wideband DPP to cope with practical THz cluster channels, where the goal is to approximate fully-digital precoders with hybrid precoders at all subcarriers. The main contributions of this paper are listed as follows:

• We propose an extended SSP algorithm (E-SSP) to realize low-complexity wideband DPP in practical cluster THz channels. The matrix decomposition optimization problem is formulated as a standard multiple-measurement-vector (MMV)-CS problem by exploiting the structures of optimal precoders, which can be solved by various classic MMV-CS algorithms such as simultaneous orthogonal matching pursuit (SOMP). To the best of our knowledge, this is the first work to solve the wideband DPP problem under the CS framework.

• Following the divide-and-conquer manner, we propose a novel two-step scheme to simplify the design process of phase shifts and delays. The first step is to design ideal measurement matrices to compensate for beam split ignoring the feasibility issue. The second step is to approximate these measurement matrices by feasible phase and delay codebooks.

• To compensate for beam split, frequency-dependent measurement matrices are designed according to the frequency dependency of wideband THz channels so that the resulted sparse supports of all measurements can be aligned in the formulated MMV-CS problem to obtain a large MMV gain. It is the first time for frequency-dependent measurement matrices to be applied to a precoding problem.

\[1\] Notice that the term beam squint is used more often in mmWave systems. Since beam split can be seen as the more severe version of beam squint, we will just use the term beam split in both mmWave and THz systems for neatness in this paper.
By exploiting the wideband features of TTD lines' delays and the coupling patterns between phase shifters and TTD lines, feasible phase and delay codebooks are efficiently and effectively designed to realize the ideal frequency-dependent measurement matrices approximately.

To further reduce the complexity of E-SSP, the properties of projections in SOMP in the angular-frequency domain are thoroughly exploited. Specifically, three tailored techniques are proposed to reduce the numbers of required iterations, atoms, and subcarriers, respectively, in the atom selection process. The resulted low-complexity E-SSP (LCE-SSP) algorithm is much more appealing in practice thanks to the dramatically reduced complexity but rarely degraded performance.

For channels with large-angular-spread clusters, it is still necessary to use iterative optimization based DPP algorithms to achieve high performance. In this case, the proposed LCE-SSP algorithm can serve as a cheap yet high-quality initializer to improve both the convergence speed and performance.

The rest of this paper is organized as follows. In Section II, the hybrid massive MIMO system, the wideband THz channel model, and the DPP optimization problem are introduced. Then, the extended SSP algorithm is elaborated in Section III, whose complexity is further reduced using the efficient atom selection techniques presented in Section IV. In Section V, the superiority of the proposed approach is validated through numerical results, and the paper is eventually concluded in Section VI.

**Notations:** Italic, bold-face lower-case and bold-face upper-case letters denote scalar, vector, and matrix, respectively, $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{\dagger}$, $\mathbb{E}()$, $\odot$, and $\otimes$ denote conjugate, transpose, conjugate transpose, Moore-Penrose inverse, expectation, Hadamard product, and Kronecker product, respectively. $\mathbf{A}^{-\frac{1}{2}}$ denotes the matrix that satisfies $\mathbf{A}^{-\frac{1}{2}} \mathbf{A}^{-\frac{1}{2}} = \mathbf{A}^{-1}$. $\| \mathbf{A} \|_F$ denotes the Frobenius norm of matrix $\mathbf{A}$ and $\| \mathbf{a} \|_p$ denotes the $p$-norm of vector $\mathbf{a}$. $\| \mathbf{a} \|$ denotes the modulus of complex scalar $\mathbf{a}$ while $| \mathbf{A} |$ denotes the determinant of matrix $\mathbf{A}$. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix while $\mathbf{1}_N$ and $\mathbf{0}_N$ denote the $N \times 1$ all-one and all-zero column vectors. $\text{diag}(\cdot)$ can denote the diagonalization operation or the inverse diagonalization operation depending on specific situations. $a \div b$ denotes the operation of taking the quotient of the dividend $a$ and the divisor $b$. Hermiting the Matlab style of matrix slicing, we use $\mathbf{A}(i, j)$, $\mathbf{A}(i, :)$, $\mathbf{A}(i, j)$, and $\mathbf{A}(i)$ to denote the element at the $i$-th row and $j$-th column, the elements at the $i$-th row, the elements at the $j$-th column, and the elements from the $i$-th column to the $j$-th column of matrix $\mathbf{A}$, respectively. $\lfloor \cdot \rfloor$ can denote the horizontal or vertical concatenation of matrices or vectors depending on specific situations. $a : b : c$ denotes the arithmetic sequence vector that starts from $a$ and ends at $c$ with common difference $b$. $\mathbb{C}^{n \times m}$ denote the $n \times m$ complex space. $\mathbb{C}N(\mu, \sigma^2)$ denotes a circularly symmetric complex Gaussian (CSCG) random variable with mean $\mu$ and variance $\sigma^2$.

II. System Model and Problem Formulation

In this section, we first introduce the hybrid massive MIMO system under the DPP architecture and the practical wideband THz cluster channel model along with beam split. Then, the DPP optimization problem and its conversion to the matrix decomposition form are elaborated.

A. System and Channel Models

We consider the hybrid massive MIMO system illustrated in Fig. 1, where a base station (BS) with an $N_{RF}$-antenna uniform linear array (ULA) and $N_{RF}$ RF chains serves a user with an $N_R$-antenna ULA. We first consider the single-user scenario while the extension to the multi-user scenario will be investigated in our simulation. The number of data streams is $N_s$. Usually, we have $N_s = N_R \ll N_{RF} \ll N_T$ in practice [25]. To combat the frequency selectivity of channels, OFDM is adopted where the total system bandwidth $f_s$ is evenly divided by $K$ subcarriers.

According to measurement results, THz channels usually have clustered characteristics [32]. Considering half-wavelength antenna spacing, we first define the $N$-dimensional ULA response vector as

$$\mathbf{a}_N(\cdot) \triangleq [1, e^{-j\pi(\cdot)}, \ldots, e^{-j\pi(N-1)(\cdot)}]^T / \sqrt{N},$$

then the $N_R \times N_T$ channel matrix at the $k$-th subcarrier from the BS to the user can be expressed as

$$\mathbf{H}_k = \sqrt{\frac{N_T N_R}{N_c N_p}} \sum_{i=1}^{N_c} \sum_{j=1}^{N_p} \alpha_{i,j} e^{-j2\pi f_k \tau_{i,j}} \mathbf{a}_{N_R}(\psi_{i,j,k}) \mathbf{a}_{N_T}^T(\psi_{i,j,k}),$$

**Fig. 1.** Hybrid massive MIMO system under the DPP architecture with cluster channels.
where $N_s$ and $N_p$ denote the number of clusters and the number of subpaths in a cluster, respectively, $f_k = f_c + (k - 1 - \frac{K - 1}{2})\eta$ is the frequency of the $k$-th subcarrier with $f_c$ and $\eta = \frac{f_c}{f_D}$ denoting the central frequency and the subcarrier frequency spacing, respectively. Besides, $\alpha_{i,j}, \tau_{i,j}, \psi_{i,k}^R, \psi_{i,k}^I$ denote the complex gain, the delay, the equivalent angle of departure (AoD) at the BS, and the equivalent angle of arrival (AoA) at the user of the $j$-th subpath in the i-th cluster, respectively.

Notice that, the equivalent path angles are dependent on subcarrier index $k$ due to beam split. Specifically, we have $\psi_{i,j,k} = \frac{\Delta}{\sqrt{f_c}}\sin(\theta_{i,j})$ at both transceiver sides, where $\sin(\theta_{i,j})$ denotes the actual physical path angle. In the i-th cluster, we have $\theta_{i,j}^R = \theta_{i,j}^R + \Delta\theta_{i,j}, \theta_{i,j}^I = \theta_{i,j}^I + \Delta\tau_{i,j}$, where $\theta_{i,j}^R, \theta_{i,j}^I, \tau_{i,j}$ denote the mean AoD, mean AoA, and mean delay, respectively, while $\Delta\theta_{i,j}, \Delta\theta_{i,j}^R, \Delta\tau_{i,j}$ follow the Laplacian distribution [32] with zero mean and standard deviation being $\sigma_{\theta}^R, \sigma_{\theta}^I, \sigma_{\tau}$. The standard deviation here is also known as the angular or delay spread.

