Article

Channel Covariance Matrix Estimation via Dimension Reduction for Hybrid MIMO MmWave Communication Systems

Rui Hu [ID]**, Jun Tong *, Jiangtao Xi, Qinghua Guo and Yanguang Yu**

School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW 2522, Australia

* Correspondence: jtong@uow.edu.au

check for updates

Abstract: Hybrid massive MIMO structures with lower hardware complexity and power consumption have been considered as potential candidates for millimeter wave (mmWave) communications. Channel covariance information can be used for designing transmitter precoders, receiver combiners, channel estimators, etc. However, hybrid structures allow only a lower-dimensional signal to be observed, which adds difficulties for channel covariance matrix estimation. In this paper, we formulate the channel covariance estimation as a structured low-rank matrix sensing problem via Kronecker product expansion and use a low-complexity algorithm to solve this problem. Numerical results with uniform linear arrays (ULA) and uniform squared planar arrays (USPA) are provided to demonstrate the effectiveness of our proposed method.

1. Introduction

Millimeter wave (mmWave) communications are promising for future-generation wireless communications for their advantages such as large bandwidths, narrow beams, and secure transmissions [1,2]. Large-scale multiple-input multiple-output (MIMO) hybrid structures equipped with only a few RF chains have generated great interests for mmWave systems due to their low complexity and near-optimal performance [3,4]. Precoders and combiners must be carefully designed to exploit the potential of large-scale MIMO hybrid systems, e.g., to achieve high data transmission rates. They can be designed based on the instantaneous channel matrix [5,6], which may be estimated by using channel estimation techniques [3,7]. However, the instantaneous channel can vary fast [8], especially at mmWave frequencies [9,10], and the precoder/combiner has to be redesigned once the instantaneous channel changes [5,6].

Although the instantaneous mmWave channel can change very fast, the long-term channel statistics, e.g., the angular power spectrum, can be stationary for tens to hundreds of coherence blocks [9]. Recently, the channel covariance information has been utilized to design the analog precoders/combiners [9,11], which remain fixed when the covariance matrix is unchanged. The effective digital system has a reduced dimensionality, which greatly reduces the cost of acquiring the instantaneous channel state information (CSI) and simplifies the optimization of the digital precoders and combiners. The channel covariance matrix should be firstly estimated to realize the designs in [9,11]. With large antenna arrays, the channel covariance matrix has a large dimensionality, which demands a large number of observations to be used when traditional covariance matrix estimators are adopted. Meanwhile, the hybrid structure only allows a reduced number of observations to be acquired at the receiver, which makes the task of channel covariance estimation challenging.

In order to address this challenge, [10,12] propose several compressive sensing (CS) based channel covariance estimators, which explore the relations between the angle of departure (AoD)/angle of arrival (AoA) and the channel covariance matrix. Their methods need a dictionary for searching the AoD/AoA, and the resulting performance improves when the resolution of the dictionary becomes higher. However, high-resolution dictionary yields high computational complexity. Moreover, these CS-based estimators require the number of paths in the channel to be known a priori. In [13], an analytical expression of the channel covariance matrix is derived and computed through the information obtained from one instantaneous channel realization, which can be estimated from low-dimensional observations. In [14], the covariance matrices of vector channels are estimated by solving a subspace estimation problem leveraging their low-rank property. Also, tensor decomposition has been used for dimension reduction for the mmWave channel estimation problem [15,16]. It has been recently used for channel covariance estimation in frequency-selective channels [17], where the channel is represented as a low-rank third-order tensor in terms of factor matrices. The channel covariance matrix is obtained from the estimated factor matrices. The methods of [13,14,17] focus on vector channels, which may not be directly applicable to matrix channels where both the transmitter and receiver employ multiple antennas.

In this paper, we investigate the mmWave channel covariance matrix estimation problem for hybrid mmWave communication systems that are equipped with uniform linear arrays (ULA) or uniform square planar arrays (USPA). Both the transmitter and the receiver have multiple antennas. The main contributions are as follows:

1. We show that the mmWave MIMO channel covariance matrix follows a Kronecker product expansion model [18]. Following [18–20], we show that this model can be used for reducing the effective dimension of the large-dimensional channel covariance matrices in mmWave MIMO systems. We further show that permutation can reduce the rank of the mmWave channel covariance matrix, which admits an expression of the summation of vector outer products. We thus formulate the channel covariance matrix estimation problem as a low-rank matrix sensing problem.
2. Although the aforementioned low-rank matrix sensing problem has a smaller size than the original problem, the complexity can still be high when the numbers of the transmitter/receiver antennas are large. In order to reduce the complexity, we further exploit the structures of the ULA or USPA to reduce the dimensionality of the problem and formulate the problem as a *structured* low-rank matrix sensing problem. We adapt the recently proposed generalized conditional gradient and alternating minimization (GCG-Alt) algorithm [21], which has low computational complexity, to find the solution. Numerical results with ULA and USPA suggest that our proposed estimator is effective in estimating the mmWave channel covariance matrix.

The rest of this paper is organized as follows. We introduce the spatial channel model and the hybrid system in Section 2. In Section 3, we formulate the channel covariance estimation problem as a structured low-rank matrix sensing problem and present the solution. We show the simulation results in Section 4 and conclude the paper in Section 5.

Notations: Bold uppercase $\mathbf{A}$ denotes a matrix and bold lowercase $\mathbf{a}$ denotes a column vector. $\mathbf{A}^*$, $\mathbf{A}^T$, and $\mathbf{A}^H$ denote the conjugate, transpose, and conjugate transpose of matrix $\mathbf{A}$, respectively. $\mathbf{a}(i)$ denotes the i-th element of vector $\mathbf{a}$. $[\mathbf{A}]_{a:b,:}$ denotes the submatrix of $\mathbf{A}$ made of its a-th to b-th rows. $[\mathbf{A}]_{a:b,c:d}$ denotes the submatrix of $\mathbf{A}$ defined by its a-th to b-th rows and c-th to d-th columns. $\|\mathbf{A}\|_F$ and $\|\mathbf{A}\|_*$ are the Frobenius norm and the nuclear norm of $\mathbf{A}$. For $\mathbf{A} \in \mathbb{C}^{M \times N}$, $\mathrm{vec}(\mathbf{A}) \in \mathbb{C}^{MN \times 1}$ is a column vector obtained through the vectorization of $\mathbf{A}$ and $\mathrm{vec}^{-1}(\mathbf{A}) \in \mathbb{C}^{M \times N}$ is a matrix obtained by the inverse of vectorization. For matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. $\mathcal{CN}(a, b^2)$ represents complex Gaussian distribution with mean a and variance b^2. $\mathcal{U}(a, b)$ represents uniform distribution with support $[a, b]$.

2. Spatial Channel Model

Consider point-to-point mmWave transmissions, where the transmitter has N_t antennas and the receiver has N_r antennas. We assume the following spatial channel [22]:

$$\mathbf{H} = \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{l=1}^{L} g_{kl} \mathbf{a}_r(\phi_{kl}^r, \theta_{kl}^r) \mathbf{a}_t^H(\phi_{kl}^t, \theta_{kl}^t) \in \mathbb{C}^{N_r \times N_t}, \tag{1}$$

where K is the number of clusters, and L is the number of rays within each cluster. As reported in [22], the number of clusters is often small, e.g., $K = 1, 2$, but the number of rays inside each cluster can be large, e.g., $L = 30$. $\mathbf{a}_r(\phi_{kl}^r, \theta_{kl}^r)$ and $\mathbf{a}_t(\phi_{kl}^t, \theta_{kl}^t)$ are the array response vectors at the receiver and transmitter, respectively, where $\phi_{kl}^r, \theta_{kl}^r, \phi_{kl}^t$, and θ_{kl}^t are the azimuth AoA, elevation AoA, azimuth AoD, and elevation AoD on the l-th ray of the k-th cluster, respectively. These angles can be characterized by cluster center angles and angular spreads: Each cluster covers a range of angles and the angular spread describes the span of each cluster. The angular spread in the mmWave propagation environment is considered to be small [13]. Measurements of the angular spread taken in the urban area of New York City are presented in [22] in terms of the root-mean-square (rms) of all the measurements. At the carrier frequency $f_c = 28$ GHz, example angular spreads of $15.5°, 6°, 10.2°$, and $0°$ are reported for $\phi_{kl}^r, \theta_{kl}^r, \phi_{kl}^t$, and θ_{kl}^t, respectively. The small-scale fading coefficient g_{kl} is assumed complex Gaussian, i.e., $g_{kl} \sim \mathcal{CN}(0, \gamma_k^2)$, where γ_k^2 is the fraction power of the k-th cluster [22] (Equation (7)).

As discussed in [9], though the small-scale fading gains $\{g_{kl}\}$ change fast, the AoDs/AoAs and γ_k^2 may remain stationary over tens to hundreds of coherence blocks. Assume that $\{g_{kl}, \forall k, \forall l\}$ are mutually independent, then channel covariance matrix can be modeled as

$$\mathbf{R} \triangleq \mathbb{E}[\text{vec}(\mathbf{H}) \text{vec}^H(\mathbf{H})]$$

$$= \frac{1}{L} \sum_{k=1}^{K} \gamma_k^2 \sum_{l=1}^{L} \widetilde{\mathbf{T}}_{kl}^t \otimes \widetilde{\mathbf{T}}_{kl}^r \in \mathbb{C}^{N_r N_t \times N_r N_t}, \tag{2}$$

where

$$\widetilde{\mathbf{T}}_{kl}^t \triangleq \mathbf{a}_t^*(\phi_{kl}^t, \theta_{kl}^t) \mathbf{a}_t^T(\phi_{kl}^t, \theta_{kl}^t) \in \mathbb{C}^{N_t \times N_t}, \tag{3}$$

and

$$\widetilde{\mathbf{T}}_{kl}^r \triangleq \mathbf{a}_r(\phi_{kl}^r, \theta_{kl}^r) \mathbf{a}_r^H(\phi_{kl}^r, \theta_{kl}^r) \in \mathbb{C}^{N_r \times N_r}. \tag{4}$$

Note that Equation (2) is the same as the channel covariance expression in [9] when $L = 1$. In the following, we first present our proposed covariance matrix estimation method for systems equipped with the ULA and then discuss its adaptation to systems that adopt the USPA.

