2022

Low-Rank Channel Estimation for Millimeter Wave and Terahertz Hybrid MIMO Systems

Khawaja Fahad Masood

Low-Rank Channel Estimation for Millimeter Wave and Terahertz Hybrid MIMO Systems

A thesis submitted in partial fulfillment of the requirements for the award of the degree

Doctor of Philosophy

from

UNIVERSITY OF WOLLONGONG

by

Khawaja Fahad Masood

School of Electrical, Computer and Telecommunications Engineering

December 2022

Statement of Originality

I, Khawaja Fahad Masood, declare that this thesis, submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy, in the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Signed

Khawaja Fahad Masood

December 02, 2022

Abstract

Massive multiple-input multiple-output (MIMO) is one of the fundamental technologies for 5G and beyond. The increased number of antenna elements at both the transmitter and the receiver translates into a large-dimension channel matrix. In addition, the power requirements for the massive MIMO systems are high, especially when fully digital transceivers are deployed. To address this challenge, hybrid analog-digital transceivers are considered a viable alternative. However, for hybrid systems, the number of observations during each channel use is reduced. The high dimensions of the channel matrix and the reduced number of observations make the channel estimation task challenging. Thus, channel estimation may require increased training overhead and higher computational complexity.

The need for high data rates is increasing rapidly, forcing a shift of wireless communication towards higher frequency bands such as millimeter Wave (mmWave) and terahertz (THz). The wireless channel at these bands is comprised of only a few dominant paths. This makes the channel sparse in the angular domain and the resulting channel matrix has a low rank. This thesis aims to provide channel estimation solutions benefiting from the low rankness and sparse nature of the channel. The motivation behind this thesis is to offer a desirable trade-off between training overhead and computational complexity while providing a desirable estimate of the channel.

Firstly, this thesis presents a narrowband channel estimation solution for hybrid MIMO systems. The channel is estimated by a three-stage approach with the first stage exploiting the low-rankness of the channel matrix, the second and third stages exploiting the angular sparsity of the channel. The low-rank estimate of the channel is obtained in the first stage by solving an inductive matrix completion (IMC) problem. Then using the low-rank estimate, the sample covariance matrices (SCMs) for the AoAs and AoDs are obtained. Then spectrum estimation techniques are applied to separately estimate the angles of arrival (AoAs) and angles of departure (AoDs) from the respective SCMs. Finally, channel path gains are estimated by solving the sparse recovery problem.

Secondly, this thesis proposes a channel estimation solution for the intelligent reflective surfaces (IRS)-aided hybrid MIMO systems. By adopting a two-stage approach, the proposed solution progressively estimates the channel parameters. The training for both stages is performed separately. In the first stage of training, the IRS phase shifts are fixed and the precoder/combiners are varied to obtain the training data. The low-rankness of the effective channel between the transmitter-IRS-receiver is exploited by formulating the problem as low-rank matrix recovery and solving it using IMC. Then capitalizing on the low-rank estimate, the transmitter AoDs and receiver AoAs are estimated by utilizing spectrum estimation techniques. During the second stage of training, hybrid precoders/combiners are aligned towards the estimated AoDs/AoAs and the IRS phase shifts are varied. Using the second stage training data the angle difference between the IRS AoAs and AoDs is estimated by applying the least squares (LS) and spectrum estimation. Finally, the composite path gains of the channel are estimated using the estimated angles and second-stage training data.

Thirdly, this thesis presents a channel estimation solution for wideband hybrid MIMO systems. The proposed solution exploits the low-rank nature of the multi-tap wideband channel matrix and the problem is formulated as a low-rank matrix sensing (LRMS) problem. The LRMS formulation offers flexibility in regards to

precoder/combiner design and training symbol transmission but requires additional computational efforts as compared to IMC. To reduce the computational effort, two low-complexity implementations of LRMS are presented. The first implementation adopts an iterative approach to solve the pseudo-inverse problems within LRMS and the second implementation adopts a very sparse sensing matrix. Furthermore, two different performance enhancement approaches are presented. As LRMS suffers performance loss when the channel matrix is either tall or fat, the first performance enhancement approach addresses this problem by adopting matrix reshaping to benefit from the shift-invariance property of uniform arrays. The second performance enhancement approach benefits from the knowledge of the array response. The AoAs/AoDs and channel path gains are estimated using spectrum estimation and sparse recovery, respectively.

Acknowledgments

Any successful journey in this world requires support and sacrifices from those around you. I would like to thank and acknowledge the people without whom I would have not completed this journey.

First of all, I would like to thank my principal supervisor, Dr. Jun Tong, for his endless support and guidance throughout my Ph.D. journey. I consider myself extremely lucky that I was able to complete my Ph.D. under Dr. Jun Tong's supervision. Since day one he has believed in my abilities and encouraged me to challenge myself. Throughout the difficult times, he has shown support and guided me in the right direction. I would like to thank my co-supervisors, Professor Jiangtao Xi and Associate Professor Qinghua Guo for their support and encouragement.

I would like to thank my group members for their valuable input into my work. Specially, Dr. Rui Hu and Dr. Tainle Liu for their valuable time and for sharing their knowledge with me. During this journey, I have found lifelong friends including Mr. Muhammad Ayoub, Mr. Faizan Shoaib, Mr. Rasikh Habib, Mr. Shabbir Ahmad, Mr. Mairaj Soomro, Mr. Muhammad Safdar, Mr. Muhammad Abid, and last but not the least Mr. Mubashir Alam.

I would like to thank HEC Pakistan and UOW for their joint scholarship to fund my Ph.D. degree. I would like to thank COMSATS university of science and technology for allowing me to grow in my field.

Finally, I would like to thank my family members. The sacrifice they have made to see me excel in my field can not be expressed in words. My late mother (Ami) always prayed for me and took care of me. She passed away during the fourth year of my Ph.D. and I will always miss her. I wish she was here to see me achieve my goals. My father (Abu), has been the rock of our family and he has taken care of us all through thick and thin. He is my role model and my decisions in life are based on the moral values he taught me. I would like to thank my Aunt for always keeping me in her prayers and supporting me through the journey. I would like to thank my brother for always guiding me and being there for me through the ups and downs of life. I am grateful to my sisters for their love and support. I am also thankful to my sister-in-law and my cute nieces and nephews including Hasham, Ayan, Hania, Musa, and Shanzy.

Contents

List of Figures

List of Tables

Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
4G	the 4-th generation
5G	the 5-th generation
6G	the 6-th generation
ADC	analog-to-digital converter
AoA	angle of arrival
AoD	angle of departure
CSI	channel state information
dB	decibel
DFT	discrete Fourier transform
EE	energy efficiency
EVD	Eigenvalue decomposition
FBSS	forward-backward spatial smoothing
flops	floating point operations
i.i.d	independent identical distributed
IMC	inductive matrix completion
IRS	intelligent reflective surface

KPI	key performance indicator
LoS	line of sight
LRMC	low-rank matrix completion
LRMS	low-rank matrix sensing
LS	least squares
MIMO	multiple-input multiple-output
MMSE	minimum mean square error
mmWave	millimeter wave
MSE	mean square error
NMSE	normalized mean square error
NP	non-deterministic polynomial-time
OFDM	orthogonal frequency division multiplexing
PNR	pilot to noise ratio
QAM	quadrature amplitude modulation
QoS	quality of service
RF	radio frequency
RIP	restricted isometry property
Rx	receiver
SCM	sample covariance matrix
SE	spectral efficiency
SVD	singular value decomposition
THz	terahertz
TR	Toeplitz rectification
Tx	transmitter
ULA	uniform linear array
UPA	uniform planner array

Chapter 1

Introduction

This chapter provides a general introduction to future wireless communications and an outline of the thesis. The comprehensive literature review is provided in Chapter 2.

The growing demands of high data rates, low latency, high energy efficiency, extended coverage area, etc. are steering the research community toward new technologies. The 5G technology is in the deployment phase in many countries and the foundations of the future technology, 6G, are being laid out. The requirements for 5G were released in International Mobile Telecommunications 2020 (IMT-2020) by the International Telecommunications Union Radio-communication sector (ITU-R). The key performance indicators (KPIs) for the IMT-2020 include peak data rates $\geq$ 10 Gb/s, user data rates $\geq$ 100 MB/s, 3 times the spectral efficiency (SE) as compared to 4G, 100 times energy efficiency (EE) as compared to 4G, 1 ms latency over the air, support for mobility up to 500 km/h and greater than 100 Mb/s cell edge throughput [1–3]. Although 5G is still being deployed, it may lack the support for several future wireless applications such as augmented reality (AR), virtual reality (VR), mixed reality (MR) and industrial automation, etc. [4]. Foreseeing the future needs, ITU has started a focus group on Technologies for Network 2030 (FG NET-2030) as an early step towards 6G [5]. The KPIs suggested for 6G include,

peak data rates $\geq$ 1 Tbs, user experience data rates $\geq$ 1Gb/s, 5 times more SE as compared to 5G, 100 times EE as compared to 5G, $10 - 100\mu$s latency over the air and mobility support up to 1000 km/h [6–8].

To achieve the performance goals of future wireless systems the key candidate technologies include communication at higher frequency bands, massive multiple-input multiple-output (MIMO), hybrid beamforming, and the use of passive reflective surfaces. Communications at higher frequency bands such as millimeter wave (mmWave) bands and terahertz (THz) bands offer large unused bandwidth to cater the needs of higher data rates. However, waves at higher frequency bands incur higher path losses and are more prone to blockage and absorption [2, 9–13]. To compensate for these losses massive MIMO can be used to benefit from beamforming gains due to the increased number of antennas [14–16]. Massive MIMO can also benefit from the small antenna size at higher frequencies and antenna arrays having large antenna elements can be easily deployed. A fully digital massive MIMO system requires a dedicated radio frequency (RF) chain for each antenna element, increasing the energy requirements of the system. To cater energy needs of the massive MIMO, hybrid analog-digital systems with a reduced number of RF chains are considered as a potential solution by the research community [17]. Another technology that will be crucial to ensure the quality of service (QoS) for future wireless networks is intelligent reflective surfaces (IRS) (also referred to as large intelligent surfaces (LIS) and reflective intelligent surfaces(RIS) [18–20]). IRS-aided systems can offer strong propagation paths via IRS in case the direct channel between the transmitter and receiver is blocked.

To fully benefit from the fruits of 5G/6G, all or some of these technologies will be used collectively. From the signal processing perspective, tasks like beamforming [21], MIMO data detection [22], interference alignment [23] etc. require reliable channel state information (CSI). However, the technologies discussed above when combined can make the channel estimation task challenging. For instance, the use of massive MIMO and IRS increases the channel matrix dimensions. This increases

the computational complexity requirements for channel estimation. Also, when the hybrid systems are deployed, the time required for training is increased due to the reduced number of baseband observations. The silver lining here is that at higher frequency bands the channel is sparse in the angular domain and this sparsity results in a low-rank channel matrix. This thesis aims to exploit the sparse and low-rank nature of the channel to provide channel estimation solutions at low complexity and low training overhead.

The remaining of the thesis is organized as follows:

- In Chapter 2, a detailed literature review and thesis motivation are presented.
- In Chapter 3, the narrowband channel estimation solution for the hybrid MIMO systems is discussed. The contributions of this chapter have been published as a journal paper: K. F. Masood, R. Hu, J. Tong, J. Xi, Q. Guo, and Y. Yu, "A Low-Complexity Three-Stage Estimator for Low-Rank mmWave Channels," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5920–5931, 2021.
- In Chapter 4, the channel estimation approach to estimate the IRS-aided hybrid MIMO systems is presented. The work reported in this chapter has been published as journal paper. K. F. Masood, J. Tong, J. Xi, J. Yuan and Y. Yu, "Inductive Matrix Completion and Root-MUSIC-Based Channel Estimation for Intelligent Reflecting Surface (IRS)-Aided Hybrid MIMO Systems," in *IEEE Transactions on Wireless Communications*, doi: 10.1109/TWC.2023.3257138.
- In Chapter 5, the proposed wideband channel estimation solution for hybrid MIMO systems is presented. The proposed work in this chapter has been submitted to *IEEE Journal of Selected Topics in Signal Processing* with the title K. F. Masood, J. Tong, J. Xi, J. Yuan, Q. Guo, and Y. Yu, "Low-Rank Matrix Sensing-Based Channel Estimation for mmWave and THz Hybrid MIMO Systems", minor revision, May. 2023.

- Chapter 6 concludes the contributions of the thesis and discusses the potential future work.

Chapter 2

Literature Review

This chapter presents an extensive literature review on channel estimation schemes for hybrid MIMO systems. The background on the system and channel model is presented in Section 2.1. The details on the existing channel estimation solutions are presented in Section 2.2 and the thesis motivation and contribution are discussed in Section 2.3.

2.1 Research Background

This section is aimed to provide an overview of the key technologies considered in this thesis. Those include massive MIMO, communication at mmWave and THz frequency bands, and IRS-aided systems.

2.1.1 Massive MIMO

For massive MIMO, different transmitter (receiver) architectures can be used, namely, analog structure, digital structure, and hybrid-analog digital structure. The overviews of these architectures are given below.

2.1.1.1 Digital and Analog Structures

For a fully digital structure, each antenna requires a dedicated RF chain. This means that $Q_T = N_T$ and $Q_R = N_R$, where Q_T and Q_R denote the number of RF chains at the transmitter and receiver, respectively and N_T and N_R are the number of antennas at the transmitter and receiver, respectively. So, for a fully digital structure up to N_T (N_R) symbols can be processed at the transmitter (receiver). As suggested by the name the fully digital structure lacks an analog beamformer. A fully digital system can offer the best signal processing capabilities at the cost of increased power requirements.

For analog beamforming, only a single RF chain is shared by all the transmit (receive) antennas i.e. $Q_R = Q_T = 1$. Although this requires low power, the signal processing capabilities are greatly affected due to the reduced number of RF chains. Therefore, for practical systems hybrid analog-digital systems are more viable to reduce the power requirements. Next, we discuss the hybrid structure that combines analog and digital structures.

2.1.1.2 Hybrid Analog-Digital Structure

A hybrid analog-digital structure offers a trade-off between the processing capabilities and power requirements. For hybrid structures, the number of RF chains is less than the number of antennas at the transmitter and receiver, i.e., $Q_R < N_R$ and $Q_T < N_T$, respectively. This reduces power consumption due to the reduced number of RF chains. An RF chain is composed of analog to digital/digital to analog converters and mixers that are in general power hungry. A hybrid structure with networks of fully connected phase shifters is considered for massive MIMO systems in [24, 25]. The term "fully connected" means that each RF chain is connected with all antennas. Other hybrid structures include switch-based and partially connected structures. This thesis focuses on the fully connected hybrid structure. This is motivated by the flexibility offered for beamforming design as compared to switch-based

or partially connected hybrid structures. Hybrid analog-digital structure are considered for both the transmitter and receiver thus making the RF chain requirements considerably less. For illustration, a transmitter and receiver equipped with a fully connected hybrid structure are shown in Fig. 2.1. The total number of phase shifters required for such a structure is $Q_T N_T$ and $Q_R N_R$ for the transmitter and receiver, respectively. The baseband precoder is denoted by $\mathbf{P}_{\text{BB}}$ and the RF precoder is denoted by $\mathbf{P}_{\text{RF}}$. Similarly, the baseband combiner is denoted by $\mathbf{W}_{\text{BB}}$, and the RF combiner is denoted by $\mathbf{W}_{\text{RF}}$. The hybrid transceiver can provide data rates close to that of fully digital but the channel estimation is more challenging due to the reduced number of baseband measurements [9, 26].

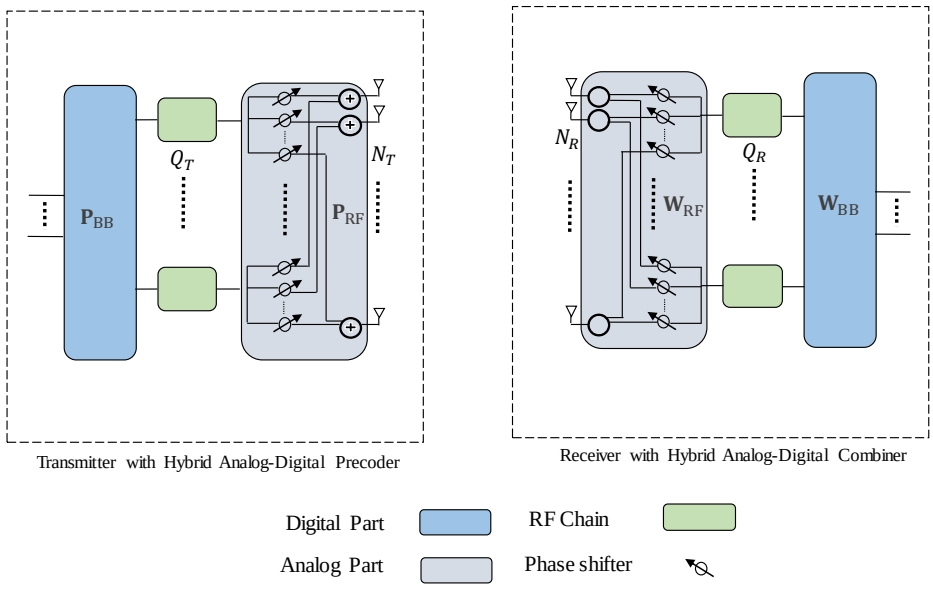


Figure 2.1: Hybrid analog-digital structure with networks of fully connected phase shifters.

2.1.2 Communication in mmWave and THz Bands

Communication at higher frequencies such as mmWave and THz bands has gained significant interests due to their rich spectrum resources. The mmWave band ranges from 30 to 300 GHz and the THz band ranges from 0.1 to 10 THz [27, 28]. It is possible to achieve the peak data rates of requirements of 5G and 6G by utilizing large unused bandwidth available at these frequency bands. Along with large unused

bandwidth, the mmWave/THz signals have shorter wavelengths to enable the use of large antenna arrays lending support for the implementation of massive MIMO systems [29]. Due to large antenna arrays with compact sizes, the beamwidth becomes narrow and this offers security against jamming and eavesdropping.

There are certain challenges for mmWave/THz communication along with the benefits mentioned above. Such high-frequency bands face higher penetration and path loss [2, 11, 30] compared to the sub-6 GHz bands. The path loss for free space is given by Friis transmission formula [11] as

$$P_R(D) = P_T \mathcal{G}_T \mathcal{G}_R (\frac{\lambda_c}{4\pi})^2 D^{-2}, \tag{2.1}$$

where $P_R(D)$, P_T, $\mathcal{G}_T$, $\mathcal{G}_R$, and λ_c represent the received signal power at a distance D, transmit power, transmit antenna gain, receive antenna gain, and carrier wavelength, respectively. Due to the shorter wavelength, mmWave/THz bands suffer from a greater path loss as compared to lower frequency bands. The path loss is also affected by the rain attenuation and air conditions [31]. Penetration loss is also higher at higher frequencies, e.g. the penetration loss for brick and tinted glass at 28 GHz is around 40 dB [32]. Forced by these losses there exist only a few dominant paths between the transmitter and receiver [33, 34]. To compensate for these losses at higher frequencies hybrid massive MIMO and IRS-aided systems can be used along with mmWave/THz communications.

This thesis assumes the widely used geometric representation of the mmWave/THz channel. The geometric representation of the narrowband channel is given as [26, 35]

$$\mathbf{H} = \sum_{l=1}^{L} \gamma_l \mathbf{a}_R(\theta_{R,l}) \mathbf{a}_T^H(\phi_{T,l}), \tag{2.2}$$

where $\theta_{R,l}$, $\phi_{T,l}$ and γ_l represent the angle of arrival (AoA) at the receiver, the angle of departure (AoD) from the transmitter and the complex path gain for the l-th path, respectively, and L denotes the total number of paths between the transmitter and

receiver. Also, $\mathbf{a}_T(\phi_{T,l})$ and $\mathbf{a}_R(\theta_{R,l})$ denote the array response for the transmitter and receiver for the l-th path.The array response vector for N-element uniform linear arrays (ULA)s for angle θ is given as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}}\left[1, \mathrm{e}^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta)}, \cdots, \mathrm{e}^{j(N-1)\frac{2\pi}{\lambda_c}\beta\cos(\theta)}\right]^T, \tag{2.3}$$

where the antenna spacing is $\beta = \lambda_c/2$. The transmitter and receiver array response vectors corresponding to the respective angles are based on (2.3). Alternatively, (2.2) can be rewritten as

$$\mathbf{H} = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}\mathbf{A}_T^H(\boldsymbol{\phi}_T), \tag{2.4}$$

where the columns of $\mathbf{A}_R(\boldsymbol{\theta}_R) \in \mathbb{C}^{N_R\times L}$ and $\mathbf{A}_T(\boldsymbol{\phi}_T) \in \mathbb{C}^{N_T\times L}$ are the array responses at the receiver and transmitter, respectively, and $\boldsymbol{\Gamma}$ is a diagonal matrix whose diagonal elements are γ_l.

This thesis also extends to the wideband scenario using a geometric mmWave/THz wideband multi-tap channel model, where the channel model for the d-th delay tap is given as [36–38]

$$\mathbf{H}_d = \sum_{l=0}^{L-1}\gamma_l p(dT_s - \tau_l)\mathbf{a}_r(\theta_l)\mathbf{a}_t^H(\phi_l), \tag{2.5}$$

where the impulse response of the pulse shaping filter at τ is given as $p(\tau)$ and T_s is the sampling period. Furthermore, (2.5) can also be written as

$$\mathbf{H}_d = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}_{\boldsymbol{d}}\mathbf{A}_T^H(\boldsymbol{\phi}_T), \tag{2.6}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{d}}$ is a diagonal matrix whose diagonal elements are $\gamma_l p(dT_s - \tau_l)$.

Due to a limited number of scatters at higher frequency bands the number of dominant paths is low, for instance, $L \in [3, 5]$ for mmWave channels [34]. This thesis also focuses on the THz channels. The sparsity level further increases when the frequency range changes from mmWave to THz bands. This is due to increased path loss and penetration loss at higher frequency. The THz channels are mostly

dominated by the LOS rays and in the absence of LOS, up to three NLOS paths are dominant [39]. Apart from reduced number of paths due to higher losses at THz frequencies the remaining channel characteristic are similar to mmWave channels. Therefore, the channel models in (2.2) and (2.5) are applicable to mmWave and THz bands. The channel is constructed using channel parameters like AoAs, AoDs, and channel gains and the task of channel estimation can be accomplished by estimating these parameters instead of estimating directly the overall channel matrix.

2.1.3 Intelligent Reflective Surfaces

Due to reduced channel paths at the higher frequency bands as compared to the sub-6 GHz bands [11], mmWave and THz channels are more susceptible to the blockage of propagation paths, which can affect the coverage and QoS of the system. Consequently, IRS has emerged as a potential solution to provide improved coverage and QoS. IRS is usually constructed using passive reflective surfaces or meta-surfaces. They can programmably alter the phase and/or amplitude of the incident signals consuming very low power [4, 40]. This offers a new degree of freedom for more controllable wireless environments and thus improves the communication performance [41]. For example, in the absence of line-of-sight (LOS) paths between the transmitter and receiver, IRS can provide LOS paths from transmitter to IRS and IRS to receiver with strong gains to boost the coverage. A MIMO system with the direct path between the transmitter and the receiver unavailable due to blockage is depicted in Fig 2.2. The first element is turned on (reflection mode) and the remaining elements are turned off (absorption mode).

Although the IRS-assisted system can provide strong paths between the transmitter and receiver via IRS, this significantly increases the channel dimension as compared to that of the direct channel between the transmitter and receiver. This dimension increase depends on the number of elements used in the IRS. The effective channel between the transmitter and receiver can be now written as $\mathbf{H} = \mathbf{G\Omega F} \in$

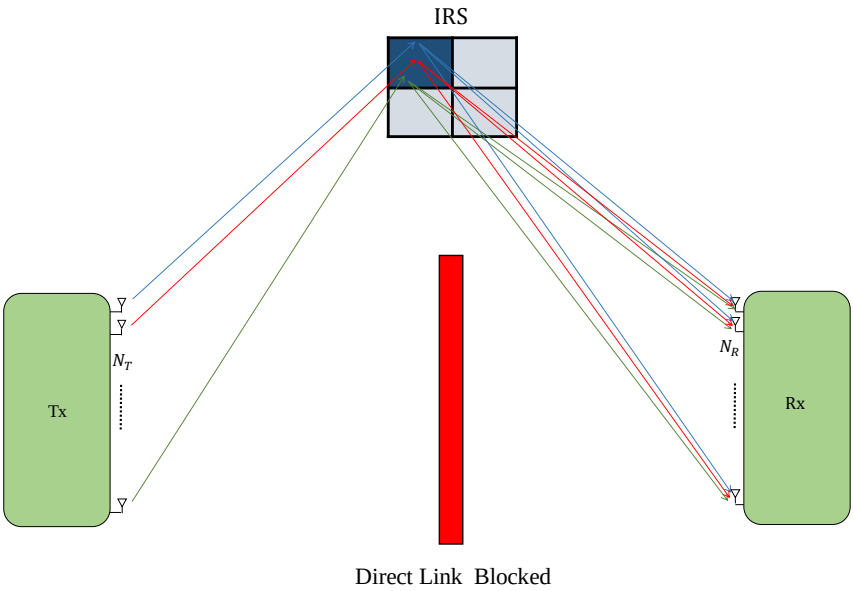


Figure 2.2: An IRS-assisted MIMO system with direct link between the transmitter and receiver blocked.

$\mathbb{C}^{N_R \times N_T}$. With $\mathbf{F} \in \mathbb{C}^{N_I \times N_T}$ and $\mathbf{G} \in \mathbb{C}^{N_R \times N_I}$ denoting the transmitter (Tx)-IRS channel and IRS-receiver (Rx) channel, respectively, $\mathbf{\Omega} \in \mathbb{C}^{N_I \times N_I}$ is a diagonal matrix containing the amplitude and phase of IRS elements and N_I denotes the number of IRS elements.

The individual channels $\mathbf{F}$ and $\mathbf{G}$ can be estimated at the IRS when IRS is equipped with baseband processing capabilities and such IRS elements are termed as active IRS. However, active IRS will increase the cost and power requirements of the system. This thesis only considers passive IRS and the channel estimation is performed at the receiver.

2.2 Existing Channel Estimation Solutions for Hybrid MIMO Systems and Research Challenges

In this section, a review of the existing channel estimation strategies for hybrid MIMO systems is presented. Those include the channel estimation solutions for narrowband mmWave/THz channels, IRS-assisted mmWave/THz channels, and wide-

band mmWave/THz channels.

2.2.1 Narrowband mmWave and THz Channels

Acquiring the CSI is crucial for optimizing the performance of mmWave/THz massive MIMO. However, channel estimation is challenging due to the increased dimensionality of the channel matrix, and the hybrid structure further limits the number of observations. On the other hand, the mmWave/THz channels are sparse in the angular domain with a limited number of dominant paths due to sparse scattering at higher frequencies [42]. This can be exploited to reduce the training overhead. Compressive sensing (CS) techniques have been used to estimate the mmWave channel by jointly finding the AoA, AoD, and channel gain for the dominant paths [35, 43, 44]. There are also solutions that iteratively estimate the AoAs/AoDs and path gains [45], [46]. In [45], an iterative algorithm is proposed for sparse signal recovery and applied to mmWave channel estimation. Sparse message passing (SMP) is used to find the locations of the non-zero entries (corresponding to the AoAs/AoDs) and linear minimum mean squared error (LMMSE) estimation is applied to estimate non-zero entries (path gains). A similar approach is proposed in [46], which combines SMP and least square (LS) to estimate the mmWave channel. In general, CS-based estimators require a dictionary enumerating the candidate AoA and AoD pairs, whose size equals the product of the grid sizes of the AoA and AoD. To tackle the issue of grid mismatch, high-resolution dictionaries of large sizes may be adopted [43]. This, however, may lead to a substantial computational complexity for the sparse recovery process. Beam tracking techniques that exploit the sparse nature of the mmWave MIMO channel and the correlation of AoAs/AoDs for multiple channel realizations are proposed in [47] and [48], where the training beams and transmitting power are optimized to enhance performance.

Low-rank matrix sensing methods that leverage the low-rank nature of massive MIMO channels have been considered in [49–53]. In [50], matrix completion (MC)

and inductive MC (IMC)-based training and estimation schemes are proposed for mmWave MIMO with hybrid transceivers. A recovery algorithm based on the generalized conditional gradient with alternating minimization (GCG-ALTMIN) is designed, which is shown to exhibit lower complexity and also provide more robustness against model mismatch than the CS estimators that often involve high-dimensional dictionaries. However, such MC estimators do not exploit the knowledge of the array response and do not directly provide estimates of the AoAs, AoDs, and gains of the paths. Therefore, they may under-utilize the available *a priori* information about the channel and face challenges when the CSI needs to be fed back to the transmitter. Schemes exploiting both the low-rankness and array responses of the mmWave channel have also been proposed. In [51], a two-stage estimator is proposed to first apply low-rank matrix recovery and then refine the estimation using CS with the assumption of known array response. In [52], the low-rankness and knowledge about the array response are exploited simultaneously and an alternating direction method of multipliers (ADMM)-based approach is used for estimating the channel. Both approaches utilize predefined grids of the angles and high-dimensional dictionaries, whose performance and complexity depend on the grid employed. In [53], a MC-based semi-blind approach is proposed, which exploits payload data to reduce the training requirement and recovers the channel by using regularized alternating least squares and bilinear generalized approximate message passing. In [33], the iterative reweighted super-resolution (IR-SR) estimator is proposed to address the issue of grid mismatch. Initial estimates of the AoAs/AoDs are obtained first and then a gradient descent method is used to iteratively refine the estimates such that super-resolution estimation is achieved.

Traditional spectrum estimation techniques have also been successfully applied to massive MIMO channel estimation [54–59]. In [55], the AoAs and AoDs are jointly estimated using two-dimensional (2D) beamspace multiple signal classification (MUSIC). For frequency-selective channels, [56] proposes to jointly estimate the AoAs, AoDs, and delays for the propagation paths using three-dimensional (3D) rotational

invariant techniques (ESPRIT) in the discrete Fourier transform (DFT) beamspace. This approach is extended in [57] to a two-stage training scheme which first obtains coarse estimates of the direction sectors of interests and then refines the parameter estimates. In [58], specially optimized precoders and combiners are utilized to find the mainbeam, then the peak of the mainbeam is regarded as the LOS path within quantization error, and finally, two approaches are presented to mitigate the quantization error. In [59], a multi-stage training scheme with optimized beamformer design for AoA/AoD estimation followed by angle pairing is proposed, where training data spanning multiple channel realizations are involved. In general, the above spectrum estimation-based estimators need to tailor the precoder and combiner for the hybrid MIMO transceivers and search over high-dimensional parameters.

2.2.2 IRS-Aided mmWave and THz Channels

The channels of IRS-aided systems may be estimated in different manners, depending on whether the IRS possesses baseband signal processing capabilities. With separate channel estimation, the Tx-to-IRS and IRS-to-Rx subchannels are both estimated explicitly. This often requires a certain number of active elements to be deployed at the IRS, leading to semi-passive IRS, so that digital observations can be captured at the IRS. A relatively low training overhead may suffice but this costs increased complexity and power consumption of the IRS. Various schemes for separate channel estimation have been proposed, see [60–63] for examples.