Without loss of generality, we assume $\alpha_{i,j} \in C \mathcal{N}(0,1), \forall i,j$, $\theta_{i,j}^R, \theta_{i,j}^I \sim \mathcal{U}(0,2\pi)], \forall i,j$, where $\max\{i,j\}$ denotes the maximum delay.

To effectively compensate for beam split, the DPP architecture is adopted. As illustrated in Fig. 1, a small-scale TTD line network is inserted between RF chains and the large-scale phase shifter network. To reduce the hardware cost and energy consumption, each RF chain only connects to a certain TTD line network. To reduce the hardware cost and energy consumption, each RF chain only connects to a certain TTD line network. To reduce the hardware cost and energy consumption, each RF chain only connects to a certain TTD line network. So, the $N_R \times 1$ user’s received signal at the $k$-th subcarrier can be expressed as

$$y_k = \sqrt{\rho}H_kA_kD_kx_k + n_k,$$

where $x_k$ denotes the $N_s \times 1$ transmit signal at the BS at the $k$-th subcarrier satisfying $\mathbb{E}\{x_kx_k^H]\} = \frac{1}{N_s}I_{N_s}$, and $\rho$ denotes the transmit power at the $k$-th subcarrier.

The analog and digital precoders at the $k$-th subcarrier are denoted by $A_k \in \mathbb{C}^{N_T \times N_R}$ and $D_k \in \mathbb{C}^{N_R \times N_T}$, respectively. The additive white Gaussian noise (AWGN) at the $k$-th subcarrier is denoted by $n_k$, whose elements follow $\mathcal{CN}(0,\sigma_n^2)$ with $\sigma_n^2$ denoting the noise variance. The signal-to-noise ratio (SNR) is defined as $\rho/\sigma_n^2$.

The main difference between the DPP architecture and the conventional phase-shifter-based hybrid precoding architecture is that the analog precoder at each subcarrier is frequency-dependent. Specifically, the analog precoder at the $k$-th subcarrier can be expressed as

$$A_k = A \odot (T_k \otimes 1_M),$$

where $A$ denotes the common phase matrix introduced by the phase shifter network, and the $N_{TTD} \times N_{RF}$ matrix $T_k$ denotes the frequency-dependent phases introduced by the TTD line network at the $k$-th subcarrier. The Kronecker product of $T_k$ and $1_M$ comes from the one-to-M connection mode between TTD lines and antennas at the BS. Denoting the delay matrix of the TTD line network as $T$, a certain TTD line’s phase at a certain subcarrier is proportional to both the delay parameter and the subcarrier frequency and can be expressed as

$$T_k(n_{TTD},n_{RF}) = e^{-j2\pi f_k T(n_{TTD},n_{RF})}, \forall n_{TTD}, n_{RF}, k.$$

B. Problem Formulation

We first assume that perfect downlink CSI is available at the BS and the impact of imperfect CSI will be investigated in simulation. Based on the downlink CSI, analog and digital precoders need to be designed to maximize the sum rate of all subcarriers, which can be expressed as

$$R = \sum_{k=1}^{K} \log_2 \left( I_{N_R} + \frac{\rho}{N_s\sigma_n^2}H_kA_kD_kH_k^H \right).$$

In the presence of constraints caused by the hardware used for analog precoding implementation, directly maximizing (6) is difficult due to its complicated form. Therefore, we turn to solve the following matrix decomposition problem to approximate the fully-digital precoders with the hybrid precoders at all subcarriers, which is proved to be approximately equivalent to sum rate maximization while much simpler to solve [9]:

$$P_1 : \min_{A,T,D_k} \sum_{k=1}^{K} \| F_k - A_kD_k \|^2_F,$$

subject to

$$|A(n_{TTD},n_{RF})| = 1, \forall n_{TTD}, n_{RF},$$

$$0 < T(n_{TTD},n_{RF}) < t_{max},$$

$$\forall n_{TTD}, n_{RF},$$

$$\| A_kD_k \|^2_F = N_s, \forall k.$$
III. EXTENDED SSP BASED WIDEBAND DPP

In this section, we first briefly introduce the principles of the conventional SSP algorithm designed for narrowband phase-shifter-based hybrid precoding. Then, we extend it properly to solve the considered wideband DPP problem.

A. SSP Based Narrowband Hybrid Precoding

Considering narrowband systems with only phase shifters, \( P_1 \) is simplified to the following optimization problem with \( N_{T T D} = 0 \) and the subscript \( k \) discarded:

\[
\mathcal{P}_2 : \quad \min_{\tilde{A}, D} \| F - \tilde{A}D \|_F^2,
\]

\[\text{s.t.} \quad |A(n_{T}, n_{RF})| = 1, \forall n_{T}, n_{RF}, \]

\[
\|AD\|_F^2 = N_s.
\]

As one of the most classical algorithms to solve \( \mathcal{P}_2 \), the key idea of SSP is that the fully-digital precoder can be approximated by the linear combinations of several array response vectors when the channel is sparse in the angular domain [9], e.g., in the mmWave or THz band. Based on such precoder structures, \( \mathcal{P}_2 \) can be converted to the following CS form:

\[
\mathcal{P}_3 : \quad \min_{\tilde{A}} \| F - \tilde{A}D \|_F^2,
\]

\[\text{s.t.} \quad \| \text{diag}(\tilde{D}\tilde{D}^H) \|_0 = N_{RF}, \]

\[
\| \tilde{A}D \|_F^2 = N_s,
\]

where the \( N_T \times G \) measurement matrix consisting of \( G \) atoms (e.g., columns) in the form of ULA response vectors is defined as

\[
\tilde{A} \triangleq \begin{bmatrix} a_{N_T}(\phi^1), a_{N_T}(\phi^2), \cdots, a_{N_T}(\phi^G) \end{bmatrix},
\]

with the corresponding angles evenly dividing the entire angular space such that

\[
\phi^g = -1 + (2g - 1)/G, \forall g.
\]

Now, let us describe \( \mathcal{P}_3 \) using CS terms. The columns of \( F \) are measurements, \( \tilde{A} \) is the common measurement matrix, and the columns of the \( G \times N_s \) auxiliary matrix \( D \) are sparse vectors corresponding to \( N_s \) streams that have common sparse supports. The number of non-zero rows in \( D \) equals the number of RF chains, which means that the analog precoding vector of each RF chain has to be one of the atoms of the measurement matrix. As a typical MMV-CS problem, \( \mathcal{P}_3 \) can be readily solved by the slightly modified SOMP algorithm demonstrated in Algorithm 1. In each of the \( N_{RF} \) iterations, the atom with maximal summed projections of streams on the residual is selected to update the analog precoder while the digital precoder is updated by least square (LS) based on the temporal analog precoder and the fully-digital precoder. At the end, the power of the digital precoder is normalized to meet the transmit power constraint. Notice that, on-grid OMP is more appealing in practice thanks to its low complexity and competitive performance although using more advanced CS algorithms like sparse Bayesian learning and off-grid techniques [36] can lead to slightly better performance.

**Algorithm 1** SSP Based Narrowband Hybrid Precoding [9]

**Input:** \( \tilde{A}, F, N_{RF} \);

1. Initialize \( \tilde{A}^{(0)} \) as an empty matrix, initialize \( F^{(0)}_{\text{res}} = \tilde{F} \);
2. for \( i = 1 : 1 : N_{RF} \) do
   3. \( \psi^{(i)} = \text{diag}(\Psi^{(i)} \Psi^{(i)H}) \);
   4. \( g^{(i)} = \text{argmax} \psi^{(i)}(j) \);
   5. \( A^{(i)} = [A^{(i-1)} | \tilde{A}(\cdot, g^{(i)})] \);
   6. \( D^{(i)} = [A^{(i)}F_{\text{res}}] \); \( \| F - A^{(i)}D^{(i)} \|_F \);
9. end for
10. \( A = A^{(N_{RF})} \);
11. \( D = \sqrt{N_s} A^{(N_{RF})} D^{(N_{RF})} \);

**Output:** \( A, D \);

As indicated by [25] and [26], directly applying SSP to wideband DPP will result in poor performance, which is not surprising because beam split and the delay optimization of TTD lines are not considered in the original designs. Next, we will introduce the proposed extension of SSP from narrowband to the considered wideband DPP scenario.