For the ULA, the array responses $\mathbf{a}_t(\phi_{kl}^t, \theta_{kl}^t)$ and $\mathbf{a}_r(\phi_{kl}^r, \theta_{kl}^r)$ are independent of the elevation angles. They can thus be abbreviated as $\mathbf{a}_t(\phi_{kl}^t)$ and $\mathbf{a}_r(\phi_{kl}^r)$. For an N_a-element ULA with distance d between adjacent antennas, the array response is

$$\mathbf{a}(\phi_{kl}) = \frac{1}{\sqrt{N_a}} [1, e^{j\frac{2\pi}{\lambda_c} d \sin(\phi_{kl})}, \cdots, e^{j(N_a - 1)\frac{2\pi}{\lambda_c} d \sin(\phi_{kl})}]^T,$$

where λ_c is the carrier wavelength and $N_a = N_t$ or N_r is the number of antennas at the transmitter or receiver. Accordingly, $\widetilde{\mathbf{T}}_{kl}^t$ of Equation (3) and $\widetilde{\mathbf{T}}_{kl}^r$ of Equation (4) become

$$\widetilde{\mathbf{T}}_{kl}^t = \mathbf{a}_t^*(\phi_{kl}^t) \mathbf{a}_t^T(\phi_{kl}^t) \tag{5}$$

and

$$\widetilde{\mathbf{T}}_{kl}^r = \mathbf{a}_r(\phi_{kl}^t) \mathbf{a}_r^H(\phi_{kl}^t), \tag{6}$$

respectively, which are Toeplitz–Hermitian. Since the Kronecker product of two Toeplitz–Hermitian matrices is block-Toeplitz–Hermitian [23], the channel covariance matrix **R** defined in Equation (2) is block-Toeplitz–Hermitian.

We next discuss the hybrid system. We assume phase shifter-based hybrid transceivers [21] shown in Figure 1, where the antennas and analog phase shifters at the transmitter or receiver are fully connected. Assume that there are $K_t \ll N_t$ radio frequency (RF) chains at the transmitter and $K_r \ll N_r$ RF chains at the receiver. For single-stream transmissions with one symbol s transmitted, the received signal is written as

$$\mathbf{y} = \mathbf{W}^H \mathbf{H} \mathbf{f} s + \mathbf{W}^H \mathbf{n}, \tag{7}$$

where **W** and **f** are the receiving processing matrix and transmitting processing vector, respectively, and **n** is the noise vector. Up to K_r digital symbols can be observed at the receiver after each transmission. In hybrid transceivers, we have $\mathbf{W} = \mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}$ and $\mathbf{f} = \mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB}}$, where $\mathbf{W}_{\text{RF}}$ and $\mathbf{F}_{\text{RF}}$ are the analog combiner and precoder, respectively, and $\mathbf{W}_{\text{BB}}$ and $\mathbf{f}_{\text{BB}}$ are the digital combiner and precoder, respectively. In addition, due to the constraints of the phase shifters in the RF combiner and precoder, the entries in $\mathbf{W}_{\text{RF}}$ and $\mathbf{F}_{\text{RF}}$ have constant modulus.
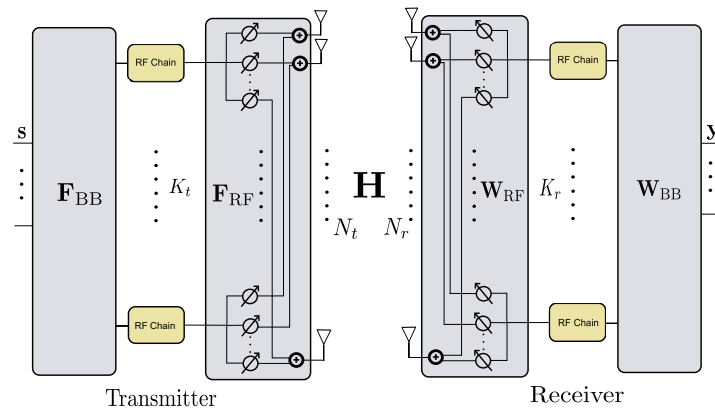

Figure 1. The phase shifter-based hybrid transceiver.

Note that using single-stream transmissions during the channel training avoids the interferences caused by transmitting multiple symbols simultaneously, and this has been widely considered [3,4,7].

When N_t and N_r are large, the dimension of the channel covariance matrix **R** is large. In this case, estimating **R** can be difficult when only a small number of observations available, which is typical in the hybrid system.

From Equation (2), **R** follows the Kronecker product expansion model [18]. In the following, we explore this property and the block-Toeplitz–Hermitian structure of **R** to reduce the dimensionality of the problem of estimating **R**, and formulate the channel covariance matrix estimation problem as a structured low-rank matrix sensing problem.

3. Structured Low-Rank Covariance Matrix Sensing

3.1. Rank Reduction By Permutation

Define

$$\mathbf{T}_{kl}^t = \frac{\gamma_k}{\sqrt{L}} \widetilde{\mathbf{T}}_{kl}^t \in \mathbb{C}^{N_t \times N_t} \tag{8}$$

and

$$\mathbf{T}_{kl}^r = \frac{\gamma_k}{\sqrt{L}} \widetilde{\mathbf{T}}_{kl}^r \in \mathbb{C}^{N_r \times N_r} \tag{9}$$

respectively, where $1 \leq l \leq L$ and $1 \leq k \leq K$. Then $\mathbf{R}$ of Equation (2) can be written compactly as

$$\mathbf{R} = \sum_{k=1}^{K} \sum_{l=1}^{L} \mathbf{T}_{kl}^t \otimes \mathbf{T}_{kl}^r \in \mathbb{C}^{N_t N_r \times N_t N_r}, \tag{10}$$

where the summation involves KL terms. Note that $\mathbf{T}_{kl}^t$ and $\mathbf{T}_{kl}^r$ are Toeplitz–Hermitian. Denote the following $N_r \times N_r$ submatrix of $\mathbf{R}$ as

$$\mathbf{R}_{mn} \triangleq [\mathbf{R}]_{((m-1)N_r+1):mN_r,((n-1)N_r+1):nN_r}, \tag{11}$$

where $1 \leq m \leq N_t$ and $1 \leq n \leq N_t$. Define a permutation operator $\mathcal{P}(\cdot)$ that permutes the $N_t N_r \times N_t N_r$ matrix $\mathbf{R}$ into a $N_t^2 \times N_r^2$ matrix

$$\mathbf{R}_p = \mathcal{P}(\mathbf{R})$$

by stacking each submatrix $\mathbf{R}_{mn}$ into a row vector as

$$[\mathcal{P}(\mathbf{R})]_{m+(n-1)N_t,:} = \text{vec}^T(\mathbf{R}_{mn}) \in \mathbb{C}^{1 \times N_r^2}.$$

We write

$$\mathbf{t}_{kl}^t = \text{vec}(\mathbf{T}_{kl}^t) \in \mathbb{C}^{N_t^2 \times 1},$$

and

$$\mathbf{t}_{kl}^r = \text{vec}(\mathbf{T}_{kl}^r) \in \mathbb{C}^{N_r^2 \times 1}.$$

Then based on the Kronecker product expansion property [23,24], $\mathbf{R}_p$ can be written as a sum of vector outer products

$$\mathbf{R}_p = \sum_{k=1}^{K} \sum_{l=1}^{L} \mathbf{t}_{kl}^t (\mathbf{t}_{kl}^r)^T \in \mathbb{C}^{N_t^2 \times N_r^2}. \tag{12}$$

Note that if we have $\mathbf{R}_p$, we can obtain $\mathbf{R}$ as $\mathcal{P}^{-1}(\mathbf{R}_p)$.

From Equation (12), we can see that the column space of $\mathbf{R}_p$ is spanned by $\{\mathbf{t}_{kl}^t\}$ and the row space of $\mathbf{R}_p$ is spanned by $\{\mathbf{t}_{kl}^r\}$. Recall that $\mathbf{t}_{kl}^t = \text{vec}(\mathbf{T}_{kl}^t)$ and by using the relation between $\mathbf{T}_{kl}^t$ and the transmitter array response $\mathbf{a}_t(\phi_{kl}^t)$ shown in Equations (5) and (8), $\mathbf{t}_{kl}^t$ can be written as

$$\mathbf{t}_{kl}^t = \frac{\gamma_k}{\sqrt{N_t L}} [\mathbf{a}_t^T(\phi_{kl}^t), e^{-j\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^t)} \mathbf{a}_t^T(\phi_{kl}^t), \ldots, e^{-j(N_t-1)\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^t)} \mathbf{a}_t^T(\phi_{kl}^t)]^T$$

$$= \frac{\gamma_k}{\sqrt{L}} \mathbf{a}_t^*(\phi_{kl}^t) \otimes \mathbf{a}_t(\phi_{kl}^t), \tag{13}$$

where ϕ_{kl}^t is the azimuth AoD. We can see that $\mathbf{t}_{kl}^t$ consists of the array response vector $\mathbf{a}_t(\phi_{kl}^t)$ and the column space of $\mathbf{R}_p$ is determined by the set

$$\mathcal{C}_t = \{\mathbf{a}_t^*(\phi_{kl}^t) \otimes \mathbf{a}_t(\phi_{kl}^t), 1 \leq k \leq K, 1 \leq l \leq L\}.$$

As introduced earlier, small angular spreads are observed in the mmWave propagation environment, which indicates that the AoDs inside a cluster are closely spaced and their corresponding array response vectors are highly correlated. Therefore, for the k-th cluster, though the number of rays L inside can be large, the space spanned by $\{\mathbf{a}_t^*(\phi_{kl}^t) \otimes \mathbf{a}_t(\phi_{kl}^t), 1 \leq l \leq L\}$ may be well approximated by a low-rank space. In addition, since the number of clusters K is generally small (e.g., $K = 1$ or 2), both $\mathcal{C}_\mathbf{t}$ and $\mathbf{R}$ can be low-rank. This is similar to the low-rankness of the mmWave channel $\mathbf{H}$, which has been validated by the experimental and simulation results in [22]. The low-rank property of

$\mathbf{R}_p$ can be shown numerically. Denote by r_{ch} the rank of $\mathbf{R}_p$ or $\mathbf{R}$, and let $\sigma_1 > \sigma_2 > \ldots > \sigma_{r_{\text{ch}}}$ be the singular values of $\mathbf{R}_p$ or $\mathbf{R}$. We may use

$$p_e \overset{\Delta}{=} \frac{\sum_{j=1}^{r_{\text{sub}}} \sigma_j^2}{\sum_{i=1}^{r_{\text{ch}}} \sigma_i^2} \tag{14}$$

to measure the energy captured by a rank-r_{sub} approximation of $\mathbf{R}_p$ or $\mathbf{R}$, where r_{sub} is the rank of the subspace of $\mathbf{R}_p$ or $\mathbf{R}$.