In this thesis, we focus on fully passive IRS-aided MIMO systems with hybrid transceivers. This requires cascaded channel estimation based only on the observations at the hybrid receiver. Due to the high dimensionality, the training overhead can be significantly increased using classical estimators such as the LS. Tremendous efforts have been made to address this crucial challenge, such as [64–76], and a thorough survey can be found in [4]. For example, [64] exploits a bilinear matrix factorization model for the training data. Capitalizing on the sparsity and low-rankness

of the factor matrices, they develop a two-stage estimator by using iterative sparse matrix factorization based on the bilinear generalized approximate message passing (BiG-AMP) algorithm followed by low-rank matrix completion via Riemannian gradient. A two-timescale approach is studied in [65] to reduce the training overhead by exploiting the quasi-static nature of the base station (BS)-to-IRS subchannel. This may require the BS operating in the full-duplex mode capable of self-interference mitigation. In [66], the redundancy in the multi-user cascaded channels is leveraged and a three-phase estimator using LS and LMMSE estimation is designed. Anchor-based solutions are investigated, e.g., in [67]. There are also solutions exploiting machine learning (ML). For example, [68–70] propose to denoise the (interpolated) LS estimates of the cascaded channel using neural networks. Their pilot overhead depends on the requirement of the LS estimation, which in turn depends on the dimensionality and the number of RF chains at the receiver. A conditional generative adversarial network-based solution is also proposed in [71]. Many of the above solutions assume a single antenna at the users and employ single-stage training, which may incur a substantial training overhead due to the lower beamforming gains achievable. Furthermore, they generally do not utilize the knowledge of the array responses at the transmitter, receiver, or IRS, but aim to directly estimate the entries of the channel matrices.

For IRS-aided mmWave and THz MIMO systems, it may also be beneficial to exploit knowledge about the array responses for channel estimation. The associated parametric representations of the cascaded channel may then be employed to reduce the dimensionality of the signal processing problems. Instead of directly estimating the cascaded channel matrix or its factors, parameters such as the AoAs, AoDs, and path gains can be estimated for acquiring the CSI. For example, in [72], the cascaded channel estimation is formulated as a sparse recovery problem, which is then solved using on-grid CS algorithms including orthognal matching pursuit (OMP) and generalized approximate message passing (GAMP). With this scheme, a multi-dimensional dictionary accounting for multiple path directions is used, which may

have a big size that affects the computational complexity. In [73], a similar sparse recovery formulation is developed while a two-stage approach is applied to solve the problem for achieving a lower complexity. First, the AoAs at the Rx and AoDs at the Tx are jointly estimated using 2D CS or super-resolution spectrum estimation techniques (e.g., the beamspace ESPRIT). With such angle information generated, the angles related to the IRS are next estimated using similar techniques. By decoupling the angle estimation into two stages, the complexity can be reduced but 2D CS or spectrum estimation is still required. In [74], multiuser MIMO systems are considered and the channel estimation problem is formulated as a matrix-calibration-based matrix factorization task. The channel sparsity over a predefined dictionary of the path angles and slow variations of the IRS-to-BS channel is exploited to develop a message-passing-based algorithm for channel recovery. In [75], another two-stage estimator based on an iterative reweighted solution to a CS problem is proposed, which can address the grid errors to enhance performance. Note that [72–75] all employ a single-stage training without exploiting prior knowledge of the path directions. By contrast, in a recent work [76], two-stage training is introduced, where the first stage aims to recover the AoAs and AoDs at the Rx and Tx, respectively, and the second stage exploits those angle information to optimize the training scheme. Gridless spectrum estimation based on atomic norm minimization (ANM) is adopted at both stages to recover the angles of interest. This scheme can benefit significantly from the gleaned angle information for achieving beamforming gains at a low feedback overhead. The parameters of the cascaded channel are estimated progressively, reducing the complexity as compared to approaches adopting joint parameter estimation. However, the overall complexity is influenced by the algorithms for solving the multiple semidefinite programming (SDP) problems involved.

2.2.3 Wideband mmWave and THz Channels

In general, the mmWave and THz channels have fewer dominant paths due to higher path loss and penetration loss. This leads to a sparse representation of the channel in the angular domain. Exploiting the sparsity, the channel estimation problem can be formulated as a sparse recovery problem that can be solved by using CS algorithms. For wideband cases, [36] provides three CS-based solutions using the OMP algorithm, in the time domain, frequency domain, or both domains. In [38], the block sparsity in the angular domain is exploited for different channel taps of the wideband channel using the block OMP algorithm. Systems equipped with few-bit ADCs are also considered in [77] and the joint sparsity in the time and angle domains is exploited. The resulting CS problem is solved using the approximate message passing (AMP) algorithm. The beam squint effect at THz bands is considered in [78] and a CS-based channel estimator with hybrid combining is designed for MIMO-OFDM. In [79], a two-step time-domain channel estimator is proposed for wideband hybrid MIMO systems. LS estimation is first applied to provide an initial estimate of the effective channel, then CS is applied to refine the estimate. In [80], the group sparsity between the sub-carriers is exploited by sparse Bayesian learning (SBL). Also, in [81], the sparsity in the delay and angle domains is exploited and an OMP-based solution is proposed for doubly selective channels. Due to the usage of predefined dictionaries, on-grid CS-based schemes are generally prone to power leakage issues caused by grid mismatch.

There have been significant efforts in addressing the grid mismatch issue of CS. For multi-user MIMO-OFDM systems, [82] exploits distributed CS for the initial estimation of the LOS AoA/AoD and then refines the estimates using grid matching pursuit and adaptive measurement matrices. In [83], the common channel sparsity between sub-carriers is exploited for hybrid MIMO-OFDM systems, where initial angle estimation from on-grid CS is refined by using a local search in their neighborhood. A mixed CS-ML (maximum likelihood) approach is proposed in [84],

where the simultaneous weighted (SW)-OMP algorithm is used to provide initial estimates of the AoAs/AoDs, then the estimates are refined by a sparsity-adaptive gradient approach. Deep learning (DL)-CS is utilized in [85] for MIMO-OFDM channels, where offline-trained denoising convolution neural networks (DnCNN) is applied to denoise a correlation matrix of the received signal and the measurement matrix, based on which refined angle estimation can be achieved using CS. SBL is applied in [86] to frequency-domain channel estimation to obtain coarse estimates of the AoDs/AoDs, followed by an iterative refinement using the distributed compressed sensing simultaneous orthogonal matching pursuit (DCS-SOMP) algorithm. Though these methods can alleviate the power leakage issues of on-grid CS, they induce extra online or offline complexity.

Tensor-based channel estimation schemes are also studied for mmWave and THz MIMO-OFDM systems. For example, in [87], the received training data is arranged in an order-3 tensor. Exploiting the low-rankness of the tensor, alternate least squares (ALS) is applied to find the CANDECOMP/PARAFAC (CP) decomposition of the tensor, based on which the AoAs, AoDs, and the delays of the paths are estimated using a correlation approach. A similar formulation of the received training data is adopted in [88], and the structural property of the uniform arrays along with tensor unfolding is utilized to sequentially estimate the path parameters. Both [87] and [88] rely on the knowledge of the array responses to retrieve the path parameters such that the overall channel can be reconstructed. Furthermore, a grid search of the path parameters is required. In [89], tensor-based minimum mean squared error (MMSE) and tensor-based OMP schemes are introduced in order to reduce the complexities as compared to the traditional vector-MMSE and vector-OMP solutions. They either require statistical knowledge of the channel correlation in the different domains or utilize grid search for the path parameters.

Spectrum estimation techniques provide another effective tool for wideband mmWave MIMO channel estimation. In [59], high-resolution estimators exploiting the ESPRIT (estimating signal parameters via rotational invariance techniques) are pro-

posed for MIMO-OFDM systems. To utilize ESPRIT to estimate the AoAs and AoDs separately (followed by angle pairing), a specially tailored multi-stage training scheme is employed. A two-stage gridless approach based on 3-D Unitary Tensor-ESPRIT in the DFT beamspace is developed in [90, 91] for MIMO-OFDM. This approach can exploit *a priori* knowledge about the path directions for refining the training and estimation process. The above ESPRIT-based solutions can achieve super-resolution estimation of the path parameters. Meanwhile, they also require the shift-invariance property of ULAs to hold true.

Most of the above sparsity-exploiting wideband MIMO channel estimators rely on prior knowledge of the array responses. Furthermore, discrete or continuous dictionaries are required, either explicitly or implicitly. In fact, exploiting the low-rankness of the mmWave or THz channel contributes an alternative mechanism for low-overhead channel estimation, which does not necessarily require knowledge of the array responses. This has been demonstrated in several low-rank matrix completion (LRMC)-based estimators/detectors for MIMO channels, such as [49, 50, 53] where robustness against the non-idealities of the array responses can also be observed. Furthermore, LRMC can be used as a building block for constructing more sophisticated channel estimators when knowledge of the array response is also available. Interesting studies include [52, 92], where the low-rankness of the element-domain channel matrix and the sparsity in the beam space are simultaneously incorporated in the channel estimation formulation. The resulting problem is solved iteratively using alternating direction method of multipliers (ADMM), which alternately projects the intermediate estimates onto distinct low-dimensional subspaces spanned, respectively, by the singular vectors and array response vectors. Being powerful in exploiting both the low rankness and sparsity, the solutions still utilize discrete grids of the path angles and require a significant number of iterations to converge.

2.3 Thesis Motivation and Contribution

In this section, we present the motivation and contributions of the thesis. Most of the existing solutions discussed above are unable to fully exploit the sparsity and low rankness of the channels with an acceptable trade-off between estimation performance, computational complexity, and training overhead. This thesis aims to provide channel estimation solutions exhibiting better trade-offs between estimation accuracy, computational effort, and training overhead. Since the channel is sparse and low-rank, this is achieved by estimating the channel parameters. In the following, we discuss the channel estimation solutions for the narrowband, IRS-aided, and wideband systems.

2.3.1 Channel Estimation for Narrowband Hybrid MIMO Systems.

For narrowband channel estimation, we propose a three-stage mmWave massive MIMO channel estimator that exploits both the low-rankness of the mmWave channel and the knowledge of the array response.

- IMC [50] is first applied to provide an initial estimate of the high-dimensional channel matrix. Then the AoAs and AoDs of the propagation paths are retrieved separately by utilizing the subspace-based spectrum estimation algorithm root-MUSIC [93–95]. Finally, the path gains are estimated by solving a sparse recovery problem using OMP [43, 96], which also automatically pairs the AoAs and AoDs for the paths.

- Compared with the MC estimators [50], the proposed estimator achieves enhanced performance by further exploiting the knowledge of the array response, which also yields a sparse representation of the channel estimate, supporting low-overhead feedback. It does not rely on a predefined dictionary, hence avoiding the grid mismatch issue often encountered in CS estimators. Though

also employing traditional spectrum estimation techniques, in contrast to [55–59], the proposed estimator exploits the denoised estimate of the full channel matrix and one-dimensional (1D) spectrum estimation. Each stage is implemented using a low-complexity algorithm and thus the overall complexity is kept low.

- The simulation results suggest that the proposed estimator provides an effective approach for estimating low-rank mmWave MIMO channels.

2.3.2 Channel Estimation for IRS-Assisted Hybrid MIMO Systems

This thesis then proposes a low-complexity, two-stage cascaded channel estimator for IRS-aided MIMO with hybrid transceivers. This approach has a similar two-stage training scheme as [76] but we aim to achieve a different tradeoff among estimation accuracy, training overhead, and computational complexity. Our contributions can be summarized as follows:

- We propose low-cost, multi-step solutions to progressively estimate the cascaded channel parameters including the transmitter AoDs, receiver AoAs, IRS angles (the differences between the AoDs and AoAs at the IRS), and composite path gains. Those include the IMC and LS approaches for obtaining samples sharing the subspaces spanned by the array responses, forward-backward spatial smoothing (FBSS) for addressing coherence issues, and angle estimation based on root-MUSIC. The multiple parameters are estimated separately, each with low complexity. They are finally automatically associated by solving a small-size CS problem for reconstructing the overall cascaded channel. The proposed solution can achieve super-resolution estimation of the channel parameters to address the grid mismatch issue of schemes employing grid-based sparse recovery. Compared to the ANM-based gridless solution [76], our approach can improve the performance by jointly estimating the IRS angles at

Stage 2 and meanwhile reduces the complexity by avoiding multiple uses of SDP.

- We introduce training schemes that are suitable for the proposed estimator. At Stage 1, we adopt the hybrid precoders and combiners that are compatible with IMC, such that an effective channel matrix made of the subchannels and IRS phase shifts can be reconstructed with a low training overhead. This facilitates the estimation of the outer angles (i.e., the transmitter AoDs and the receiver AoAs). For Stage 2, we apply subarray sampling at the IRS while performing training for estimating the IRS angles. This effectively reduces the dimensionality of the LS estimation involved and hence alleviates the training requirement. For the case with uniform planner array (UPA) at the IRS, an L-shaped subarray structure is suggested to enable low complexity, and separate estimation of the azimuth and elevation angles, while maintaining high spatial resolution at the same time.

- Simulation studies are performed to compare the proposed estimator with several recently proposed estimators. It is shown that high-accuracy estimation of the channel parameters can be achieved by the proposed solution, which may yield better performance when there are multiple paths in the channel and the numbers of antennas and training overhead are limited. We also carry out a detailed analysis of the computational complexity.

2.3.3 Channel Estimation for Wideband Hybrid MIMO Systems

This thesis finally investigates low-rank matrix sensing (LRMS)-based estimators for wideband channels in hybrid MIMO systems. By exploiting the low-rankness of the channel matrix, we develop the training schemes, recovery algorithms, and techniques for complexity reduction and performance enhancement. The contributions can be summarized as follows:

- We propose to formulate the channel estimation problem for wideband hybrid MIMO systems as a LRMS problem. The time-domain training schemes based on hybrid transceivers are introduced, which can operate under intersymbol interference and allow flexible training schemes. Then the generalized conditional gradient-alternating minimization (GCG-ALTMIN) algorithm is adapted to jointly recover the channel tap matrices. In order to reduce the computational complexity arising from the channel matrix's high dimensionality in wideband cases, we propose a preconditioned conjugate gradient (PCG) implementation of the LRMS solution, which features a low-complexity preconditioner and an effective scheduling of the computation process. We also introduce a LRMC scheme for the wideband case, as a low-complexity, special instance of the LRMS solution. These LRMS estimators *do not require knowledge of the array responses*, in contrast to many existing solutions.

- We further propose techniques for enhancing the performance of the LRMS channel estimators using partial or full knowledge of the array responses. For cases with a very tall or fat concatenated channel matrix, which often arise in wideband channels with the numbers of the transmitter and receiver antennas differing significantly, we propose to apply rank-preserving reshaping of the channel matrix. This requires only the *shift invariance property* of the antenna arrays and can effectively reduce the training complexity of the LRMS estimators. For cases with *full knowledge of the array responses*, we introduce a spectrum denoising (SD) approach to exploit the array responses shared by the channel taps to enhance the performance at low computational overhead. Treating the outputs from the LRMS/LRMC estimators as noisy samples at the receiver and transmitter arrays, subspace-based super-resolution algorithms, such as the root-MUSIC algorithm for ULAs, are employed to estimate the AoAs/AoDs of the channel paths. Finally, the wideband channel estimate is refined by using the angle information, without extra training

overhead.

- The proposed estimators are compared with several alternative solutions via complexity analysis and simulation studies. We show that the proposed estimators can achieve better estimation accuracy and complexity-performance tradeoffs as compared to the alternative solutions.

Chapter 3

A Low-Complexity Three-Stage Estimator for Low-Rank mmWave Channels

3.1 Introduction

This chapter investigates channel estimation for low-rank mmWave MIMO systems. Hybrid MIMO transceivers equipped with ULAs and phase shifter networks are considered. We propose a novel three-stage channel estimator by exploiting the low-rankness of the channel matrix and knowledge of the array response: We first obtain a low-rank estimate of the channel matrix using IMC; then estimate the AoA and AoD for the propagation paths by solving two 1-D spectrum estimation problems using Toeplitz rectification (TR) and the root-MUSIC algorithm, and finally pair the AoAs and AoDs and estimate the channel gains by solving a sparse recovery problem. Each stage is implemented at low complexity and thus the overall complexity is kept low. *A priori* knowledge about the channel is exploited progressively to enhance the performance. The simulation results suggest significant gains in the channel estimation performance along with sparse representations of the estimated channel.

The rest of the chapter is organized as follows. The system model is introduced

in Section 3.2. The proposed solution is presented in Section 3.3. The simulation results are shown in Section 3.4 and the chapter is summarized in Section 3.5.

3.2 System Model

3.2.1 Hybrid Transceiver

Following [43], we consider a hybrid MIMO system with N_T transmitter antennas and N_R receiver antennas equipped with fully connected networks of phase shifters as shown in Fig. 3.1. The numbers of phase shifters are $Q_T N_T$ and $Q_R N_R$, respectively, at the transmitter and receiver, where Q_T and Q_R are the numbers of radio frequency (RF) chains at the transmitter and receiver, respectively. We assume ULAs for both

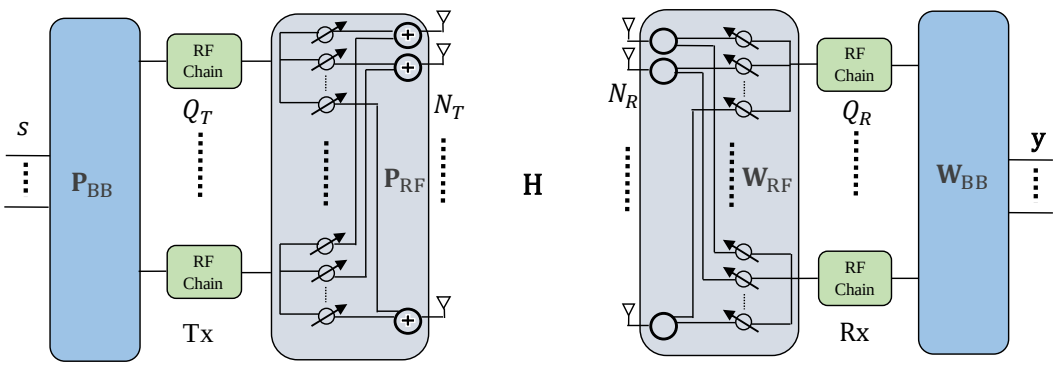


Figure 3.1: System model.

the transmitter and receiver. The channel is modeled as [33], [34]:

$$\mathbf{H} = \sqrt{\frac{N_T N_R}{\rho}} \sum_{l=1}^{L} \gamma_l \mathbf{a}_R(\theta_{R,l}) \mathbf{a}_T^H(\phi_{T,l}), \tag{3.1}$$

where L represents the number of paths, ρ denotes the average path loss, γ_l represents the small-scale channel gain of the l-th path, $\mathbf{a}_T(\phi_{T,l})$ and $\mathbf{a}_R(\theta_{R,l})$ denote the array response vectors of the transmitter and receiver for the l-th path, respectively. Here $\theta_{R,l}$ and $\phi_{T,l}$ represent the AoA and AoD for the l-th path, respectively. The

array response vector for $\theta_{R,l}$ is given as

$$\mathbf{a}_R(\theta_{R,l}) = \frac{1}{\sqrt{N_R}} \left[1, \mathrm{e}^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}, \cdots, \mathrm{e}^{j(N_R-1)\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}\right]^T, \tag{3.2}$$

where λ_c is the carrier wavelength and $\beta = \lambda_c/2$ is the antenna spacing. The array response vector corresponding to $\phi_{T,l}$ is similar. The angles $\phi_{T,l}$ and $\theta_{R,l}$ are modeled to be uniformly distributed in $[0, \pi]$. As L is small [33], the rank of $\mathbf{H}$ is low relative to the dimensionality of the channel matrix in mmWave MIMO systems with large numbers of antennas.

3.2.2 Training Scheme

We consider a training scheme that is compatible with the hybrid transceiver in Fig. 3.1. There are in total M training stages, each with S steps. At the s-th step of the m-th training stage, the received signal is given as

$$\mathbf{y}_{m,s} = \mathbf{W}_{m,s}^H \mathbf{H} \mathbf{p}_m s_{m,s} + \mathbf{W}_{m,s}^H \mathbf{n}_{m,s}, \tag{3.3}$$

where the hybrid precoder $\mathbf{p}_m$ remains fixed during the S steps of each training stage, whereas the hybrid combiner $\mathbf{W}_{m,s}$ changes for every step. The number of stages is set equal to the number of transmit antennas, whereas the number of steps within each stage depends on the number of RF chains. Here we assume $S > 1$. One symbol $s_{m,s}$ is transmitted and Q_R signals are received. The hybrid precoder for the m-th stage is given as $\mathbf{p}_m = \mathbf{P}_{RF_m}\mathbf{p}_{BB_m}$, with $\mathbf{P}_{RF_m} \in \mathbb{X}^{N_T \times Q_T}$ as the RF precoder and $\mathbf{p}_{BB_m} \in \mathbb{C}^{Q_T \times 1}$ as the baseband precoder, where $\mathbb{X}$ represents the set of phase shifts. Similarly, the hybrid combiner for the s-th step of the m-th stage is given as $\mathbf{W}_{m,s} = \mathbf{W}_{RF_{m,s}}\mathbf{W}_{BB_{m,s}}$ with $\mathbf{W}_{RF_{m,s}} \in \mathbb{X}^{N_R \times Q_R}$ as the RF combiner and $\mathbf{W}_{BB_{m,s}} \in \mathbb{C}^{Q_R \times Q_R}$ as the baseband combiner. The noise vector at the receiver array is assumed to having i.i.d entries distributed as $\mathcal{CN}(0, \sigma_n^2)$. Without loss of generality, we assume $s_{m,s} = \sqrt{P}$, $\|\mathbf{f}_m\|_F^2 = 1$ and the total transmit

power $\|\mathbf{f}_m s_{m,s}\|_F^2 = P$. After completing the S steps of the m-th training stage, the baseband received signal can be written as

$$\mathbf{y}_m = \sqrt{P}\mathbf{W}_m^H \mathbf{H}\mathbf{p}_m + \mathbf{n}_m, \tag{3.4}$$

where

$$\mathbf{y}_m = [\mathbf{y}_{m,1}^T, \mathbf{y}_{m,2}^T, \ldots, \mathbf{y}_{m,S}^T]^T,$$

$$\mathbf{W}_m = [\mathbf{W}_{m,1}, \mathbf{W}_{m,2}, \ldots, \mathbf{W}_{m,S}],$$

$$\mathbf{n}_m = [\mathbf{n}_{m,1}^T \mathbf{W}_{m,1}^*, \mathbf{n}_{m,2}^T \mathbf{W}_{m,2}^*, \ldots, \mathbf{n}_{m,S}^T \mathbf{W}_{m,S}^*]^T.$$

The received signal once M training stages are completed is given as

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M]. \tag{3.5}$$

The number of stages is set to $M = N_T$ and during each stage, N_s/N_T distinct observations are acquired. In total, N_s scalar observations are obtained at the receiver and the sampling ratio is defined as $r = N_s/(N_T N_R)$. Alternatively, each training stage provides SQ_R observations and the total number of distinct observations is equal to MSQ_R.

The hybrid precoder and combiner are designed in such a way that (noisy) observations of the entries of a transformed channel matrix

$$\mathbf{C} = \mathbf{X}_R^H \mathbf{H} \mathbf{X}_T \tag{3.6}$$

are obtained using the uniform spatial sampling scheme (USS) [97]. In (3.6), the matrices $\mathbf{X}_R \in \mathbb{C}^{N_R \times N_R}$ and $\mathbf{X}_T \in \mathbb{C}^{N_T \times N_T}$ are the receiver and transmitter feature matrices, respectively. In order to guarantee the performance of recovering $\mathbf{C}$ and $\mathbf{H}$, they can be chosen to approximate unitary matrices satisfying certain incoherence properties by using the matrix decomposition method in [50, Section D]. When

the feature matrices are given, the precoder $\mathbf{p}_m$ and combiner $\mathbf{W}_{m,s}$ can be set as randomly chosen subsets of the columns of $\mathbf{X}_T$ and $\mathbf{X}_R$, respectively, to realize the USS sampling. For instance, the first two entries of the first column of $\mathbf{C}$ with noise $\widetilde{\mathbf{C}}$ can be written as

$$\begin{bmatrix} [\widetilde{\mathbf{C}}]_{1,1} \\ [\widetilde{\mathbf{C}}]_{2,1} \end{bmatrix} = \mathbf{y}_{m,s} = \mathbf{W}_{m,s}^H \mathbf{H} \mathbf{p}_m s_{m,s} + \mathbf{W}_{m,s}^H \mathbf{n}_{m,s}, \tag{3.7}$$

where the combiner for the s-th step of the m-th stage is given as $\mathbf{W}_{m,s} = \mathbf{X}_R(:, 1:2)$ and the precoder for the m-th stage is given as $\mathbf{p}_m = \mathbf{X}_T(:, 1)$. The precoder $\mathbf{p}_m$ and combiner $\mathbf{W}_{m,s}$ can be implemented using the hybrid system in Fig. 3.1. For the combiner, the following optimization problem can be formulated:

$$\begin{aligned} \underset{\mathbf{W}_{RF_{m,s}}, \mathbf{W}_{BB_{m,s}}}{\text{minimize}} \quad & \left\| \mathbf{W}_{m,s} - \mathbf{W}_{RF_{m,s}} \mathbf{W}_{BB_{m,s}} \right\|_F^2 \\ \text{subject to} \quad & \mathbf{W}_{RF_{m,s}} \in \mathbb{X}^{N_R \times Q_R}, \end{aligned} \tag{3.8}$$

where $\mathbf{W}_{RF_{m,s}}$ and $\mathbf{W}_{BB_{m,s}}$ denote the RF and baseband combiner, respectively. In order to approximately solve this problem, we can adapt the PE-ALTMIN algorithm [98] to update $\mathbf{W}_{RF_{m,s}}$ and $\mathbf{W}_{\mathrm{BB}_{m,s}}$ iteratively until convergence. In particular, at each iteration, the discrete phase shift constraint on $\mathbf{W}_{RF_{m,s}}$ is relaxed first to the unit-modulus constraint as in [98] and then the analog combiner coefficients obtained are projected onto the set $\mathbb{X}$ using the nearest distance rule. Once the hybrid combiners are designed, $\mathbf{X}_R$ is updated as the effective combiner $\mathbf{W}_{RF_{m,s}} \mathbf{W}_{BB_{m,s}}$. The precoder $\mathbf{p}_m$ is implemented similarly, followed by a normalization step to ensure $\|\mathbf{p}_m\|_F^2 = 1$, i.e., to meet the transmission power constraint. The precoder/combiner design is summarized in Algorithm 1.

The training process probes a subset of the entries of the transformed channel matrix $\mathbf{C}$. The resulting sampling pattern is recorded as a sampling operator $P_\Omega(.)$.

After training, the noisy sample of the transformed matrix $\mathbf{C}$ is denoted by

$$[P_\Omega(\widetilde{\mathbf{C}})]_{i,j} = \begin{cases} [\widetilde{\mathbf{C}}]_{i,j}, & (i,j) \in \Omega \\ 0, & \text{otherwise} \end{cases}, \tag{3.9}$$

where $[\widetilde{\mathbf{C}}]_{i,j}$ denotes the (i,j)-th entry of $\widetilde{\mathbf{C}}$ and Ω represents the sampling domain. The training scheme is also included in Algorithm 1 for ease of implementation.

With the above training scheme, each sample in $P_\Omega(\widetilde{\mathbf{C}})$ is a rank-1 sample of the original channel matrix $\mathbf{H}$. We can recover $\mathbf{C}$ first by using MC techniques and then compute $\mathbf{H}$. This IMC approach is an instance of low-rank matrix sensing. When the feature matrices are set to $\mathbf{X}_T = \mathbf{I}_{N_T}$ and $\mathbf{X}_R = \mathbf{I}_{N_R}$, we have $\mathbf{C} = \mathbf{H}$ and IMC reduces to MC where the entries of the original channel matrix are sensed. IMC can be more easily incorporated into the hybrid transceiver than MC. This is because the latter requires that only a subset of the transmitting or receiving antennas can be switched on at each training step, which requires special design of the precoder and combiners for phase shifter-based transceivers [50]. The IMC approach allows all antennas to be activated simultaneously, which leads to a lower dynamic range of the transmission power of different antennas. Therefore, we adopt the IMC approach in this chapter, but other matrix sensing approaches can be used alternatively.

3.3 The Three-Stage Channel Estimator

We introduce in this section the three-stage estimator for estimating the channel $\mathbf{H}$ in the system described in Section 3.2. We aim to achieve low-complexity, high-performance estimation with a low training overhead by exploiting both the low-rankness and array response knowledge of the mmWave channel. The proposed estimator successively employs low-rank matrix sensing, spectrum estimation, and sparse recovery at the three stages to exploit various *a priori* knowledge about

the channel. Low-complexity techniques, i.e., the GCG-ALTMIN algorithm, the root-MUSIC algorithm, and the OMP algorithm are used. The integration of these algorithms addresses their respective limitations and hence results in improved performance. In particular, the array response knowledge, which is missed in Stage 1, is exploited in Stage 2 to learn the path angles. Meanwhile, the initial low-rank estimate of the channel matrix from Stage 1 provides Stage 2 with denoised inputs for applying 1D root-MUSIC for super-resolution AoA and AoD estimation. In Stage 3, the high complexity and grid mismatch challenge generally encountered in OMP estimators are addressed by the estimates of the path angles obtained from Stage 2. The rank knowledge gleaned from Stage 1 also facilitates the recovery of the path information in the subsequent stages.

3.3.1 Stage 1: Low-Rank Channel Estimation

With full-rank $\mathbf{X}_R$ and $\mathbf{X}_T$ in (3.6), $\mathbf{C}$ is a low-rank matrix sharing rank with $\mathbf{H}$. This is exploited to recover $\mathbf{C}$ from its sampled entries in (3.9) by formulating the following low-rank matrix sensing problem:

$$\min_{\widehat{\mathbf{C}}} \operatorname{rank}(\widehat{\mathbf{C}}), \quad \text{s.t.} \quad \|P_\Omega(\widehat{\mathbf{C}}) - P_\Omega(\widetilde{\mathbf{C}})\|_F^2 \leq \delta^2, \tag{3.10}$$

where δ^2 is a tolerance to account for the noise and $\|\cdot\|_F$ represents the Frobenius norm. The above problem is NP-hard. We can adopt nuclear norm regularization to reformulate the problem and estimate the transformed channel matrix as [50]

$$\widehat{\mathbf{C}}^\star \triangleq \arg\min_{\widehat{\mathbf{C}}} \quad \frac{1}{2}\|P_\Omega(\widehat{\mathbf{C}}) - P_\Omega(\widetilde{\mathbf{C}})\|_F^2 + \mu\|\widehat{\mathbf{C}}\|_*, \tag{3.11}$$

where $\mu > 0$ is the regularization parameter and $\|\cdot\|_*$ represents the nuclear norm.