B. Frequency-Dependent Measurement Matrices

Without beam split and TTD lines, the extension of SSP from narrowband to wideband would be straightforward. Since the equivalent path angles are the same at all subcarriers, the sparse supports of projections would be aligned along the subcarrier dimension. However, with beam split, such common sparsity structure among subcarriers no longer holds if a *frequency-independent* measurement matrix is still used. As a result, target atoms that have good performance at all subcarriers cannot be accurately captured by simple averaging. Besides, using same atoms as analog precoders of all subcarriers cannot exploit TTD lines’ ability of providing *frequency-dependent* phases. To address this issue, we use *frequency-dependent* measurement matrices,\(^5\) where the measurement matrix used at the \( k \)-th subcarrier is defined as

\[
\tilde{A}_{k}^{\text{real}} \triangleq \begin{bmatrix} a_{N_T}(\phi_k^1), a_{N_T}(\phi_k^2), \cdots, a_{N_T}(\phi_k^G) \end{bmatrix},
\]

with *frequency-dependent* angles corresponding to atoms being

\[
\phi_k^g = \frac{f_k}{f_c} \phi^g, \forall g.
\]

We remark that the factor \( \frac{f_k}{f_c} \) in (20) plays the role of *predistortion* such that the sparse supports of projections at all subcarriers would be aligned after experiencing the distortion of beam split. Fig. 2 illustrates the effectiveness of the proposed frequency-dependent measurement matrices in terms of support alignment, where the grayscale images are summed projections of streams at different subcarriers and angles while the dashed curves are obtained by further averaging over all

\(^5\) Notice that such an idea is previously used in [19] for wideband THz channel estimation. Thanks to the similar CS formulation, it can be naturally borrowed for the wideband THz precoding problem considered in this paper.
subcarriers. As we can see, with the frequency-independent measurement matrix, the sparse supports of projections at different subcarriers are not aligned. In contrast, with the proposed frequency-dependent measurement matrices, perfect common sparsity structure is indicated by the vertical straight lines on the grayscale image and the sharp peaks of the dashed curve.

We would like to mention that we have also tried to use broader beams [20] as atoms to cover wider angular ranges of clusters and optimize the measurement matrices based on channel samples using dictionary learning [37]. Nevertheless, no performance gain is observed since the energy distribution in the angular domain of cluster channels changes from sample to sample, thus cannot be dynamically matched by any fixed measurement matrices. As a result, although these designs have different sample-wise performance, their average performance on a large number of samples is similar.

C. Feasible Phase and Delay Codebooks

So far, feasibility issues are not considered yet. In practice, a group of analog precoders at all subcarriers are feasible only when specific phase and delay parameters satisfying respective constraints can be found. Therefore, we need to further optimize a feasible phase codebook, \( \mathbf{A} \), and a feasible delay codebook \( \mathbf{T} \), that can be implemented by actual hardware, such that the corresponding feasible measurement matrices are as close to the ideal measurement matrices as possible. Mathematically, the optimization problem is formulated as follows:

\[
\begin{align*}
\mathcal{P}_4 : \quad & \min_{\mathbf{A}, \mathbf{T}} \quad \sum_{k=1}^{K} \left\| \mathbf{A}_k^{\text{ideal}} - \mathbf{A} \odot (\mathbf{T}_k \otimes \mathbf{1}_M) \right\|_F^2, \\
\text{s.t.} \quad & |\mathbf{A}(n_T, g)| = 1, \forall n_T, g, \\
& \mathbf{T}_k(n_{TTD}, g) = e^{-j2\pi f_k T(n_{TTD}, g)}, \forall n_{TTD}, g, k, \\
& 0 \leq T(n_{TTD}, g) \leq t_{\max}, \forall n_{TTD}, g. 
\end{align*}
\]

(21)

As a special case of \( \mathcal{P}_1 \) when \( N_s = G \) and the digital precoders are ignored, most existing algorithms for \( \mathcal{P}_1 \) can be used to solve \( \mathcal{P}_4 \) with slight modifications, such as the AM algorithm proposed in [26]. Nevertheless, we propose a simple yet effective approach for codebook design, which can achieve low matrix approximation error with low complexity. To simplify the solution process, the following two Lemmas will be used, which are proved in Appendices A and B.

**Lemma 1**: The impact of adding any bias to all subcarriers’ frequencies can be compensated by modifying the phase shifts properly according to the delays of TTD lines.

**Lemma 2**: Adding any bias to the delays of all TTD lines connected to an RF chain will not change the sum rate.

Let us first consider \( N_{TTD} = N_T \), in which case the ideal measurement matrices are exactly achievable and the matrix approximation error can be zero. Since phase shifts of phase shifters are frequency-independent, it is straightforward to use the same phase codebook as the narrowband scenario in (17). Denote the initial phase codebook as \( \mathbf{A}' \), we set \( \mathbf{A}' = \mathbf{A} \). Then, TTD lines are exploited to compensate for frequency-dependent phase differences between \( \mathbf{A}' \) and \( \mathbf{A}_k^{\text{ideal}}, \forall k \). Considering the \( g \)-th atom, \( n_T \)-th antenna, and \( k \)-th subcarrier, we should let

\[
\begin{align*}
e^{-j2\pi (f_k - f_e) T(n_{TTD}, g)} = \mathbf{A}_k^{\text{ideal}}(n_T, g) / \mathbf{A}_k(n_T, g) \\
e^{-j\pi(nT-1)(f_k - f_e)\phi_{\mathbf{A}'}/f_c}, \quad (25)
\end{align*}
\]

where a bias of \( -f_e \) is added to all subcarriers’ frequencies according to **Lemma 1** and \( T(n_{TTD}, g) \) denotes the unconstrained delay of the \( n_{TTD} \)-th TTD line. Solving (25), we get the initial delay codebook as

\[
\mathbf{T}'(n_T, g) = \frac{(n_T - 1)\phi_{\mathbf{A}'} / 2f_c}{g}, \forall n_T, g. 
\]

(26)

However, the delay in (26) is negative when \( \phi_{\mathbf{A}'} < 0 \). To meet the non-negative constraint of delays, **Lemma 2** is used to obtain the final delay codebook as

\[
\mathbf{T}(n_{TTD}, g) = \begin{cases} 
\mathbf{T}'(n_T, g) + t_{\max}, & 0 \leq g \leq G/2, \\
\mathbf{T}'(n_T, g), & G/2 < g \leq G.
\end{cases}
\]

(27)

Next, we consider the practical case where \( N_{TTD} \ll N_T \) and the ideal measurement matrices can only be realized approximately. To minimize the approximation error, the computed delay of the middle antenna of each subarray is shared by all antennas in the subarray such that

\[
\begin{align*}
\mathbf{A}^{\prime}(n_{TTD}, g) & = \mathbf{A}^{\prime}(n_{TTD} - 1, g) + t_{\max}, & 0 \leq g \leq G/2, \\
\mathbf{A}^{\prime}(n_{TTD} - 1, g) & = \frac{2f_c}{n_{TTD} - 1} / \phi_{\mathbf{A}^{\prime}}, & G/2 < g \leq G.
\end{align*}
\]

(28)

where \( n_{TTD}' = (2Mn_{TTD} - M)/2 + 1 \) denotes the index of the \( n_{TTD} \)-th subarray’s middle antenna in the entire antenna array. When \( N_{TTD} = N_T, M = 1, \) (28) reduces to (27).