Figure 2 shows an example of a ULA system with $K \in \{1, 2, 3, 4\}, L = 30, N_t = 64$, and $N_r = 16$. The covariance matrix $\mathbf{R}$ and its permuted version $\mathbf{R}_p$ have sizes of 1024×1024 and 4096×256, respectively, which shows that $\mathbf{R}_p$ is a taller matrix. The horizontal AoDs

$$\phi_{kl}^t \sim \mathcal{U}(\phi_k^t - v_h^t, \phi_k^t + v_h^t), l = 1, 2, \cdots, L, \tag{15}$$

where the center angles ϕ_k^t are distributed uniformly in $[0, 2\pi]$ and separated by at least one angular spread $v_h^t = 10.2°$. Similarly, the horizontal AoAs

$$\phi_{kl}^r \sim \mathcal{U}(\phi_k^r - v_h^r, \phi_k^r + v_h^r), l = 1, 2, \cdots, L, \tag{16}$$
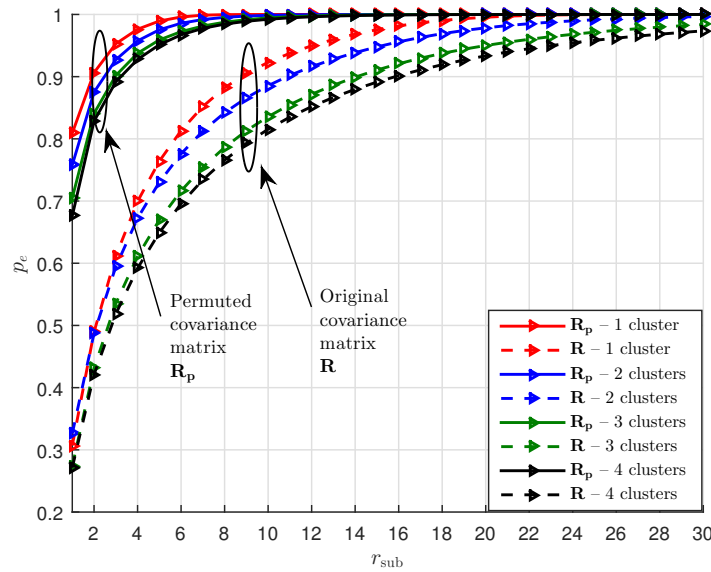
where $v_h^r = 15.5°$.

Figure 2. Energy captured by a rank-r_{sub} approximation of $\mathbf{R}_p$ and $\mathbf{R}$.

The cluster powers are generated following [22] (Table I). It can be seen from Figure 2 that for capturing a majority of the total energy, e.g., with $p_e = 0.95, 0.99$, the required r_{sub} for $\mathbf{R}_p$ is generally much smaller than $\min(N_t^2, N_r^2, KL)$ and is also much smaller than that for $\mathbf{R}$. In the following, we use r_p as the r_{sub} of $\mathbf{R}_p$ and $r_{\mathbf{R}}$ as the r_{sub} of $\mathbf{R}$ for a certain p_e. As such, $\mathbf{R}_p$ may be well approximated as a rank-r_p matrix. One may use low-rank matrix recovery methods, e.g., matrix completion methods, to estimate the best rank-r_p approximation of $\mathbf{R}_p$ from a small amount of observations. However, when N_t and N_r are large, which is the case in mmWave communications, the number of parameters required by the rank-r_p approximation of $\mathbf{R}_p$, i.e., $(N_t^2 + N_t^2) \times r_p$, is still large. Therefore, estimating the subspaces of $\mathbf{R}_p$ can be computationally expensive.

3.2. Dimension Reduction by Exploiting the Toeplitz–Hermitian Structure

Recall that $\mathbf{R}$ is block-Toeplitz–Hermitian and $\mathbf{R}_p = \mathcal{P}(\mathbf{R})$ is a permutation of $\mathbf{R}$. From Equations (12) and (13), we can see that $\mathbf{R}_p$ is also specially structured: $\mathbf{R}_p$ is the summation of the outer products of $\mathbf{t}_{kl}^t$ and $\mathbf{t}_{kl}^r$, where $\mathbf{t}_{kl}^t$ and $\mathbf{t}_{kl}^r$ are the vectorizations of Toeplitz–Hermitian matrices $\mathbf{T}_{kl}^t$ and $\mathbf{T}_{kl}^r$, respectively. Since the Toeplitz–Hermitian matrix $\mathbf{T}_{kl}^t \in \mathbb{C}^{N_t^2 \times N_t^2}$ is determined by its first column and first row (its first row is the conjugate transpose of its first column), we can represent $\mathbf{t}_{kl}^t$ in terms of the entries in the first column and first row of $\mathbf{T}_{kl}^t$. We can represent $\mathbf{t}_{kl}^r$ in the same way. Therefore, the total numbers of unknowns in $\mathbf{t}_{kl}^t$ and $\mathbf{t}_{kl}^r$ are $2N_t - 1$ and $2N_r - 1$, respectively. Then we can reduce the problem size of $(N_t^2 + N_r^2) \times r_p$ to $2(N_t + N_r - 1) \times r_p$. In the following, we show how the problem size can be reduced.

First, let us use an example with $N_t = 3$ to illustrate the structure of $\mathbf{t}_{kl}^t$. The array response

$$\mathbf{a}_t(\phi_{kl}^t) = \frac{1}{\sqrt{3}}[1, e^{j\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)}, e^{j2\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)}]^T.$$

Then according to Equation (13), we have

$$\mathbf{t}_{kl}^t = \frac{\sqrt{L}}{\gamma_k}\mathbf{a}_t^*(\phi_{kl}^t) \otimes \mathbf{a}_t(\phi_{kl}^t) = \frac{\sqrt{L}}{3\gamma_k}\begin{bmatrix} 1 \\ e^{j\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)} \\ e^{j2\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)} \\ e^{-j\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)} \\ 1 \\ e^{j\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)} \\ e^{-j2\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)} \\ e^{-j\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)} \\ 1 \end{bmatrix}.$$

We can see that all the 9 elements in $\mathbf{t}_{kl}^t$ can be represented by the elements in $\mathbf{a}_t(\phi_{kl}^t)$ and $\mathbf{a}_t^*(\phi_{kl}^t)$. Now construct a vector

$$\mathbf{a}_{kl} = \frac{\sqrt{L}}{3\gamma_k}[\mathbf{a}_t^T(\phi_{kl}^t), e^{-j\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)}, e^{-j2\frac{2\pi}{\lambda_c}d\sin(\phi_{kl}^k)}]^T \in \mathbb{C}^{5\times 1},$$

then $\mathbf{t}_{kl}^t$ can be rewritten as

$$\mathbf{t}_{kl}^t = \begin{bmatrix} [\mathbf{I}_3, \mathbf{0}_{3\times 2}] \\ [\mathbf{0}_{1\times 3}, 1, 0] \\ [\mathbf{I}_2, \mathbf{0}_{2\times 3}] \\ [\mathbf{0}_{1\times 4}, 1] \\ [\mathbf{0}_{1\times 3}, 1, 0] \\ [1, \mathbf{0}_{1\times 4}] \end{bmatrix}\mathbf{a}_{kl}.$$

In fact, $\mathbf{a}_{kl}(4) = (\mathbf{a}_{kl}(2))^*$ and $\mathbf{a}_{kl}(5) = (\mathbf{a}_{kl}(3))^*$. Therefore, $\mathbf{t}_{kl}^t$ can be expressed as a product of a weight matrix and a vector $\mathbf{a}_{kl}$. Furthermore, the weight matrix depends only on the structure of the antenna array and is independent of the path angles.

Similarly, for the general cases, we can express $\mathbf{t}_{kl}^t$ with a weight matrix $\mathbf{\Gamma}_u \in \mathbb{C}^{N_t^2 \times (2N_t-1)}$ and a vector $\mathbf{a}_{kl} \in \mathbb{C}^{(2N_t-1)\times 1}$, and express $\mathbf{t}_{kl}^r$ with a weight matrix $\mathbf{\Gamma}_v \in \mathbb{C}^{N_r^2 \times (2N_r-1)}$ and a vector $\mathbf{b}_{kl} \in \mathbb{C}^{(2N_r-1)\times 1}$. We require

$$\mathbf{a}_{kl}(x + N_t - 1) = (\mathbf{a}_{kl}(x))^*, 2 \leq x \leq N_t,$$

and

$$\mathbf{b}_{kl}(y + N_r - 1) = (\mathbf{b}_{kl}(y))^*, 2 \leq y \leq N_r.$$

We then have

$$\mathbf{t}_{kl}^t = \mathbf{\Gamma}_u \mathbf{a}_{kl}, \quad \text{and} \quad \mathbf{t}_{kl}^r = \mathbf{\Gamma}_v \mathbf{b}_{kl}, \tag{17}$$

where $\mathbf{\Gamma}_u = [\mathbf{\Gamma}_{u1}, \mathbf{\Gamma}_{u2}]$ with $\mathbf{\Gamma}_{u1} \in \mathbb{C}^{N_t^2 \times N_t}$, $\mathbf{\Gamma}_{u2} \in \mathbb{C}^{N_t^2 \times (N_t-1)}$, and

$$\mathbf{\Gamma}_{u1} = \begin{bmatrix} \mathbf{I}_{N_t} \\ \mathbf{0}_{1 \times N_t} \\ [\mathbf{I}_{N_t-1}, \mathbf{0}_{(N_t-1) \times 1}] \\ \mathbf{0}_{2 \times N_t} \\ [\mathbf{I}_{N_t-2}, \mathbf{0}_{(N_t-2) \times 2}] \\ \mathbf{0}_{3 \times N_t} \\ \vdots \\ [1, \mathbf{0}_{1 \times (N_t-1)}] \end{bmatrix}, \quad \mathbf{\Gamma}_{u2} = \begin{bmatrix} \mathbf{0}_{N_t \times (N_t-1)} \\ \mathbf{e}_1^T \\ \mathbf{0}_{N_t-1 \times (N_t-1)} \\ \mathbf{e}_2^T \\ \mathbf{e}_1^T \\ \mathbf{0}_{(N_t-2) \times (N_t-1)} \\ \mathbf{e}_3^T \\ \mathbf{e}_2^T \\ \mathbf{e}_1^T \\ \vdots \\ \mathbf{0}_{1 \times (N_t-1)} \end{bmatrix}$$

with $\mathbf{e}_i \in \mathbb{C}^{(N_t-1) \times 1}$ being a vector whose i-th entry is 1 and other entries are zero. $\mathbf{\Gamma}_v$ is constructed similarly as $\mathbf{\Gamma}_u$, and $\mathbf{\Gamma}_u$ and $\mathbf{\Gamma}_v$ are both full-rank. This is because $\mathbf{\Gamma}_u$ and $\mathbf{\Gamma}_v$ consist of 1's and 0's, and there is only one 1 in each row of $\mathbf{\Gamma}_u$ and $\mathbf{\Gamma}_v$. Therefore, Equation (12) can be rewritten as

$$\begin{aligned} \mathbf{R}_p &= \sum_{k=1}^{K} \sum_{l=1}^{L} \mathbf{\Gamma}_u \mathbf{a}_{kl} \mathbf{b}_{kl}^T \mathbf{\Gamma}_v^T \\ &= \mathbf{\Gamma}_u \left(\sum_{k=1}^{K} \sum_{l=1}^{L} \mathbf{a}_{kl} \mathbf{b}_{kl}^T \right) \mathbf{\Gamma}_v^T \\ &= \mathbf{\Gamma}_u \mathbf{C} \mathbf{\Gamma}_v^T, \end{aligned} \tag{18}$$

where $\mathbf{C} = \sum_{k=1}^{K} \sum_{l=1}^{L} \mathbf{a}_{kl} \mathbf{b}_{kl}^T$.