Various algorithms can be used to solve the above problem. We adopt the GCG-ALTMIN algorithm [50] which is based on the generalized conditional gradient framework. The GCG-ALTMIN algorithm iteratively refines the low-rank estimate

of $\mathbf{C}$ by using the top singular vectors of the residual error matrix and alternately updating the small-size bilinear factors for local minimization. The algorithm features low complexity at each iteration and fast convergence, as demonstrated in [50]. Once the transformed matrix $\widehat{\mathbf{C}}^{\star}$ is estimated, the low-rank channel matrix can be estimated as

$$\widehat{\mathbf{H}} = (\mathbf{X}_R^H)^{-1}\widehat{\mathbf{C}}^{\star}(\mathbf{X}_T)^{-1}. \tag{3.12}$$

In summary, this stage applies IMC to produce an estimate of the full channel matrix $\mathbf{H}$ from partial observations of the transformed channel matrix $\mathbf{C}$, addressing the challenge of training shortage in hybrid MIMO systems. It does not exploit array response or angle grids and can thus be used alone as a low-complexity robust mmWave channel estimator [50]. However, this scheme may perform worse than channel estimators that exploit the knowledge of the channel response. The rank $\widehat{L}$ of the estimate $\widehat{\mathbf{H}}$ equals to the number of GCG iterations, which is typically low.

3.3.2 Stage 2: AoA and AoD Estimation

Stage 1 exploits the low-rankness of the channel to obtain an initial channel estimate $\widehat{\mathbf{H}}$. Based on $\widehat{\mathbf{H}}$, this stage exploits the knowledge of the array response to further estimate the AoAs $\boldsymbol{\theta}_R = \{\theta_{R,1}, \theta_{R,2}, \ldots, \theta_{R,L}\}$ and AoDs $\boldsymbol{\phi}_T = \{\phi_{T,1}, \phi_{T,2}, \ldots, \phi_{T,L}\}$ of the propagation paths. This will be used in Stage 3 to improve the channel estimation accuracy and also produce a parametric representation of the channel matrix. There are various spectrum estimation techniques [93–95, 99–102]. Among those, root-MUSIC has a low computational complexity for ULAs as it does not require spectrum search and only relies on polynomial roots. We thus consider 1D root-MUSIC [93–95] here. The standard 1D root-MUSIC algorithm for angle estimation requires highly precise knowledge of the 'signal subspace' for the observation of an array, which is typically obtained from training samples of an abundant number or low noise level. This knowledge, however, is not directly available at the hybrid receiver when the training scheme in Section 3.2 is applied. In the following, we

discuss an approach exploiting $\widehat{\mathbf{H}}$.

The channel estimate $\widehat{\mathbf{H}}$ in (3.12) can be written as

$$\widehat{\mathbf{H}} = \mathbf{H} + \mathbf{E}, \tag{3.13}$$

where $\mathbf{E}$ represents the estimation error in Stage 1. Note that (3.1) can be rewritten as

$$\mathbf{H} = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}\mathbf{A}_T^H(\boldsymbol{\phi}_T), \tag{3.14}$$

where the columns of $\mathbf{A}_R(\boldsymbol{\theta}_R) \in \mathbb{C}^{N_R \times L}$ and $\mathbf{A}_T(\boldsymbol{\phi}_T) \in \mathbb{C}^{N_T \times L}$ are the array responses at the receiver and transmitter, respectively, and $\boldsymbol{\Gamma}$ is a diagonal matrix whose diagonal elements are $\sqrt{\frac{N_T N_R}{\rho}}\gamma_l$. Replacing $\mathbf{H}$ with (3.14), we have

$$\widehat{\mathbf{H}} = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}\mathbf{A}_T^H(\boldsymbol{\phi}_T) + \mathbf{E}. \tag{3.15}$$

We now estimate the AoAs and AoDs of the channel paths separately from the above channel estimate, utilizing the knowledge of the array response. When the channel estimation is perfect, i.e., $\mathbf{E} = \mathbf{0}$, it can be easily seen that the subspaces spanned by the columns and rows of $\widehat{\mathbf{H}}$ are the same as those spanned by the array response vectors corresponding to the AoAs and AoDs, respectively. Classical subspace methods such as root-MUSIC can be applied to recover these angles. However, they are sensitive to estimation errors. To see this, let us consider the AoA estimation. We treat the columns of $\widehat{\mathbf{H}}$ as samples of the received signal of an imagined ULA:

$$\mathbf{x}_n = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\lambda}_n + \mathbf{e}_n, n = 1, 2, \ldots, N_T, \tag{3.16}$$

where $\boldsymbol{\lambda}_n$ is the n-th column of $\boldsymbol{\Gamma}\mathbf{A}_T^H(\boldsymbol{\phi}_T)$ that serves the source for generating the observation $\mathbf{x}_n$. The root-MUSIC algorithm utilizes the signal and noise subspaces

estimated from the sample covariance matrix (SCM) of $\mathbf{x}_n$:

$$\begin{aligned}\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R} =& \frac{1}{N_T}\sum_{n=1}^{N_T}\mathbf{x}_n\mathbf{x}_n^H = \frac{1}{N_T}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H \\ =& \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Delta}\mathbf{A}_R^H(\boldsymbol{\theta}_R) + \boldsymbol{\Sigma},\end{aligned} \tag{3.17}$$

where the covariance matrix of the source is given by

$$\boldsymbol{\Delta} \triangleq \frac{1}{N_T}\boldsymbol{\Gamma}\mathbf{A}_T^H(\boldsymbol{\phi}_T)\mathbf{A}_T(\boldsymbol{\phi}_T)\boldsymbol{\Gamma}^H, \tag{3.18}$$

and the error of covariance matrix estimation is

$$\boldsymbol{\Sigma} \triangleq \frac{1}{N_T}\left(\mathbf{H}\mathbf{E}^H + \mathbf{E}\mathbf{H}^H + \mathbf{E}\mathbf{E}^H\right). \tag{3.19}$$

From the above, when the estimation error $\mathbf{E}$ vanishes, the signal and noise subspaces can be correctly estimated from the eigenvectors of $\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}$. However, the non-vanishing $\mathbf{E}$ can significantly degrade the estimation of the subspaces: With a small number of samples N_T, $\boldsymbol{\Sigma}$ is generally far from (scaled) identity matrix and the subspace methods can break down [103, 104]. There are also, however, a number of useful observations. First, when $N_T \to \infty$ and the entries of $\mathbf{E}$ are i.i.d. with zero mean, $\boldsymbol{\Sigma}$ becomes a (scaled) identity matrix. Furthermore, $\boldsymbol{\Delta}$ becomes diagonal. In this case, $\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}$ becomes a Toeplitz matrix whose principal eigensubspace corresponds to the signal subspace. This is because as $N_T \to \infty$, the spectral resolution of the ULA becomes infinitely high and the steering vectors in $\mathbf{A}_T(\boldsymbol{\phi}_T)$ corresponding to different AoAs become nearly orthogonal. In order to compensate for the non-vanishing errors and finite number of samples, inspired by the above observations, we exploit the asymptotic characteristic of $\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}$ and apply TR, which has been successfully applied to address similar issues [104–107], to improve the estimation of the covariance matrix and also the subspaces. The TR of SCM can be obtained

using

$$\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}) = \sum_{m=-(N_R-1)}^{N_R-1} \frac{1}{N_R - |m|} \text{Tr}\ (\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R} \mathbf{Q}^m) \mathbf{Q}^{-m}, \tag{3.20}$$

where $\mathbf{Q} \in \mathbb{C}^{N_R \times N_R}$ is a shift matrix having superdiagonal entries equal to 1 and all other entries equal to 0 and $\text{Tr}(\cdot)$ represents the matrix trace operation. Here $\mathbf{Q}^{-m} = (\mathbf{Q}^T)^m$ and $\mathbf{Q}^0 = \mathbf{I}_{N_R}$. In other words, TR refines the estimate of the covariance matrix by averaging the diagonal and subdiagonal entries of the SCM. After obtaining the TR of SCM, root-MUSIC is applied to $\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R})$ to estimate the AoAs. This can be done by using the eigenvalue decomposition (EVD) of $\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R})$ given as

$$\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}) = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^H, \tag{3.21}$$

where $\mathbf{U}$ represents the eigenvector matrix and $\boldsymbol{\Lambda}$ is diagonal with eigenvalues in the descending order as the diagonal entries. Let the rank of $\widehat{\mathbf{H}}$ be $\widehat{L}$, $\mathbf{U}_s \in \mathbb{C}^{N_R \times \widehat{L}}$ consist of the eigenvectors of $\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R})$ that correspond to the $\widehat{L}$ largest eigenvalues, and $\mathbf{U}_n \in \mathbb{C}^{N_R \times (N_R - \widehat{L})}$ correspond to the remaining $N_R - \widehat{L}$ eigenvalues. It can be seen that when $\mathbf{E} = \mathbf{0}$, $\mathbf{A}_R(\boldsymbol{\theta}_R)$, $\mathbf{U}_s$ and the columns of $\widehat{\mathbf{H}}$ span the same 'signal' subspace. Let the correlation matrix be

$$\mathbf{V} \triangleq \mathbf{U}_n \mathbf{U}_n^H. \tag{3.22}$$

Then a pseudo-spectrum used for locating the AoAs can be written as [94], [95]:

$$P(\theta) = \frac{1}{|\mathbf{a}^H(\theta) \mathbf{V} \mathbf{a}(\theta)|}. \tag{3.23}$$

Clearly, when θ is the AoA of a propagation path and $\mathbf{E} = \mathbf{0}$, $P(\theta)$ tends to infinity and $P(\theta)$ exhibits a peak. Meanwhile, we can rewrite the denominator of (3.23) as

$$\begin{aligned}\mathbf{a}^H(\theta)\mathbf{V}\mathbf{a}(\theta) &= \sum_{k=0}^{N_R-1}\sum_{m=0}^{N_R-1} e^{-jku\cos(\theta)} V_{(k,m)} e^{jmu\cos(\theta)} \\ &= \sum_{q=-N_R+1}^{N_R-1} v_q e^{-jqu\cos(\theta)},\end{aligned} \tag{3.24}$$

where $u = \frac{2\pi}{\lambda_c}\alpha$, $V_{(k,m)}$ denotes the (k,m)-th entry of $\mathbf{V}$ and

$$v_q = \sum_{\{k,m:k-m=q\}} V_{(k,m)}$$

is the sum of entries of the qth diagonal of $\mathbf{V}$. It can be verified that $v_q = v_{-q}^*$. Note that v_0 corresponds to the main diagonal of $\mathbf{V}$. We can further rewrite (3.24) as a polynomial

$$B(z) = \sum_{q=-N_R+1}^{N_R-1} v_q z^q, \tag{3.25}$$

where $z \triangleq \mathrm{e}^{-ju\cos(\theta)}$. There are $2(N_R-1)$ roots of $B(z)$. The $\widehat{L}$ roots, denoted by $\{z_1, z_2, \cdots, z_{\widehat{L}}\}$, that are closest to the unit circle, are chosen to yield the AoAs of the paths as

$$\widehat{\theta}_{R,l} = \cos^{-1}\left(\frac{1}{u}\angle z_l\right), \quad l = 1, 2, \cdots, \widehat{L}, \tag{3.26}$$

where $\angle z_l$ is the phase angle of z_l.

Similarly, the SCM for estimating the AoD can be written as

$$\widehat{\mathbf{R}}_{\phi_T} = \frac{1}{N_R}\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}, \tag{3.27}$$

and the AoDs for the paths $\widehat{\boldsymbol{\phi}}_T$ can be estimated by applying the TR to $\widehat{\mathbf{R}}_{\phi_T}$ followed by the root-MUSIC algorithm.

In summary, this stage estimates $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\phi}}_T$ separately from an initial channel

estimate using the standard 1D root-MUSIC procedure involving knowledge of the receiver array and transmitter array, respectively. Compared to the CS-based channel estimators such as [43] that jointly estimate the AoA and AoD for each path, the angle estimation here has reduced dimensionality and thus can be implemented with low complexity. Furthermore, root-MUSIC does not rely on a predefined angle grid and is able to achieve super-resolution for angle estimation. The 1D spectrum estimation here also distinguishes the proposed estimator from the treatments in [55–59], where 2D or 3D spectrum estimation is used.

3.3.3 Stage 3: Path Gain Estimation

The sets of the AoA estimates $\widehat{\boldsymbol{\theta}}_R$ and AoD estimates $\widehat{\boldsymbol{\phi}}_T$ obtained from Stage 2 are now used together with the initial estimate $\widehat{\mathbf{H}}$ obtained from Stage 1 to refine the channel estimation. We model the channel matrix as

$$\widehat{\mathbf{H}} = \mathbf{A}_R(\widehat{\boldsymbol{\theta}}_R)\boldsymbol{\Gamma}\mathbf{A}_T^H(\widehat{\boldsymbol{\phi}}_T). \tag{3.28}$$

Assuming perfect estimates of the angles, $\boldsymbol{\Gamma}$ is a sparse channel gain matrix. Since the correspondence between $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\phi}}_T$ is unknown yet, the matrix $\boldsymbol{\Gamma}$ is not necessarily diagonal but is assumed to have $\widehat{L}$ nonzero entries, where $\widehat{L}$ is the rank of $\widehat{\mathbf{H}}$. After vectorization we have

$$\begin{aligned}\widehat{\mathbf{h}} &\triangleq \mathrm{vec}(\widehat{\mathbf{H}}) = \left(\mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T) \otimes \mathbf{A}_R(\widehat{\boldsymbol{\phi}}_R)\right)\mathrm{vec}(\boldsymbol{\Gamma}) \\ &= \left(\mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T) \otimes \mathbf{A}_R(\widehat{\boldsymbol{\phi}}_R)\right)\boldsymbol{\gamma} \\ &\triangleq \boldsymbol{\Psi}\boldsymbol{\gamma},\end{aligned} \tag{3.29}$$

where $\boldsymbol{\Psi} = \mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T) \otimes \mathbf{A}_R(\widehat{\boldsymbol{\phi}}_R) \in \mathbb{C}^{N_T N_R \times \widehat{L}^2}$, $\boldsymbol{\gamma} \in \mathbb{C}^{\widehat{L}^2 \times 1}$ and $\otimes$ represents the Kronecker product. We estimate the channel gain vector by solving the following

sparse recovery problem:

$$\boldsymbol{\gamma}^\star = \arg\min_{\boldsymbol{\gamma}} ||\text{vec}(\widehat{\mathbf{H}}) - \boldsymbol{\Psi}\boldsymbol{\gamma}||^2, \qquad ||\boldsymbol{\gamma}||_0 = \widehat{L}. \tag{3.30}$$

In this chapter, we use OMP [43, 96] to solve the above problem but other sparse recovery methods can also be used. The sparsity pattern of $\boldsymbol{\gamma}^\star$ also reveals the correspondence of the AoAs, AoDs, and path gains. Finally, the channel matrix is reconstructed as

$$\widehat{\mathbf{H}} = \text{vec}^{-1}(\boldsymbol{\Psi}\boldsymbol{\gamma}^\star). \tag{3.31}$$

The above OMP estimator, though in the same form as in [43], generally exhibits a low complexity because the dictionary $\boldsymbol{\Psi}$ here has a much smaller size $\widehat{L}^2$ as compared to the size of $N_T N_R$ in the OMP estimator with non-redundant dictionaries in [43]. This is attributed to the super-resolution estimates of the AoAs and AoDs already obtained from Stage 2. The OMP algorithm is set to terminate after $\widehat{L}$ iterations according to the rank knowledge available from Stage 1. The three-stage channel estimator is summarized in Algorithm 2 for completeness.

3.3.4 Complexity Analysis

We here analyze the complexity of the proposed three-stage channel estimator. The main steps contributing to the complexity of all the stages and the numbers of floating point operations of real numbers (flops) required are given in Table 3.1. Note that each complex number multiplication requires six flops and each complex number addition requires two flops. For the GCG-ALTMIN algorithm used in Stage 1, the main steps include finding the top singular vectors in Step 4 and computing the step size in Step 6 of the GCG algorithm as in[50, Algorithm 1]. The main steps contributing to the ALTMIN process include [50, Eq. (61), (64)]. Overall, the complexity of Stage 1 is $\mathcal{O}(N_R N_T)$ per iteration, which is lower than many MC algorithms and CS algorithms such as the OMP, as it works on the transformed channel matrix directly

rather than a large dictionary and avoids full singular value decomposition (SVD). In Stage 2, the root-MUSIC algorithm requires an eigenvalue decomposition (EVD) of the Toeplitz-rectified covariance matrices $\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R})$ and $\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\phi}_T})$ followed by finding the roots of polynomials. The TR only requires $\mathcal{O}(N_R^2 + N_T^2)$ operations. The root-MUSIC applied to find the AoAs and AoDs has a complexity of $\mathcal{O}(N_R^3 + N_T^3)$. In Stage 3, the complexity is mainly due to the application of the dictionary in Step 3 and solving the least square problem in Step 5 of the OMP algorithm of [43, Algorithm 1]. Note that the dictionary $\mathbf{\Psi}$ admits a Kronecker product representation. Thus, by exploiting the identity $(\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{C}) = \text{vec}(\mathbf{ACB})$, Stage 3 has a complexity $\mathcal{O}(\widehat{L}^2 N_R(N_T + \widehat{L}))$ which is low because $\widehat{L}$ is small. Therefore, the overall complexity of the proposed three-stage estimator is low.

The computational complexity of several existing estimators considered in this chapter is analyzed and summarized in Table 3.2. Similar to Stage 3 of the proposed estimator, the identity $(\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{C}) = \text{vec}(\mathbf{ACB})$ has been used while analyzing the complexity of relevant operations which require the application of the Kronecker-structured dictionary. The OMP estimator [43] requires an overall complexity of $\mathcal{O}(N_X G_R(N_Y + G_T))$ per iteration, where N_X and N_Y are the numbers of the transmitted and received beams, respectively, and G_R and G_T represent the size of the grid for the AoA and AoD, respectively. Note that as $\min\{G_T, G_R\} \gg \widehat{L}$, directly applying OMP as in [43] exhibits a higher complexity as compared to Stage 3 of the proposed estimator. Furthermore, the number of iterations Q is generally higher than the actual sparsity level due to grid mismatch. This leads to the increased complexity of the OMP estimator.

For the ADMM estimator [52], which jointly estimates the low-rank channel and its sparse beamspace representation, the main contribution to complexity is the SVD in [52, Step 2 of Algorithm 1] for each iteration and the application of large dictionary matrices. The overall complexity per iteration of the ADMM estimator is cubic in the size of the channel matrix. The computational complexity of the IR-SR estimator [33] depends mainly on its gradient step in [33, Step 5 of Algorithm 1]. Additionally,

the IR-SR estimator requires SVD for the initial estimate of the number of paths. As IR-SR iteratively refines the angle estimates for super-resolution performance, the number of iterations is relatively high, hence further adding to the complexity. Numerical results will be provided in Section 3.4 to demonstrate the difference in the complexity of different estimators.

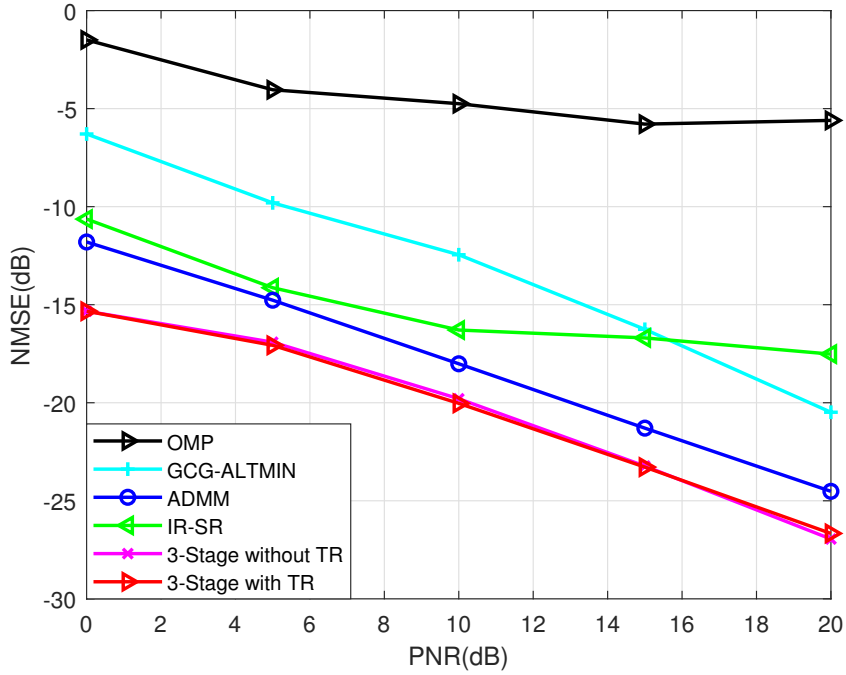
3.4 Simulation Results

In this section, we present simulation results to demonstrate the performance of the proposed three-stage estimator. The channel estimators employing the OMP algorithm [43], the GCG-ALTMIN algorithm [50], the ADMM algorithm[1] [52], and the IR-SR algorithm[2] [33] are also included. The channel model follows [33], i.e., the number of paths $L = 3$, the average path loss $\rho = 1$, the path gains $\{\gamma_l\}$ are Gaussian-distributed, the AoAs are AoDs are uniformly distributed in $[0, \pi]$ and the line-of-sight (LOS) channel has a Rician K-factor of 20 dB. The pilot-to-noise ratio (PNR) is defined as PNR $= \frac{P}{\sigma_n^2}$. The number of samples acquired per training step is equal to Q_R for the GCG-ALTMIN, OMP, IR-SR estimators, and the proposed estimator, and is $Q_R - 1$ for the ADMM estimator. Therefore, to achieve the same sampling ratio, the ADMM estimator[3] requires longer training time. The sampling ratio r gives the ratio of the total scalar observations obtained at the receiver and the total number of entries $N_R N_T$ of the channel matrix. The normalized mean square error (NMSE) for channel estimation is defined as the mean of $\frac{||\widehat{\mathbf{H}}-\mathbf{H}||_F^2}{||\mathbf{H}||_F^2}$. Similarly, the NMSE for angle estimation is defined as the mean of $\frac{||\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}||_F^2}{||\boldsymbol{\phi}||_F^2}$, where the subscripts for the AoA/AoD are omitted for brevity. When the number of estimated paths is less than the actual number of paths, 0° is assumed for the AoAs and AoDs of the

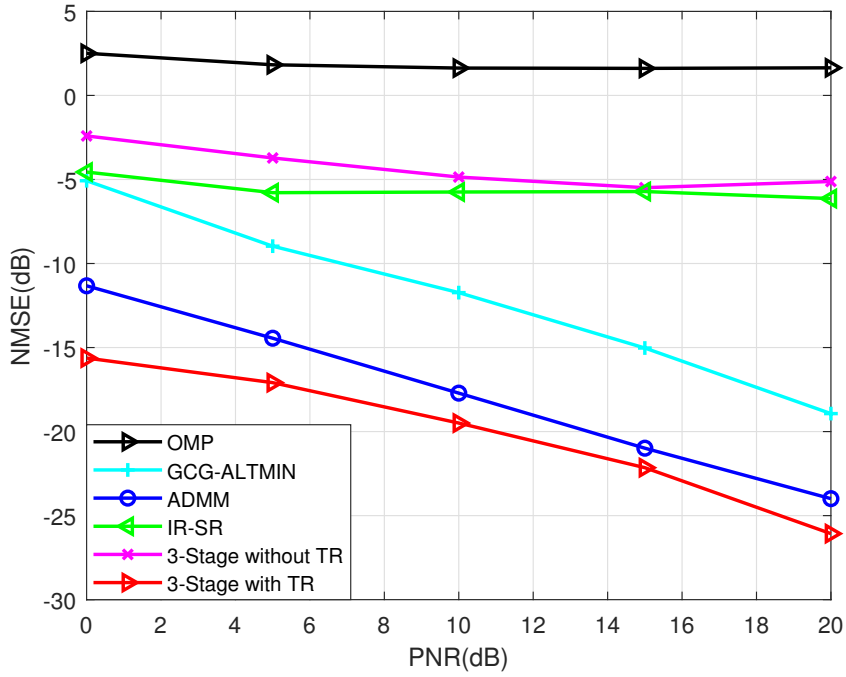
[1]Simulation codes used are available online:
https://github.com/vlaxose/spl18

[2]Simulation codes used are available online:
http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html

[3]The ADMM-based estimator [52] is a MC-based scheme. For the phase shifter-based hybrid MIMO, a training scheme designed in [50] is adopted here to achieve the sampling for ADMM-based estimator. The proposed estimator uses a different training process devised for the IMC approach [50].

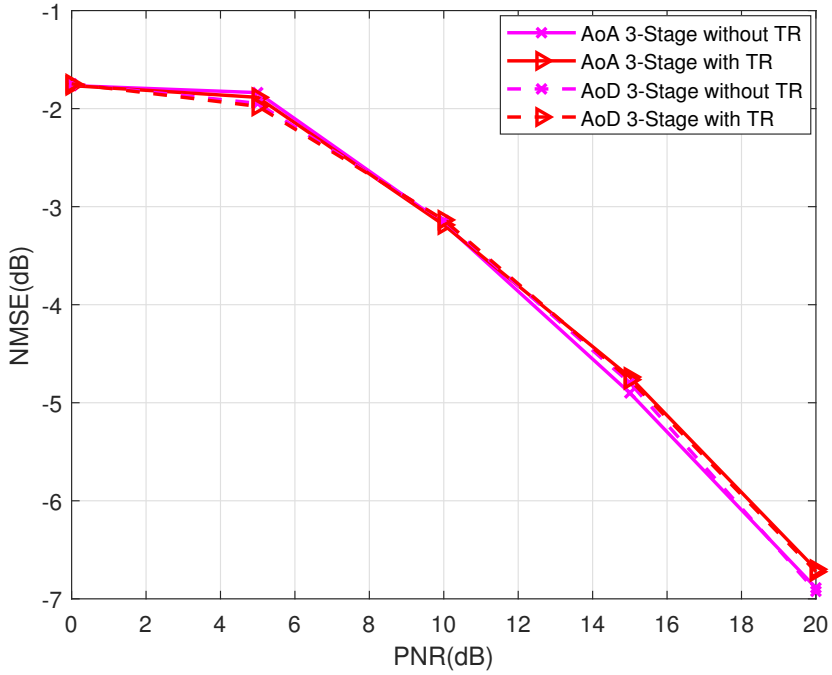


(a) $N_T = N_R = 64$

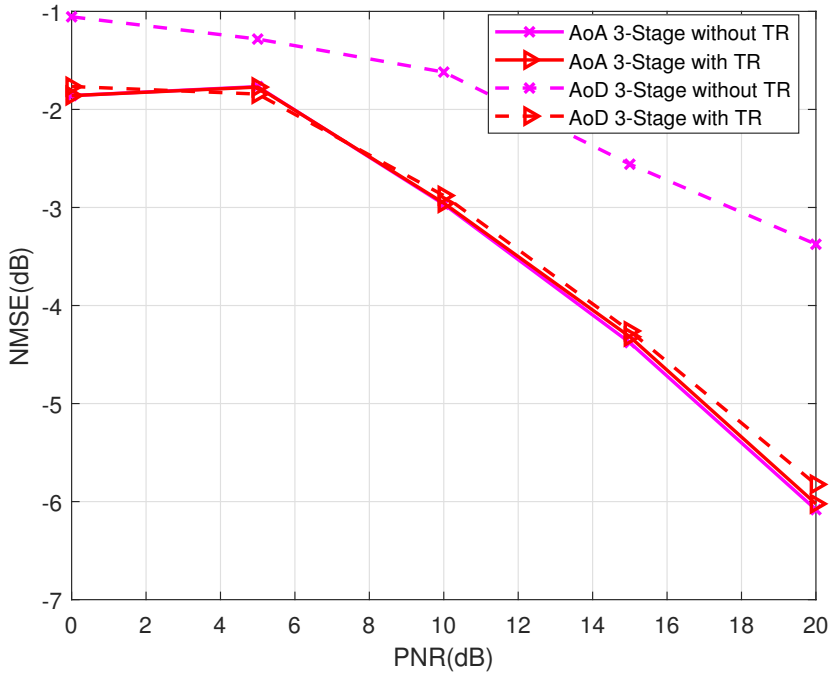


(b) $N_T = 128, N_R = 32$

Figure 3.2: Channel estimation performance versus PNR at a sampling ratio of $r = 0.25$.

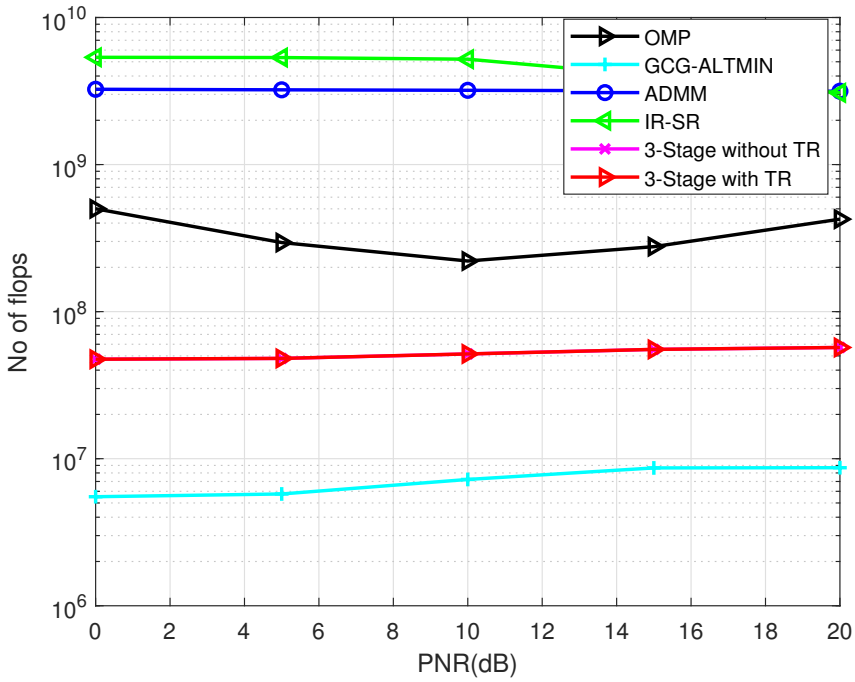


(a) $N_T = N_R = 64$

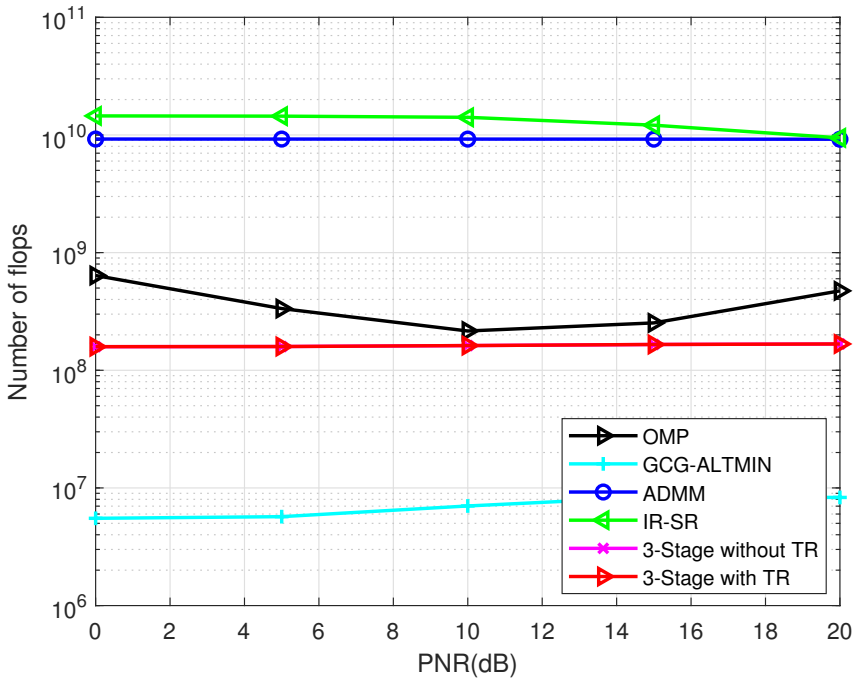


(b) $N_T = 128, N_R = 32$

Figure 3.3: AoA/AoD estimation performance versus PNR at a sampling ratio of $r = 0.25$.