Once the delay codebook is determined, the initial phase codebook needs to be modified accordingly based on **Lemma 1** to compensate for the impact of the previously added bias to all subcarriers’ frequencies. With slight notation abuse, we still use \( \mathbf{A} \) to denote the final phase codebook in wideband DPP, which can be expressed as

\[
\mathbf{A} = \mathbf{A} \odot (\triangle \mathbf{A} \otimes \mathbf{1}_M),
\]

(29)

where \( \triangle \mathbf{A}^{\prime}(n_{TTD}, g) = e^{j2\pi f_k T(n_{TTD}, g)}, \forall n_{TTD}, g \). Eventually, the feasible measurement matrices can be computed by

\[
\mathbf{A}_k = \mathbf{A} \odot (\mathbf{T}_k \otimes \mathbf{1}_M), \forall k,
\]

(30)

where \( \mathbf{T}_k(n_{TTD}, g) = e^{-j2\pi f_k T(n_{TTD}, g)} \) is proportional to the true subcarrier frequency. Using \( \mathbf{A}_k, \forall k \), \( \mathcal{P}_1 \) can be converted to the following MMV-CS problem \( \mathcal{P}_5 \), where the constraint (32) is based on the fact that the sparse supports at all
subcarriers are basically aligned even with beam split. \[ \mathcal{P}_5 : \min_{D_k} \sum_{k=1}^{K} \left\| F_k - \tilde{A}_k D_k \right\|_F^2, \] \[ \text{s.t.} \quad \left\| \tilde{A}_k D_k \right\|_F^2 = N_k, \forall k, \] \[ = N_k, \forall k, \] (31) (32) (33)

D. DPP Algorithm

Given the above measurement matrices and codebooks, we extend Algorithm 1 to Algorithm 2 to solve \( \mathcal{P}_5 \), which is named as Extended SSP (E-SSP). Compared to SSP, E-SSP has similar basic process while several major differences are emphasized as follows. First, the projections need to be computed at all subcarriers, so both the number of measurements and the number of sparse vectors are increased by a factor of \( K \). Second, the summed projections of streams for atom selection require further averaging over all subcarriers. Last but not least, finer digital precoders are computed given the analog precoders. The previously used LS operation in SSP is optimal only in terms of matrix decomposition. To achieve our ultimate goal of sum rate maximization, SVD is executed on the equivalent channel of each subcarrier. The normalization term, \( \left( A_k^H A_k \right)^{-\frac{1}{2}} \), in lines 18 – 19 can transfer the power constraint from the hybrid precoder to the digital precoder [20], thus facilitating the calculation of the water-filling based optimal power allocation vector, \( p_{eq,k} \), in line 19.

**Algorithm 2** E-SSP Based Wideband DPP

**Input:** \( \tilde{A}, \tilde{T}, \tilde{A}_k, F_k, H_k, \forall k, N_{RF}, G, K; \)

1. Initialize \( A^{(0)}, T^{(0)}, A_k^{(0)}, \forall k \) as empty matrices, initialize \( F_{res,k}^{(0)} = F_k, \forall k; \)
2. for \( i = 1 : 1 : N_{RF} \) do
3. \( \psi^{(i)} = 0_G; \)
4. for \( k = 1 : 1 : K \) do
5. \( \psi^{(i)}_k = A_k F_{res,k}^{(i-1)}; \)
6. \( \psi^{(i)} = \psi^{(i)} + \psi^{(i)}_k; \)
7. end for
8. \( g^{(i)} = \arg\max \psi^{(i)}(j); \)
9. \( A^{(i)} = A^{(i-1)}[A(\cdot, g^{(i)})]; \)
10. \( T^{(i)} = T^{(i-1)}T(\cdot, g^{(i)}); \)
11. \( A_k^{(i)} = A_k^{(i-1)}[A_k(\cdot, g^{(i)})], \forall k; \)
12. for \( k = 1 : 1 : K \) do
13. \( D_k^{(i)} = A_k^{(i)} F_k; \)
14. \( F_{res,k}^{(i)} = \frac{F_k - A_k^{(i)} D_k^{(i)}}{\left\| F_k - A_k^{(i)} D_k^{(i)} \right\|_F}; \)
15. end for
16. end for
17. \[ A = A^{(N_{RF})}, T = T^{(N_{RF})}, A_k = A_k^{(N_{RF})}, \forall k; \]
18. for \( k = 1 : 1 : K \) do
19. \( H_{eq,k} = H_k A_k \left( A_k^H A_k \right)^{-\frac{1}{2}} = U_{eq,k} \Sigma_{eq,k} V_{eq,k}^H; \)
20. \( D_k = \left( A_k^H A_k \right)^{-\frac{1}{2}} V_{eq,k}(\cdot, 1 : N_k) \text{diag}(p_{eq,k}); \)
21. end for

**Output:** \( A, T, D_k, \forall k; \)

IV. LOW-COMPLEXITY EXTENDED SSP BASED WIDEBAND DPP

The E-SSP algorithm proposed in the previous section can still have high complexity when the numbers of RF chains, subcarriers, and atoms are large. To enhance the practicality of the proposed approach, in this section, we reduce the complexity of E-SSP dramatically without sacrificing much performance using three efficient atom selection techniques.

A. Non-Iterative Atom Selection

The complexity of SSP is proportional to the number of iterations. In this subsection, we propose to discard the iterative process and select all atoms at once by exploiting the properties of projections in the angular domain.

In the original SOMP algorithm, the necessity of iteratively subtracting previously selected atoms’ contributions comes from the correlation among different atoms. Ignoring the small
matrix approximation error, the atoms of the ideal measurement matrices are in the form of ULA response vectors according to (19). It can be proved that two ULA response vectors corresponding to different angles are asymptotically orthogonal as the length of vectors tends to infinity [5]. Since the number of antennas is relatively large in practical THz massive MIMO systems, the correlation among atoms would be relatively small. Therefore, it is reasonable to directly select the top-$N_{RF}$ atoms based on the initial projections without requiring multiple iterations. Nevertheless, when the angular distance between two atoms is very small, e.g., smaller than the angular Rayleigh resolution of the array [6], their mutual correlation is still non-negligible. To exclude atoms whose large projection comes from leaked power of neighbouring atoms, we further narrow down candidate atoms to only peaks on the initial projection curve, which can be found by various peak-finding algorithms.

To promote intuitive understanding, one example of the initial projection curve and the atoms selected by different methods is given in Fig. 3. As we can see, directly selecting top-$N_{RF}$ atoms results in the selection of multiple nearby atoms with strong mutual correlation, as indicated by the pluses. In contrast, with the additional peak-finding based filtering, the same atoms as the original iterative method are selected, as indicated by the coincident crosses and circles.

![Fig. 3. An exemplary initial projection curve and the selected atoms with different methods. To make the figure clear, we set $\sigma_\theta = 0^\circ, G = 1024$. Other parameters see the default setting in simulation.](image)

**B. Hierarchical Atom Selection and Refinement**

Although more atoms can improve the performance of E-SSP thanks to higher angular resolution, the complexity grows as well. To enjoy the benefit of high angular resolution with low complexity, in this subsection, we borrow the hierarchical beam sweeping idea from codebook-based precoding approaches [28] to reduce the number of atoms that require projection computation.

Thanks to the continuity of the angles corresponding to atoms, nearby atoms have similar projections so that the projection curve is smooth in the angular domain. Therefore, the relative positions of peaks on the projection curve rarely change with $G$. For instance, if the 18-th atom is a peak when $G = 128$, it is very likely that there is also a peak around the 9-th and the 36-th atom when $G = 64$ and $G = 256$, respectively. Therefore, in the first stage, coarse-grained measurement matrices with only $G_c$ atoms ($G_c \ll G$), $\mathbf{A}_k^{\prime}, \forall k$, will be used to select $N_{RF}$ coarse-grained atoms, where the columns of each $\mathbf{A}_k^{\prime}$ are evenly sampled from the corresponding fine-grained $\mathbf{A}_k$ with $G$ atoms. Then, $2G_a$ fine-grained atoms will be added around each selected coarse-grained atom, to obtain new measurement matrices, $\mathbf{A}_k^{\prime}, \forall k$. In the second stage, atom selection procedures are executed once again using $\mathbf{A}_k^{\prime}, \forall k$, to determine the final fine-grained atoms. Using this two-stage hierarchical atom selection and refinement method, the number of atoms that require projection computation is at most $(G_c + 2N_{RF}G_a)$ while this number is $G$ if the fine-grained $\mathbf{A}_k, \forall k$, are directly used. Define the angular resolution reduction ratio as $\triangle G \triangleq G/G_c$, we will see that the performance degradation is marginal with a large $\triangle G$ and a small $G_a$ under typical system settings in simulation, thus reducing the complexity dramatically.