As shown above, $\mathbf{R}_p$ is approximately low-rank. Since the fixed weight matrices $\mathbf{\Gamma}_u$ and $\mathbf{\Gamma}_v$ are full-rank, $\mathbf{C}$ is approximately low-rank. Hence estimating a low-rank approximation of $\mathbf{R}_p$ is equivalent to estimating a low-rank approximation of $\mathbf{C}$. Note that the size of $\mathbf{C} \in \mathbb{C}^{(2N_t-1) \times (2N_r-1)}$ is much smaller than the size of $\mathbf{R}_p \in \mathbb{C}^{N_t^2 \times N_r^2}$, and this can greatly reduce the complexity of the problem.

3.3. Training

We assume that the channel matrix $\mathbf{H}$ remains static during a snapshot and suppose we have T snapshots. For different snapshots, we assume that the AoAs/AoDs and the fraction power γ_k^2 remain unchanged, but the small-scale fading gain $g_{kl} \sim \mathcal{CN}(0, \gamma_k^2)$ can change [9]. Suppose the transmitter sends out S training beams during each snapshot. For the s-th training beam of the t-th snapshot, we employ the transmitting vector $\mathbf{f}_{t,s} \in \mathbb{C}^{N_t}$ and the receiving matrix $\mathbf{W}_{t,s} \in \mathbb{C}^{N_r \times K_r}$. Therefore, in each snapshot, after the transmitter sends out S training beams, the receiver receives SK_r symbols, and the sampling ratio is $SK_r / N_r N_t$. We design $\mathbf{f}_{t,s}$ and $\mathbf{W}_{t,s}$ and their corresponding $\mathbf{F}_{RF}, \mathbf{f}_{BB}, \mathbf{W}_{RF}$, and $\mathbf{W}_{BB}$ realizations for the hybrid structure according to the training scheme in [21] (Section III.D). For the s-th training beam of the t-th snapshot, the received signal is

$$\begin{aligned} \mathbf{y}_{t,s} &= \mathbf{W}_{t,s}^H \mathbf{H}_t \mathbf{f}_{t,s} s + \mathbf{W}_{t,s}^H \mathbf{n}_{t,s} \\ &= (\mathbf{f}_{t,s}^T \otimes \mathbf{W}_{t,s}^H) \text{vec}(\mathbf{H}_t) s + \mathbf{W}_{t,s}^H \mathbf{n}_{t,s}, \end{aligned} \tag{19}$$

where $\mathbf{n}_{t,s}$ is the noise vector and $\mathbf{H}_t$ is the channel matrix at snapshot t. Without loss of generality, assume identical training symbols $s = \sqrt{P}$. By setting $\|\mathbf{f}_{t,s}\|_F^2 = 1$, the total transmitting power is $\|\mathbf{f}_{t,s}s\|_F^2 = P$, and the pilot-to-noise ratio (PNR) is defined as

$$\text{PNR} = \frac{\|\mathbf{f}_{t,s}s\|_F^2}{\sigma^2}, \tag{20}$$

where the noise is assumed to be an additive white Gaussian noise (AWGN) with variance σ^2. In the t-th snapshot and after the transmitter sends out all the S training beams, stack the received signals as

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{f}_{t,1}^T \otimes \mathbf{W}_{t,1}^H \\ \mathbf{f}_{t,2}^T \otimes \mathbf{W}_{t,2}^H \\ \vdots \\ \mathbf{f}_{t,S}^T \otimes \mathbf{W}_{t,S}^H \end{bmatrix} \text{vec}(\mathbf{H}_t) + \begin{bmatrix} \mathbf{W}_{t,1}^H \mathbf{n}_{t,1} \\ \mathbf{W}_{t,2}^H \mathbf{n}_{t,2} \\ \vdots \\ \mathbf{W}_{t,S}^H \mathbf{n}_{t,S} \end{bmatrix}, \tag{21}$$

$$= \mathbf{P}_t \text{vec}(\mathbf{H}_t) + \mathbf{n}_t \in \mathbb{C}^{SK_r \times 1}, \tag{22}$$

where

$$\mathbf{P}_t = \begin{bmatrix} \mathbf{f}_{t,1}^T \otimes \mathbf{W}_{t,1}^H \\ \mathbf{f}_{t,2}^T \otimes \mathbf{W}_{t,2}^H \\ \vdots \\ \mathbf{f}_{t,S}^T \otimes \mathbf{W}_{t,S}^H \end{bmatrix}, \quad \text{and} \quad \mathbf{n}_t = \begin{bmatrix} \mathbf{W}_{t,1}^H \mathbf{n}_{t,1} \\ \mathbf{W}_{t,2}^H \mathbf{n}_{t,2} \\ \vdots \\ \mathbf{W}_{t,S}^H \mathbf{n}_{t,S} \end{bmatrix}.$$

Suppose the trainings are the same for different snapshots, i.e., $\mathbf{f}_{1,s} = \mathbf{f}_{2,s} = \ldots = \mathbf{f}_{T,s} = \mathbf{f}_s$ and $\mathbf{W}_{1,s} = \mathbf{W}_{2,s} = \ldots = \mathbf{W}_{T,s} = \mathbf{W}_s$. We then have

$$\mathbf{P} = \mathbf{P}_1 = \ldots = \mathbf{P}_T = \begin{bmatrix} \mathbf{f}_1^T \otimes \mathbf{W}_1^H \\ \mathbf{f}_2^T \otimes \mathbf{W}_2^H \\ \vdots \\ \mathbf{f}_S^T \otimes \mathbf{W}_S^H \end{bmatrix}, \tag{23}$$

and

$$\boldsymbol{\Sigma} = \mathbf{P}\mathbf{R}\mathbf{P}^H + \boldsymbol{\Sigma}_n, \tag{24}$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$ represent the covariance matrices of the received signal $\mathbf{y}_t$ and the noise $\mathbf{n}_t$, respectively.

After T snapshots, we can compute the dimension-reduced sample covariance matrix (SCM) of $\mathbf{y}_t$ as

$$\mathbf{S} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t^H \in \mathbb{C}^{SK_r \times SK_r}. \tag{25}$$

We permute $\mathbf{S}$ into $\mathbf{S}_p \in \mathbb{C}^{S^2 \times K_r^2}$ in a similar procedure as $\mathbf{R}$ is permuted into $\mathbf{R}_p$.

3.4. Low-Rank Matrix Sensing Problem

We can now formulate the channel covariance estimation problem as a low-rank matrix sensing problem [25]:

$$\min_{\widehat{\mathbf{R}}_p} \text{rank}(\widehat{\mathbf{R}}_p) \quad \text{s.t.} \quad \|\mathcal{A}(\widehat{\mathbf{R}}_p) - \text{vec}(\mathbf{S}_p)\|_F^2 \leq \zeta^2, \tag{26}$$

where $\widehat{\mathbf{R}}_p$ is the estimate of $\mathbf{R}_p$, $\mathcal{A} : \mathbb{C}^{N_t^2 \times N_r^2} \to \mathbb{C}^{S^2 K_r^2 \times 1}$ is an appropriate linear map, and ζ^2 is a constant to account for the fitting error. Replacing $\widehat{\mathbf{R}}_p$ with Equation (18), we can reformulate Equation (26) as

$$\min_{\widehat{\mathbf{C}}} \text{rank}(\widehat{\mathbf{C}}) \quad \text{s.t.} \quad \|\mathcal{A}(\boldsymbol{\Gamma}_u \widehat{\mathbf{C}} \boldsymbol{\Gamma}_v^T) - \text{vec}(\mathbf{S}_p)\|_F^2 \leq \zeta^2, \tag{27}$$

where $\widehat{\mathbf{C}}$ is the estimate of $\mathbf{C}$.

In general, problem Equation (27) is a nonconvex optimization problem and difficult to solve. In this paper, we solve the relaxed version of problem Equation (27) [26]:

$$\min_{\widehat{\mathbf{C}}} \phi(\widehat{\mathbf{C}}) = f(\widehat{\mathbf{C}}) + \mu\|\widehat{\mathbf{C}}\|_* \tag{28}$$

where

$$f(\widehat{\mathbf{C}}) = \frac{1}{2}\|\mathcal{A}(\mathbf{\Gamma}_u\widehat{\mathbf{C}}\mathbf{\Gamma}_v^T) - \mathrm{vec}(\mathbf{S}_p)\|_F^2 \tag{29}$$

and $\mu > 0$ is a regularization coefficient. After some manipulations, we have $\mathcal{A}(\mathbf{\Gamma}_u\widehat{\mathbf{C}}\mathbf{\Gamma}_v^T) = \mathbf{Q}\mathrm{vec}(\widehat{\mathbf{C}})$, where

$$\mathbf{Q} = \begin{bmatrix} (\mathbf{f}_1^H \otimes \mathbf{f}_1^T)\mathbf{\Gamma}_u \otimes (\mathbf{W}_1^* \otimes \mathbf{W}_1)\mathbf{\Gamma}_v \\ (\mathbf{f}_1^H \otimes \mathbf{f}_2^T)\mathbf{\Gamma}_u \otimes (\mathbf{W}_1^* \otimes \mathbf{W}_2)\mathbf{\Gamma}_v \\ \vdots \\ (\mathbf{f}_S^H \otimes \mathbf{f}_S^T)\mathbf{\Gamma}_u \otimes (\mathbf{W}_S^* \otimes \mathbf{W}_S)\mathbf{\Gamma}_v \end{bmatrix}, \tag{30}$$

and $\mathbf{Q} \in \mathbb{C}^{S^2 K_r^2 \times (2N_t-1)(2N_r-1)}$. The direct evaluation of $\|\widehat{\mathbf{C}}\|_*$, which is the nuclear norm (i.e., the summation of the singular values) of $\widehat{\mathbf{C}}$, is computationally expensive. Following [21], $\|\widehat{\mathbf{C}}\|_*$ can be written as

$$\|\widehat{\mathbf{C}}\|_* = \frac{1}{2}\min_{\mathbf{U},\mathbf{V}}\{\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 : \widehat{\mathbf{C}} = \mathbf{U}\mathbf{V}^T\}. \tag{31}$$