(a) $N_T = N_R = 64$



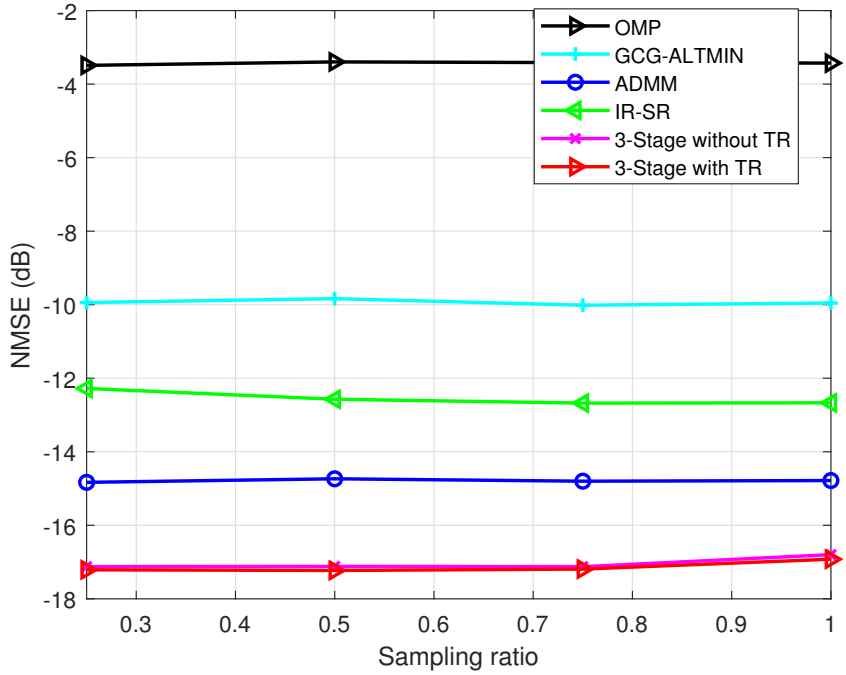
(b) $N_T = 128, N_R = 32$

Figure 3.4: Computational complexity at different PNR levels corresponding to NMSE performance in Fig. 3.2.

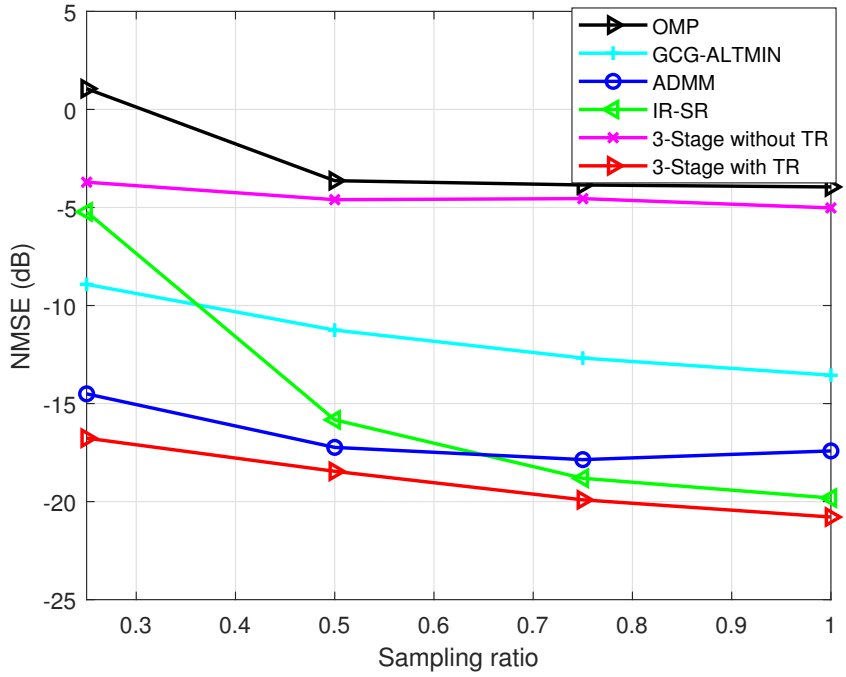
missing paths in $\widehat{\boldsymbol{\phi}}$.

We consider two cases for the hybrid transceivers, with $N_T = N_R = 64, Q_T = Q_R = 8$, and $N_T = 128, N_R = 32, Q_T = 16, Q_R = 4$, respectively. Six-bit phase shifters are used at both the transmitter and receiver. The OMP estimator uses grid sizes of $G_T = N_T$ and $G_R = N_R$ for the AoD and AoA, respectively. Fig. 3.2 shows the channel estimation performance for a sampling ratio $r = 0.25$. The proposed three-stage estimator shows significant gains at different PNR levels as compared to the alternative estimators. This is because the GCG-ALTMIN estimator at Stage 1 achieves a good initial estimate of the channel matrix while the estimation of the AoAs and AoDs of the channel paths at Stage 2 provides further refinements with a low increase of computational complexity. In contrast to the grid-based schemes such as the OMP and ADMM estimators, the proposed approach does not rely on predefined grids and thus avoids the potential issue of grid mismatch. Also, though not improving the performance when $N_T = N_R$, TR becomes critical when $N_T = 128, N_R = 32$ because of the limited number of snapshots when estimating the AoDs using root-MUSIC. The performance of estimating AoAs/AoDs is shown in Fig. 3.3. In Fig. 3.3b there is a degradation in the AoD estimation when TR is not used because the number of snapshots $N_R = 32$ used to generate the SCM for root-MUSIC is significantly smaller than the number of transmitter antennas $N_T = 128$.

Fig. 3.4 shows the computational effort required for different estimators, which correspond to the NMSE results in Fig. 3.2. As described in Section 3.3.4, the least expensive is the GCG-ALTMIN estimator. The three-stage estimator has higher complexity than the GCG-ALTMIN, mainly due to the root-MUSIC algorithms employed in Stage 2. TR adds little to the computational complexity and thus the difference in the complexity of the proposed estimator with and without TR is negligible in Fig. 3.4. Overall the proposed estimator achieves the lowest NMSE with low computational complexity. The OMP estimator exhibits moderate complexity but the performance is limited. The ADMM and IR-SR estimators can achieve good



(a) $N_T = N_R = 64$



(b) $N_T = 128, N_R = 32$

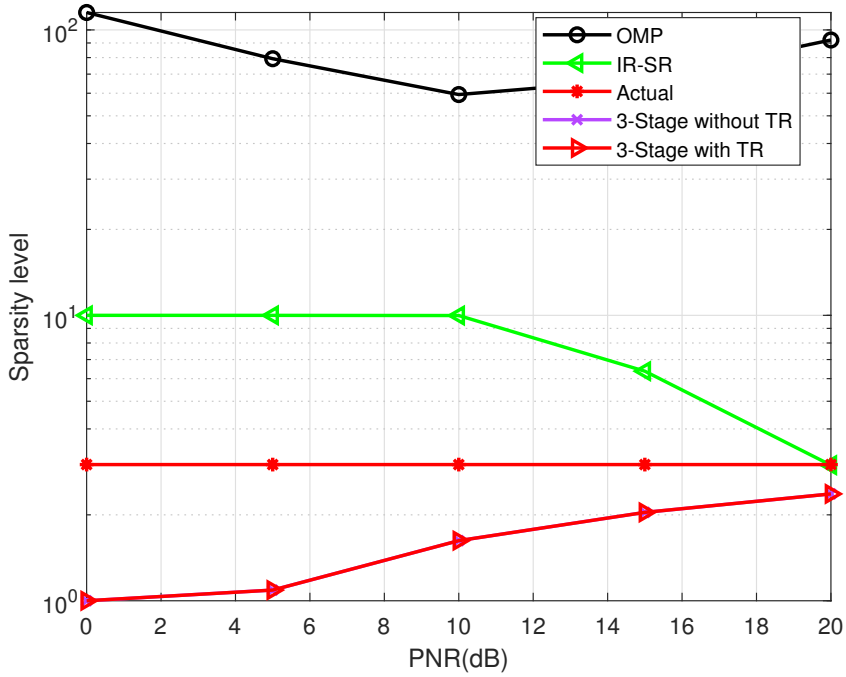
Figure 3.5: Channel estimation performance versus sampling ratio at PNR = 5 dB.

estimation accuracy as they simultaneously exploit the low rankness of the channel and knowledge of the array response. However, apart from higher per iteration cost, both these estimators converge in a relatively large number of iterations. Therefore, their overall complexity is higher than other alternatives. Additionally, the run time for different algorithms, for the settings in Fig. 3.4a and PNR = 20dB are shown in Table 3.3. The run time for the proposed solutions is way less than the existing algorithm, with ADMM-based solution being the most expensive.

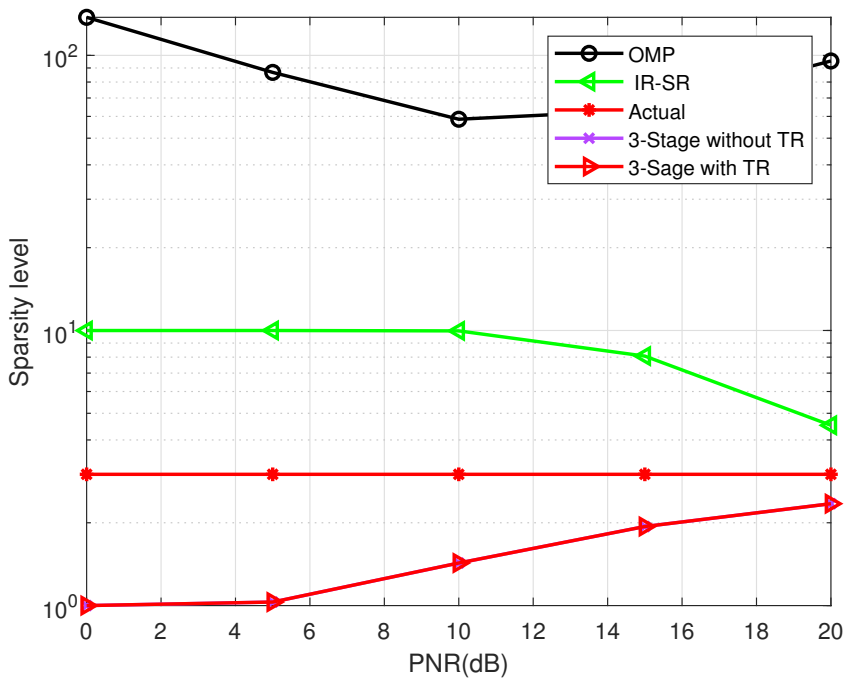
The performance of the proposed estimator with different sampling ratios at PNR = 5 dB is shown in Fig. 3.5. It can be seen that the proposed estimator shows more than 2 dB gains with different sampling ratios as compared to the ADMM estimator. Note that the latter simultaneously exploits the low rankness and sparsity of the channel while our proposed estimator exploits them at different stages with a lower implementation complexity. The IR-SR approaches the proposed estimator at higher sampling ratios at a higher computational complexity.

Fig. 3.6 shows the sparsity level of the estimated channel, i.e., the number of channel paths recovered. The results are compared with those of the OMP and IR-SR estimators which also provide path information. It is seen that the proposed estimator provides a more sparse solution while achieving similar or lower MSE. This indicates that the proposed estimators are able to more accurately estimate the dominant paths of the channel, which may allow efficient CSI feedback when needed.

Fig. 3.7 shows the channel estimation performance for typical number of paths ($L \in [1, 5]$) for mmWave channels [35] at $r = 0.25$ and PNR = 5 dB. The proposed estimator performs best when the number of paths is less. As the number of paths increases, the performance degrades due to the degradation of the estimation performance by the GCG-ALTMIN estimator at Stage 1. Therefore, the proposed estimator is more suitable for low-rank channels where IMC can perform effectively. Meanwhile, stages 2 and 3 of the proposed solution may still be applied when other estimators are used at Stage 1.

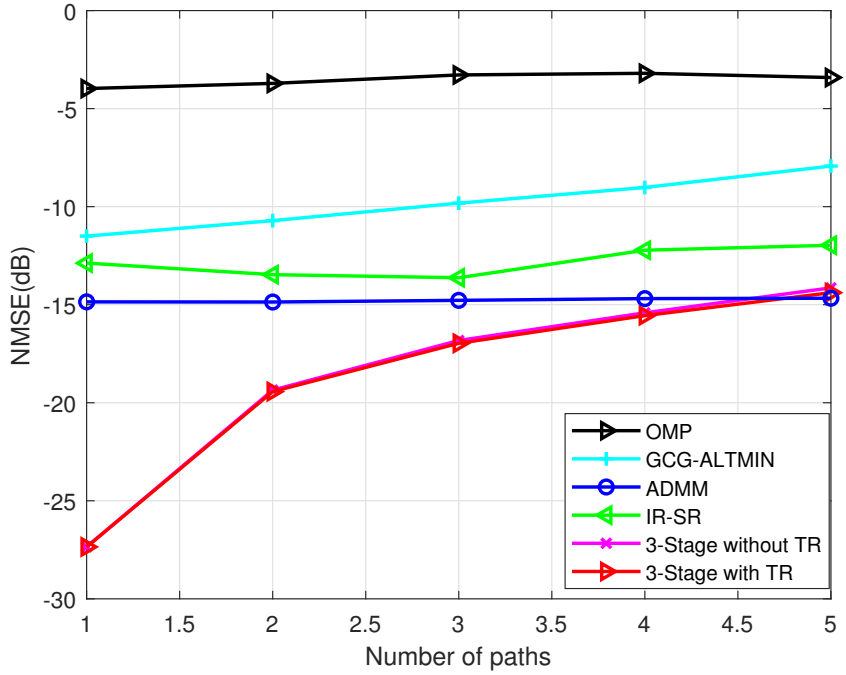


(a) $N_T = N_R = 64$

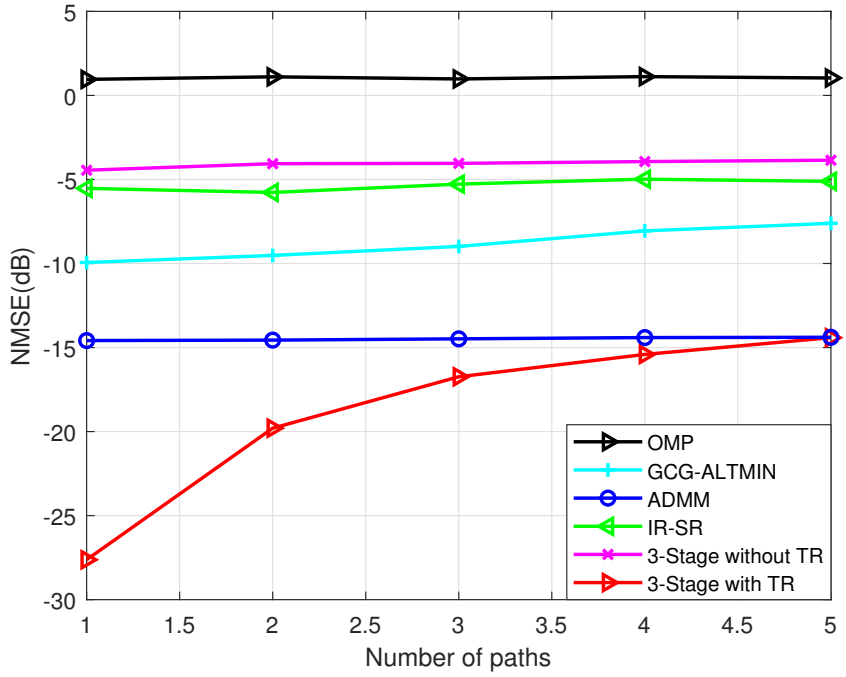


(b) $N_T = 128, N_R = 32$

Figure 3.6: Average sparsity level (number of paths recovered) versus PNR at a sampling ratio of $r = 0.25$.



(a) $N_T = N_R = 64$



(b) $N_T = 128, N_R = 32$

Figure 3.7: Channel estimation performance versus the number of paths at PNR = 5 dB.

3.5 Summary

We have presented a three-stage approach to estimate low-rank mmWave MIMO channels, which benefits from the low-rankness of the channel and the knowledge of the array response. The estimator also yields a sparse representation of the channel which may be desired when CSI feedback is required. Low-complexity techniques are employed at each stage to exploit their strengths such as low training overhead and super-resolution spectrum estimation. Consequently, the overall computational complexity of the proposed estimator is low. Meanwhile, the integration of the adopted techniques addresses their respective challenges such as the under-exploitation of the array response knowledge with matrix sensing, compatibility of one-dimensional spectrum estimation with hybrid MIMO transceivers, and grid mismatch with compressive sensing. Consequently, the proposed estimator can achieve lower estimation error for the channel matrix at low complexity.

Algorithm 1: Precoder/Combiner Design and Training Scheme.

Input: N_R, N_T, Q_R and S

Precoder/combiner design:

Generate initial feature matrices $\mathbf{X}_R$ and $\mathbf{X}_T$ following [50, Section D].

Set $E = N_R/Q_R$ and $\mathcal{Z} = \{1, 2, \cdots, N_R\}$.

for $e = 1, 2, \cdots, E$ **do**

  Set $\mathcal{Z}_e \subset \mathcal{Z}$ be a size-Q_R random subset of $\mathcal{Z}$.

  Set $\mathbf{W}_e = \mathbf{X}_R(:, \mathcal{Z}_e)$.

  Find $\widehat{\mathbf{W}}_e = \mathbf{W}_{RF_e}\mathbf{W}_{BB_e}$ from $\mathbf{W}_e$ by solving (3.8) using the PE-ALTMIN algorithm [98].

  Set $\mathbf{X}_R(:, \mathcal{Z}_e) = \widehat{\mathbf{W}}_e$.

  Set $\mathcal{Z} = \mathcal{Z} - \mathcal{Z}_e$.

end

for $m = 1, 2, \cdots, N_T$ **do**

  Set $\mathbf{p}_m = \mathbf{X}_T(:, m)$.

  Find $\widehat{\mathbf{p}}_m = \mathbf{P}_{RF_m}\mathbf{p}_{BB_m}$ from $\mathbf{p}_m$ using the PE-ALTMIN algorithm similar to (3.8).

  Set $\mathbf{X}_T(:, m) = \widehat{\mathbf{p}}_m$.

end

Training Process:

Set $\Omega = \{\}$.

$P_\Omega(\widetilde{\mathbf{C}}) = \mathbf{0}_{N_R \times N_T}$

for $m = 1, 2, \cdots, N_T$ **do**

  Set $\mathbf{p}_m = \mathbf{X}_T(:, m)$.

  Set $\mathcal{E} = \{1, 2, \cdots, E\}$.

  for $s = 1, 2, \cdots, S$ **do**

    Set e as a random entry of $\mathcal{E}$.

    Set $\mathbf{W}_{m,s} = \mathbf{X}_R(:, \mathcal{Z}_e)$.

    Obtain $\mathbf{y}_{m,s}$ using (3.3).

    for $n = 1, 2, \cdots, Q_R$ **do**

      Set $\Omega = \{\Omega, \{\mathcal{Z}_e(n), m\}\}$.

      Set $[P_\Omega(\widetilde{\mathbf{C}})]_{\mathcal{Z}_e(n),m} = \mathbf{y}_{m,s}(n)$.

    end

    Set $\mathcal{E} = \mathcal{E} - \{e\}$.

  end

end

Output: $\mathbf{X}_R$, $\mathbf{X}_T$ and $P_\Omega(\widetilde{\mathbf{C}})$.

Table 3.1: Computational Complexity of the Three-Stage Estimator. The exponent parameter $q = 2$, r represents the sampling ratio, oversampling parameter $g = 10$, $l \in \{1, 2, \cdots, \widehat{L}\}$ represents the number of the GCG iteration, Q is the number of updates of the alternate minimization (ALTMIN) process, and $m \in \{1, 2, \cdots, \widehat{L}\}$ represents the number of the OMP iteration in Stage 3.

Stage	Algorithms	Main steps	Flops per iteration	Total
1	GCG ($\widehat{L}$ iterations)	Step 4 Algorithm 1[50]	$8(2q+3)(g+1)N_TN_R$	$8\widehat{L}BN_TN_R + \frac{1}{3}Q\widehat{L}(\widehat{L}+1)\big(rN_RN_T(16\widehat{L}+32) + (N_T+N_R)(3\widehat{L}^2+19\widehat{L}+18)\big)$ where $B = (2q+3)(g+1)+(4r+16)$
		Step 6 Algorithm 1[50]	$(4r+16)N_TN_R$	
	ALTMIN (Q iterations per GCG iteration)	Eq. (61) [50]	$8l^2rN_RN_T + 4l^3N_T + 16l^2N_T + 8lrN_RN_T$	
		Eq. (64) [50]	$8l^2rN_RN_T + 4l^3N_R + 16l^2N_R + 8lrN_RN_T$	
2	Root-MUSIC (AoA)	SCM Eq.(17) and TR Eq. (20)	$N_T(N_T+1)(4N_R-1) + 2N_TN_R + 2N_R^2$	$\frac{209}{3}(N_R^3+N_T^3) - (N_R^2+N_T^2)(4\widehat{L}+123) - (N_R+N_T)(2\widehat{L}-4N_RN_T-134) + 8N_RN_T - 2\widehat{L}(\widehat{L}+1) - \frac{316}{3}$
	(1 iteration)	EVD and EV multiplication Eq. (21) and (22)	$23N_R^3 + (N_R-\widehat{L})(N_R-\widehat{L}+1)(4N_R-1)$	
		Polynomial root finding for Eq. (25)	$\frac{16}{3}(2N_R-2)^3 + 2(N_R-1)^2 + 6(2N_R-2)$	
	Root-MUSIC (AoD)	SCM and TR	$N_R(N_R+1)(4N_T-1) + 2N_RN_T + 2N_T^2$	
	(1 iteration)	EVD and EV multiplication	$23N_T^3 + (N_T-\widehat{L})(N_T-\widehat{L}+1)(4N_T-1)$	
		Polynomial root finding	$\frac{16}{3}(2N_T-2)^3 + 2(N_T-1)^2 + 6(2N_T-2)$	
3	OMP($\widehat{L}$ iterations)	Step 3 Algorithm 1 [43]	$8\widehat{L}N_T(N_R+\widehat{L}) - 2\widehat{L}(N_T+\widehat{L})$	$\widehat{L}\big(8\widehat{L}N_T(N_R+\widehat{L}) - 2\widehat{L}(N_T+\widehat{L}) + 12\widehat{L}N_TN_R + 2N_TN_R - 2\widehat{L}\big)$
		Step 5 Algorithm 1 [43]	$24mN_TN_R - 10N_TN_R - 4m + 2$	

Table 3.2: Computational Complexity of Several Existing Estimators. N_X and N_Y: the numbers of the transmitted and received beams, respectively; G_R and G_T: the size of the grid for the AoA and AoD, respectively; $m \in \{1, 2, \cdots, Q\}$: the number of the OMP iteration; $\widehat{L}_a$: the sparsity level predicted by the ADMM algorithm; $\widehat{L}_s$: the sparsity level predicted by SVD preconditioning of the IR-SR estimator.

Algorithms	Main Steps	Flops per iteration	Total
OMP (Q iterations)	Step 3, Algorithm 1 [43]	$8G_RN_X(N_Y + G_T) - 2G_R(N_X + G_T)$	$Q\big(8G_RN_X(N_Y + G_T) - 2G_R(N_X + G_T) + 12QN_TN_R + 2N_XN_Y - 2Q\big)$
	Step 5 Algorithm 1 [43]	$24mN_XN_Y - 10N_XN_Y - 4m + 2$	
ADMM (Q iterations)	Step 2, Algorithm 1 [52]	$29N_T^3 + 16N_R^2N_T + 24N_RN_T^2 + 2\widehat{L}_aN_R + 8\widehat{L}_aN_RN_T - 2N_RN_T$	$Q\big(29N_T^3 + 40N_T^2N_R + 48N_R^2N_T + 8\widehat{L}_aN_RN_T + 2N_R\widehat{L}_a + 2N_RN_T\big)$
	Step 3 Algorithm 1 [52]	$8N_RN_T(N_T + N_R) + 6N_RN_T$	
	Step 4 Algorithm 1 [52]	$8N_RN_T(N_T + N_R) + 2N_RN_T$	
	Step 5 Algorithm 1 [52]	$8N_RN_T(N_T + N_R) - 4N_RN_T$	
IR-SR (Q iterations for Algorithm 1, 1 iteration for Algorithm 2)	Step 1 Algorithm 2 [33]	$29N_X^3 + 16N_Y^2N_X + 24N_X^2N_Y$	$Q\big(16N_XN_Y\widehat{L}_s^2 - 4N_X\widehat{L}_s^2 + 24N_RN_Y\widehat{L}_s + 24N_TN_X\widehat{L}_s + 64N_XN_Y\widehat{L}_s + 8\widehat{L}_s^3 - 12N_X\widehat{L}_s - 6N_Y\widehat{L}_s + 110\widehat{L}_s^2 - 2\hat{L}_s - 4\big) + 29N_X^3 + 16N_Y^2N_X + 24N_X^2N_Y + 8\widehat{L}_sN_RN_T - 2\widehat{L}_sN_Y + 5\widehat{L}_sN_Y\log_2 N_Y + 8\widehat{L}_sN_TN_X - 2\widehat{L}_sN_X + 5\widehat{L}_sN_X\log_2 N_X$
	Step 4, 5, Algorithm 2 [33]	$8\widehat{L}_sN_RN_Y - 2\widehat{L}_sN_Y + 5\widehat{L}_sN_Y\log_2 N_Y + 8\widehat{L}_sN_TN_X - 2\widehat{L}_sN_X + 5\widehat{L}_sN_X\log_2 N_X$	
	Step 5, Algorithm 1 [33] (Q iterations)	$12N_XN_Y\widehat{L}_s^2 - 3N_X\widehat{L}_s^2 + 16N_RN_Y\widehat{L}_s + 16N_TN_X\widehat{L}_s + 46N_XN_Y\widehat{L}_s + 4\widehat{L}_s^3 - 9N_X\widehat{L}_s - 4N_Y\widehat{L}_s + 88\widehat{L}_s - 4$	
	Step 6, Algorithm 1 [33] (Q iterations)	$4N_XN_Y\widehat{L}_s^2 + 8N_YN_R\widehat{L}_s + 8N_XN_T\widehat{L}_s + 18N_XN_Y\widehat{L}_s + 4\widehat{L}_s^3 - N_X\widehat{L}_s^2 - 2N_Y\widehat{L}_s - 3N_X\widehat{L}_s + 22\widehat{L}_s^2 - 2\hat{L}_s$	

Algorithm 2: Three-Stage Estimator

Input: $P_\Omega(\widetilde{\mathbf{C}}), \mathbf{X}_R$ and $\mathbf{X}_T$.

Stage 1:

1. Find $\widehat{\mathbf{C}}^\star$ by solving (3.11) using [50, Algorithm 1].
2. Obtain the low-rank estimate $\widehat{\mathbf{H}}$ using (3.12).

Stage 2:

3. Construct $\widehat{\mathbf{R}}_\theta$ for $\widehat{\boldsymbol{\theta}}_R$ using (3.17).
4. Find $\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R})$ by applying TR on $\widehat{\mathbf{R}}_\theta$ using (3.20).
5. Find $\widehat{\boldsymbol{\theta}}_R$ using root-MUSIC (3.21)-(3.26) on $\mathcal{T}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R})$.
6. Construct $\widehat{\mathbf{R}}_{\phi_T}$ for $\widehat{\boldsymbol{\phi}}_T$ using (3.27).
7. Find $\mathcal{T}(\widehat{\mathbf{R}}_{\phi_T})$ by applying TR on $\widehat{\mathbf{R}}_{\phi_T}$ similar to (3.20).
8. Find $\widehat{\boldsymbol{\phi}}_T$ using root-MUSIC (3.21)-(3.26) on $\mathcal{T}(\widehat{\mathbf{R}}_{\phi_T})$.

Stage 3:

9. Construct dictionary $\boldsymbol{\Psi} = \mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T) \otimes \mathbf{A}_R(\widehat{\boldsymbol{\theta}}_R)$ using $\widehat{\boldsymbol{\phi}}_T$ and $\widehat{\boldsymbol{\theta}}_R$.
10. Estimate $\boldsymbol{\gamma}^\star$ by solving (3.30) using the OMP Algorithm.
11. Construct the estimated channel matrix $\widehat{\mathbf{H}}$ using (3.31).

Output: $\widehat{\mathbf{H}}, \widehat{\boldsymbol{\theta}}_R, \widehat{\boldsymbol{\phi}}_T$ and $\boldsymbol{\gamma}^\star$

Table 3.3: Run-time comparison for different algorithms. System specifications: Windows 10, Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz Processor, 16 GB RAM and Matlab R2019b.

Name of the Algorithm	Run-Time (Seconds)
GCG-ALTMIN	0.1864
3-Stage without TR	0.2449
3-Stage with TR	0.3059
OMP	0.6523
IR-SR	3.1519
ADMM	12.12

Chapter 4

Inductive Matrix Completion and Root-MUSIC-Based Channel Estimation for Intelligent Reflecting Surface (IRS)-Aided Hybrid MIMO Systems

4.1 Introduction

This chapter studies the estimation of cascaded channels in passive IRS-aided MIMO systems employing hybrid precoders and combiners. We propose a low-complexity solution that estimates the channel parameters progressively. The AoDs and AoAs at the transmitter and receiver, respectively, are first estimated using IMC followed by root-MUSIC-based super-resolution spectrum estimation. Forward-backward spatial smoothing (FBSS) is applied to address the coherence issue. Using the estimated AoAs and AoDs, the training precoders and combiners are then optimized

and the angle differences between the AoAs and AoDs at the IRS are estimated using the LS method followed by FBSS and the root-MUSIC algorithm. Finally, the composite path gains of the cascaded channel are estimated using on-grid sparse recovery with a small-size dictionary. The simulation results suggest that the proposed estimator can achieve improved channel parameter estimation performance with lower complexity as compared to several recently reported alternatives, thanks to the exploitation of the knowledge of the array responses and low-rankness of the channel using low-complexity algorithms at all the stage

The rest of the Chapter is organized as follows. Section 4.2 introduces the system model. Section 4.3 presents the proposed solution. The simulation results are discussed in Section 4.4 and the chapter summary is presented in Section 4.5.