**C. Partial Subcarrier Exploitation**

Since the number of subcarriers is usually large in practical wideband THz systems due to huge bandwidth, the complexity of E-SSP would be pretty high if projections at all subcarriers need to be computed. In this subsection, we propose to only exploit partial subcarriers’ information for atom selection to reduce the complexity similar to [38] which considers wideband channel estimation.

When multiple sparse vectors have the perfect common sparsity structure, averaging the projections of their measurements can reduce the equivalent noise, thus improving performance. However, this benefit would saturate when the equivalent noise is low enough and no longer a performance bottleneck. For instance, the bottleneck could be the unrecovable information loss in the imperfect fully-digital precoders in this case. Observing the CS model in (31), we find that the measurements, i.e., the fully-digital precoders at different subcarriers, are computed based on different subchannels. Therefore, the source of measurement noise is actually channel error, which can be caused by either imperfect estimation or feedback in practice. With lower channel error, fewer subcarriers are required to achieve a low enough equivalent noise level. In the extreme case where perfect CSI is available, only a single subcarrier is sufficient to achieve accurate atom selection with noiseless measurements. Nevertheless, the common sparsity structure among subcarriers is still imperfect due to finite angular resolution of atoms and the approximation error between the ideal and feasible measurement matrices. As result, multiple measurements are still beneficial with perfect CSI while fewer measurements are required compared to the imperfect CSI case, as will be shown in simulation.

Denote the number of exploited subcarriers as $K^\prime$, which can be evenly sampled from all subcarriers, the complexity of atom selection can be reduced by $\Delta K$ times, where $\Delta K \triangleq K/K^\prime$ denotes the subcarrier reduction ratio. After the atoms are selected based on only $K^\prime$ subcarriers’ fully-digital precoders, the analog precoders at all $K$ subcarriers are determined while the digital precoders at all $K$ subcarriers can still be calculated in the same way. Notice that this technique
is tailored for E-SSP and will lead to severe performance degradation if applied to other DPP algorithms, such as the AM algorithm proposed in [26]. The reason is that without utilizing the highly structured codebooks, the optimized phases and delays that have good performance at a few sampled subcarriers are not necessarily good at the remaining subcarriers.

D. Low-Complexity DPP Algorithm

Overall, with the above three efficient atom selection techniques, the LCE-SSP algorithm for wideband DPP is given in Algorithm 3. For compactness, several functions are used in the algorithm. Specifically, the PeakFinder(·, g, N\text{peak}) function used in line 8 and line 20 returns the indexes of top-\(N\text{peak}\) peaks given index vector \(g\). In line 13, the IndexCleaner(·, a, b) function cleans the index set by deleting duplicated elements and limiting the range of elements to \([a, b]\) while the len(·) function returns the length of the input vector.

E. Baseline DPP Algorithms and Complexity Analysis

In this subsection, several major baseline DPP algorithms are introduced and the complexity of different DPP algorithms is analyzed and compared, as listed in Table I. Only complexity-dominant operations are counted to make the \(O\) complexity expressions concise. To be fair, the digital precoders in all baseline algorithms are calculated in the same way as E-SSP and LCE-SSP, so this part of complexity is the same in all algorithms. The complexity difference mainly comes from different ways of obtaining the analog precoders.

For all matrix decomposition based DPP algorithms, we first need \(K\) matrix SVDs to obtain the fully-digital precoders. Then, in E-SSP demonstrated in Algorithm 2, \(3K\) matrix multiplications (line 5, 12, 13) and \(K\) Moore-Penrose matrix inverses (line 12) are executed in each of the \(N_{RF}\) iterations of atom selection. Finally, \(2K\) matrix multiplications are needed to calculate digital precoders (line 18). Notice that the complexity of codebook optimization is not included since it only occurs once offline for each system setting.

Compared to E-SSP, the number of subcarriers involved in the calculation of fully-digital precoders in LCE-SSP reduces from \(K\) to \(K'\). Besides, the numbers of iterations, subcarriers, and atoms involved for atom selection in Algorithm 3 reduce from \(N_{RF}, K, G\) to 1, \(K'\), and at most \((G_c + 2N_{RF}G_0)\), respectively.

In the AM algorithm proposed in [26], the phases, delays, and digital precoders are updated alternately when the other two are fixed in each iteration. The number of iterations is denoted by \(N_{iter}\) and the update of delay parameters is based on the one-dimensional search among \(S\) discrete delay grids from 0 to \(t'_{\text{max}}\). Besides, since AM’s performance also depends on the initializer, the initialization algorithm’s complexity denoted by \(O_{\text{Init}}\) also needs to be included in its overall complexity.

In the heuristic DPP algorithm proposed in [25], fully-digital precoders are not needed since the analog precoders are computed based on angles of channel paths with very low complexity. Nevertheless, this algorithm cannot be directly applied to cluster channels where there are far more subpaths than RF chains. Simply using the mean angles of clusters (if assumed available) leads to poor performance. To include heuristic DPP as a meaningful baseline algorithm and enrich

| Algorithm                        | Complexity                                                                 |
|----------------------------------|-----------------------------------------------------------------------------|
| E-SSP                            | \(O(KN_TN_K^2 + KN_TN_{RF}(N_{RF}^2 + N_{RF}N_s + N_sG + N_{RF} + N_R))\) |
| LCE-SSP                          | \(O(K'N_TN_K^2 + KN_TN_{RF}(2N_{RF} + N_R + N_s) + K'N_TN_s(G_c + 2N_{RF}G_0))\) |
| AM                               | \(O_{\text{Init}} + O(KN_TN_K^2 + KN_TN_{RF}(N_{iter}(N_{RF} + N_s + S/M) + N_{RF} + N_R))\) |
| Heuristic DPP                    | \(O_{FA} + O(KN_TN_{RF}(N_{RF} + N_R))\)                                    |

Algorithm 3 LCE-SSP Based Wideband DPP

Input: \(A, T, \tilde{A}_k, F_k, H_k, \forall k, N_{RF}, G, G_0, \Delta G, K', \Delta K;\)

1. \(g^c = 1 : \Delta G : 1 + (G_c - 1)\Delta G, \psi^c = 0_{G_c};\)
2. for \(k' = 1 : \Delta K : 1 + (K' - 1)\Delta K\) do
3. \(\tilde{A}_{k'}^H = A_{k'}^H(\cdot, g^c);\)
4. \(\Psi_{k'}^H = A_{k'}^H F_{k'};\)
5. \(\psi_{k'}^H = \text{diag}(\Psi_{k'}^H, \Psi_{k'}^{H');}\)
6. \(\psi_{k'}^H = \psi^c + \psi_{k'}^c;\)
7. end for
8. \(g^f = \text{PeakFinder}(\psi^f, 1 : 1 : G_c, N_{RF});\)
9. \(g^f = \emptyset;\)
10. for \(i = 1 : 1 : N_{RF}\) do
11. \(g^f = g^f[\Delta G \times g^f(i) - G_0 : 1 : \Delta G \times g^f(i) + G_0];\)
12. end for
13. \(g^f = \text{IndexCleaner}(g^f, 1, G);\)
14. for \(k' = 1 : \Delta K : 1 + (K' - 1)\Delta K\) do
15. \(\tilde{A}_{k'}^H = A_{k'}^H(\cdot, g^f);\)
16. \(\Psi_{k'}^H = A_{k'}^H F_{k'};\)
17. \(\psi_{k'}^H = \text{diag}(\Psi_{k'}^H, \Psi_{k'}^{H');}\)
18. \(\psi_{k'}^H = \psi_{k'}^H + \psi_{k'}^{H';}\)
19. end for
20. \(q = \text{PeakFinder}(\psi^f, g^f, N_{RF});\)
21. \(A = \tilde{A}(\cdot, g^f), T = \tilde{T}(\cdot, g^f), \tilde{A}_k = \tilde{A}_k(\cdot, g^f), \forall k;\)
22. for \(k = 1 : 1 : K\) do
23. \(H_{eq,k} = H_{k}A_k(\tilde{A}_{k}^H A_k)^{-\frac{1}{2}} = U_{eq,k}\Sigma_{eq,k}V_{eq,k}^H;\)
24. \(D_k = \left(A_{k}^H A_k\right)^{-\frac{1}{2}} V_{eq,k}(\cdot, 1 : N_s)\text{diag}(p_{eq,k});\)
25. end for