Therefore, finding a $\widehat{\mathbf{C}}$ to minimize the objective function in Equation (28) becomes finding a pair of $(\mathbf{U}, \mathbf{V})$ to minimize

$$\begin{aligned} \widetilde{\phi}(\mathbf{U},\mathbf{V}) &\triangleq f(\mathbf{U}\mathbf{V}^T) + \frac{1}{2}\mu(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) \\ &= \frac{1}{2}\|\mathbf{Q}\mathrm{vec}(\mathbf{U}\mathbf{V}^T) - \mathrm{vec}(\mathbf{S}_p)\|_F^2 \\ &\quad + \frac{1}{2}\mu(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2). \end{aligned} \tag{32}$$

A similar low-rank recovery problem is recently studied in [21] for instantaneous mmWave channel estimation, where a training scheme is designed such that the channel can be estimated by solving a matrix completion (MC) problem. A generalized conditional gradient and alternating minimization (GCG-Alt) algorithm is developed, which is shown to be able to provide accurate low-rank solutions at low complexity. In this work, we adapt the GCG-Alt algorithm to solve Equation (32) for our covariance matrix estimation problem. In the following, we discuss the key steps of the GCG-Alt algorithm for solving Equation (32) and refer readers to [21] for more detailed treatments.

The GCG-Alt algorithm consists of a relaxed GCG algorithm and an AltMin algorithm. Let $\widehat{\mathbf{C}}_{k-1}$ be the solution to $\mathbf{C}$ at the $(k-1)$-th GCG iteration. The relaxed GCG algorithm first produces an output

$$\widehat{\mathbf{C}}_k = (1 - \eta_k)\widehat{\mathbf{C}}_{k-1} + \theta_k \mathbf{Z}_k, \tag{33}$$

where $\mathbf{Z}_k$ is the outer product of the top singular vector pair of $\mathrm{vec}^{-1}(-\nabla f(\widehat{\mathbf{C}}_k))$. The calculations of $\mathrm{vec}^{-1}(-\nabla f(\widehat{\mathbf{C}}_k))$ and the parameter θ_k here are different from those in [21]. For problem Equation (32), we calculate $\mathrm{vec}^{-1}(-\nabla f(\widehat{\mathbf{C}}_k))$ as

$$-\nabla f(\widehat{\mathbf{C}}_k) = -(\mathbf{Q}^H\mathbf{Q}\mathrm{vec}(\widehat{\mathbf{C}}_{k-1}) - \mathbf{Q}\mathrm{vec}(\mathbf{S}_p)),$$

and the parameter θ_k as

$$\theta_k = \frac{\mathcal{R}\left(\mathbf{q}_{z_k}^H \text{vec}(\mathbf{S}_p) - (1 - \eta_k)\mathbf{q}_{z_k}^H \mathbf{Q}\text{vec}(\widehat{\mathbf{C}}_k)\right) - \mu}{\mathbf{q}_{z_k}^H \mathbf{q}_{z_k}}, \tag{34}$$

where $\mathbf{q}_{z_k} = \mathbf{Q}\text{vec}(\mathbf{Z}_k)$ and $\mathcal{R}(\cdot)$ denotes the real part of a number. Since $\widehat{\mathbf{C}}_k = \mathbf{U}_k\mathbf{V}_k^T$, updating $\widehat{\mathbf{C}}_k$ is equivalent to updating $\mathbf{U}_k = [\sqrt{1 - \eta_k}\mathbf{U}_{k-1}, \sqrt{\theta_k}\mathbf{u}_k]$ and $\mathbf{V}_k = [\sqrt{1 - \eta_k}\mathbf{V}_{k-1}, \sqrt{\theta_k}\mathbf{v}_k]$. Then the obtained $\mathbf{U}_k$ and $\mathbf{V}_k$ are used as the initial input of the AltMin algorithm, i.e., $\mathbf{U}_k^0 \leftarrow \mathbf{U}_k, \mathbf{V}_k^0 \leftarrow \mathbf{V}_k$. After I_a iterations of the AltMin algorithm, update $\mathbf{U}_k = \mathbf{U}_k^{I_a}$ and $\mathbf{V}_k = \mathbf{V}_k^{I_a}$. For completeness, we summarize the GCG-Alt algorithm in Algorithm 1. After obtaining $\widehat{\mathbf{C}}$, we have $\widehat{\mathbf{R}}_p^r = \mathbf{\Gamma}_u\widehat{\mathbf{C}}\mathbf{\Gamma}_u^T$ and $\widehat{\mathbf{R}} = \mathcal{P}^{-1}(\widehat{\mathbf{R}}_p^r)$.

Algorithm 1 The GCG-Alt Algorithm for Estimating $\widehat{\mathbf{C}}$ of Equation (28)

1: **Input:** $\text{vec}(\mathbf{S}_p), \mathbf{Q}, \mathbf{Q}^H\mathbf{Q}, \mu, \epsilon, \epsilon_a$

2: **Initialization:** $\mathbf{U}_0 = \varnothing, \mathbf{V}_0 = \varnothing, k = 0, \epsilon_0 = \infty$

3: **while** $\epsilon_k > \epsilon$ **do**

4:    $(\mathbf{u}_k, \mathbf{v}_k) \leftarrow$ singular vector pair of $\mathbf{Z}_k$

5:    $k = k + 1$

6:    $\eta_k \leftarrow 2/(k+1)$ and determine θ_k using Equation (34)

7:    $\mathbf{U}_k \leftarrow [\sqrt{1 - \eta_k}\mathbf{U}_{k-1}, \sqrt{\theta_k}\mathbf{u}_k]$

8:    $\mathbf{V}_k \leftarrow [\sqrt{1 - \eta_k}\mathbf{V}_{k-1}, \sqrt{\theta_k}\mathbf{v}_k]$

9:    **Initialization**: $i = 0, \epsilon_k^0 = \infty, (\mathbf{U}_k^0, \mathbf{V}_k^0) \leftarrow (\mathbf{U}_k, \mathbf{V}_k)$

10:    **while** $\epsilon_k^i > \epsilon_a$ **do**

11:        $i = i + 1$

12:        update $\mathbf{U}_k^i$ and $\mathbf{V}_k^i$ via the AltMin algorithm [21]

13:        calculate $\epsilon_k^i = \frac{\widetilde{\phi}(\mathbf{U}_k^{i-1}, \mathbf{V}_k^{i-1}) - \widetilde{\phi}(\mathbf{U}_k^i, \mathbf{V}_k^i)}{\widetilde{\phi}(\mathbf{U}_k^{i-1}, \mathbf{V}_k^i)}$

14:    **end while**

15:    $(\mathbf{U}_k, \mathbf{V}_k) \leftarrow (\mathbf{U}_k^i, \mathbf{V}_k^i)$

16:    calculate $\epsilon_k = \frac{\widetilde{\phi}(\mathbf{U}_{k-1}, \mathbf{V}_{k-1}) - \widetilde{\phi}(\mathbf{U}_k, \mathbf{V}_k)}{\widetilde{\phi}(\mathbf{U}_{k-1}, \mathbf{V}_{k-1})}$

17: **end while**

18: **Output:** $\widehat{\mathbf{C}} = \widehat{\mathbf{C}}_k = \mathbf{U}_k\mathbf{V}_k^T$

3.5. Computational Complexity

Define a flop as an operation of real-valued numbers. Let $M = SK_r$ be the number of received symbols during each snapshot. Following the computational complexity analysis in [21], the computational complexity of the GCG-Alt estimator is about $8r_{\text{est}}(I_a r_{\text{est}} + I_a + 1)(2N_t - 1)^2(2N_r - 1)^2 + 8/3I_a r_{\text{est}}(r_{\text{est}} + 1)(2r_{\text{est}} + 1)(2N_t - 1)(2N_r - 1)(N_r + N_t - 1) + I_a r_{\text{est}}^2(r_{\text{est}} + 1)^2((2N_r - 1)^3 + (2N_t - 1)^3) + 16r_{\text{est}}(2N_r - 1)(2N_t - 1)M^2$, where I_a is the number of iterations of the AltMin algorithm and r_{est} is the estimated rank of $\widehat{\mathbf{C}}$ by the GCG-Alt estimator. Later in Section 4, we show the computational complexity of the GCG-Alt estimator with specific examples.

3.6. Extension to the USPA System

We now follow the same process introduced in Section III. A-D to estimate the channel covariance matrix for USPA systems. To account for the different array structure of the USPA, the weight matrices of Equation (17) are redesigned. For a $\sqrt{N_a} \times \sqrt{N_a}$ USPA placed on the yz plane with distance d between adjacent antennas, the array response is

$$\mathbf{a}(\phi_{kl}, \theta_{kl}) = \mathbf{a}_y(\phi_{kl}, \theta_{kl}) \otimes \mathbf{a}_z(\theta_{kl}), \tag{35}$$

where

$$\mathbf{a}_y(\phi_{kl}, \theta_{kl}) = \frac{1}{N_a^{\frac{1}{4}}} [1, e^{j\frac{2\pi}{\lambda_c} d \sin(\phi_{kl}) \sin(\theta_{kl})}, \cdots, e^{j(\sqrt{N_a}-1)\frac{2\pi}{\lambda_c} d \sin(\phi_{kl}) \sin(\theta_{kl})}]^T$$

is the array response along the y axis and

$$\mathbf{a}_z(\theta_{kl}) = \frac{1}{N_a^{\frac{1}{4}}} [1, e^{j\frac{2\pi}{\lambda_c} d \cos(\theta_{kl})}, \cdots, e^{j(\sqrt{N_a}-1)\frac{2\pi}{\lambda_c} d \cos(\theta_{kl})}]^T$$

is the array response along the z axis. We design the weight matrices by examining the structure of $\widetilde{\mathbf{T}}_{kl}^t$ defined in Equation (3) which is written as