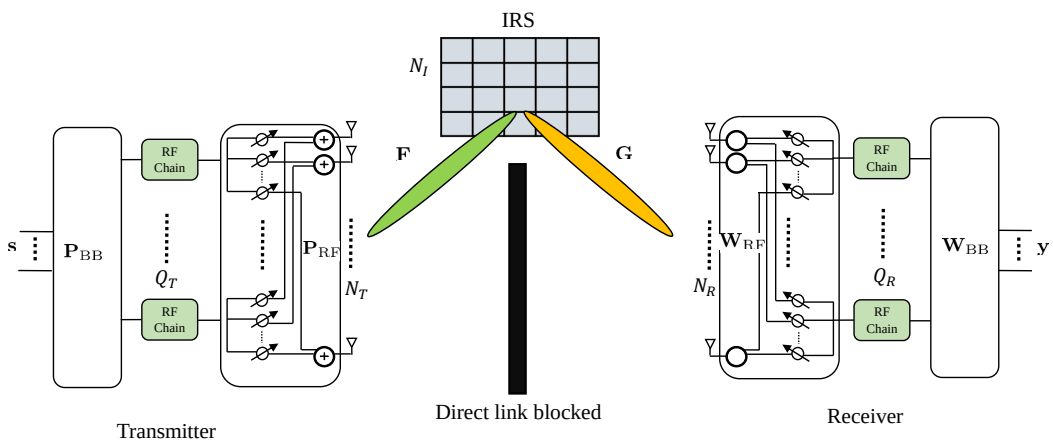


Figure 4.1: System model for IRS-aided hybrid MIMO system.

4.2 System Model

As illustrated in Fig. 4.1, we consider a point-to-point MIMO system equipped with hybrid transceivers aided by a fully passive IRS. The transmitter, receiver, and IRS employ ULAs with N_T, N_R and N_I array elements, respectively[1]. There are $Q_T \leq N_T$ and $Q_R \leq N_R$ RF chains at the transmitter and receiver, respectively.

[1]The techniques proposed by this chapter can be extended to multiple-user systems with ULA and UPA applied at the transmitter, receiver, and IRS. In Section 4.3.4 we will discuss the treatment for UPA at the IRS.

Fully connected phase shifter networks are assumed for the transmitter and receiver but the techniques can also be easily extended to other hybrid transceivers [108]. The direct channel between the transmitter and receiver is assumed to be blocked for simplicity, but there exist many viable solutions including Chapter 3 and [43, 50, 52] to estimate it.

The channel between the transmitter and the IRS is geometrically modeled as [73, 76]

$$\mathbf{F} = \sqrt{\frac{N_T N_I}{L_F}} \sum_{l=1}^{L_F} \gamma_{F,l} \mathbf{a}_I(\theta_{I,l}) \mathbf{a}_T^H(\phi_{T,l}) = \mathbf{A}_I(\boldsymbol{\theta}_I) \boldsymbol{\Gamma}_F \mathbf{A}_T^H(\boldsymbol{\phi}_T) \in \mathbb{C}^{N_I \times N_T}, \tag{4.1}$$

where $\theta_{I,l}$, $\phi_{T,l}$ and $\gamma_{F,l}$ represent the AoA at the IRS, the AoD at the transmitter and the complex path gain for the l-th path, respectively, and L_F denotes the total number of paths between the transmitter and IRS. Furthermore, $\mathbf{A}_I(\boldsymbol{\theta}_I)$, $\mathbf{A}_T(\boldsymbol{\phi}_T)$ and $\boldsymbol{\Gamma}_F$ denote the IRS array response matrix, transmitter array response matrix, and diagonal path gain matrix, respectively. Similarly, the channel between the IRS and the receiver is modeled as

$$\mathbf{G} = \sqrt{\frac{N_I N_R}{L_G}} \sum_{l=1}^{L_G} \gamma_{G,l} \mathbf{a}_R(\theta_{R,l}) \mathbf{a}_I^H(\phi_{I,l}) = \mathbf{A}_R(\boldsymbol{\theta}_R) \boldsymbol{\Gamma}_G \mathbf{A}_I^H(\boldsymbol{\phi}_I) \in \mathbb{C}^{N_R \times N_I} \tag{4.2}$$

where notation similar to those in (4.1) is used. In the above, the array response vector for a ULA with N elements can be written as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} [1, e^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta)}, \ldots, e^{j(N-1)\frac{2\pi}{\lambda_c}d\cos(\theta)}]^T \tag{4.3}$$

where λ_c is the wavelength, $\beta = \lambda_c/2$ is the inter-element spacing, and θ is the steering angle. We focus on mmWave and THz bands where the number of paths L_F and L_G are typically small.

The effective channel between the transmitter and receiver via IRS is given as

$$\boldsymbol{\mathcal{H}} = \mathbf{G}\boldsymbol{\Omega}\mathbf{F} \in \mathbb{C}^{N_R \times N_T}, \tag{4.4}$$

where $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\omega})$ and $\boldsymbol{\omega}$ contains the phase shifts for all the IRS elements:

$$\boldsymbol{\omega} = [\Upsilon_1 e^{j\zeta_1}, \Upsilon_2 e^{j\zeta_2}, \ldots, \Upsilon_{N_I} e^{j\zeta_{N_I}}]^T \in \mathbb{C}^{N_I \times 1},$$

where Υ_i and ζ_i denote the reflection coefficient and the phase shift for the i-th IRS element, respectively. Setting $\Upsilon_i = 1$ or 0 suggests that the i-th element is turned on or off, respectively.

Let $\diamond$ denote the Khatri-Rao product. By using the identity $\text{vec}(\mathbf{A}\text{diag}(\mathbf{b})\mathbf{C}) = (\mathbf{C}^T \diamond \mathbf{A})\mathbf{b}$, the effective channel can be rewritten as

$$\text{vec}(\boldsymbol{\mathcal{H}}) = \text{vec}(\mathbf{G}\boldsymbol{\Omega}\mathbf{F}) = (\mathbf{F}^T \diamond \mathbf{G})\boldsymbol{\omega} = \mathbf{H}\boldsymbol{\omega}, \tag{4.5}$$

where $\mathbf{H} \triangleq \mathbf{F}^T \diamond \mathbf{G} \in \mathbb{C}^{N_R N_T \times N_I}$ is the cascaded channel. During data transmissions, the precoder, combiner, and IRS phase shifts need to be optimized according to $\mathbf{H}$. However, it is challenging to estimate $\mathbf{H}$ due to its high dimensionality and limited observations at the hybrid receivers. In order to address this problem, we propose below a solution with low training overhead and low computational complexity.

4.3 The Proposed Channel Estimator

4.3.1 Parametric Representation of the Cascaded Channel

Let $\otimes$ denote the Kronecker product. By using the identities $(\mathbf{AB}) \diamond (\mathbf{CD}) = (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \diamond \mathbf{D})$ and $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$, the cascaded channel from

(4.5) can be modeled as

$$\begin{aligned}\mathbf{H} &= (\mathbf{A}_I(\boldsymbol{\theta}_I)\boldsymbol{\Gamma}_F\mathbf{A}_T^H(\boldsymbol{\phi}_T))^T \diamond \left(\mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}_G\mathbf{A}_I^H(\boldsymbol{\phi}_I)\right)\\ &= (\mathbf{A}_T^*(\boldsymbol{\phi}_T)\boldsymbol{\Gamma}_F\mathbf{A}_I^T(\boldsymbol{\theta}_I)) \diamond (\mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}_G\mathbf{A}_I^H(\boldsymbol{\phi}_I))\\ &= ((\mathbf{A}_T^*(\boldsymbol{\phi}_T)\boldsymbol{\Gamma}_F) \otimes (\mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}_G)) \left(\mathbf{A}_I^T(\boldsymbol{\theta}_I) \diamond \mathbf{A}_I^H(\boldsymbol{\phi}_I)\right)\\ &= \mathbf{A}_{TR}(\boldsymbol{\phi}_T, \boldsymbol{\theta}_R)\boldsymbol{\Gamma}\mathbf{A}_I^H(\boldsymbol{\psi}_I)\end{aligned} \tag{4.6}$$

where

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_F \otimes \boldsymbol{\Gamma}_G = \mathrm{diag}(\boldsymbol{\gamma}), \tag{4.7}$$

with $\boldsymbol{\gamma} \in \mathbb{C}^{L_F L_G \times 1}$ containing the composite of the channel path gains of $\mathbf{G}$ and $\mathbf{F}$,

$$\mathbf{A}_{TR}(\boldsymbol{\phi}_T, \boldsymbol{\theta}_R) = \mathbf{A}_T^*(\boldsymbol{\phi}_T) \otimes \mathbf{A}_R(\boldsymbol{\theta}_R), \tag{4.8}$$

and

$$\mathbf{A}_I(\boldsymbol{\psi}_I) = \left(\mathbf{A}_I^H(\boldsymbol{\theta}_I) \diamond \mathbf{A}_I^T(\boldsymbol{\phi}_I)\right)^T = [\mathbf{a}_I(\psi_{I,1,1}), \ldots, \mathbf{a}_I(\psi_{I,i,j}), \ldots, \mathbf{a}_I(\psi_{I,L_F,L_G})] \tag{4.9}$$

with

$$\psi_{I,i,j} = \cos^{-1}(\cos(\phi_{I,j}) - \cos(\theta_{I,i})) \tag{4.10}$$

being the effective angle difference between the i-th AoA and j-th AoD at the IRS. From (4.6), though the cascaded channel $\mathbf{H}$ may have a very large dimension, it can be parameterized with a small number of path directions and gains. This can be exploited to reduce the training overhead and computational complexity. However, joint estimation of these parameters may involve high computational complexities. We, therefore, propose a low-complexity multi-stage solution below.

As shown in Fig. 4.2, the proposed solution estimates the outer angles $\boldsymbol{\phi}_T$ and $\boldsymbol{\theta}_R$ in the first stage using varying hybrid precoder/combiners at the transmitter and receiver but fixed phase shifts at the IRS. IMC and spectrum estimation using the

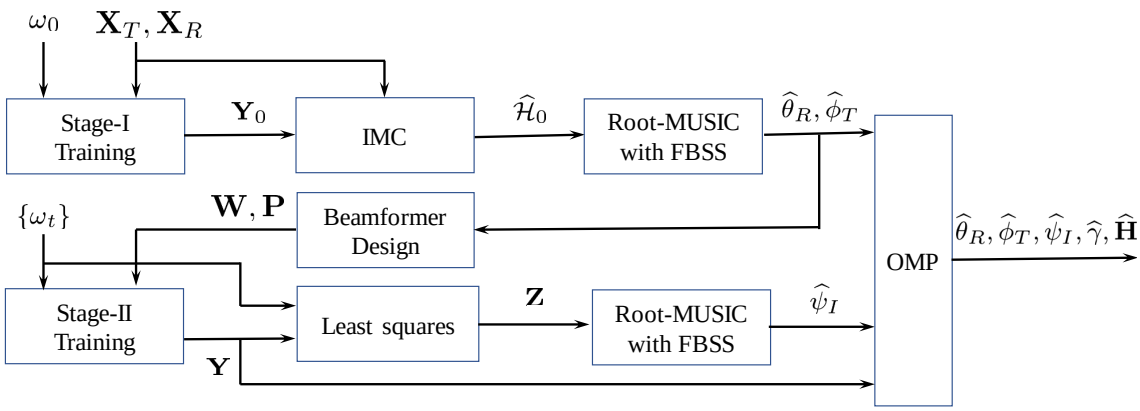


Figure 4.2: Flowchart of the proposed scheme for estimating the cascaded channel of IRS-aided MIMO systems.

root-MUSIC algorithm is applied to estimate the angles. In the second stage, the IRS angles $\boldsymbol{\psi}_I$ are estimated using fixed hybrid precoders/combiners constructed using the estimated outer angles and varying IRS phase shifts. Similarly, the root-MUSIC algorithm is applied at this stage. Finally, the estimated angles are associated by solving a small-size on-grid CS problem using OMP, which also yields the composite path gains. The proposed solution has similarities to [76] in terms of two-stage training and progressive estimation of the channel parameter. However, different training and estimation schemes are deployed, which can improve the complexity-performance tradeoff.

4.3.2 Stage 1: Estimation of Outer Angles

4.3.2.1 Training

In order to estimate the outer angles $(\boldsymbol{\theta}_R, \boldsymbol{\phi}_T)$, the IRS phase shifts are randomly chosen as $\boldsymbol{\Omega}_0 = \text{diag}(\boldsymbol{\omega}_0)$ from the feasible set and remain unchanged. This gives the effective channel

$$\boldsymbol{\mathcal{H}}_0 = \mathbf{G}\boldsymbol{\Omega}_0\mathbf{F} \in \mathbb{C}^{N_R \times N_T}. \tag{4.11}$$

For mmWave and THz channels, $\mathbf{G}$ and $\mathbf{F}$ are generally low-rank due to the sparsity in the angular domain. Therefore, $\boldsymbol{\mathcal{H}}_0$ is also low-rank with rank no higher than

$\min(\text{rank}(\mathbf{F}), \text{rank}(\mathbf{G}))$. Furthermore, $\boldsymbol{\mathcal{H}}_0$ can be modeled as

$$\boldsymbol{\mathcal{H}}_0 = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}_G\mathbf{A}_I^H(\boldsymbol{\phi}_I)\boldsymbol{\Omega}_0\mathbf{A}_I(\boldsymbol{\theta}_I)\boldsymbol{\Gamma}_F\mathbf{A}_T^H(\boldsymbol{\phi}_T). \tag{4.12}$$

From (4.12), if $\boldsymbol{\mathcal{H}}_0$ is known, then subspace methods such as the root-MUSIC may be used to estimate the angles $(\boldsymbol{\theta}_R, \boldsymbol{\phi}_T)$, similar to the treatments in Chapter 3. In order to reduce the training overhead for estimating $\boldsymbol{\mathcal{H}}_0$ using the hybrid receiver, we propose to estimate $\boldsymbol{\mathcal{H}}_0$ using low-rank matrix recovery methods. We adopt the IMC scheme [50] for its low complexity and high performance.

We now describe the training scheme. Assume a training length of S channel uses. During the s-th channel use, the transmitter sends a single pilot symbol x_s. The receiver observes Q_R symbols through its RF chains:

$$\mathbf{y}_s = \mathbf{W}_s^H\boldsymbol{\mathcal{H}}_0\mathbf{p}_s x_s + \mathbf{W}_s^H\mathbf{n}_s \in \mathbb{C}^{Q_R\times 1}, \tag{4.13}$$

where $\mathbf{p}_s = \mathbf{P}_{\text{RF},s}\mathbf{p}_{\text{BB},s} \in \mathbb{C}^{N_T\times 1}$ is the hybrid precoder with the RF precoder $\mathbf{P}_{\text{RF},s} \in \mathbb{C}^{N_T\times Q_T}$ and baseband precoder $\mathbf{p}_{\text{BB},s} \in \mathbb{C}^{Q_T\times 1}$. Similarly, the hybrid combiner $\mathbf{W}_s = \mathbf{W}_{\text{RF},s}\mathbf{W}_{\text{BB},s} \in \mathbb{C}^{N_R\times Q_R}$ with $\mathbf{W}_{\text{RF},s} \in \mathbb{C}^{N_R\times Q_R}$ as the RF combiner and $\mathbf{W}_{\text{BB},s} \in \mathbb{C}^{Q_R\times Q_R}$ the basedband combiner. Without loss of generality, we assume $x_s = 1$, $||\mathbf{p}_s||_F^2 = 1$ and $||\mathbf{W}_s||_F^2 = Q_R$. The noise $\mathbf{n}_s \in \mathbb{C}^{N_R\times 1} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I})$, where σ_n^2 is the average noise power. The received signal after S training steps is given as

$$\mathbf{y}_0 = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_S^T]^T \in \mathbb{C}^{SQ_R\times 1}. \tag{4.14}$$

We apply the IMC approach that first estimates the following transformed matrix

$$\boldsymbol{\mathcal{C}}_0 = \mathbf{X}_R^H\boldsymbol{\mathcal{H}}_0\mathbf{X}_T \in \mathbb{C}^{N_R\times N_T}, \tag{4.15}$$

and then recover $\boldsymbol{\mathcal{H}}_0$ as $\boldsymbol{\mathcal{H}}_0 = (\mathbf{X}_R^H)^{-1}\boldsymbol{\mathcal{C}}_0(\mathbf{X}_T)^{-1}$ where $\mathbf{X}_T$ and $\mathbf{X}_R$ are the feature matrices. The hybrid precoders and combiners are designed in a way such that

entries of $\boldsymbol{\mathcal{C}}_0$ are observed from the above training process. This is implemented by selecting $\mathbf{W}_s$ and $\mathbf{p}_s$ from the columns of $\mathbf{X}_R \in \mathbb{C}^{N_R \times N_R}$ and $\mathbf{X}_T \in \mathbb{C}^{N_T \times N_T}$, respectively, following the uniform spatial sampling [97]. Discussion on the coherence properties of $\mathbf{X}_R$ and $\mathbf{X}_T$ can be found in [50, Section D] and the implementation of the precoders/combiners follows Section 3.2.2 of Chapter 3 .

4.3.2.2 Estimation of Outer Angles

Rewrite the received signal by $\mathbf{Y}_0 = P_\Omega(\widetilde{\boldsymbol{\mathcal{C}}}_0)$, where P_Ω denotes the sampling operator with a sampling pattern Ω and $\widetilde{\boldsymbol{\mathcal{C}}}_0$ is a noisy version of $\boldsymbol{\mathcal{C}}_0$ i.e. $\widetilde{\boldsymbol{\mathcal{C}}}_0 = \boldsymbol{\mathcal{C}}_0 + \mathbf{X}_R^H \mathbf{N}$, where $\mathbf{N} \in \mathbb{C}^{N_R \times N_T}$ is the noise matrix. We can now estimate $(\boldsymbol{\phi}_T, \boldsymbol{\phi}_R)$ from $\mathbf{Y}_0 = P_\Omega(\widetilde{\boldsymbol{\mathcal{C}}}_0)$ using IMC followed by spectrum estimation. The transformed matrix $\boldsymbol{\mathcal{C}}_0 = \mathbf{X}_R^H \boldsymbol{\mathcal{H}}_0 \mathbf{X}_T$ is low-rank as $\boldsymbol{\mathcal{H}}_0$ is low-rank. Therefore, $\boldsymbol{\mathcal{C}}_0$ can be estimated first by solving the following low-rank matrix recovery problem

$$\min_{\boldsymbol{\mathcal{C}}_0} \operatorname{rank}(\boldsymbol{\mathcal{C}}_0), \quad \text{s.t.} \quad \|P_\Omega(\boldsymbol{\mathcal{C}}_0) - \mathbf{Y}_0\|_F^2 \leq \delta^2, \tag{4.16}$$

where δ^2 is a tolerance to account for the noise and $\|\cdot\|_F$ represents the Frobenius norm. Nuclear norm regularization is applied to reformulate the NP-hard problem above to estimate $\boldsymbol{\mathcal{C}}_0$ as

$$\widehat{\boldsymbol{\mathcal{C}}}_0 \triangleq \arg\min_{\boldsymbol{\mathcal{C}}_0} \quad \frac{1}{2}\|P_\Omega(\boldsymbol{\mathcal{C}}_0) - \mathbf{Y}_0\|_F^2 + \mu\|\boldsymbol{\mathcal{C}}_0\|_*, \tag{4.17}$$

where $\mu > 0$ is a regularization parameter and $\|\cdot\|_*$ represents the nuclear norm. By using a Frobenius norm characterization of the nuclear norm, we can let $\boldsymbol{\mathcal{C}}_0 \triangleq \mathbf{U}\mathbf{V}^H$ and recover $\boldsymbol{\mathcal{C}}_0$ by solving

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\|P_\Omega(\mathbf{U}\mathbf{V}^H) - \mathbf{Y}_0\|_F^2 + \frac{1}{2}\mu(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2). \tag{4.18}$$

This is a regularized least squares problem if the sizes of $\mathbf{U}$ and $\mathbf{V}$ are fixed according to the rank of $\boldsymbol{\mathcal{C}}_0$. However, this rank is unknown in practice. We adopt the

GCG-ALTMIN algorithm [50] that progressively increases the sizes of $\mathbf{U}$ and $\mathbf{V}$ and exhibits a low complexity and fast convergence. The GCG-ALTMIN iteratively refines the rank and estimate of $\boldsymbol{\mathcal{C}}_0$ by using the top singular vectors of a residual error matrix and alternately updating $\mathbf{U}$ and $\mathbf{V}$ using local minimization.

Once the transformed matrix $\widehat{\boldsymbol{\mathcal{C}}}_0$ is estimated, the low-rank channel matrix can be estimated as

$$\widehat{\boldsymbol{\mathcal{H}}}_0 = (\mathbf{X}_R^H)^{-1}\widehat{\boldsymbol{\mathcal{C}}}_0(\mathbf{X}_T)^{-1}. \tag{4.19}$$

We can then estimate the outer angles $\boldsymbol{\phi}_T$ and $\boldsymbol{\theta}_R$. Following, [108] we estimate them separately to reduce the computational cost. In this chapter, we apply the root-MUSIC algorithm [93–95], which avoids peak search and offers high-resolution estimates of the angles with low complexity. Note that the channel estimate $\widehat{\boldsymbol{\mathcal{H}}}_0$ in (4.19) can be modeled as

$$\widehat{\boldsymbol{\mathcal{H}}}_0 = \boldsymbol{\mathcal{H}}_0 + \mathbf{E} = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Gamma}_0\mathbf{A}_T^H(\boldsymbol{\phi}_T) + \mathbf{E}, \tag{4.20}$$

where $\mathbf{E}$ represents the estimation error and $\boldsymbol{\Gamma}_0 \triangleq \boldsymbol{\Gamma}_G\mathbf{A}_I^H(\boldsymbol{\phi}_I)\boldsymbol{\Omega}_0\mathbf{A}_I(\boldsymbol{\theta}_I)\boldsymbol{\Gamma}_F$. It is clear that the row and column subspaces of $\boldsymbol{\mathcal{H}}_0$ are spanned by the receiver and transmitter steering vectors. This can be utilized to estimate the AoAs and AoDs separately by using subspace methods.

We first estimate the AoAs $\boldsymbol{\theta}_R$ at the receiver. The estimation of the AoDs $\boldsymbol{\phi}_T$ at the transmitter is similar. We model the columns of $\widehat{\boldsymbol{\mathcal{H}}}_0$ as samples of the received signal of an $N_R \times 1$ ULA as

$$\mathbf{x}_n = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\lambda}_n + \mathbf{e}_n, n = 1, 2, \ldots, N_T, \tag{4.21}$$

where $\boldsymbol{\lambda}_n$ is the n-th column of $\boldsymbol{\Gamma}_0\mathbf{A}_T^H(\boldsymbol{\phi}_T)$ that serves as the "source" for generating the observation $\mathbf{x}_n$ at the receiver array. The root-MUSIC algorithm can be applied using the signal and noise subspaces estimated from the sample covariance matrix

(SCM) of $\mathbf{x}_n$:

$$\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R} = \frac{1}{N_T}\sum_{n=1}^{N_T}\mathbf{x}_n\mathbf{x}_n^H = \frac{1}{N_T}\widehat{\boldsymbol{\mathcal{H}}}_0\widehat{\boldsymbol{\mathcal{H}}}_0^H = \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Delta}\mathbf{A}_R^H(\boldsymbol{\theta}_R) + \boldsymbol{\Sigma}, \tag{4.22}$$

where the "source" covariance matrix is given by

$$\boldsymbol{\Delta} \triangleq \frac{1}{N_T}\boldsymbol{\Gamma}_0\mathbf{A}_T^H(\boldsymbol{\phi}_T)\mathbf{A}_T(\boldsymbol{\phi}_T)\boldsymbol{\Gamma}_0^H, \tag{4.23}$$

and the error of the covariance matrix estimation is

$$\boldsymbol{\Sigma} \triangleq \frac{1}{N_T}\left(\boldsymbol{\mathcal{H}}_0\mathbf{E}^H + \mathbf{E}\boldsymbol{\mathcal{H}}_0^H + \mathbf{E}\mathbf{E}^H\right). \tag{4.24}$$

Eigenvalue decomposition of $\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}$ can be used to find the signal and noise subspaces required by root-MUSIC.

In general, the "source" covariance matrix $\boldsymbol{\Delta}$ in (4.22) is non-diagonal. This suggests that the "source" signals in the model of (4.21) are correlated. It is known that with correlated sources, standard subspace methods based on the SCM may perform poorly. We thus adopt the FBSS [109–111] here to improve angle estimation.

In order to estimate L angles using the N_R-element array (4.21), the FBSS constructs $U = L + 1$ forward and backward uniform sub-arrays, each with $\mathcal{S} = N_R - L$ elements. Neighboring subarrays differ by only one element. Consider a reference ULA subarray with $\mathcal{S}$ antennas. Its array response matrix can be written as

$$\widetilde{\mathbf{A}}_R(\boldsymbol{\theta}_R) = [\widetilde{\mathbf{a}}_R(\theta_{R,1}), \widetilde{\mathbf{a}}_R(\theta_{R,2}), \ldots, \widetilde{\mathbf{a}}_R(\theta_{R,L})] \in \mathbb{C}^{\mathcal{S}\times L}, \tag{4.25}$$

with array response for the l-th "source" given as

$$\widetilde{\mathbf{a}}_R(\theta_{R,l}) = \frac{1}{\sqrt{N_R}}[1, e^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}, \ldots, e^{j(\mathcal{S}-1)\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}]^T. \tag{4.26}$$

Let $\mathbf{D}$ be a diagonal matrix with entries

$$\mathbf{D} = \mathrm{diag}\{e^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,1})}, \mathrm{e}^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,2})}, \ldots, e^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,L})}\}. \tag{4.27}$$

Then the received signals of the u-th forward subarray, whose array response vectors are shifted versions of those of the reference subarray, can be written as

$$\mathbf{x}_{n,u}^f \triangleq [x_{n,u}, x_{n,u+1}, \ldots, x_{n,u+\mathcal{S}-1}]^T = \widetilde{\mathbf{A}}_R(\boldsymbol{\theta}_R)\mathbf{D}^{(u-1)}\boldsymbol{\lambda}_n + \mathbf{e}_{n,u}^f, \tag{4.28}$$

where $x_{n,i}$ denotes the i-th entry of $\mathbf{x}_n$ in (4.21) and $\mathbf{e}_{n,u}^f$ denotes the corresponding subvector of $\mathbf{e}_n$. Letting the covariance matrices of the "sources" $\boldsymbol{\lambda}_n$ and "errors" $\mathbf{e}_{n,u}^f$ be $\boldsymbol{\Sigma}_s$ and $\boldsymbol{\Sigma}_{e,u}^f$, respectively, we have

$$\widetilde{\mathbf{R}}_{\boldsymbol{\theta}_R}^{f,u} \triangleq \mathrm{E}[\mathbf{x}_{n,u}^f \mathbf{x}_{n,u}^{f^H}] = \widetilde{\mathbf{A}}_R(\boldsymbol{\theta}_R)\mathbf{D}^{(u-1)}\boldsymbol{\Sigma}_s(\mathbf{D}^{(u-1)})^*\widetilde{\mathbf{A}}_R^H(\boldsymbol{\theta}_R) + \boldsymbol{\Sigma}_{e,u}^f. \tag{4.29}$$

Define the forward covariance matrix as

$$\widetilde{\mathbf{R}}_{\boldsymbol{\theta}_R}^f \triangleq \frac{1}{U}\sum_{u=1}^{U}\widetilde{\mathbf{R}}_{\boldsymbol{\theta}_R}^{f,u} = \widetilde{\mathbf{A}}_R(\boldsymbol{\theta}_R)\boldsymbol{\Sigma}_s^f\widetilde{\mathbf{A}}_R^H(\boldsymbol{\theta}_R) + \boldsymbol{\Sigma}_e^f \tag{4.30}$$

with

$$\boldsymbol{\Sigma}_s^f \triangleq \frac{1}{U}\sum_{u=1}^{U}\mathbf{D}^{(u-1)}\boldsymbol{\Sigma}_s(\mathbf{D}^{(u-1)})^*, \quad \boldsymbol{\Sigma}_e^f \triangleq \frac{1}{U}\sum_{u=1}^{U}\boldsymbol{\Sigma}_{e,u}^f.$$

In general, $\boldsymbol{\Sigma}_s^f$ has a higher rank than $\boldsymbol{\Sigma}_s$ when the "sources" are correlated. This is beneficial for applying subspace methods for finding the angles. Similarly we can construct the u-th backward subarray as

$$\begin{aligned}\mathbf{x}_{n,u}^b &\triangleq \left[x_{n,N_R-u+1}^*, x_{n,N_R-u}^*, \cdots, x_{n,N_R-u-\mathcal{S}}^*\right]^T \\ &= \widetilde{\mathbf{A}}_R(\boldsymbol{\theta}_R)\mathbf{D}^{-N_R+u}\boldsymbol{\lambda}_n^* + \mathbf{e}_{n,u}^b.\end{aligned} \tag{4.31}$$

Following the same assumption as for the forward subarray, we can verify

$$\begin{aligned}\widetilde{\mathbf{R}}_{\boldsymbol{\theta}_R}^{b,u} \triangleq & \mathrm{E}[\mathbf{x}_{n,u}^{b}\mathbf{x}_{n,u}^{b^H}] \\ = & \widetilde{\mathbf{A}}_R(\boldsymbol{\theta}_R)\mathbf{D}^{(-N_R+u)}\boldsymbol{\Sigma}_s^*(\mathbf{D}^{(-N_R+u)})^*\widetilde{\mathbf{A}}_R^H(\boldsymbol{\theta}_R) + \boldsymbol{\Sigma}_{e,u}^b\end{aligned} \tag{4.32}$$

where $\boldsymbol{\Sigma}_{e,u}^b$ denotes the covariance matrix of $\mathbf{e}_{n,u}^b$. The backward covariance matrix can be defined similarly as

$$\widetilde{\mathbf{R}}_{\boldsymbol{\theta}_R}^{b} \triangleq \frac{1}{U}\sum_{u=1}^{U}\widetilde{\mathbf{R}}_{\boldsymbol{\theta}_R}^{b,u} = \widetilde{\mathbf{A}}_R(\boldsymbol{\theta}_R)\boldsymbol{\Sigma}_s^b\widetilde{\mathbf{A}}_R^H(\boldsymbol{\theta}_R) + \boldsymbol{\Sigma}_e^b \tag{4.33}$$

with

$$\boldsymbol{\Sigma}_s^b = \frac{1}{U}\sum_{u=1}^{U}\mathbf{D}^{(-N_R+u)}\boldsymbol{\Sigma}_s^*(\mathbf{D}^{(-N_R+u)})^*, \quad \boldsymbol{\Sigma}_e^b \triangleq \frac{1}{U}\sum_{u=1}^{U}\boldsymbol{\Sigma}_{e,u}^b.$$

Inspired by that (4.30) and (4.33) share the same signal subspace in the error-free case, the SCM smoothed by applying FBSS can be used to estimate a subarray covariance matrix from the samples as

$$\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}^{SS} \triangleq \frac{1}{2N_TU}\sum_{n=1}^{N_T}\sum_{u=1}^{U}\left(\mathbf{x}_{n,u}^{f}\mathbf{x}_{n,u}^{f^H} + \mathbf{x}_{n,u}^{b}\mathbf{x}_{n,u}^{b^H}\right). \tag{4.34}$$

Let $\mathbf{J}$ be the anti-diagonal identity matrix. Then the received signals of the u-th forward and $(U-u+1)$-th backward subarrays can be related as

$$\mathbf{x}_{n,U-u+1}^{b} = \mathbf{J}(\mathbf{x}_{n,u}^{f})^*, \forall u,$$

and

$$\begin{aligned}\mathbf{x}_{n,U-u+1}^{b}(\mathbf{x}_{n,U-u+1}^{b})^H = & \mathbf{J}(\mathbf{x}_{n,u}^{f})^*(\mathbf{x}_{n,u}^{f})^T\mathbf{J} \\ = & \mathbf{J}[(\mathbf{x}_{n,u}^{f})(\mathbf{x}_{n,u}^{f})^H]^*\mathbf{J}, \forall u.\end{aligned}$$

This suggests that the FBSS covariance matrix can be alternatively obtained from

(4.22) as

$$\begin{aligned}\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}^{SS} =& \text{FBSS}(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}) \\ \triangleq& \frac{1}{2U}\sum_{u=1}^{U}\Big(\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}(u:u+\mathcal{S}-1, u:u+\mathcal{S}-1) \\ &+ \mathbf{J}\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}^{*}(u:u+\mathcal{S}-1, u:u+\mathcal{S}-1)\mathbf{J}\Big).\end{aligned} \tag{4.35}$$

After that, the root-MUSIC algorithm is applied on $\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}^{SS}$ to estimate $\boldsymbol{\theta}_R$. Though FBSS sacrifices the array aperture and the resolution of angle estimation, the overall accuracy is generally improved, especially in large arrays. For estimating $\boldsymbol{\phi}_T$ using root-MUSIC, the construction of the required covariance matrix $\widehat{\mathbf{R}}_{\boldsymbol{\phi}_T}^{SS}$ follows the same way as $\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R}^{SS}$ based on $\widehat{\boldsymbol{\mathcal{H}}}_0^H$ and the discussion is omitted for brevity.