Output: \(A, T, D_k, \forall k;\)
the comparison, we propose to find representative angles for it using the LCE-SSP algorithm with slight modifications. Specifically, we set $N_s$ to $N_R$ and replace the fully-digital precoders with channels in Algorithm 3. Besides, ideal rather than feasible measurement matrices can be used since we are doing pure signal processing. In this way, $N_{RF}$ representative angles can be found such that the constructed pseudo path channels are close to the original cluster channels. We remark that such an extension of heuristic DPP to cluster channels is not trivial since the proposed algorithm is still used. The complexity of finding angles denoted by $O_{\text{F}}$ also needs to be included in the overall complexity of heuristic DPP.

With typical values of system parameters and algorithm hyperparameters, the complexities of E-SSP and AM are comparable while the complexities of LCE-SSP and heuristic DPP are much lower. Intuitively, the average running time of LCE-SSP (several tens of milliseconds) can be dozens or even hundreds of times shorter than that of AM (several seconds) using the same programming language and CPU.

We further remark that in FDD systems, the proposed E-SSP and LCE-SSP also have much lower feedback overhead [9] than the AM algorithm. Thanks to the predetermined measurement matrices and codebooks, only $N_{RF}$ atom indexes are required to configure the hardware to construct the structured analog precoders at the BS, where each index needs $\log_2 G$ bits to represent. In contrast, in the AM algorithm, each of the $N_{RF} N_{TTD}$ delay parameters needs $\log_2 S$ bits to represent while the $N_{RF} N_T$ phase parameters are continuous (or quantized in practice) between 0 and $2\pi$, thus resulting in a much higher feedback overhead.

V. SIMULATION RESULTS

In this section, extensive numerical results\textsuperscript{6} are provided to validate the superiority of the proposed approach in terms of performance, complexity, and robustness. Unless specified, the following typical system and channel parameters [26], [32] will be used as the default setting in simulation: $N_T = 256$, $N_R = N_{RF} = N_s = 4$, $N_{TTD} = 16$, $f_c = 100 \text{ GHz}$, $f_s = 10 \text{ GHz}$, $K = 128$, $N_c = 4$, $N_p = 10$, $\tau_{max} = 20 \text{ ns}$, $\sigma_{\tau_i} = 0.06 \text{ ns}$, $\sigma_{\theta_i} = 5^\circ$, $\forall i$, $\text{SNR} = 10 \text{ dB}$. The achievable data rate per subcarrier and the matrix approximation mean-squared error (MSE) defined as $\frac{1}{N_T N_s} \sum_{k=1}^{K} \| F_k - A_k D_k \|_F^2$ are used as performance metrics.

In E-SSP and LCE-SSP, hyperparameters with the best performance-complexity tradeoffs are searched. First, we gradually increase $G$ until the performance of E-SSP saturates. Then, using a large $G_o$, we gradually decrease $G_c$ until the performance degradation of LCE-SSP is non-negligible. Then, the same process is repeated to determine the smallest $G_o$ and $K'$. In this way, we get $G = 1024$, $G_c = 256$, $G_a = 8$, $K' = 4$. For heuristic DPP, mean angles of clusters can be used. Or, the representative angles can be found by CS algorithms such as the proposed LCE-SSP, as elaborated in

\textsuperscript{6}The source code for reproduction is available at https://github.com/EricGJB/Extended_SSP_Wideband_DPP.

Section IV-D. For the AM algorithm, we set $S = 1024$ and $t'_{\text{max}} = N_T / f_c$ as [26], while $N_{\text{iter}}$ is set to 20 and 30 in the single-user case and the multi-user case, respectively. All the simulation results are obtained by averaging over 200 Monte Carlo channel realizations.

Fig. 4 illustrates the achievable rates of different DPP algorithms, as well as the state-of-the-art AM-based phase-shifting-only hybrid precoding algorithm [26], versus SNR. First of all, we can see that the performance of various DPP algorithm is better than the phase-shifting-only algorithm, illustrating the superiority and necessity of using TTD lines in the face of beam split. Among DPP algorithms, the heuristic DPP algorithm with angles found by the CS algorithm performs much better than simply using mean angles of clusters thanks to our extension. Nevertheless, it is still inferior to matrix approximation based algorithms since they are not restricted to physical path angles but exploit the fully-digital precoders with richer information to guide the design of hybrid precoders. LCE-SSP has almost the same performance as E-SSP in various SNR regimes, demonstrating the effectiveness of the proposed low-complexity techniques. In the AM algorithm, since the analog precoders do not have to be in specific forms, the generated beams would have larger degrees of freedom to flexibly cover each channel sample, thus resulting in higher potentially achievable performance. However, the practical application of AM is hindered by its high complexity. On the other hand, the actual performance of AM is also dependent on the initializer, whose role is to construct initial analog precoders that are close to the optimal analog precoders to simplify the iterative optimization process and reduce the probability of falling into bad local optima. As we can see, LCE-SSP beats other initializers and helps AM achieve about 90% of the performance upper bound, therefore can be used as a cheap yet effective initializer for AM in practical cluster THz channels.

Fig. 5 illustrates the impacts of channel cluster spreads, including the angular spread and the delay spread. As we can see in Fig. 5(a), when the angular spread is small, LCE-SSP alone can outperform AM with the random initializer. The simultaneous good performance and low complexity makes LCE-SSP very appealing to such channel conditions. With the growth of angular spread, the performance of E-SSP and
LCE-SSP gradually decreases since the fully-digital precoders for more complicated channels with larger angular spreads are less likely to be well approximated by only a few atoms in the predetermined measurement matrices. Nevertheless, they are still indispensable to guarantee the high performance of AM as good initializers. At last, without exploiting the fully-digital precoders, heuristic DPP is always inferior to LCE-SSP even with CS angles, either applied alone or as AM’s initializer. Then, according to Fig. 5(b), all algorithms’ performance decreases as the delay spread increases since the difference among subchannels is larger, while the narrower performance gap between LCE-SSP and AM alleviates its performance disadvantage when used alone.

Fig. 6 illustrates the impact of the number of channel clusters $N_c$. The performance of the heuristic DPP algorithm and the (LC)E-SSP algorithm gradually increases with the number of clusters thanks to the spatial multiplexing gain and saturates when the number of clusters exceeds the number of RF chains. For the AM algorithm, however, the performance plateau does not appear until $N_c = 6$ since one RF chain does not necessarily end in exploiting only one cluster, demonstrating its broad applicability. The narrow gap between LCE-SSP and the AM algorithm in small $N_c$ regimes also reveals that it is more favorable to apply LCE-SSP in sparser channels, e.g., the THz channels.

To further promote understanding of the importance of beam split compensation, Fig. 7 illustrates the achievable rates with different remaining beam split levels, which are determined by both the number of TTD lines and the fractional bandwidth defined as $f_s/f_c$. On the one hand, with larger fractional bandwidth, the initial beam split level is higher, as indicated by different heights of the starting points of the curves. On the other hand, with more TTD lines, beam split can be compensated better, as indicated by the upward trend of the curves. The performance will not improve further when beam split is compensated thoroughly while the minimal number of TTD lines required to achieve this is proportional to the fractional bandwidth [20]. For instance, the performance of LCE-SSP saturates with 32 TTD lines when the fractional bandwidth is 0.1 while only 8 TTD lines are required when the fractional bandwidth is 0.025. In contrast, the curve is still rising at $N_{TTD} = 32$ when the fractional bandwidth is 0.4.