$$\begin{aligned}
\widetilde{\mathbf{T}}_{kl}^t &= \mathbf{a}_t^*(\phi_{kl}^t, \theta_{kl}^t) \mathbf{a}_t^T(\phi_{kl}^t, \theta_{kl}^t) \\
&= \left(\mathbf{a}_{t_y}(\phi_{kl}^t, \theta_{kl}^t) \otimes \mathbf{a}_{t_z}(\theta_{kl}^t) \right)^* \left(\mathbf{a}_{t_y}(\phi_{kl}^t, \theta_{kl}^t) \otimes \mathbf{a}_{t_z}(\theta_{kl}^t) \right)^T
\end{aligned}$$

where $\mathbf{a}_{t_y}(\phi_{kl}^t, \theta_{kl}^t)$ and $\mathbf{a}_{t_z}(\theta_{kl}^t)$ are the transmitter array response vectors along the y axis and z axis, respectively. Note that $\widetilde{\mathbf{T}}_{kl}^t$ is block-Toeplitz–Hermitian. Let

$$\widetilde{\mathbf{T}}_{kl}^y = \mathbf{a}_{t_y}^*(\phi_{kl}^t, \theta_{kl}^t) \mathbf{a}_{t_y}^T(\phi_{kl}^t, \theta_{kl}^t) \in \mathbb{C}^{\sqrt{N_t} \times \sqrt{N_t}}$$

and

$$\widetilde{\mathbf{T}}_{kl}^z = \mathbf{a}_{t_z}^*(\theta_{kl}^t) \mathbf{a}_{t_z}^T(\theta_{kl}^t) \in \mathbb{C}^{\sqrt{N_t} \times \sqrt{N_t}},$$

we can verify that

$$\widetilde{\mathbf{T}}_{kl}^t = \widetilde{\mathbf{T}}_{kl}^y \otimes \widetilde{\mathbf{T}}_{kl}^z, \tag{36}$$

and $\widetilde{\mathbf{T}}_{kl}^y$ and $\widetilde{\mathbf{T}}_{kl}^z$ are Toeplitz–Hermitian. Then for the USPA, $\mathbf{T}_{kl}^t$ of Equation (8) can be written as

$$\mathbf{T}_{kl}^t = \frac{\gamma_k}{\sqrt{L}} \widetilde{\mathbf{T}}_{kl}^y \otimes \widetilde{\mathbf{T}}_{kl}^z. \tag{37}$$

In Section III. B, we have expressed $\mathbf{t}_{kl}^t = \text{vec}^{-1}(\mathbf{T}_{kl}^t)$, where $\mathbf{T}_{kl}^t$ of Equation (8) is Toeplitz–Hermitian matrix, in terms of a weight matrix and a vector. We have similar expressions for the vectorizations of the Toeplitz–Hermitian matrices $\widetilde{\mathbf{T}}_{kl}^y$ and $\widetilde{\mathbf{T}}_{kl}^z$. Let

$$\begin{aligned}
\widetilde{\mathbf{t}}_{kl}^y &= \text{vec}(\widetilde{\mathbf{T}}_{kl}^y) \\
&= \mathbf{\Gamma}_y \mathbf{a}_{kl}^y
\end{aligned} \tag{38}$$

where $\mathbf{\Gamma}_y \in \mathbb{C}^{N_t \times (2\sqrt{N_t}-1)}$ is the weight matrix and $\mathbf{a}_{kl}^y \in \mathbb{C}^{(2\sqrt{N_t}-1)\times 1}$, and

$$\begin{aligned}
\widetilde{\mathbf{t}}_{kl}^z &= \text{vec}(\widetilde{\mathbf{T}}_{kl}^z) \\
&= \mathbf{\Gamma}_z \mathbf{a}_{kl}^z
\end{aligned} \tag{39}$$

where $\mathbf{\Gamma}_z \in \mathbb{C}^{N_t \times (2\sqrt{N_t}-1)}$ is the weight matrix and $\mathbf{a}_{kl}^z \in \mathbb{C}^{(2\sqrt{N_t}-1) \times 1}$. Let

$$\mathbf{\Gamma}_y^{(a)} = [\mathbf{\Gamma}_y]_{1+(a-1)\sqrt{N_t}:a\sqrt{N_t},:}, \ 1 \le a \le \sqrt{N_t},$$

and

$$\mathbf{\Gamma}_z^{(b)} = [\mathbf{\Gamma}_z]_{1+(b-1)\sqrt{N_t}:b\sqrt{N_t},:}, \ 1 \le b \le \sqrt{N_t}.$$

By exploring the matrix vectorization process, we have

$$
\begin{aligned}
\mathbf{t}_{kl}^t &= \text{vec}(\mathbf{T}_{kl}^t) \\
&= \mathbf{\Gamma}_u \mathbf{a}_{kl},
\end{aligned}
\tag{40}
$$

where

$$
\mathbf{\Gamma}_u =
\begin{bmatrix}
\mathbf{\Gamma}_y^{(1)} \otimes \mathbf{\Gamma}_z^{(1)} \\
\mathbf{\Gamma}_y^{(1)} \otimes \mathbf{\Gamma}_z^{(2)} \\
\vdots \\
\mathbf{\Gamma}_y^{(1)} \otimes \mathbf{\Gamma}_z^{(\sqrt{N_t})} \\
\mathbf{\Gamma}_y^{(2)} \otimes \mathbf{\Gamma}_z^{(1)} \\
\mathbf{\Gamma}_y^{(2)} \otimes \mathbf{\Gamma}_z^{(2)} \\
\vdots \\
\mathbf{\Gamma}_y^{(\sqrt{N_t})} \otimes \mathbf{\Gamma}_z^{(\sqrt{N_t})}
\end{bmatrix}
\in \mathbb{C}^{N_t^2 \times (2\sqrt{N_t}-1)^2}
\tag{41}
$$

is the weight matrix and

$$\mathbf{a}_{kl} = \frac{\gamma_k}{\sqrt{L}} (\mathbf{a}_{kl}^y \otimes \mathbf{a}_{kl}^z) \in \mathbb{C}^{(2\sqrt{N_t}-1)^2 \times 1}$$

is a vector. Then for the USPA system, $\mathbf{\Gamma}_u$ of Equation (17) becomes Equation (41) and $\mathbf{\Gamma}_v$ of Equation (17) is constructed similarly as Equation (41); the sizes of vectors $\mathbf{a}_{kl}$ and $\mathbf{b}_{kl}$ of Equation (17) have changed: $\mathbf{a}_{kl} \in \mathbb{C}^{(2\sqrt{N_t}-1)^2 \times 1}$ and $\mathbf{b}_{kl} \in \mathbb{C}^{(2\sqrt{N_r}-1)^2 \times 1}$, and consequently, the size for matrix $\mathbf{C}$ of Equation (18) has changed: $\mathbf{C} \in \mathbb{C}^{(2\sqrt{N_t}-1)^2 \times (2\sqrt{N_r}-1)^2}$. After obtaining the weight matrices, we can follow the process in Section III. C-D to estimate $\mathbf{C}$ and then have the channel covariance matrix estimated as $\widehat{\mathbf{R}} = \mathcal{P}^{-1}(\mathbf{\Gamma}_u \widehat{\mathbf{C}} \mathbf{\Gamma}_v^T)$.

4. Simulations

We now evaluate the performance of our proposed design for fully connected hybrid transceivers with the ULA and USPA.

4.1. The ULA System

We assume a carrier frequency of $f_c = 28$ GHz. For the ULA system, $N_t = 64, N_r = 16, K_t = 16$, and $K_r = 4$. The number of clusters $K \in \{1, 2\}$, and there are $L = 30$ rays in each cluster. The horizontal AoDs and AoAs are generated as Equation (15) with $v_h^t = 10.2°$ and as Equation (16) with $v_h^r = 15.5°$, respectively. The cluster powers are generated following [22] (Table I). We compare the GCG-Alt estimator with the DCOMP estimator in [10], which has varying receiving processing matrices $\mathbf{W}_{t,s}$ and transmitting processing vectors $\mathbf{f}_{t,s}$ during training and has the best performance among other estimators in [10]. The DCOMP estimator needs a dictionary matrix with G_t grid points that is associated with AoD and a dictionary matrix with G_r grid points that is associated with AoA. Let L_p be the number of paths in the channel, the DCOMP estimator assumes that L_p is known. For the DCOMP estimator, we set $G_t = 2N_t = 128, G_r = 2N_r = 32$, and $L_p = r_R$. Based on Figure 2, for $p_e = 0.99$,

$r_R = 18$ and 24 for $K = 1$ and 2, respectively. For the GCG-Alt estimator, we set $\mu = \sigma^2$, $\epsilon = 0.003$, and $\epsilon_a = 0.1$. The performance metric η [10]

$$\eta = \frac{\text{tr}(\widehat{\mathbf{M}}^H \mathbf{R} \widehat{\mathbf{M}})}{\text{tr}(\mathbf{M}^H \mathbf{R} \mathbf{M})}$$

is used to measure how close the subspace of $\widehat{\mathbf{R}}$ is to the subspace of $\mathbf{R}$, where $\widehat{\mathbf{M}} \in \mathbb{C}^{N_t N_r \times r_R}$ and $\mathbf{M} \in \mathbb{C}^{N_t N_r \times r_R}$ are the singular vector matrices of $\widehat{\mathbf{R}}$ and $\mathbf{R}$, respectively. We also use the average of the normalized mean square error

$$\text{NMSE} = \frac{\|\widehat{\mathbf{R}} - \mathbf{R}\|_F^2}{\|\mathbf{R}\|_F^2}$$

to measure their performance.