4.3.3 Stage 2: Estimation of IRS Angles and Composite Path Gains

4.3.3.1 Training

We next estimate the IRS angles $\boldsymbol{\psi}_I$ and the composite path gains of the cascaded channel in (4.6). The estimates $(\widehat{\boldsymbol{\theta}}_R, \widehat{\boldsymbol{\phi}}_T)$ of the outer angles are used here to construct the hybrid precoder and combiner, respectively, for achieving beamforming gains. The desired precoder and combiner are given as

$$\widehat{\mathbf{W}} = \mathbf{A}_R(\widehat{\boldsymbol{\theta}}_R) \in \mathbb{C}^{N_R\times L_G}, \widehat{\mathbf{P}} = \mathbf{A}_T(\widehat{\boldsymbol{\phi}}_T) \in \mathbb{C}^{N_T\times L_F}. \tag{4.36}$$

They are implemented approximately as $\mathbf{P}$ and $\mathbf{W}$ using the fully connected hybrid transceivers and the PE-Altmin algorithm [98].

Without loss of generality, we assume $L_G \leq Q_R$, but the treatment can be extended for $L_G > Q_R$. For the cases when $L_G > Q_R$, for each precoding vector multiple combiner matrices $\widehat{\mathbf{W}}$ are used for the AoAs. For example, for $L_G = 4$ and $Q_R = 2$, there will be two different combiner matrices, $\widehat{\mathbf{W}}_1 = \mathbf{A}_{R_{:,1:Q_R}}(\widehat{\boldsymbol{\theta}}_R)$ and $\widehat{\mathbf{W}}_2 = \mathbf{A}_{R_{:,Q_R+1:L_G}}(\widehat{\boldsymbol{\theta}}_R)$. This will double the total number of channel uses for Stage-

2 training. Therefore, the training overhead for $L_G > Q_R$ is higher than $L_G \leq Q_R$. There are in total D steps in IRS angle training and each step spans L_F channel uses. The precoder $\mathbf{P}$, combiner $\mathbf{W}$ and also the IRS phase shifts $\mathbf{\Omega}_d$ remain fixed for each step. The resulting effective channel $\boldsymbol{\mathcal{H}}_d = \mathbf{G}\mathbf{\Omega}_d\mathbf{F}$. During the l-th channel use of the d-th step, the received signal is given as

$$\mathbf{y}_{d,l} = \mathbf{W}^H \boldsymbol{\mathcal{H}}_d \mathbf{P}\mathbf{s}_{d,l} + \mathbf{W}^H \mathbf{n}_{d,l} \in \mathbb{C}^{L_G \times 1}, \tag{4.37}$$

where $\mathbf{s}_{d,l} \in \mathbb{C}^{L_F \times 1}$ is the training symbol. We choose $\{\mathbf{s}_{d,l}, l = 1, 2, \cdots, L_F\}$ as columns of an $L_F \times L_F$ unitary matrix and without loss of generality, as the identity matrix. Furthermore, we assume $||\mathbf{W}||_F^2 = L_G$, $||\mathbf{P}||_F^2 = L_F$ and $||\mathbf{P}\mathbf{s}_{d,l}||_F^2 = 1$. As a result, the observation at the receiver at the d-th step is equivalent to

$$\mathbf{Y}_d = [\mathbf{y}_{d,1}, \mathbf{y}_{d,2}, \cdots, \mathbf{y}_{d,L_F}] = \mathbf{W}^H \boldsymbol{\mathcal{H}}_d \mathbf{P} + \mathbf{N}'_d \in \mathbb{C}^{L_G \times L_F}. \tag{4.38}$$

Let $\mathbf{y}_d = \text{vec}(\mathbf{Y}_d)$. It can be verified that

$$\begin{aligned} \mathbf{y}_d &= \text{vec}(\mathbf{W}^H \boldsymbol{\mathcal{H}}_d \mathbf{P}) + \mathbf{n}'_d \\ &= \left((\mathbf{F}\mathbf{P})^T \diamond (\mathbf{W}^H \mathbf{G})\right) \boldsymbol{\omega}_d + \mathbf{n}'_d \\ &= \mathbf{Z}\boldsymbol{\omega}_d + \mathbf{n}'_d, \end{aligned} \tag{4.39}$$

where

$$\mathbf{Z} = \mathbf{\Psi}\mathbf{\Gamma}\mathbf{A}_I^H(\boldsymbol{\psi}) \in \mathbb{C}^{L_F L_G \times N_I}, \tag{4.40}$$

$$\mathbf{\Psi} = \left(\mathbf{P}^T \mathbf{A}_T^*(\boldsymbol{\phi}_T)\right) \otimes \left(\mathbf{W}^H \mathbf{A}_R(\boldsymbol{\theta}_R)\right) \in \mathbb{C}^{L_F L_G \times L_F L_G}. \tag{4.41}$$

Recall that $\mathbf{\Gamma}$ in (4.40), defined in (4.7), is diagonal. In the ideal case with infinite numbers of antennas, perfect outer angle estimation, and infinite resolution of the transceiver phase shifters, $\mathbf{\Psi}$ is an identity matrix. In this case, the IRS angles can be separately estimated from the corresponding rows of $\mathbf{Z}$, as in [76]. However, $\mathbf{\Psi}$ is not identity or diagonal in practical systems. We therefore consider the joint

estimation of the IRS angles based on $\mathbf{Z}$, which may improve performance.

Variable phase shifts $\{\mathbf{\Omega}_d\}$ at the IRS are used during the D training steps, yielding the overall observation

$$\begin{aligned}\mathbf{Y} &= [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_D] \\ &= \mathbf{Z}\bar{\mathbf{\Omega}} + \mathbf{N}' \in \mathbb{C}^{L_G L_F \times D},\end{aligned} \tag{4.42}$$

where $\bar{\mathbf{\Omega}} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_D] \in \mathbb{C}^{N_I \times D}$. If $D \geq N_I$, we can apply the LS method to estimate $\mathbf{Z}$, based on which the IRS angles $\boldsymbol{\psi}_I$ can be estimated using subspace methods. However, this may require substantial training when N_I is large. In order to alleviate the training overhead, we choose $\bar{\mathbf{\Omega}}$ as

$$\bar{\mathbf{\Omega}} = \begin{bmatrix} \mathbf{\Theta} \\ \mathbf{0} \end{bmatrix} \in \mathbb{C}^{N_I \times D}, \tag{4.43}$$

where $\mathbf{\Theta} \in \mathbb{C}^{D \times D}$ is a DFT matrix with entries of unit magnitude and $\mathbf{0}$ denotes an all-zero matrix. This is equivalent to turning off $N_I - D$ elements of the IRS during the IRS angle training and sampling only D elements of the IRS. For simplicity, we set the remaining IRS elements to form a smaller ULA of size D. This reduces the aperture of the IRS and sacrifices the spatial resolution but may provide an economic way for training. A similar training design is used in [112] to reduce the training overhead. In general, the resolution of angle estimation techniques improves when the array aperture is larger. Turning off elements of IRS decreases the spatial resolution specially for small arrays. However, for small arrays, the training overhead is already low and such strategies are not required. For larger arrays, turning off some of the elements does not greatly effect the angle estimation accuracy but can be beneficial in reducing the training overhead.

4.3.3.2 Estimation of IRS Angles

In order to estimate $\boldsymbol{\psi}_I$ we first obtain

$$\widetilde{\mathbf{Z}} = \frac{1}{D}\mathbf{Y}\boldsymbol{\Theta}^H = \boldsymbol{\Psi}\boldsymbol{\Gamma}\widetilde{\mathbf{A}}_I^H(\boldsymbol{\psi}_I) + \widetilde{\mathbf{N}} \in \mathbb{C}^{L_F L_G \times D}, \tag{4.44}$$

where $\widetilde{\mathbf{A}}_I(\boldsymbol{\psi}_I) \in \mathbb{C}^{D \times L_F L_G}$ contains the steering matrix for the sub-array of the IRS corresponding to its D switched-on elements and $\widetilde{\mathbf{N}} = \frac{1}{D}\mathbf{N}'\boldsymbol{\Theta}^H$. Now the IRS angles can be estimated based on

$$\widehat{\mathbf{R}}_{\boldsymbol{\psi}_I} = \frac{1}{L_F L_G}\widetilde{\mathbf{Z}}^H\widetilde{\mathbf{Z}}. \tag{4.45}$$

Similarly to the outer angle estimation in Stage 1, in order to alleviate the issue of coherence of the "source", FBSS can be applied to produce $\widehat{\mathbf{R}}_{\boldsymbol{\psi}_I}^{\mathrm{SS}} = \mathrm{FBSS}(\widehat{\mathbf{R}}_{\boldsymbol{\psi}_I})$ before applying root-MUSIC to produce the IRS angle estimate $\widehat{\boldsymbol{\psi}}_I$, where $\mathrm{FBSS}(\cdot)$ follows (4.35).

4.3.3.3 Estimation of Composite Path Gains

Once $\boldsymbol{\theta}_R$, $\boldsymbol{\phi}_T$, and $\boldsymbol{\psi}_I$ are estimated, the composite path gains $\boldsymbol{\Gamma}$ can be estimated by fitting the received training signal to the model

$$\mathbf{y} \triangleq \mathrm{vec}(\mathbf{Y}) = \boldsymbol{\Phi}\boldsymbol{\gamma} + \mathbf{n}' \tag{4.46}$$

using least squares, where

$$\begin{aligned} \boldsymbol{\Phi} &= \left(\bar{\boldsymbol{\Omega}}^T\mathbf{A}_I^*(\boldsymbol{\psi}_I)\right) \otimes \left(\mathbf{P}^T\mathbf{A}_T^*(\boldsymbol{\phi}_T)\right) \otimes \left(\mathbf{W}^H\mathbf{A}_R(\boldsymbol{\theta}_R)\right) \\ \boldsymbol{\gamma} &\triangleq \mathrm{vec}(\boldsymbol{\Gamma}). \end{aligned} \tag{4.47}$$

Note that $\mathbf{W}$, $\mathbf{P}$ and $\bar{\boldsymbol{\Omega}}$ are known from IRS angle training, and the estimates of $\boldsymbol{\psi}_I$, $\boldsymbol{\phi}_T$ and $\boldsymbol{\theta}_R$ are also available. However, we have estimated the path angles separately for low complexity and thus they are not associated, which leads to an unknown

sparsity pattern of $\boldsymbol{\gamma}$. For associating the angles and also for obtaining the composite path gains, we resort to a CS approach. Since $\boldsymbol{\gamma} \in \mathbb{C}^{L_F^2 L_G^2 \times 1}$ is a sparse vector and has only $L_G L_F$ non-zeros entries, the above problem can be solved efficiently using OMP as

$$\widehat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} ||\mathbf{y} - \widehat{\boldsymbol{\Phi}}\boldsymbol{\gamma}||^2, \qquad ||\boldsymbol{\gamma}||_0 = L_G L_F. \tag{4.48}$$

where the "dictionary" $\widehat{\boldsymbol{\Phi}}$ is obtained by replacing the angles in (4.47) by their estimates. Once the gains are estimated, we can obtain the estimate of the cascaded channel matrix (4.6) as

$$\widehat{\mathbf{H}} = (\mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T) \otimes \mathbf{A}_R(\widehat{\boldsymbol{\theta}}_R))\mathrm{Mat}(\widehat{\boldsymbol{\gamma}})\mathbf{A}_I^H(\widehat{\boldsymbol{\psi}}_I). \tag{4.49}$$

Note that the estimated $\mathrm{Mat}(\widehat{\boldsymbol{\gamma}})$ is not necessarily diagonal due to random permutation. The overall channel estimation process for the ULA case discussed in Section 4.3.1-4.3.3 is summarized in Algorithm 3.

4.3.4 Extension to UPA at IRS

In the above, we have assumed ULA at the transmitter, receiver, and IRS. We now extend the scheme to IRS equipped with UPA. When an $N_y \times N_z$ UPA is located on the yz plane, as illustrated in Fig. 4.3, the steering vector for a path with azimuth angle α_a and elevation angle α_e is given by

$$\mathbf{a}(\alpha_a, \alpha_e) = \mathbf{a}_y(\alpha_a, \alpha_e) \otimes \mathbf{a}_z(\alpha_e),$$

where

$$\mathbf{a}_y(\alpha_a, \alpha_e) = \frac{1}{\sqrt{N_y}}[1, e^{j\frac{2\pi}{\lambda_c}\beta\sin(\alpha_a)\sin(\alpha_e)}, \ldots, e^{j(N_y-1)\frac{2\pi}{\lambda_c}\beta\sin(\alpha_a)\sin(\alpha_e)}]^T, \tag{4.50}$$

Algorithm 3: Cascaded channel estimation for IRS-aided hybrid MIMO systems employing ULAs at the transmitter, IRS, and receiver.

Stage 1:
Input: $\mathbf{Y}_0, \mathbf{X}_T$, and $\mathbf{X}_R$.
1. Find $\widehat{\boldsymbol{\mathcal{C}}}_0^{\star}$ by solving (4.17) using the GCG-ALTMIN algorithm [50, Algorithm 1].
2. Obtain the low-rank estimate $\widehat{\boldsymbol{\mathcal{H}}}_0$ using (4.19).
3. Construct $\widehat{\mathbf{R}}^{\text{SS}}_{\boldsymbol{\theta}_R}$ from $\widehat{\boldsymbol{\mathcal{H}}}_0$ using (4.22) and (4.35).
4. Find $\widehat{\boldsymbol{\theta}}_R$ from $\widehat{\mathbf{R}}^{\text{SS}}_{\boldsymbol{\theta}_R}$ using root-MUSIC.
5. Construct $\widehat{\mathbf{R}}^{\text{SS}}_{\boldsymbol{\phi}_T}$ from $\widehat{\boldsymbol{\mathcal{H}}}_0^H$ similar to (4.22) and (4.35).
6. Find $\widehat{\boldsymbol{\phi}}_T$ from $\widehat{\mathbf{R}}^{SS}_{\boldsymbol{\phi}_T}$ using root-MUSIC.

Output: $\widehat{\boldsymbol{\phi}}_T, \widehat{\boldsymbol{\theta}}_R$.

Stage 2:
Input: $\mathbf{Y}, \mathbf{W}, \mathbf{P}$, and $\bar{\boldsymbol{\Omega}}$.
7. Obtain $\widetilde{\mathbf{Z}}$ using (4.44).
8. Construct $\widehat{\mathbf{R}}^{\text{SS}}_{\psi_I}$ from $\widetilde{\mathbf{Z}}$ using (4.45) and FBSS.
9. Find $\widehat{\boldsymbol{\psi}}_I$ from $\widehat{\mathbf{R}}^{\text{SS}}_{\boldsymbol{\psi}_I}$ using root-MUSIC.
10. Find $\widehat{\boldsymbol{\gamma}}$ using (4.48).
11. Construct the cascaded channel $\widehat{\mathbf{H}}$ using (4.49).

Output: $\widehat{\boldsymbol{\psi}}_I$, $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\mathbf{H}}$.

and

$$\mathbf{a}_z(\alpha_e) = \frac{1}{\sqrt{N_z}}[1, e^{j\frac{2\pi}{\lambda_c}\beta\cos(\alpha_e)}, \ldots, e^{j(N_z-1)\frac{2\pi}{\lambda_c}\beta\cos(\alpha_e)}]^T. \tag{4.51}$$

Let

$$u = \sin(\alpha_a)\sin(\alpha_e), \quad v = \cos(\alpha_e). \tag{4.52}$$

Then the steering vector can be rewritten as

$$\mathbf{a}(u, v) = \mathbf{a}_y(u) \otimes \mathbf{a}_z(v), \tag{4.53}$$

where

$$\begin{aligned} \mathbf{a}_y(u) &= \frac{1}{\sqrt{N_y}}[1, e^{j\frac{2\pi}{\lambda_c}\beta u}, \ldots, e^{j(N_y-1)\frac{2\pi}{\lambda_c}\beta u}]^T, \\ \mathbf{a}_z(v) &= \frac{1}{\sqrt{N_z}}[1, e^{j\frac{2\pi}{\lambda_c}\beta v}, \ldots, e^{j(N_z-1)\frac{2\pi}{\lambda_c}\beta v}]^T. \end{aligned} \tag{4.54}$$

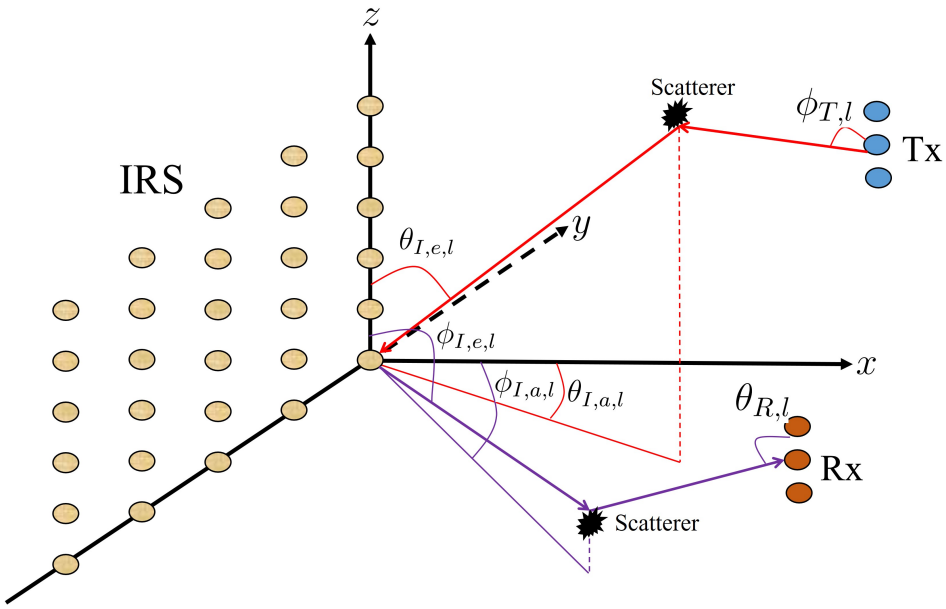


Figure 4.3: System model for the case with UPA at the IRS and ULAs at the transmitter and receiver.

Denote by $(\theta_{I,a,l}, \theta_{I,e,l})$ the pair of the azimuth and elevation components of the AoA for the l-th path impinging on the IRS and similarly for those of the AoD of the l-th path departing the IRS by $(\phi_{I,a,l}, \phi_{I,e,l})$. Following the change of variables as in (4.52), we can rewrite the steering matrix in (4.40) for the UPA as

$$\begin{aligned}\mathbf{A}_I(\mathbf{u},\mathbf{v}) &= \left(\mathbf{A}_I^H(\mathbf{u}_F,\mathbf{v}_F) \diamond \mathbf{A}_I^T(\mathbf{u}_G,\mathbf{v}_G)\right)^T \\ &= [\mathbf{a}(u_{1,1},v_{1,1}),\ldots,\mathbf{a}(u_{i,j},v_{i,j}),\ldots,\mathbf{a}(u_{L_F,L_G},v_{L_F,L_G})],\end{aligned}$$

where

$$\begin{aligned}u_{i,j} &= \sin(\phi_{I,a,i})\sin(\phi_{I,e,i}) - \sin(\theta_{I,a,j})\sin(\theta_{I,e,j}), \\ v_{i,j} &= \cos(\phi_{I,e,i}) - \cos(\theta_{I,e,j}).\end{aligned} \tag{4.55}$$

In this case, the estimation of the outer angles follows that for the ULA case in Section 4.3.2. The IRS angle estimation now amounts to estimating $(\mathbf{u},\mathbf{v})$. In order to achieve high accuracy with low training overhead and low computational complexity, we propose to switch on only an L-shaped subarray as illustrated in Fig. 4.4 during the IRS angle estimation. The same DFT phase shifts matrix $\bar{\mathbf{\Omega}}$ as in (4.43) is applied to the subarray. The observation after completing the Stage-2

training can then be modeled similarly to (4.44) as

$$\widetilde{\mathbf{Z}} = \mathbf{\Psi}\mathbf{\Gamma}\widetilde{\mathbf{A}}_I^H(\mathbf{u},\mathbf{v}) + \widetilde{\mathbf{N}} \in \mathbb{C}^{L_F L_G \times D}, \tag{4.56}$$

where $\widetilde{\mathbf{A}}_I(\mathbf{u},\mathbf{v}) \in \mathbb{C}^{D\times L_F L_G}$ is the array response matrix of the L-shaped subarray. It is clear that the IRS angle information is embedded in the row subspace of $\widetilde{\mathbf{Z}}$.

Note that the L-shaped subarray consists of partially overlapped ULAs along the y and z axes. Denote by $\mathcal{I}_{y,j}$ the indices of the IRS elements of the j-th ULA parallel to the y-axis in Fig. 4.4. We choose the observations in $\widetilde{\mathbf{Z}}$ corresponding to the J_y ULAs and stack them as

$$\widetilde{\mathbf{Z}}_y = \begin{bmatrix} \widetilde{\mathbf{Z}}(:,\mathcal{I}_{y,1}) \\ \widetilde{\mathbf{Z}}(:,\mathcal{I}_{y,2}) \\ \vdots \\ \widetilde{\mathbf{Z}}(:,\mathcal{I}_{y,J_y}) \end{bmatrix} \in \mathbb{C}^{J_y L_F L_G \times N_y}. \tag{4.57}$$

In order to estimate $\mathbf{u}$, we first compute

$$\mathbf{R}_u = \frac{1}{J_y L_F L_G}\widetilde{\mathbf{Z}}_y^H \widetilde{\mathbf{Z}}_y, \tag{4.58}$$

then obtain FBSS($\mathbf{R}_\mathrm{u}$) and finally apply root-MUSIC. We can apply a similar procedure to find $\mathbf{v}$ by using the observations corresponding to the ULAs parallel to the z axis at the IRS. The adopted L-shaped subarray, which has been examined in [113–115], achieves a larger spatial aperture along the y and z axis, in contrast to rectangular subarrays with the same number of elements. Unlike joint 2D spectrum estimation [116], we here estimate $\mathbf{u}$ and $\mathbf{v}$ separately for reduced computational complexity. This, however, does not give associated estimates of $\mathbf{u}$ and $\mathbf{v}$.

Again the estimation of the composite path gains and association of the estimates of the path angles can be achieved by solving the CS problem in (4.48), with the

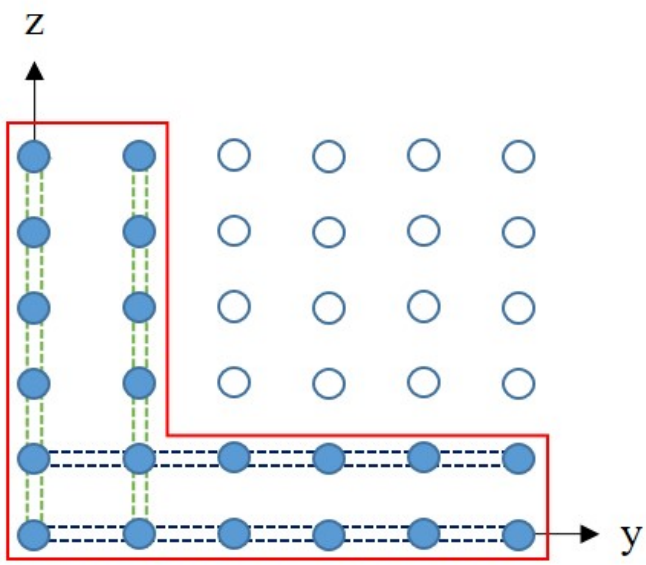


Figure 4.4: An example of IRS with an L-shaped subarray switched on, $N_y = N_z = 6$, $J_y = J_z = 2$, $N_I = 36$ and $D = 20$.

dictionary updated for the UPA case as

$$\widehat{\mathbf{\Phi}} = \left(\bar{\mathbf{\Omega}}^T \mathbf{A}_I^*(\widehat{\mathbf{u}}, \widehat{\mathbf{v}})\right) \otimes \left(\mathbf{P}^T \mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T)\right) \otimes \left(\mathbf{W}^H \mathbf{A}_R(\widehat{\boldsymbol{\theta}}_R)\right). \tag{4.59}$$

Finally, the cascaded channel is reconstructed as

$$\widehat{\mathbf{H}} = (\mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T) \otimes \mathbf{A}_R(\widehat{\boldsymbol{\theta}}_R))\mathrm{Mat}(\widehat{\boldsymbol{\gamma}})\mathbf{A}_I^H(\widehat{\mathbf{u}}, \widehat{\mathbf{v}}). \tag{4.60}$$

Treatments the same as this subsection can be applied to the case with UPA at the transmitter and/or receiver. The details are omitted for conciseness.

4.3.5 Computational Complexity

The computational cost of the proposed estimator is kept low. For Stage 1 (outer angle estimation), the GCG-ALTMIN algorithm has a complexity of $\mathcal{O}(IL_m^3 N_T N_R + L_m^2(N_T + N_R) + L_m N_T N_R)$ where $L_m = \min(L_F, L_G)$ and I is the number of iterations of the alternate minimization step of GCG-ALTMIN, which is low when the channel is low-rank with small L_F and L_G. In large systems, the overall complexity is

dominated by the root-MUSIC algorithm with complexity $\mathcal{O}({N_R}^3 + {N_T}^3)$. Since $\bar{\mathbf{\Omega}}$ is constructed using the DFT matrix, (4.44) has a low cost.

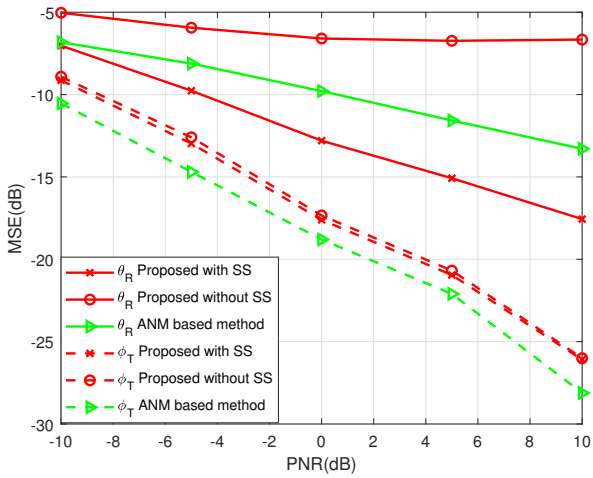
In Stage 2, the inner angle estimation with ULA at the IRS using root-MUSIC has a complexity of $\mathcal{O}(D^3)$. The path gains are estimated using the OMP algorithm with a complexity of $\mathcal{O}(L_F^3 L_G^3 D)$ which is also low when the channel has a small number of paths. For the case with UPA at the IRS, the proposed estimator has a complexity of $\mathcal{O}(N_v^3 + N_u^3)$ for estimating the IRS angles in Stage 2. The composite path gain estimation involves a larger dictionary for the OMP algorithm due to the increased number of angles, but still requires the same order of complexity $\mathcal{O}(L_F^3 L_G^3 D)$.