Next, we show the impact of initializers on AM’s convergence performance and speed in Fig. 8. According to Fig. 8(a), both the convergence performance and speed of AM initialized by LCE-SSP are better than that initialized by the other two initializers in two different system scales, indicating its superiority. Notice that, the MSEs will first increase at the first iteration, then monotonically decrease to...
Fig. 8. Impacts of initialization methods on the convergence speed and performance of AM. In subfigure (a), the solid curves and the dashed curves denote $N_T = 128, N_{TTD} = 8, N_R = N_{RF} = N_s = 2$ and $N_T = 256, N_{TTD} = 16, N_R = N_{RF} = N_s = 4$, respectively. In subfigure (b), the solid curves and the dashed curves denote convergence performance and speed, respectively.

Fig. 9. Impacts of channel error on achievable rate and the selection of $K'$ in LCE-SSP. In subfigure (a), the solid and dashed curves denote the performance with and without channel error, respectively. The converged values that are lower than the initial values due to different discretized time delays in the initialization algorithms and the AM algorithm. Same conclusions can be drawn from Fig. 8(b), where the early exit strategy is used to accelerate convergence. Define the exit threshold of matrix approximation MSE in AM as $\epsilon$, we can see that $\epsilon = 10^{-6}$ is a good choice to decrease the number of iterations without sacrificing performance. When $\epsilon = 10^{-6}$, the average number of iterations of AM with the LCE-SSP initializer (11) is smaller than that with the heuristic DPP initializer (15) or with the random initializer (18).

So far we have been assuming perfect CSI is available. However, channel estimation or feedback error is inevitable in practice. So, we investigate the impact of channel error, which is defined as $\mathbb{E} \left( \frac{1}{N} \sum_{k=1}^{K} \frac{\|H_k - \hat{H}_k\|_F^2}{\|H_k\|_F^2} \right)$ with $\hat{H}_k$ denoting the $k$-th estimated subchannel. In Fig. 9. To generate imperfect CSI, we add Gaussian noise to the perfect CSI, whose variance can be adjusted to realize different simulation settings. From Fig. 9(a), the performance degradation speed of heuristic DPP and LCE-SSP is slower than AM as channel error increases thanks to the restriction on analog precoders’ structures. Specifically, the performance of AM starts to drop quickly when channel error is larger than $-15$ dB while the performance of the other two algorithms keeps almost unchanged until the channel error is larger than $-5$ dB. When the channel error is 0 dB, LCE-SSP has similar performance as AM with the heuristic DPP initializer. Such strong robustness to imperfect CSI makes LCE-SSP appealing in challenging scenarios with large noises and limited resources during channel estimation and feedback. Initialized by LCE-SSP, AM still has the best performance thanks to the robust initialization point. As elaborated in Section IV-C, the selection of $K'$ depends on the channel error level. From Fig. 9(b), we can observe that 4 subcarriers are enough to achieve the saturated performance with perfect CSI, while this number increases to 32 when the channel error is 0 dB. With a proper selection of $K'$, the complexity of LCE-SSP is reduced dramatically without performance degradation compared to E-SSP.

The extension from the single-user scenario to the multi-user scenario is straightforward in matrix decomposition based DPP algorithms, as long as the fully-digital precoders are available [26]. For the calculation of the fully-digital precoders and the digital precoders at the end, we perform zero-forcing (ZF) with water-filling based power allocation on...
the original channels and the equivalent channels, respectively. For heuristic DPP, $N_{RF}/N_U$ representative angles are contributed by each user’s channels to make up all $N_{RF}$ angles, where $N_U$ denotes the number of users. We consider each user has a single antenna and one data stream as in [26]. Two RF chain configurations are investigated, where $N_{RF}$ is set to $N_U$ to guarantee that each user is served in the first configuration and set to $N_U N_c$ to thoroughly exploit the spatial multiplexing gain in the second configuration. The performance when $N_U = 4$ is illustrated in Fig. 10. We can see that among all candidate initializers, LCE-SSP always leads to the best performance of AM. Specifically, it achieves 84% performance of fully-digital precoding in the first configuration and this ratio increases to 95% in the second configuration thanks to sufficient RF chains. Apart from better performance, the LCE-SSP initializer also leads to faster convergence of AM. For instance, when $N_{RF} = N_U, \epsilon = 10^{-6}$, and SNR = 10 dB, the average number of iterations with the LCE-SSP initializer is about half of that with the random initializer, indicating twice the speedup. In practice, the proposed LCE-SSP can be flexibly applied in different modes according to different requirements and RF chain configurations. When high performance is required and RF chains are limited, LCE-SSP can serve as a cheap yet effective initializer of AM. When low complexity is required and RF chains are sufficient, LCE-SSP alone can already achieve satisfactory performance. We also notice that the advantage of LCE-SSP over heuristic DPP is smaller than the single-user scenario since ZF is not the optimal fully-digital precoder in the multi-user scenario. Nevertheless, the good performance of heuristic DPP is still based on the representative angles found by LCE-SSP.

VI. CONCLUSION

In this paper, we have proposed an extended SSP based DPP algorithm for THz massive MIMO systems. By exploiting the precoder structures, the matrix decomposition problem is converted to an MMV-CS problem, which can be readily solved by the slightly modified SOMP algorithm. To compensate for beam split, ideal frequency-dependent measurement matrices are designed for the alignment of sparse supports at all subcarriers, which can be approximately realized by optimized feasible phase and delay codebooks. To further reduce complexity, several efficient atom selection techniques are developed without sacrificing much performance. According to simulation results, by applied alone or in combination with existing DPP approaches, the proposed approach demonstrates superiority in performance, complexity, and robustness over existing approaches, thus is promising in future 6G THz massive MIMO systems. In the future, we will consider to extend the proposed approaches to the near-field scenario, where the array response vector is related to not only the channel path angle, but also the distance to the scatter.

APPENDIX

A. Proof of Lemma 1

Consider a TTD line connected to an RF chain has delay $t$, and the $m$-th phase shifter connected to it has phase shift $e^{j \theta_m}, \forall m$, then the superimposed phase of the TTD line and the $m$-th phase shifter at the $k$-th subcarrier is $e^{j (\theta_m - 2 \pi f_k t)}, \forall m, k$. If we add a bias of $- \Delta f$ to all subcarriers’ frequencies, the TTD line’s phase at the $k$-th subcarrier will change from $e^{-j 2 \pi f_k t}$ to $e^{-j 2 \pi (f_k - \Delta f) t}$ to keep the superimposed phases unchanged, we need to modify the phase shifters’ phase shifts to $e^{j (\theta_m - 2 \pi f_k \Delta f)} \forall m$. The reasoning process holds for any valid $\Delta f$, thus proving Lemma 1.