We set PNR $= 10$ dB and the number of training beams $S = 32$, and compare the GCG-Alt estimator with the DCOMP estimator under different T. With $S = 32$ per snapshot, the sampling ratio at each snapshot is $SK_r / N_r N_t = 12.5\%$. The comparison result shown in Figure 3 suggests that when the sampling ratio per snapshot is 12.5%, our proposed estimator requires fewer snapshots to obtain an $\widehat{\mathbf{R}}$ whose subspace is close to that of $\mathbf{R}$, as compared to the DCOMP estimator. The NMSE result shown in Figure 4 suggests that our proposed GCG-Alt estimator can obtain a more accurate covariance matrix estimate.
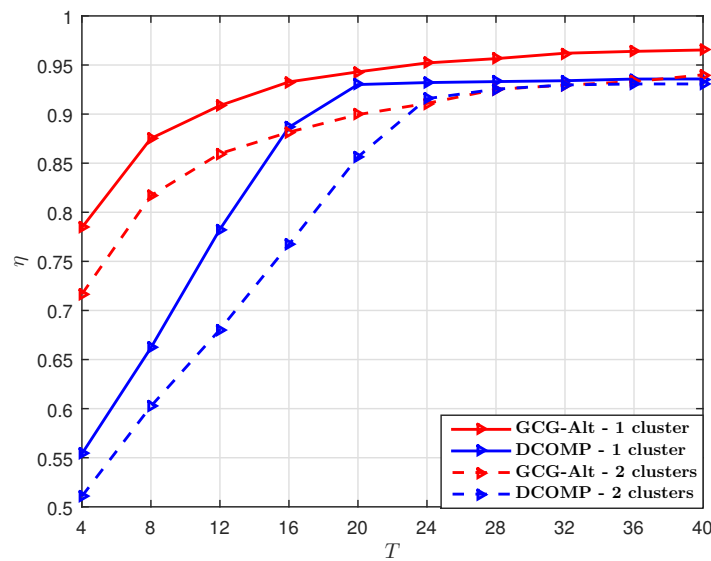

Figure 3. Comparison of η of the GCG-Alt estimator and the DCOMP estimator under the ULA system, where $N_t = 64$, $N_r = 16$, PNR $= 10$ dB, and $S = 32$.
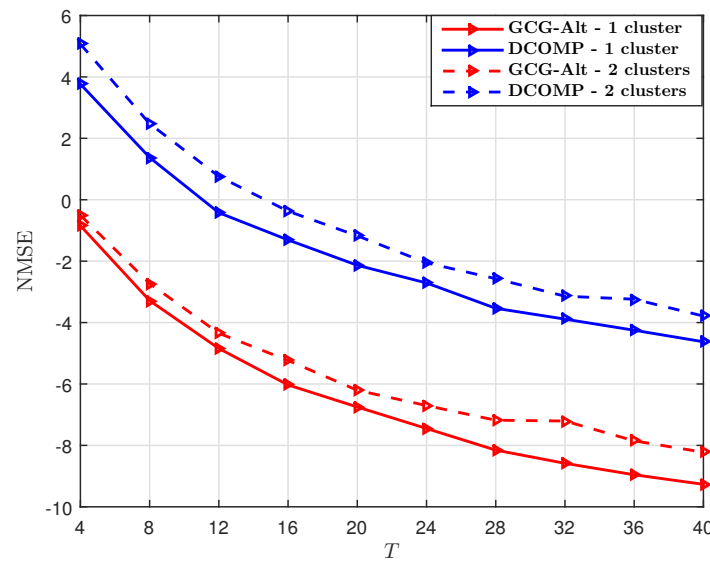
Figure 4. Comparison of NMSE of the GCG-Alt estimator and the DCOMP estimator under the ULA system, where $N_t = 64, N_r = 16, \mathrm{PNR} = 10$ dB, and $S = 32$.

We also compare the computational complexity of the GCG-Alt estimator and the DCOMP estimator. The computational complexity of the DCOMP estimator is about $8TL_pG_tG_r(M^2 + M)$ flops, where $M = SK_r$. For the GCG-Alt estimator, based on our observations, the number of iterations of the AltMin algorithm $I_a \leq 2$, the estimated rank $r_{\mathrm{est}} \approx 4$ when $K = 1$ and $r_{\mathrm{est}} \approx 5$ when $K = 2$. Figure 5 shows the comparison results with different T. We can see that the computational complexity of the GCG-Alt estimator is lower than that of the DCOMP estimator. Also, the computational complexity of the GCG-Alt estimator does not increase as T increases. This is because we use $\mathbf{S}_p \in \mathbb{C}^{S^2 \times K_r^2}$, which is the permutation of the SCM of $\mathbf{y}_t$ shown in Equation (25), and its size is irrelevant to T.
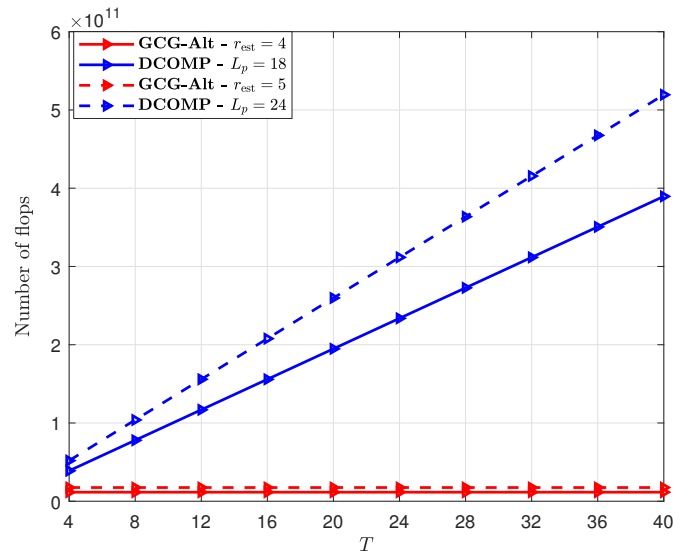

Figure 5. Complexity comparison of the GCG-Alt estimator and the DCOMP estimator under the ULA system, where $N_t = 64, N_r = 16, \mathrm{PNR} = 10$ dB, and $S = 32$.

Then we set the number of snapshots $T = 40$, and compare the GCG-Alt estimator with the DCOMP estimator under different S. The result shown in Figure 6 suggests that when $T = 40$, the GCG-Alt estimator can obtain a more accurate subspace estimation than the DCOMP estimator when the number of training beams $S \geq 24$ per snapshot. Note that $S = 24$ corresponds to a sampling ratio of 9.375% per snapshot.
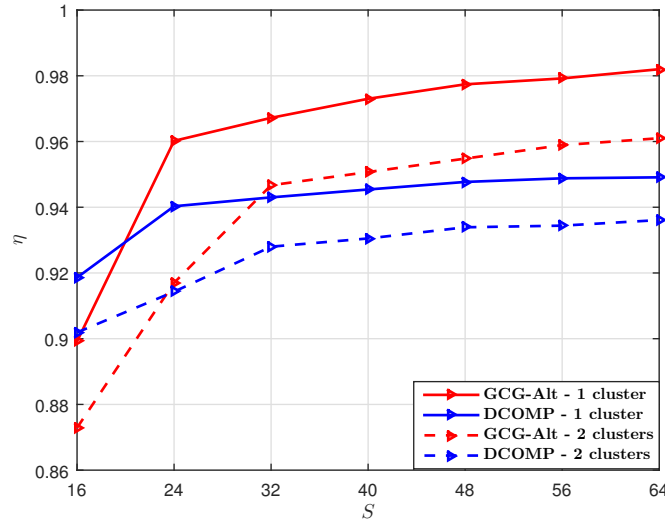
Figure 6. Comparison of η of the GCG-Alt estimator and the DCOMP estimator under the ULA system, where $N_t = 64, N_r = 16, \mathrm{PNR} = 10$ dB, and $T = 40$.

The GCG-Alt estimator explores both the Kronecker structure and the block-Toeplitz–Hermitian structure of **R** while the DCOMP estimator only considers the Hermitian structure of **R**, so the GCG-Alt estimator can reach an accurate subspace estimation of **R** with fewer snapshots. We use the same training for different snapshots while the DCOMP estimator uses different trainings per snapshot (i.e., varying $\mathbf{W}_{t,s}$ and $\mathbf{f}_{t,s}$). When S is small (e.g., $S \leq 16$), the DCOMP estimator outperforms the GCG-Alt estimator. However, the GCG-Alt estimator performs better when S becomes larger (e.g., $S \geq 24$). Note that for the DCOMP estimator, estimating more paths (i.e., L_p is large) yields better performance, but its computational complexity also increases.

4.2. The USPA System

We next consider the system with the USPA at the transmitter and receiver. The parameters f_c, K, L, ϕ_{kl}^t, and ϕ_{kl}^r are assumed the same as in the ULA system. The transmitter has an 8×8 USPA (i.e., $N_t = 64$) and $K_t = 16$ RF chains, and the receiver has a 4×4 USPA (i.e., $N_r = 16$) and $K_r = 4$ RF chains. We assume the elevation AoD angular spread $v_v^t = 0°$ and the elevation AoA angular spread $v_v^r = 6°$ based on the measurement results in [22]. The elevation AoDs and AoAs are distributed as

$$\theta_{kl}^t \sim \mathcal{U}(\theta_k^t - v_v^t, \theta_k^t + v_v^t),$$

$$\theta_{kl}^r \sim \mathcal{U}(\theta_k^r - v_v^r, \theta_k^r + v_v^r),$$

with the elevation center angles θ_k^t and θ_k^r being generated in the same manner as the azimuth center angles in the ULA system. For the DCOMP estimator, we set $G_t = 2\sqrt{N_t} \times 2\sqrt{N_t} = 256$ and $G_r = 2\sqrt{N_r} \times 2\sqrt{N_r} = 64$. The parameters L_p, μ, ϵ, and ϵ_a for the GCG-Alt estimator and DCOMP estimator are the same as in the ULA system.

We set $\mathrm{PNR} = 10$ dB. The performance comparison with $S = 32$ under different T is shown in Figure 7 and the performance comparison with $T = 40$ under different S is shown in Figure 8. We can see that both of the GCG-Alt estimator and the DCOMP estimator achieve higher η for the USPA system. One reason for this is that the USPA system has lower resolution than the ULA system in the azimuth direction even though they have the same number of transmitter and receiver antennas. For the USPA system, the azimuth AoD is resolved by an $\sqrt{N_t} = 8$-element antenna array and the azimuth AoA is resolved by a $\sqrt{N_r} = 4$-element antenna array; while for the ULA system, the azimuth AoD is resolved by a $N_t = 64$-element antenna array and the azimuth AoA is resolved by a $N_r = 16$-element antenna

array. Therefore, for the same angular spread, the USPA system resolves fewer paths than the ULA system, which results in a lower rank.
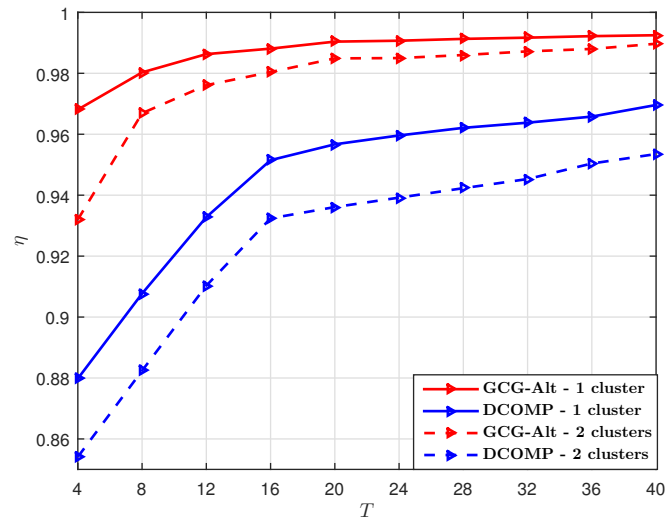