4.4 Simulation Results

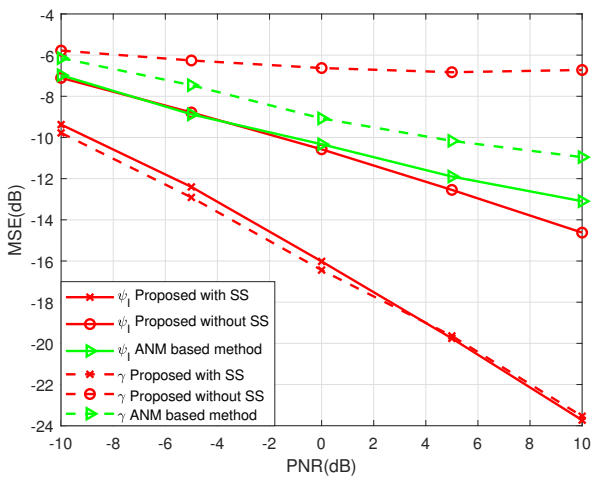
This section presents the simulation results. For the case with a ULA at the IRS, the path gains $\gamma_{F,l}$ and $\gamma_{G,l}$ in (4.1) and (4.2) follow $\mathcal{CN}(0,1)$ while all the AoAs and AoDs are uniformly distributed in $[30°, 150°]$. For the case with a UPA at the IRS, the azimuth angles and elevation angles are uniformly distributed in $[-90°, 90°]$ and $[30°, 150°]$, respectively. Define the pilot-to-noise-ratio as (PNR) $= 10\log_{10}(\frac{1}{\sigma_n^2})$. The normalized mean squared error (NMSE) for estimating the cascaded channel is evaluated as the average of $\frac{||\mathbf{H}-\widehat{\mathbf{H}}||_F^2}{||\mathbf{H}||_F^2}$. The mean squared error (MSE) for angles $\boldsymbol{\theta}_R$, $\boldsymbol{\phi}_T$, $\boldsymbol{\psi}_I$ and gains $\boldsymbol{\gamma}$ are estimated, respectively, by averaging $\frac{||\cos(\boldsymbol{\theta}_R)-\cos(\widehat{\boldsymbol{\theta}}_R)||_F^2}{L_G}$, $\frac{||\cos(\boldsymbol{\phi}_T)-\cos(\widehat{\boldsymbol{\phi}}_T)||_F^2}{L_F}$, $\frac{||\cos(\boldsymbol{\psi}_I)-\cos(\widehat{\boldsymbol{\psi}}_I)||_F^2}{L_F L_G}$, and $\frac{||\boldsymbol{\gamma}-\widehat{\boldsymbol{\gamma}}||_F^2}{L_F L_G}$. For the UPA case, $\frac{||\mathbf{u}-\widehat{\mathbf{u}}||_F^2}{L_F L_G}$ and $\frac{||\mathbf{v}-\widehat{\mathbf{v}}||_F^2}{L_F L_G}$ are averaged to measure the MSE of estimating $\mathbf{u}$ and $\mathbf{v}$, respectively. The proposed approach applies

$$T = S + DL_F \tag{4.61}$$

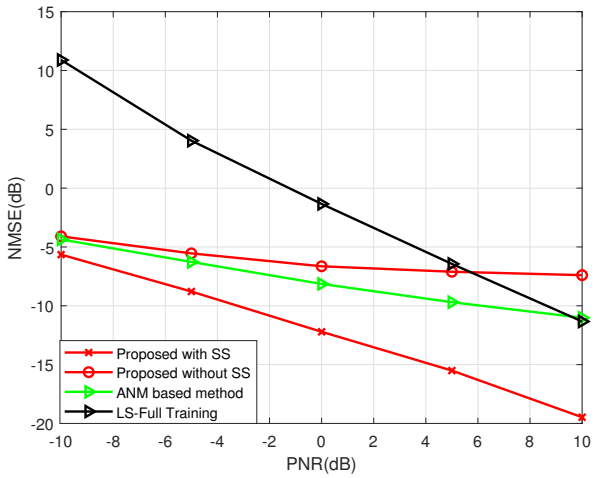
channel uses for training, with S channel uses for estimating the outer angles in Stage 1 and DL_F channel uses for estimating the IRS angles in Stage 2.



(a) MSE of outer angle estimation versus PNR



(b) MSE of IRS angle and composite path gain estimation versus PNR



(c) NMSE of cascaded channel matrix estimation versus PNR

Figure 4.5: Performance of the path angle, gain and cascaded channel estimation with $L_F = 1, L_G = 2$, ULAs at the transmitter, receiver, and IRS, $N_T = N_R = 16, N_I = 32, Q_R = Q_T = 2$.

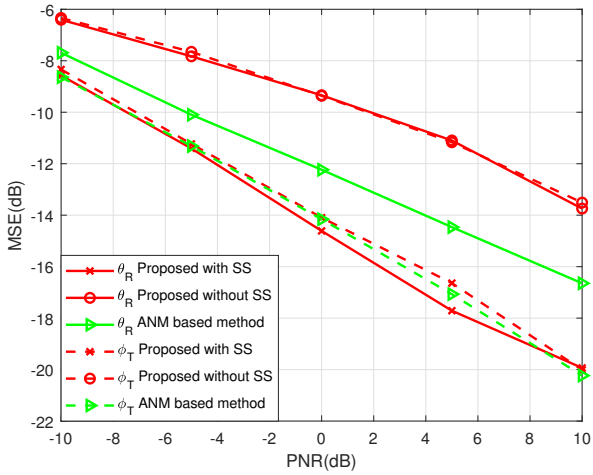
4.4.1 ULA at the IRS

We first consider the case with ULAs at the transmitter, receiver, and IRS. We compare the proposed estimator with the following two alternatives:

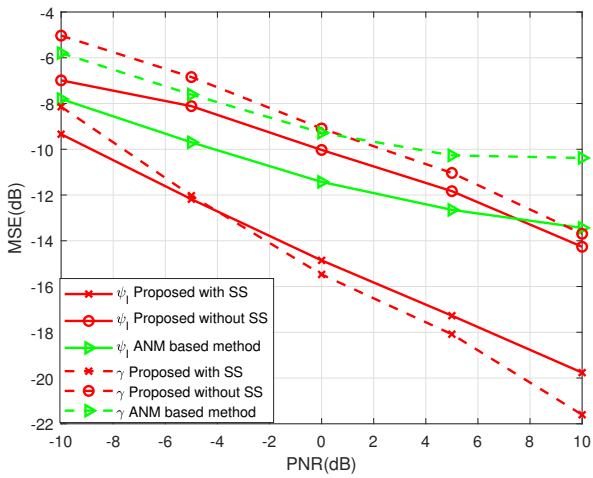
- *The ANM-based two-stage estimator of [76]*: At stage 1, randomly generated precoding and combining matrices $\widetilde{\mathbf{P}}$ and $\widetilde{\mathbf{W}}$ are applied to produce the training observation $\widetilde{\mathbf{Y}}_0 = \widetilde{\mathbf{W}}^H \mathbf{G}\boldsymbol{\Omega}_0\mathbf{F}\widetilde{\mathbf{P}} + \mathbf{N} \in \mathbb{C}^{\frac{SQ_R}{N_T}\times N_T}$ for a fixed, randomly generated $\boldsymbol{\Omega} = \mathrm{diag}(\boldsymbol{\omega}_0)$, followed by ANM for estimating $\boldsymbol{\theta}_R$ and $\boldsymbol{\phi}_T$. In order to compare under the same training overhead, we set $\widetilde{\mathbf{P}} \in \mathbb{C}^{N_T\times N_T}$ and $\widetilde{\mathbf{W}} \in \mathbb{C}^{N_R\times \frac{SQ_R}{N_T}}$ at Stage 1. The estimation accuracy of $\boldsymbol{\phi}_T$ and $\boldsymbol{\theta}_R$ depends on the numbers of precoder and combiner vectors, respectively. At Stage 2, the precoder and combiner are redesigned using $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\phi}}_T$ in the same way as in Section 4.3.3.1. However, the IRS phase shifts vary randomly at D steps with all the IRS elements switched on. ANM is applied to estimate each IRS angle separately. This ANM-based estimator has a complexity of $\mathcal{O}((\max\{N_T + SQ_R/N_T, N_R + N_T, N_I + 1\})^{3.5})$ when semidefinite programming (SDP) is applied to solve the ANM problems. This is generally more complex than the proposed estimator, especially for large systems.

- *The LS estimator*: Approximately unitary precoders $\widetilde{\mathbf{P}} \in \mathbb{C}^{N_T\times N_T}$ and combiners $\widetilde{\mathbf{W}} \in \mathbb{C}^{N_R\times N_R}$ are implemented using the PE-Altmin algorithm for the hybrid transmitter and receiver, and the IRS phase shifts are selected as the columns of the DFT matrix $\bar{\boldsymbol{\Omega}} \in \mathbb{C}^{N_I\times N_I}$. The training overhead is $T_{\mathrm{LS}} = N_T N_I N_R/Q_R$ channel uses.

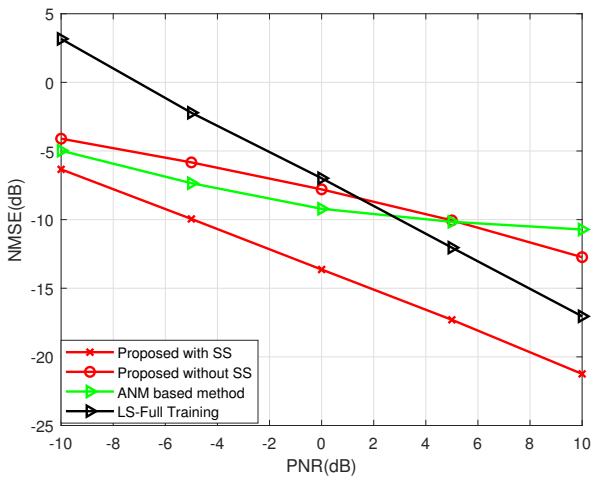
Fig. 4.5 compares the performance for $L_F = 1, L_G = 2$. The proposed and the ANM-based methods apply the two-stage training with $S = 64, D = 16$ and $T = S + DL_F = 80$ while the LS estimator uses a much higher training overhead of $T_{\mathrm{LS}} = 4096$. It can be seen that the proposed estimator achieves good accuracies for estimating the channel parameter and the overall cascaded channel. FBSS effectively



(a) MSE of outer angle estimation versus PNR



(b) MSE of IRS angle and composite path gain estimation versus PNR



(c) NMSE of cascaded channel matrix estimation versus PNR

Figure 4.6: Performance of the path angle, gain and cascaded channel estimation for the system same as that in Fig. 5 except $L_F = 2$, $T = 96$.

improves the angle estimation, especially when there are multiple paths, as in the case for $\boldsymbol{\theta}_R$ with $L_G = 2$. Compared with the ANM-based estimator, the proposed estimator achieves a significantly more accurate estimation of the inner angles $\boldsymbol{\psi}_I$ and path gains. This is because the ANM-based estimator assumes both $\mathbf{A}^H(\widehat{\boldsymbol{\phi}}_F)\mathbf{P}$ and $\mathbf{W}^H\mathbf{A}(\widehat{\boldsymbol{\theta}}_G)$ as identity matrices and then solves separate ANM problems to find the IRS angles. As mentioned in Section 4.3.3.1, this requires perfect outer angle estimation and infinite sizes for the transmitter and receiver arrays. The proposed estimator does not rely on such an assumption and thus achieves a more robust performance. This also translates into the improvement of the performance of the cascaded channel estimation. Both the proposed and the ANM-based estimators significantly outperform the LS estimator. The latter requires a much higher training overhead but does not benefit from the channel sparsity and the knowledge of the array responses. Fig. 4.6 shows the performance for a multi-path scenario with $L_F = L_G = 2$. It is seen that the proposed estimator with FBSS shows the best performance among the candidate estimators. Furthermore, more significant gains are observed for the proposed estimator when L_F is increased from that of Fig. 4.5.

The channel estimation performance versus training overhead T is demonstrated in Fig. 4.7, where T is varied by varying D in Stage 2. The results suggest that the performance of the proposed method generally improves when T increases. When FBSS is used, around 4dB gain can be achieved for the proposed solution when T varies from 88 to 128. The ANM method shows a more stable performance with respect to T because the separate estimation of the inner angles is influenced by the leakage between the paths which can not be effectively mitigated by increasing the training data.

4.4.2 UPA at the IRS

A case with UPA at the IRS is demonstrated in Fig. 4.8 where a 16×16 UPA with $N_I = 256$ is employed at the IRS. For comparison with the proposed esti-

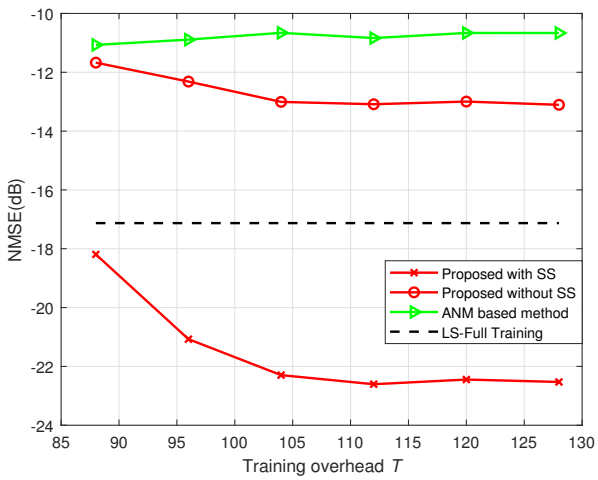


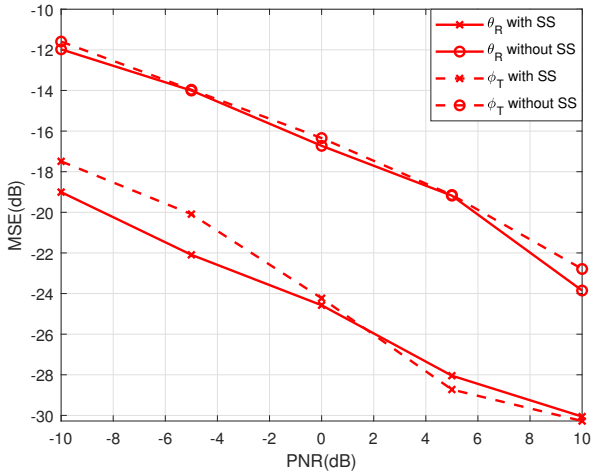
Figure 4.7: Channel estimation performance versus training overhead $T = S + L_F D$ at PNR = 10 dB with $L_F = L_G = 2$, ULAs at the transmitter, receiver and IRS, $N_T = N_R = 16, N_I = 32$, and $Q_R = Q_T = 2$. The proposed method and the ANM-based method apply the two-stage training with $S = 64$ fixed for Stage 1 and D varying from 12 to 32 for Stage 2, while the LS estimator has a fixed training overhead of $T_{\text{LS}} = 4096$.

mator, we consider the LS estimator and the CS-based TRICE estimator of [73]. For the proposed estimator, $S = 128, D = 31, N_y = N_z = 16, J_y = J_z = 1$ and $T = S + L_F D = 190$. For the TRICE estimator $\bar{N_T} = \bar{N_R} = 16$, $\bar{N} = 32$, $T_{\text{TRICE}} = 2048, G_t = G_r = 64, G = 128 \times 128$. For the LS estimator, $T_{\text{LS}} = 65536$. Similarly to the proposed estimator, the TRICE estimator first estimates the outer angles followed by estimating the IRS angles and composite path gains using CS. However, it has a single stage of training and does not exploit the estimated outer angles while designing the training precoder and combiner. Due to the lower beamforming gains, the TRICE estimator requires a higher training overhead. Denote by $\bar{\mathbf{P}} \in \mathbb{C}^{N_T \times \bar{N_T}}$, $\bar{\mathbf{W}} \in \mathbb{C}^{N_R \times \bar{N_R}}$ and the IRS phase shift matrix $\bar{\mathbf{\Omega}} \in \mathbb{C}^{N_I \times \bar{N_I}}$, where $\bar{N_T}$, $\bar{N_R}$ and $\bar{N_I}$ represent the numbers of precoders, combining vectors and IRS states, respectively. They are used by the TRICE estimator to produce $\bar{N_T}\bar{N_R}\bar{N_I}$ observations using $\bar{N_T}\bar{N_R}\bar{N_I}/Q_R$ channel uses. The TRICE estimator has a complexity of $\mathcal{O}(L_G L_F(\bar{N_T}\bar{N_R}G_r G_t + \bar{N_I}G))$, where G_t, G_r and G represent the grid sizes for the AoD, AoA and IRS angle, respectively. In our simulations, $\bar{\mathbf{P}}$ and $\bar{\mathbf{W}}$ are taken from approximated unitary matrices (randomly constructed in the same way as $\mathbf{X}_T$ and $\mathbf{X}_R$ in (4.15)) and the IRS phase shift matrix $\bar{\mathbf{\Omega}}$ has phase shifts

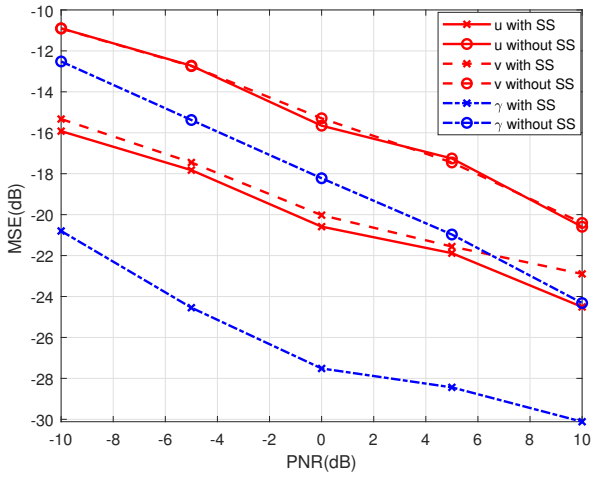
uniformly distributed over $[0, 2\pi)$. From Fig. 4.8, the outer angle estimation with the proposed estimator has a similar performance as for the case with ULA at the IRS. High accuracy for estimating $\mathbf{u}$ and $\mathbf{v}$ is also achieved, and the FBSS is effective for improving the performance with the L-shaped array adopted at the IRS. With FBSS, the proposed estimator achieves an overall performance of cascaded channel estimation significantly better than the alternative estimators, even at a much lower training overhead. This is due to the two-stage training which benefits from the beamforming gains at Stage 2, and the super-resolution estimation of the outer and IRS angles.

4.5 Summary

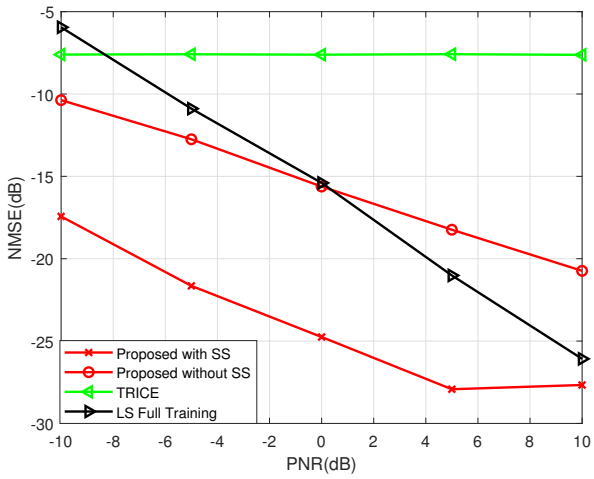
In summary, we have presented a parametric method for estimating the cascaded channel for fully passive IRS-aided MIMO systems with hybrid transceivers. The proposed estimator benefits from the low-rank nature of the channel and the knowledge of the array responses. It provides a low-complexity, multiple-stage solution using simple yet effective tools, including IMC, FBSS, and the root-MUSIC algorithm. This solution can progressively obtain the channel parameters and the training and estimation process adapts to the knowledge generated, which not only provides better estimation performance but also reduces the training overhead. As seen from the simulation results, the proposed estimator outperforms several recently studied solutions in estimation accuracy. The overall computational complexity of the proposed estimator is also kept low.



(a) MSE of outer angle estimation versus PNR



(b) MSE of IRS angle and composite path gain estimation versus PNR



(c) NMSE of cascaded channel matrix estimation versus PNR

Figure 4.8: Performance of the path angle, gain and cascaded channel estimation for a system with a UPA at the IRS, ULAs at the transmitter and receiver, and $L_F = 2, L_G = 2$. $N_T = N_R = 32, N_I = 256, Q_R = Q_T = 4$.

Chapter 5

Low-Rank Matrix Sensing-Based Wideband Channel Estimation for mmWave and THz Hybrid MIMO Systems

5.1 Introduction

This chapter studies the channel estimation for wideband hybrid MIMO systems. We present solutions exploiting the low rankness of the concatenated channel matrix of the delay taps. The channel estimation problem is formulated as a low-rank matrix sensing (LRMS) problem and solved using a low-complexity GCG-ALTMIN algorithm. This LRMS-based solution can accommodate different channel dimensions, precoder/combiner, and training structures. Furthermore, it can be applied without knowledge of the array responses of the transceiver arrays. A preconditioned conjugate gradient (PCG) algorithm-based implementation and a low-rank matrix completion (LRMC) formulation are also introduced to further reduce the computational complexity. To enhance the performance of fat and tall channel ma-

trices, we introduce a matrix reshaping strategy that can preserve the channel rank by exploiting the shift-invariance property of uniform arrays. This can effectively enhance the performance at a given training overhead. We also introduce spectrum denoising (SD) approach for further improving the performance when knowledge of the array response is available. Simulation results suggest that the proposed solutions can improve the channel estimation performance and reduce the computational complexity as compared to several representative channel estimation schemes.

The remaining chapter is organized as follows. The system model is introduced in Section 5.2. The proposed LRMS-based channel estimation techniques are presented in Section 5.3. Computational complexity analysis and simulation results are presented in Section 5.4 and conclusions are drawn in Section 5.5.

5.2 System Model

The system model is the same as shown in Fig. 3.1, we consider a single-carrier hybrid wideband MIMO system with N_T transmitter antennas and N_R receiver antennas. which operates in the mmWave/THz bands. Similar to [81], hybrid beamforming with fully connected networks of phase shifters is utilized at both the transmitter and receiver to achieve better energy efficiency. The numbers of RF chains at the transmitter and receiver are given by $Q_T \leq N_T$ and $Q_R \leq N_R$, respectively. ULAs are assumed for both the transmitter and receiver[1].

Following [36, 38, 52, 87, 117–120] , a geometric wideband multi-tap channel model with N_c delay taps is considered, where the d-th tap matrix of the sampled channel model is given as

$$\mathbf{H}_d = \sqrt{\frac{N_T N_R}{L}} \sum_{l=0}^{L-1} \gamma_l p(dT_s - \tau_l) \mathbf{a}_R(\theta_{R,l}) \mathbf{a}_T^H(\phi_{T,l}), \tag{5.1}$$

[1]The proposed training and channel estimation solutions can also be easily extended for transmitters and receivers equipped with uniform planer arrays (UPAs) and/or switches-based hybrid transceivers.

where T_s is the sampling period,γ_l represents the complex channel gain, τ_l the path delay, $\phi_{T,l}$ and $\theta_{R,l}$ the AoD and AoA, respectively, and $\mathbf{a}_T(\phi_{T,l})$ and $\mathbf{a}_R(\theta_{R,l})$ are the transmitter and receiver array responses, respectively, all for the l-th path. There are in total L propagation paths with randomly distributed delays $\{\tau_l\}$. The impulse response $p(\cdot)$ accounts for the joint effect of the transmitter pulse shaping and receiver matched filtering and it can lead to the spread of a path l to multiple delay taps $\{\mathbf{H}_d\}$ within the support of $p(dT_s - \tau_l)$ as τ_l/T_s is generally not integer in practice. Without loss of generality, we assume that $p(\cdot)$ is a raised-cosine function. In contrast to the narrowband channel model considered in Chapters 3 and 4, the introduction of path delays further complicates the channel estimation process for wideband channels. The channel of different delay taps can be very different from each other. This is due to the fact that channel paths contributing to different delay taps can be different. Hence the concatenated channel matrix becomes N_C times large as compared to narrowband channel matrix. The receiver array response vector for the AoA $\theta_{R,l}$ is given as

$$\mathbf{a}_R(\theta_{R,l}) = \frac{1}{\sqrt{N_R}} \left[1, e^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}, \cdots, e^{j(N_R-1)\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}\right]^T, \tag{5.2}$$

where the carrier wavelength is λ_c and the antenna spacing is $\beta = \lambda_c/2$. The transmitter array response vector corresponding to $\phi_{T,l}$ is similar. The channel matrix for the d-th delay tap can be rewritten as

$$\mathbf{H}_d = \mathbf{A}_R(\boldsymbol{\theta}_R)\mathbf{G}_d\mathbf{A}_T^H(\boldsymbol{\phi}_T) \in \mathbb{C}^{N_R \times N_T} \tag{5.3}$$

where

$$\mathbf{G}_d = \mathrm{diag}(g_{d,1}, g_{d,2}, \ldots, g_{d,L}) \in \mathbb{C}^{L \times L} \tag{5.4}$$

is a diagonal matrix containing the channel path gains scaled with the impulse

response of the pulse shaping filter for the d-th tap, with entries

$$g_{d,l} = \sqrt{\frac{N_T N_R}{L}} \alpha_l p(dT_s - \tau_l).$$

Furthermore, $\mathbf{A}_T(\boldsymbol{\phi}_T) \in \mathbb{C}^{N_T \times L}$ and $\mathbf{A}_R(\boldsymbol{\theta}_R) \in \mathbb{C}^{N_R \times L}$ represent the array response matrices for the transmitter and receiver, respectively, consisting of the corresponding steering vectors.

This thesis focuses on the single-carrier wideband systems. However, extension of the proposed frameworks to the multi-carrier systems can be done with additional steps. The channel model for multi-carrier wideband system for the k-th sub-carrier can be written as [121] $\mathbf{H}[k] = \sum_{d=0}^{N_C-1} \mathbf{H}_d e^{-j\frac{2\pi kd}{K}}$, where K is the total number of sub-carriers. From the above it can be seen when the channels for all delay taps are estimated, the channel for the multi-carrier system can be estimated.

The above channel model applies to both mmWave and THz bands when the bandwidth and the size of the transceiver antenna arrays are moderate. For THz bands, due to higher path losses the channel is mostly dominated by the line-of-sight (LOS) component with very few non-LOS components [117, 120, 122]. Therefore, THz channels are generally sparser in the angular domain as compared to mmWave channels. The treatments for ultra-large arrays and ultra-wideband systems, which may face the challenges of spatial wideband effects and beam squint [78, 122, 123], are left for future work.

5.3 LRMS-Based Wideband MIMO Channel Estimation

Accurate channel estimation is essential for optimizing the transceivers for MIMO communication systems. This section presents the proposed LRMS-based wideband channel estimation solutions. First, the training scheme for the proposed channel estimators is introduced, followed by the recovery method based on the GCG-

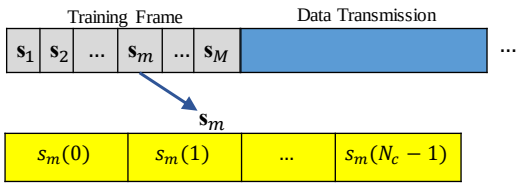


Figure 5.1: Transmission Structure

ALTMIN algorithm. Then PCG and LRMC-based techniques are introduced for computational complexity reduction. Finally, the matrix reshaping and spectrum denoising approaches are reported for performance enhancement.

5.3.1 Wideband Channel Training and LRMS-Based Channel Estimation

We consider time-invariant wideband MIMO channels in this chapter. The training frame consists of M subframes and each is transmitted over N_C time instances, where N_C equals the number of channel taps. Similar to [81], the precoder and combiner are assumed to remain unchanged within a single subframe. During the n-th time instance within the m-th subframe, a single symbol $s_m(n)$ is transmitted through the hybrid precoder and Q_R symbols are observed from the output of the hybrid combiner. The received signal for the n-th time instance within the m-th subframe is written as

$$\mathbf{y}_m(n) = \mathbf{W}_m^H \sum_{d=0}^{N_C-1} \mathbf{H}_d \mathbf{x}_m(n-d) + \mathbf{W}_m^H \mathbf{n}_m(n) \in \mathbb{C}^{Q_R \times 1}, \tag{5.5}$$

where the hybrid combiner for the m-th subframe is given as $\mathbf{W}_m = \mathbf{W}_{\mathrm{RF}_m} \mathbf{W}_{\mathrm{BB}_m}$ with the RF combiner $\mathbf{W}_{\mathrm{RF}_m} \in \mathbb{X}^{N_R \times Q_R}$ and baseband combiner $\mathbf{W}_{\mathrm{BB}_m} \in \mathbb{C}^{Q_R \times Q_R}$, where $\mathbb{X}$ represents the set of feasible phase shifts. The noise $\mathbf{n}_m(n) \in \mathbb{C}^{N_R \times 1}$ follows the complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$, with σ_n^2 being the average

noise power. In (5.5), the transmitted signal is defined as

$$\mathbf{x}_m(n) = \begin{cases} \mathbf{p}_m s_m(n), & 0 \leq n \leq N_C - 1, \\ \mathbf{p}_{m-1} s_{m-1}(n + N_C), & n < 0, \end{cases} \tag{5.6}$$

where the hybrid precoder for the m-th subframe is given as $\mathbf{p}_m = \mathbf{P}_{\mathrm{RF}_m}\mathbf{p}_{\mathrm{BB}_m}$ with $\mathbf{P}_{\mathrm{RF}_m} \in \mathbb{X}^{N_T \times Q_T}$ as the RF precoder and $\mathbf{p}_{\mathrm{BB}_m} \in \mathbb{C}^{Q_T \times 1}$ as the baseband precoder. The received signal for the n-th time instance of the m-th subframe can be alternatively written as

$$\mathbf{y}_m(n) = \mathbf{W}_m^H \mathbf{H} \boldsymbol{\xi}_{m,n} + \mathbf{W}_m^H \mathbf{n}_m(n) \in \mathbb{C}^{Q_R \times 1}, \tag{5.7}$$

where

$$\mathbf{H} \triangleq [\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{N_C - 1}] \in \mathbb{C}^{N_R \times N_C N_T} \tag{5.8}$$

is the channel matrix with the coefficients of all the delay taps concatenated together, and

$$\boldsymbol{\xi}_{m,n} \triangleq [\mathbf{x}_m^T(n), \mathbf{x}_m^T(n-1), \ldots, \mathbf{x}_m^T(n - N_C + 1)]^T \in \mathbb{C}^{N_C N_T \times 1}. \tag{5.9}$$

The received signal of the M subframes can be written as $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_M] \in \mathbb{C}^{Q_R \times N_C M}$ where $\mathbf{Y}_m \triangleq [\mathbf{y}_m(0), \mathbf{y}_m(1), \ldots, \mathbf{y}_m(N_C - 1)]$.

The training structure is shown in Fig. 5.1, where MN_C time instances are allocated for training. Without loss of generality, we assume $\|\mathbf{p}_m \mathbf{s}_m\|_F^2 = 1$ and $\|\mathbf{W}_m\|_F^2 = Q_R$ to fix the transmit and receive power for the m-th subframe, where $\|\cdot\|_F$ denotes the Frobenius norm and $\mathbf{s}_m \triangleq [s_m(0), s_m(1), \ldots, s_m(N_C - 1)]$.

From (5.1), the AoAs and AoDs are shared by all delay taps and so $\{\mathbf{H}_d\}$ and $\mathbf{H}$ share the same column subspace. We thus have

$$\mathrm{rank}(\mathbf{H}_d) \leq \mathrm{rank}(\mathbf{H}) \leq L, \forall d.$$

The number of paths, L, for the mmWave and THz bands is typically low [92, 124].