B. Proof of Lemma 2

Let us first consider a specific subcarrier and discard the subscript $k$. Denote the $N_{RF} \times 1$ vector of delay biases added to all TTD lines connected to $N_{RF}$ RF chains as $\Delta t$, and denote the original analog precoder and the analog precoder with delay biases added as $A$ and $A'$, respectively, we have $A' = A \Sigma$ and $A = A' \Sigma^{-1}$, where $\Sigma = \text{diag} (e^{-j 2 \pi f_k \Delta t})$ is an invertible diagonal matrix. Based on the properties of Moore-Penrose inverse such that $AA' = A$ and $(AB)' = B', A'$, we have

$$AA' = A(A' \Sigma^{-1})' = A \Sigma A'^\dagger A = A.$$  \hspace{1cm} (34)

Right-multiply both sides of (34) by matrix $\Sigma A'^\dagger$, we have

$$A \Sigma A'^\dagger (A \Sigma) A'^\dagger = (A \Sigma) A'^\dagger,$$  \hspace{1cm} (35)

$$A \Sigma (A'^\dagger A' A'^\dagger) = A' A'^\dagger,$$  \hspace{1cm} (36)

$$A (\Sigma A'^\dagger) = A' A'^\dagger.$$  \hspace{1cm} (37)
where the property $A^{\dagger}A^{\dagger} = A^{\dagger}$, is used to obtain (37). Eventually, based on $A^{\dagger}A^{\dagger} = A^{\dagger}$, we prove that $AA^{\dagger} = A^{\dagger}A^{\dagger}$ holds for any valid $A$, for the $k$-th subcarrier, we have $A_{k}D_{k} = A_{k}^\dagger F_{k} = A_{k}^\dagger A_{k}^\dagger F_{k} = A_{k}^\dagger D_{k}$, i.e., $A_{k}D_{k}$ does not change with $\Delta t$. Therefore, the sum rating $R = \sum_{k=1}^{K} \log_2 \left( I_{N_{t}} + \frac{\rho}{N_{t}} H_{k} A_{k} D_{k} D_{k}^H A_{k}^H H_{k}^H \right)$ also won’t change. The reasoning process holds for any valid $\Delta t$, thus proving Lemma 2.

REFERENCES

[1] X. You et al., “Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shift,” Sci. China Inf. Sci., vol. 64, pp. 1–77, Nov. 2020.

[2] I. F. Akyildiz, J. M. Jornet, and C. Han, “Terahertz band: Next frontier for wireless communications,” Phys. Commun., vol. 12, pp. 16–32, Sep. 2014.

[3] H. Song and T. Nagatsuma, “Present and future of terahertz communications,” IEEE Trans. Terahertz Sci. Technol., vol. 1, no. 1, pp. 256–263, Sep. 2011.

[4] I. F. Akyildiz, C. Han, and S. Nie, “Combating the distance problem in the millimeter wave and terahertz frequency bands,” IEEE Commun. Mag., vol. 56, no. 6, pp. 102–108, Jun. 2018.

[5] L. Lu, G. Y. Li, and A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 742–758, Oct. 2014.

[6] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[7] F. Sohrabi and W. Yu, “Hybrid digital and analog beamforming design for large-scale antenna arrays,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 501–513, Apr. 2016.

[8] E. Zhang and C. Huang, “On achieving optimal rate of digital precoder by RF-baseband codesign for MIMO systems,” in Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall), Sep. 2014, pp. 1–5.

[9] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” IEEE Trans. Wireless Commun., vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[10] C. Chen, C. Tsai, Y. Liu, W. Hung, and A. Wu, “Compressive sensing (CS) assisted low-complexity beamspace hybrid precoding for millimeter-wave MIMO systems,” IEEE Trans. Signal Process., vol. 65, no. 6, pp. 1412–1424, Mar. 2017.

[11] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, “Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 485–500, Apr. 2016.

[12] A. Kaushik, J. Thompson, E. Vlachos, C. Tsinos, and S. Chatzinotas, “Dynamic RF chain selection for energy efficient and low complexity hybrid beamforming in millimeter wave MIMO systems,” IEEE Trans. Green Commun. Netw., vol. 3, no. 4, pp. 886–900, Dec. 2019.

[13] A. Kaushik, E. Vlachos, C. Tsinos, J. Thompson, and S. Chatzinotas, “Joint bit allocation and hybrid beamforming optimization for energy efficient millimeter wave MIMO systems,” IEEE Trans. Green Commun. Netw., vol. 5, no. 1, pp. 119–132, Mar. 2021.

[14] W. Mu, C. Qi, Z. Zhang, and J. Cheng, “Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO,” IEEE Trans. Commun., vol. 68, no. 5, pp. 2838–2849, May 2020.

[15] A. M. Elbir and A. K. Papazafeiropoulos, “Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach,” IEEE Trans. Veh. Technol., vol. 69, no. 1, pp. 552–563, Jan. 2020.

[16] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, “Codebook-based hybrid precoding for millimeter wave multiuser systems,” IEEE Trans. Ph. Signal Process., vol. 65, no. 20, pp. 5289–5304, Oct. 2017.

[17] Y. Chen, D. Chen, and T. Jiang, “Non-uniform quantization codebook-based hybrid precoding to reduce feedback overhead in millimeter wave MIMO systems,” IEEE Trans. Commun., vol. 67, no. 4, pp. 2779–2791, Apr. 2019.

[18] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, “Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems,” IEEE Trans. Signal Process., vol. 66, no. 13, pp. 3393–3406, Jul. 2018.
Caijun Zhong (Senior Member, IEEE) received the B.S. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2004, and the M.S. degree in information security and the Ph.D. degree in telecommunications from University College London, London, U.K., in 2006 and 2010, respectively. From September 2009 to September 2011, he was a Research Fellow with the Institute for Electronics, Communications and Information Technologies (ECIT), Queen's University Belfast, Belfast, U.K. Since September 2011, he has been with Zhejiang University, Hangzhou, China, where he is currently a Professor. His current research interests include reconfigurable intelligent surfaces assisted communications and artificial intelligence-based wireless communications. He was a recipient of the 2013 IEEE ComSoc Asia–Pacific Outstanding Young Researcher Award. He and his coauthors have been awarded the Best Paper Award at the 2020 IEEE GLOBECOM, IEEE ICC in 2019, and IEEE ICC in 2022. He was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS. He is an Editor of Science China Information Science and China Communications.

Geoffrey Ye Li (Fellow, IEEE) is currently a Chair Professor with Imperial College London, U.K. Before joining Imperial College London in 2020, he was a Professor with the Georgia Institute of Technology, USA, for 20 years and a Principal Technical Staff Member with AT&T Labs Research (previous Bell Labs), NJ, USA, for five years. His current research interests include orthogonal frequency division multiplexing (OFDM) for wireless communications, establishing a framework for resource cooperation in wireless networks, and introducing deep learning (DL) for communications. In these areas, he has published more than 600 journals and conference papers in addition to more than 40 granted patents. His publications have been cited more than 60,000 times with an H-index of 114. He was awarded IET Fellow for his contributions to signal processing for wireless communications. He has been recognized as the World Most Influential Scientific Mind also known as a Highly Cited Researcher by Clarivate. He won the 2024 IEEE Eric E. Sumner Award and several prestigious awards from the IEEE Signal Processing Society, the IEEE Vehicular Technology Society, and the IEEE Communications Society, including the 2019 IEEE ComSoc Edwin Howard Armstrong Achievement Award.

Joseph B. Soriaga received the B.S. degree in electrical engineering from Caltech and the Ph.D. degree in electronics and communication engineering from UCSD, with a focus on coding theory, communications, and information theory. He is currently the Senior Director of technology and the Lead of the Systems Team, Corporate Research and Development, Qualcomm AI Research. The team is broadly responsible for driving hardware and software platform innovations, advancing applications in areas, such as multimedia and wireless, and contributing to fundamental research in machine learning. Prior to this role, he led the 5G Core Physical Layer Research Team, Qualcomm Wireless Research and Development, and helped drive the standardization of 5G as both a 3GPP RAN1 delegate and a Technical Specifications Editor of TS 38.202. He has also had previous experience as a Systems Team Lead at Qualcomm 3G EV-DO Commercial Cell-Site Modem (CSM). He has been granted more than 280 U.S. patents. He was awarded the IEEE Data Storage Best Paper in 2007 and the EURASIP Journal on Wireless Communications and Networking Best Paper in 2012.

Arash Behboodi (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2005 and 2007, respectively, the master's degree in philosophy from Pantheon-Sorbonne University in 2011, and the Ph.D. degree from Ecole Supérieure d’électricité (now CentraleSupélec), Gif-sur-Yvette, France, in 2012. He was a Senior Researcher with the Technical University of Berlin from 2013 to 2015 and RWTH Aachen University from 2015 to 2019. He is currently a Machine Learning Scientist and a Principal Engineer with Qualcomm AI Research, Qualcomm Technologies Netherlands B.V. His current research interests include machine learning design for wireless communication, learning theory, machine learning for inverse problems, information theory, compressed sensing, and mathematical signal processing.