Figure 7. Comparison of η of the GCG-Alt estimator and the DCOMP estimator under the USPA system, where $N_t = 64$, $N_r = 16$, $\mathrm{PNR} = 10$ dB, and $S = 32$.
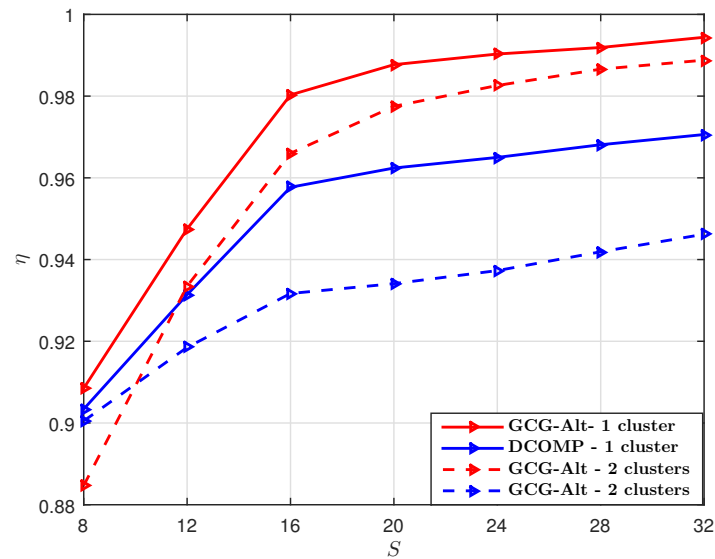

Figure 8. Comparison of η of the GCG-Alt estimator and the DCOMP estimator under the USPA system, where $N_t = 64$, $N_r = 16$, $\mathrm{PNR} = 10$ dB, and $T = 40$.

We also show the effects of angular spreads on the performance of the estimators. We set $v_v^t = 0°$, $v_v^r = 6°$, $K = 1$, $\mathrm{PNR} = 10$ dB, $S = 16$, and $T = 16$. The estimators' performance under different angular spreads for the azimuth AoD/AoA (i.e., different v_h^t and v_h^r) shown in Figure 9 suggests that the estimators achieve lower η when v_h^t and v_h^r are larger.
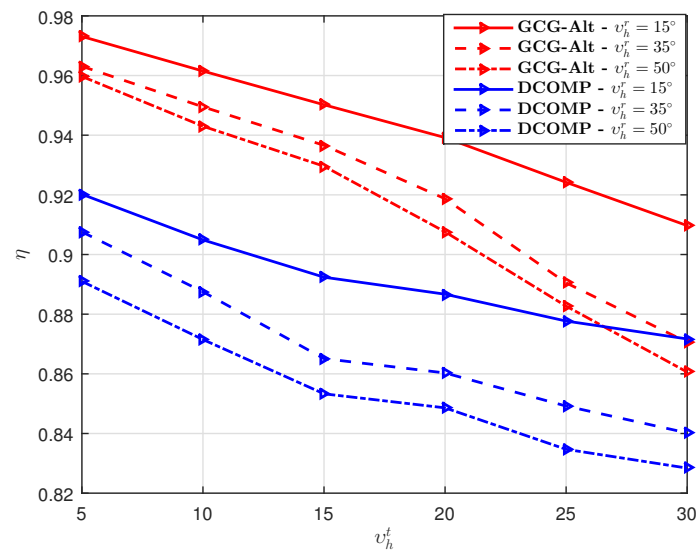
Figure 9. Comparison of η of the GCG-Alt estimator and the DCOMP estimator under the USPA system, where $N_t = 64, N_r = 16, K = 1, \mathrm{PNR} = 10$ dB, $S = 16, T = 16, v_v^t = 0°$, and $v_v^r = 6°$.

5. Conclusions

We have formulated the channel covariance estimation problem for hybrid mmWave systems as a structured low-rank matrix sensing problem by exploiting Kronecker product expansion and the structures of the ULA/USPA. The formulated problem has a reduced dimensionality and is solved by using a low-complexity GCG-Alt algorithm. The computational complexity analysis and numerical results suggest that our proposed method is effective in estimating the mmWave channel covariance matrix.

Author Contributions: Conceptualization, R.H.; Software, R.H.; Writing–original draft preparation validation, R.H.; Supervision, J.T., J.X., Q.G., Y.Y.; Validation, J.T.; Writing–review and editing, J.T.; Resources, J.X.

Abbreviations

The following abbreviations are used in this manuscript:

mmWave	Millimeter wave
MIMO	Multiple-input multiple-output
CSI	Channel state information
CS	Compressive sensing
AoA	Angle of arrival
AoD	Angle of departure
ULA	Uniform linear arrays
USPA	Uniform squared planar arrays
RF	Radio frequency
NMSE	Normalized mean square error
PNR	Pilot-to-noise ratio

References

1. Niu, Y.; Li, Y.; Jin, D.; Su, L.; Vasilakos, A.V. A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges. *Wirel. Netw.* **2015**, *21*, 2657–2676. [CrossRef]
2. Wang, X.; Kong, L.; Kong, F.; Qiu, F.; Xia, M.; Arnon, S.; Chen, G. Millimeter wave communication: A comprehensive survey. *IEEE Commun. Surv. Tuts.* **2018**, *20*, 1616–1653. [CrossRef]

3. Alkhateeb, A.; Ayach, O.E.; Leus, G.; Heath, R.W. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE J. Sel. Top. Signal Process.* **2018**, *8*, 831–846. [CrossRef]

4. Méndez-Rial, R.; Rusu, C.; González-Prelcic, N.; Alkhateeb, A.; Heath, R.W. Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches? *IEEE Access* **2016**, *4*, 247–267. [CrossRef]

5. Ayach, O.E.; Rajagopal, S.; Abu-Surra, S.; Pi, Z.; Heath, R.W. Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 1499–1513. [CrossRef]

6. Yu, X.; Shen, J.-C.; Zhang, J.; Letaief, K.B. Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 485–500. [CrossRef]

7. Lee, J.; Gil, G.-T.; Lee, Y.H. Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications. *IEEE Trans. Commun.* **2016**, *64*, 2370–2386. [CrossRef]

8. Li, Y.; Yin, Q.; Sun, L.; Chen, H.; Wang, H.M. A channel quality metric in opportunistic selection with outdated CSI over Nakagami-*m* fading channels. *IEEE Trans. Veh. Technol.* **2012**, *61*, 1427–1432. [CrossRef]

9. Li, Z.; Han, S.; Sangodoyin, S.; Wang, R.; Molisch, A.F. Joint optimization of hybrid beamforming for multi-user massive MIMO downlink. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 3600–3614. [CrossRef]

10. Park, S.; Heath, R.W. Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing based approach. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 8047–8062. [CrossRef]

11. Adhikary, A.; Nam, J.; Ahn, J.Y.; Caire, G. Joint spatial division and multiplexing–The large-scale array regime. *IEEE Trans. Inf. Theory* **2013**, *59*, 6441–6463. [CrossRef]

12. Park, S.; Heath, R.W. Spatial channel covariance estimation for mmWave hybrid MIMO architecture. In Proceedings of the 2016 50th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 6–9 November 2016.

13. Xie, H.; Gao, F.; Jin, S.; Fang, J.; Liang, Y.C. Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 4206–4218. [CrossRef]

14. Haghighatshoar, S.; Caire, G. Massive MIMO Channel Subspace Estimation From Low-Dimensional Projections. *IEEE Trans. Signal Process.* **2017**, *65*, 303–318. [CrossRef]

15. Zhou, Z.; Fang, J.; Yang, L.; Li, H.; Chen, Z.; Blum, R.S. Low-Rank Tensor Decomposition-Aided Channel Estimation for Millimeter Wave MIMO-OFDM Systems. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1524–1538. [CrossRef]

16. Wen, F.; Garcia, N.; Kulmer, J.; Witrisal, K.; Wymeersch, H. Tensor Decomposition Based Beamspace ESPRIT for Millimeter Wave MIMO Channel Estimation. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, UAE, 9–13 December 2018.

17. Park, S.; Ali, A.; González-Prelcic, N.; Heath, R.W. Spatial channel covariance estimation for the hybrid architecture at a base station: A tensor-decomposition-based approach. In Proceedings of the 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Anaheim, CA, USA, 26–29 November 2018.

18. Chi, Y. Kronecker covariance sketching for spatial-temporal data. In Proceedings of the 2016 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 29 August–2 September 2016.

19. Tsiligkaridis, T.; Hero, A.O. Covariance estimation in high dimensions via kronecker product expansions. *IEEE Trans. Signal Process.* **2013**, *61*, 5347–5360. [CrossRef]

20. Greenewald, K.; Tsiligkaridis, T.; Hero, A.O. Kronecker Sum Decompositions of Space-Time Data. In Proceedings of the 2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), St. Martin, France, 15–18 December 2013.

21. Hu, R.; Tong, J.; Xi, J.; Guo, Q.; Yu, Y. Matrix completion-based channel estimation for mmWave communication systems with array-inherent impairments. *IEEE Access* **2018**, *6*, 62915–62931. [CrossRef]

22. Akdeniz, M.; Liu, Y.; Samimi, M.; Sun, S.; Rangan, S.; Rappaport, T.; Erkip, E. Millimeter wave channel modeling and cellular capacity evaluation. *IEEE J. Sel. Areas Commun* **2014**, *32*, 1164–1179. [CrossRef]

23. Loan, C.F.V. The ubiquitous Kronecker product. *J. Comput. Appl. Math* **2000**, *123*, 85–100. [CrossRef]

24. Sun, G.; He, Z.; Tong, J.; Zhang, X. Knowledge-aided covariance matrix estimation via Kronecker product expansions for airborne STAP. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 527–531. [CrossRef]

25. Zhong, K.; Jain, P.; Dhillon, I.S. Efficient matrix sensing using rank-1 gaussian measurements. In Proceedings of the International Conference on Algorithmic Learning Theory, Banff, AB, Canada, 4–6 October 2015.

26. Yu, A.W.; Ma, W.; Yu, Y.; Carbonell, J.; Sra, S. Efficient structured matrix rank minimization. In Proceedings of the Neural Information Processing Systems Conference, Montreal, QC, Canada, 8–13 December 2014.