This suggests that the concatenated channel matrix consisting of all the delay taps is low-rank. We exploit such low-rankness of the channel matrix and formulate the channel estimation problem as a low-rank matrix sensing (LRMS) problem as

$$\min_{\widehat{\mathbf{H}}} \operatorname{rank}(\widehat{\mathbf{H}}), \quad \text{s.t.} \quad \|\mathcal{A}(\widehat{\mathbf{H}}) - \operatorname{vec}(\mathbf{Y})\|_F^2 \leq \delta^2, \tag{5.10}$$

where δ^2 is a threshold accounting for the noise and the sensing operator

$$\mathcal{A}(\widehat{\mathbf{H}}) \triangleq \mathbf{K}\operatorname{vec}(\widehat{\mathbf{H}}) \tag{5.11}$$

with the sensing matrix given as

$$\mathbf{K} \triangleq \begin{bmatrix} \boldsymbol{\Xi}_1^T \otimes \mathbf{W}_1^H \\ \boldsymbol{\Xi}_2^T \otimes \mathbf{W}_2^H \\ \vdots \\ \boldsymbol{\Xi}_M^T \otimes \mathbf{W}_M^H \end{bmatrix} \in \mathbb{C}^{MN_CQ_R \times N_CN_RN_T}, \tag{5.12}$$

where

$$\boldsymbol{\Xi}_m = [\boldsymbol{\xi}_{m,0}, \boldsymbol{\xi}_{m,1}, \cdots, \boldsymbol{\xi}_{m,N_C-1}] \in \mathbb{C}^{N_CN_T \times N_C}. \tag{5.13}$$

Remark 1: An impulsive structure may be adopted for the training symbols by choosing $\mathbf{s}_m = [1, 0, \ldots, 0], \forall m$, following [81]. Such zero padded training symbols can mitigate inter-subframe-interference, which may reduce the channel estimation complexity since it leads to a simplified transmitted training block

$$\boldsymbol{\Xi}_m = \mathbf{I} \otimes \mathbf{p}_m. \tag{5.14}$$

Note that, however, more general training sequences can be used for the proposed LRMS-based channel estimator.

The problem in (5.10) is NP-hard. We relax the problem to estimate the channel

matrix $\mathbf{H}$ from the received signal by solving

$$\arg\min_{\widehat{\mathbf{H}}} \quad f(\widehat{\mathbf{H}}) + \mu\|\widehat{\mathbf{H}}\|_*, \tag{5.15}$$

where

$$f(\widehat{\mathbf{H}}) \triangleq \frac{1}{2}\|\mathbf{K}\text{vec}(\widehat{\mathbf{H}}) - \text{vec}(\mathbf{Y})\|_F^2 \tag{5.16}$$

and $\mu > 0$ is a regularization parameter. We propose to solve the problem in (5.15) using the GCG-ALTMIN algorithm [125] to exploit its low complexity in solving similar problems. For completeness, we first outline the GCG-ALTMIN algorithm. We will then study the low-complexity implementations of LRMS and performance improvement by using matrix reshaping and SD.

The GCG-ALTMIN algorithm iteratively refines the channel estimate by rank-one updates and alternate minimization. At the l-th iteration, rank-one update via the GCG operation is applied to a rank-$(l-1)$ solution to produce a higher-rank initialization. The ALTMIN algorithm is then used to refine the solution. Let $\widehat{\mathbf{H}}_l$ denote the estimate of $\mathbf{H}$ after the l-th GCG iteration. Following [125], the GCG algorithm first updates the solution as follows:

$$\widehat{\mathbf{H}}_l = (1-\eta_l)\widehat{\mathbf{H}}_{l-1} + \zeta_l \mathbf{R}_l, \tag{5.17}$$

where $\eta_l \in [0,1]$ is the step size, $\mathbf{R}_l = \mathbf{u}_l\mathbf{v}_l^H$ denotes the outer product of the top singular vector pair of $\text{vec}^{-1}(-\nabla f(\widehat{\mathbf{H}}_l))$, and ∇ denotes the gradient. Then

$$\nabla f(\widehat{\mathbf{H}}_l) = \mathbf{K}^H\mathbf{K}\text{vec}(\widehat{\mathbf{H}}_{l-1}) - \mathbf{K}^H\text{vec}(\mathbf{Y}). \tag{5.18}$$

Furthermore,

$$\zeta_l = \frac{\mathcal{R}\left(\mathbf{k}_l^H\text{vec}(\mathbf{Y}) - (1-\eta_l)\mathbf{k}_l^H\mathbf{K}\text{vec}(\widehat{\mathbf{H}}_l)\right) - \mu}{\mathbf{k}_l^H\mathbf{k}_l} \tag{5.19}$$

where $\mathcal{R}(\cdot)$ denotes the real part of the number and $\mathbf{k}_l = \mathbf{K}\text{vec}(\mathbf{R}_l)$. The direct calculation of $\mathbf{K}^H\mathbf{K}\text{vec}(\widehat{\mathbf{H}}_{l-1})$ and $\mathbf{K}^H\text{vec}(\mathbf{Y})$ can be computationally expensive.

As will be discussed shortly in Section III-B, we can exploit the factored form of $\mathbf{K}$ in (5.12) to significantly reduce the computational effort.

Based on the initial rank-l estimate provided by the GCG algorithm, $\widehat{\mathbf{H}}_l$ is further refined by a local search using the ALTMIN algorithm. Exploiting a Frobenius norm characterization of the nuclear norm, the ALTMIN algorithm solves iteratively the following regularized least squares problem

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\|\mathbf{K}\mathrm{vec}(\mathbf{U}_l\mathbf{V}_l^H) - \mathrm{vec}(\mathbf{Y})\|_F^2 + \frac{1}{2}\mu(\|\mathbf{U}_l\|_F^2 + \|\mathbf{V}_l\|_F^2), \tag{5.20}$$

by assuming $\widehat{\mathbf{H}}_l = \mathbf{U}_l\mathbf{V}_l^H$ with $\mathbf{U}_l \in \mathbb{C}^{N_R\times l}$ and $\mathbf{V}_l \in \mathbb{C}^{N_C N_T\times l}$. Let $\mathbf{U}_l^i$ and $\mathbf{V}_l^i$ denote the refinement of $\mathbf{U}_l$ and $\mathbf{V}_l$ during the i-th ALTMIN iteration. We can refine $\mathbf{V}_l$ as $\mathbf{V}_l^i = \mathrm{vec}^{-1}(\mathbf{v}_l^i)$ with

$$\mathbf{v}_l^i = \left(\left(\boldsymbol{\mathcal{U}}_l^{i-1}\right)^H \boldsymbol{\mathcal{U}}_l^{i-1} + \mu\mathbf{I}_{lN_TN_C}\right)^{-1} \left(\boldsymbol{\mathcal{U}}_l^{i-1}\right)^H \mathrm{vec}(\mathbf{Y}) \tag{5.21}$$

and

$$\boldsymbol{\mathcal{U}}_l^{i-1} = \mathbf{K}(\mathbf{I}_{N_TN_C} \otimes \mathbf{U}_l^{i-1}) \in \mathbb{C}^{MN_CQ_R\times lN_TN_C}. \tag{5.22}$$

Similarly we can update $\mathbf{U}_l^i$ as $\mathbf{U}_l^i = \mathrm{vec}^{-1}(\mathbf{u}_l^i)$ with

$$\mathbf{u}_l^i = \left(\left(\boldsymbol{\mathcal{V}}_l^i\right)^H \boldsymbol{\mathcal{V}}_l^i + \mu\mathbf{I}_{N_Rl}\right)^{-1} \left(\boldsymbol{\mathcal{V}}_l^i\right)^H \mathrm{vec}(\mathbf{Y}) \tag{5.23}$$

and

$$\boldsymbol{\mathcal{V}}_l^i = \mathbf{K}\left((\mathbf{V}_l^i)^* \otimes \mathbf{I}_{N_R}\right) \in \mathbb{C}^{MN_CQ_R\times lN_R}. \tag{5.24}$$

The final estimate is obtained by iteratively updating $\mathbf{U}_l^i$ and $\mathbf{V}_l^i$. The above GCG-ALTMIN algorithm has the advantage that only the factored forms of $\{\widehat{\mathbf{H}}_l\}$ are updated, which results in low computation and storage requirements. A stopping criterion can be used to terminate its iterations [125].

5.3.2 Complexity Reduction via PCG Implementation

Computation of the high-dimensional Gramian matrices and inverses in (5.21) and (5.23) can incur a high computational complexity of $\mathcal{O}(l^3N_R^3+l^3N_T^3N_C^3+l^2N_R^2MN_CQ_R+l^2N_T^2N_C^3MQ_R)$, which can be challenging for real-time applications. In order to address this challenge, we apply the iterative preconditoned conjugate gradient (PCG) algorithm [126–128] to calculate (5.21) and (5.23). Although PCG has been widely used for solving large linear equations, its effectiveness depends on the preconditioner used to accelerate the convergence and also the scheduling of the matrix-vector multiplications which affect its per-iteration complexity. In the following, we introduce solutions to enable PCG-based, low-cost LRMS estimators. .

We focus on the computation of (5.21) and note that (5.23) can be treated in the same way. For conciseness, let us drop the subscripts and let

$$\begin{aligned} \boldsymbol{\mathcal{B}} &= \boldsymbol{\mathcal{U}}^H\boldsymbol{\mathcal{U}} + \mu\mathbf{I}, \\ \mathbf{d} &= \boldsymbol{\mathcal{U}}^H\mathbf{y}. \end{aligned} \tag{5.25}$$

We can then write (5.21) as the solution of

$$\boldsymbol{\mathcal{B}}\mathbf{v} = \mathbf{d}. \tag{5.26}$$

Since $\boldsymbol{\mathcal{B}}$ is positive-definite, Hermitian, PCG can be applied to solve (5.26). For simplicity, we propose to apply the diagonal Jacobi precondtioner $\boldsymbol{\mathcal{P}} = (\mathrm{diag}(\boldsymbol{\mathcal{B}}))^{-\frac{1}{2}}$, which is computationally efficient. This preconditioner can significantly reduce the condition number of $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{P}}^H$ and hence improve the convergence of the conjugate gradient (CG) algorithm. Applying preconditioning, instead of (5.26) we solve

$$\widetilde{\boldsymbol{\mathcal{B}}}\widetilde{\mathbf{v}} = \widetilde{\mathbf{d}}, \tag{5.27}$$

with $\widetilde{\boldsymbol{\mathcal{B}}} = \boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{P}}^H$, $\widetilde{\mathbf{d}} = \boldsymbol{\mathcal{P}}^H\mathbf{d}$ and $\widetilde{\mathbf{v}} = (\boldsymbol{\mathcal{P}}^H)^{-1}\mathbf{v}$. The PCG algorithm updates the estimate of $\widetilde{\mathbf{v}}$ during each iteration as

$$\begin{aligned}
\mathbf{t}_i &= \begin{cases} \mathbf{r}_0, & i = 1, \\ \mathbf{r}_{i-1} + \frac{\mathbf{r}_{i-1}^H\mathbf{r}_{i-1}}{\mathbf{r}_{i-2}^H\mathbf{r}_{i-2}}\mathbf{t}_{i-1}, & i > 1, \end{cases} \\
\alpha_i &= \frac{\mathbf{r}_{i-1}^H\mathbf{r}_{i-1}}{\mathbf{t}_i^H\widetilde{\boldsymbol{\mathcal{B}}}\mathbf{t}_i}, \\
\widetilde{\mathbf{v}}_i &= \widetilde{\mathbf{v}}_{i-1} + \alpha_i\mathbf{t}_i, \\
\mathbf{r}_i &= \mathbf{r}_{i-1} - \alpha_i\widetilde{\boldsymbol{\mathcal{B}}}\mathbf{t}_i,
\end{aligned} \tag{5.28}$$

with initial values set as $\mathbf{r}_0 = \widetilde{\mathbf{d}}$ and $\widetilde{\mathbf{v}}_0 = \mathbf{0}$. Clearly, with PCG, the computation of the Gramian matrices and matrix inversions are avoided. The computational complexity is now dominated by the matrix-vector product of the form $\widetilde{\boldsymbol{\mathcal{B}}}\mathbf{t}$ at each iteration. Its direct calculation still has a high complexity due to the high dimensionality of the matrix $\widetilde{\boldsymbol{\mathcal{B}}}$. We suggest addressing this issue by exploiting the factored form of $\widetilde{\boldsymbol{\mathcal{B}}}$ and step-wise multiplication. To see this, note that

$$\begin{aligned}
\widetilde{\boldsymbol{\mathcal{B}}}\mathbf{t} =& \boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{P}}^H\mathbf{t} \\
=& \boldsymbol{\mathcal{P}}\left(\boldsymbol{\mathcal{U}}^H\boldsymbol{\mathcal{U}} + \mu\mathbf{I}\right)\boldsymbol{\mathcal{P}}^H\mathbf{t} \\
=& \boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{U}}^H\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{P}}^H\mathbf{t} + \mu\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{P}}^H\mathbf{t}.
\end{aligned} \tag{5.29}$$

As $\boldsymbol{\mathcal{P}}$ is diagonal, matrix-vector multiplications involving $\boldsymbol{\mathcal{P}}$ have negligible complexities. We thus focus on the first term and compute it in a step-wise manner as

$$\widetilde{\mathbf{t}} = \boldsymbol{\mathcal{P}}^H\mathbf{t},\ \mathbf{c}_1 = \boldsymbol{\mathcal{U}}\widetilde{\mathbf{t}},\ \mathbf{c}_2 = \boldsymbol{\mathcal{U}}^H\mathbf{c}_1,\ \mathbf{c}_3 = \boldsymbol{\mathcal{P}}\mathbf{c}_2. \tag{5.30}$$

Recalling (5.22), we then have $\boldsymbol{\mathcal{U}} = \mathbf{K}(\mathbf{I}_{N_T N_C} \otimes \mathbf{U})$ and

$$\mathbf{c}_1 = (\mathbf{K}(\mathbf{I}_{N_T N_C} \otimes \mathbf{U}))\,\widetilde{\mathbf{t}} = \begin{bmatrix} \left(\boldsymbol{\Xi}_1^T \otimes \mathbf{W}_1^H\right)(\mathbf{I}_{N_T N_C} \otimes \mathbf{U})\widetilde{\mathbf{t}} \\ \left(\boldsymbol{\Xi}_2^T \otimes \mathbf{W}_2^H\right)(\mathbf{I}_{N_T N_C} \otimes \mathbf{U})\widetilde{\mathbf{t}} \\ \vdots \\ \left(\boldsymbol{\Xi}_M^T \otimes \mathbf{W}_M^H\right)(\mathbf{I}_{N_T N_C} \otimes \mathbf{U})\widetilde{\mathbf{t}} \end{bmatrix}. \tag{5.31}$$

Now consider the m-th block in (5.31)

$$\begin{aligned} \left(\boldsymbol{\Xi}_m^T \otimes \mathbf{W}_m^H\right)(\mathbf{I}_{N_T N_C} \otimes \mathbf{U})\widetilde{\mathbf{t}} &= \left(\boldsymbol{\Xi}_m^T \otimes (\mathbf{W}_m^H \mathbf{U})\right)\widetilde{\mathbf{t}} \\ &= \mathrm{vec}(\mathbf{W}_m^H \mathbf{U} \mathrm{vec}^{-1}(\widetilde{\mathbf{t}})\boldsymbol{\Xi}_m), \forall m, \end{aligned} \tag{5.32}$$

where we recall $\mathbf{W}_m \in \mathbb{C}^{N_R \times Q_R}$, $\mathbf{U} \in \mathbb{C}^{N_R \times l}$, $\mathrm{vec}^{-1}(\widetilde{\mathbf{t}}) \in \mathbb{C}^{l \times N_C N_T}$, and $\boldsymbol{\Xi}_m \in \mathbb{C}^{N_C N_T \times N_C}$. The above item can be computed in a step-wise manner as

$$\mathbf{C}_{1_1} = \mathbf{W}_m^H \mathbf{U} \in \mathbb{C}^{Q_R \times l},\ \mathbf{C}_{1_2} = \mathrm{vec}^{-1}(\widetilde{\mathbf{t}})\boldsymbol{\Xi}_m \in \mathbb{C}^{l \times N_C},\ \mathbf{C}_{1_3} = \mathbf{C}_{1_1}\mathbf{C}_{1_2} \in \mathbb{C}^{N_R \times N_C}. \tag{5.33}$$

The complexities of these steps are $\mathcal{O}(lQ_R N_R)$, $\mathcal{O}(lN_T N_C^2)$ and $\mathcal{O}(lQ_R N_C)$, respectively. Since there are M blocks, the total complexity of calculating $\mathbf{c}_1$ is $\mathcal{O}(Ml(Q_R N_R + N_T N_C^2 + Q_R N_C))$. For computing $\mathbf{c}_2 = \boldsymbol{\mathcal{U}}^H \mathbf{c}_1$, we use

$$\begin{aligned} \mathbf{c}_2 = (\mathbf{K}(\mathbf{I}_{N_T N_C} \otimes \mathbf{U}))^H \mathbf{c}_1 &= \left(\begin{bmatrix} \boldsymbol{\Xi}_1^T \otimes \mathbf{W}_1^H \\ \boldsymbol{\Xi}_2^T \otimes \mathbf{W}_2^H \\ \vdots \\ \boldsymbol{\Xi}_M^T \otimes \mathbf{W}_M^H \end{bmatrix} (\mathbf{I}_{N_T N_C} \otimes \mathbf{U}) \right)^H \mathbf{c}_1 \\ &= \sum_{m=1}^{M} \left(\left(\boldsymbol{\Xi}_m^T \otimes \mathbf{W}_m^H\right)(\mathbf{I}_{N_T N_C} \otimes \mathbf{U})\right)^H \mathbf{c}_{1,m} = \sum_{m=1}^{M} \left((\mathbf{I}_{N_T N_C} \otimes \mathbf{U}^H)(\boldsymbol{\Xi}_m^* \otimes \mathbf{W}_m)\right)\mathbf{c}_{1,m} \\ &= \sum_{m=1}^{M} \left(\boldsymbol{\Xi}_m^* \left(\mathbf{U}^H \mathbf{W}_m\right)\right)\mathbf{c}_{1,m} = \sum_{m=1}^{M} \mathrm{vec}(\left(\mathbf{U}^H \mathbf{W}_m\right)\mathrm{vec}^{-1}(\mathbf{c}_{1,m})\boldsymbol{\Xi}_m^H) \\ &= \sum_{m=1}^{M} \mathrm{vec}(\left(\mathbf{C}_{1_1}^H\right)\mathrm{vec}^{-1}(\mathbf{c}_{1,m})\boldsymbol{\Xi}_m^H), \end{aligned} \tag{5.34}$$

where $\mathbf{C}_{1_1} \in \mathbb{C}^{Q_R \times l}$ is already computed, $\text{vec}^{-1}(\mathbf{c}_{1,m}) \in \mathbb{C}^{Q_R \times N_C}$, and $\mathbf{\Xi}_m \in \mathbb{C}^{N_C N_T \times N_C}$. Using a similar strategy as for $\mathbf{c}_1$, the complexity of calculating $\mathbf{c}_2$ is $\mathcal{O}(lM(Q_R N_C + N_T N_C^2))$. With the above computations based on the factored forms, the complexity per PCG iteration is $\mathcal{O}(lM(Q_R N_R + N_T N_C^2 + Q_R N_C))$. This can be significantly lower than the complexity of $\mathcal{O}(lMN_T N_C^2 Q_R)$ for the direct calculation for $\mathbf{c}_1$ and $\mathbf{c}_2$ even using already calculated $\boldsymbol{\mathcal{U}}$. The PCG treatments of (5.23) require the same order of per-iteration complexity as that for (5.21).

Recall that $\mathbf{\Xi}_m = \mathbf{I} \otimes \mathbf{p}_m$ as shown in (5.14) when impulsive training is considered. Due to the block diagonal structure of $\mathbf{\Xi}_m$, the complexity of calculating $\mathbf{c}_1$ and $\mathbf{c}_2$ can be further reduced to $\mathcal{O}(lM(Q_R N_R + N_T N_C + Q_R N_C))$ using the factored forms of the relevant matrices.

5.3.3 Complexity Reduction via LRMC

The above LRMS solution allows arbitrary training schemes $\{\mathbf{W}_m, \mathbf{p}_m, \mathbf{s}_m(n)\}$ to be deployed. However, their sparsity levels affect the complexity of the GCG-ALTMIN-based estimator. We now present an alternative formulation based on LRMC, which can significantly reduce computational complexity. This is achieved by designing $\{\mathbf{W}_m, \mathbf{p}_m, \mathbf{s}_m(n)\}$ in such a way that noisy observations of the entries of a transformed matrix

$$\mathbf{C} = \mathbf{X}_R^H \mathbf{H} \overline{\mathbf{X}}_T \in \mathbb{C}^{N_R \times N_C N_T} \tag{5.35}$$

are obtained, where $\overline{\mathbf{X}}_T$ and $\mathbf{X}_R$ are the transmitter and receiver feature matrices, respectively. In order to accommodate this formulation for the hybrid transceiver and multi-tap wideband channel, we construct the transmitter feature matrix as

$$\overline{\mathbf{X}}_T \triangleq \left[\mathbf{I}_{N_C} \otimes \mathbf{x}_{T,1}, \mathbf{I}_{N_C} \otimes \mathbf{x}_{T,2}, \ldots, \mathbf{I}_{N_C} \otimes \mathbf{x}_{T,N_T}\right] \in \mathbb{C}^{N_C N_T \times N_C N_T} \tag{5.36}$$

with $\mathbf{x}_{T,i}$ representing the i-th column of a transmitter feature matrix $\mathbf{X}_T \in \mathbb{C}^{N_T \times N_T}$.

We aim to obtain first an estimate $\widehat{\mathbf{C}}$ of the transformed matrix $\mathbf{C}$, and then estimate the channel matrix as

$$\widehat{\mathbf{H}} = (\mathbf{X}_T^H)^{-1}\widehat{\mathbf{C}}(\overline{\mathbf{X}}_R)^{-1}. \tag{5.37}$$

Clearly, in order to achieve good performance, the feature matrices $\mathbf{X}_R$ and $\overline{\mathbf{X}}_T$ must be full-rank. Furthermore, they should be well-conditioned such that the inversions do not significantly amplify the errors in $\widehat{\mathbf{C}}$. We still estimate $\mathbf{C}$ by leveraging its low rankness, which originates from the low rankness of $\mathbf{H}$, by using LRMC. This requires that $\mathbf{C}$ satisfies the general incoherence properties for LRMC to be effective. Intuitively, the singular vectors of $\mathbf{C}$ should not be too sparse and should be well spread out (i.e., uncorrelated with standard basis) [129]. This also suggests that the nonzero entries of $\mathbf{C}$ should not be concentrated to only a few rows or columns, such that random sampling can be used for effective matrix completion. Based on the above considerations of the feature matrices, we apply the construction of [50, Section IV] to obtain $\mathbf{X}_T$ and $\mathbf{X}_R$ which are close to unitary and then compute $\overline{\mathbf{X}}_T$ as (5.36). We observe empirically that they meet well the requirements for the feature matrices.

The hybrid precoder and combiner are designed to form these feature matrices. For a given subframe, $\mathbf{p}_m$ is selected as one column of $\mathbf{X}_T$, while $\mathbf{W}_m$ as a group of Q_R columns of $\mathbf{X}_R$. They are implemented using the hybrid transceivers by applying the PE-ALTMIN algorithm [98]. Meanwhile, the impulsive training as (5.14) is adopted. As such, from (5.7), the received signal $\mathbf{Y}$ obtained after training gives noisy samples of $\mathbf{C}$, i.e.,

$$\mathrm{vec}(\mathbf{Y}) \triangleq \mathbf{K}_\Omega \mathrm{vec}(\widetilde{\mathbf{C}}) \in \mathbb{C}^{MN_CQ_R\times 1}, \tag{5.38}$$

where $\widetilde{\mathbf{C}}$ is a noisy version of $\mathbf{C}$. Correspondingly, $\mathbf{K}_\Omega$ is the $MN_CQ_R \times N_RN_TN_C$ sparse matrix with only one nonzero entry in each row, whose position is determined

by the sampling pattern Ω. The training during each subframe corresponds to sampling Q_R rows of N_C columns of $\widetilde{\mathbf{C}}$. The above sampling process can be chosen following uniform spatial sampling (USS) [97].

Note that the training process requires all of $\{\mathbf{W}_m\}$ and $\{\mathbf{p}_m\}$ to be selected from the columns of $\mathbf{X}_R \in \mathbb{C}^{N_R \times N_R}$ and $\mathbf{X}_T \in \mathbb{C}^{N_T \times N_T}$, even for a very large number of training subframes M. Furthermore, in order to ensure feasible implementations using hybrid receivers, $\mathbf{W}_m$ should be selected from fixed sets of the columns of $\mathbf{X}_R$, which shall also be disjoint. This differs from the more general training scheme as discussed in Section 5.3.1, where $\mathbf{W}_m$ and $\mathbf{p}_m$ are not subject to these constraints.

Both $\mathbf{H}$ and $\mathbf{C}$ are low-rank. After training, the transformed matrix $\mathbf{C}$ is first estimated from $\mathbf{Y}$ using LRMC, which is a special instance of the LRMS of (5.10) as

$$\min_{\widehat{\mathbf{C}}} \operatorname{rank}(\widehat{\mathbf{C}}), \quad \text{s.t.} \quad \|\mathcal{A}_\Omega(\widehat{\mathbf{C}}) - \operatorname{vec}(\mathbf{Y})\|_F^2 \leq \delta^2, \tag{5.39}$$

with $\mathcal{A}_\Omega(\widehat{\mathbf{C}}) \triangleq \mathbf{K}_\Omega \operatorname{vec}(\widehat{\mathbf{C}})$ as the sampling operator. The GCG-ALTMIN algorithm can be used, whose complexity is significantly lower than the general LRMS approach due to the low-complexity matrix products involving the extremely sparse matrix $\mathbf{K}_\Omega$. Note that $\mathbf{K}$ is generally a dense matrix in the LRMS formulation in (5.10)-(5.13). After solving (5.39), (5.37) can be applied to obtain the channel estimate.

5.3.4 Performance Enhancement via Matrix Reshaping

In the above, we have discussed methods for complexity reduction. This subsection introduces an effective approach for enhancing the performance with a given training overhead. For nuclear norm minimization-based LRMS with the linear operator $\mathcal{A}(\cdot)$ following the restricted isometry property (RIP) or rank-one Gaussian measurements, the training complexity (number of observations required for reliable recovery) for an $n \times p$ matrix with rank L is given as $N_s \propto Lq$, with $q = n + p$ [130]. Therefore, given a fixed rank L, a lower training complexity is required when q is smaller. As an example, consider the channel matrix model in this chapter for

$N_R = 512$, $N_T = 8$ and $N_C = 4$. Then the dimension of $\mathbf{H}$ is 512×32, giving $q = 544$ with $n = N_R = 512$ and $p = N_C N_T = 32$. If $\mathbf{H}$ can be reshaped into a matrix $\overline{\mathbf{H}}$ of dimension 128×128, then q can be reduced to 256 with $n = N_R/4$ and $p = 4N_C N_T$. If $\overline{\mathbf{H}}$ further has the same rank as $\mathbf{H}$, then the training complexity can be reduced by such reshaping. Therefore, for applications with $\mathbf{H}$ being fat or tall matrix, rank-preserving matrix reshaping may be applied to reduce the training complexity, or equivalently, improve the performance of LRMS with a given training overhead.

To apply the rank-preserving matrix reshaping to the channel matrix, the shift-invariance property of uniform arrays may be utilized, which states that two uniform arrays that are shifted versions of each other have the same signal subspace. This property is commonly exploited in ESPRIT-like spectrum estimation techniques [99]. Here, it provides a useful tool for performance enhancement via matrix reshaping.

Consider first the case where the original channel matrix $\mathbf{H}$ in (5.8) is a tall matrix with $N_R \gg N_C N_T$. Assume that N_R has positive integer factors K_R and M_R, i.e., $N_R = K_R M_R$. Then we can reshape $\mathbf{H}$ into

$$\overline{\mathbf{H}} = \left[\overline{\mathbf{H}}_0, \overline{\mathbf{H}}_1, \ldots, \overline{\mathbf{H}}_{N_C-1}\right] \in \mathbb{C}^{M_R \times N_T N_C K_R}, \tag{5.40}$$

where $\overline{\mathbf{H}}_d \in \mathbb{C}^{M_R \times N_T K_R}$ is obtained by reshaping $\mathbf{H}_d$ as

$$\overline{\mathbf{H}}_d = \left[\mathrm{vec}^{-1}(\mathbf{H}_{d,:,1}), \mathrm{vec}^{-1}(\mathbf{H}_{d,:,2}), \ldots, \mathrm{vec}^{-1}(\mathbf{H}_{d,:,N_T})\right], \tag{5.41}$$

where $\mathrm{vec}^{-1}(\mathbf{H}_{d,:,n}) \in \mathbb{C}^{M_R \times K_R}$ denotes the matricization of the n-th column of $\mathbf{H}_d$. Note that there can be multiple choices for reshaping but it is preferable to adopt the option that results in a matrix close to square, as can be seen from the earlier analysis.

To understand how the shift-invariance property of the ULAs can be leveraged

to preserve the rank of $\mathbf{H}_d$, let us rewrite the steering vector in (5.2) as

$$\begin{aligned}
\mathbf{a}_R(\theta_{R,l}) &= \mathbf{a}_{R,K_R}(\theta_{R,l}) \otimes \mathbf{a}_{R,M_R}(\theta_{R,l}) \\
\mathbf{a}_{R,K_R}(\theta_{R,l}) &\triangleq \frac{1}{\sqrt{K_R}} \left[1, \mathrm{e}^{jM_R \frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}, \cdots, \mathrm{e}^{jM_R(K_R-1)\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}\right]^T \\
\mathbf{a}_{R,M_R}(\theta_{R,l}) &\triangleq \frac{1}{\sqrt{M_R}} \left[1, \mathrm{e}^{j\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}, \cdots, \mathrm{e}^{j(M_R-1)\frac{2\pi}{\lambda_c}\beta\cos(\theta_{R,l})}\right]^T.
\end{aligned} \tag{5.42}$$

It can be easily demonstrated that

$$\overline{\mathbf{H}}_d = \sqrt{\frac{N_T N_R}{L}} \sum_{l=0}^{L-1} \alpha_l p(dT_s - \tau_l) \mathbf{a}_{R,M_R}(\theta_{R,l}) \left(\mathbf{a}_T(\phi_{T,l}) \otimes \mathbf{a}^*_{R,K_R}(\theta_{R,l})\right)^H, \tag{5.43}$$

where $\mathbf{a}_{R,M_R}(\theta_{R,l}) \left(\mathbf{a}_T(\phi_{T,l}) \otimes \mathbf{a}^*_{R,K_R}(\theta_{R,l})\right)^H$ is a reshaped version of $\mathbf{a}_R(\theta_{R,l})\mathbf{a}_T^H(\phi_{T,l})$ and both form rank-1 matrices. This demonstrates that the rank is preserved via reshaping.

When $N_R \ll N_C N_T$, similar rank-preserving reshaping may be applied. Consider a case with $N_T = K_T M_T$. The channel matrix can be reshaped into

$$\overline{\mathbf{H}} = \left[\overline{\mathbf{H}}_0, \overline{\mathbf{H}}_1, \ldots, \overline{\mathbf{H}}_{N_C-1}\right] \in \mathbb{C}^{N_R K_T \times \frac{N_T N_C}{K_T}} \tag{5.44}$$

where K_T is the reshaping factor and $\overline{\mathbf{H}}_d \in \mathbb{C}^{N_R K_T \times \frac{N_T}{K_T}}$ is written as

$$\overline{\mathbf{H}}_d = \left[\mathrm{vec}(\mathbf{H}_{d,:,1:K_T}), \mathrm{vec}(\mathbf{H}_{d,:,K_T+1:2K_T}), \ldots, \mathrm{vec}(\mathbf{H}_{d,:,(M_T-1)K_T+1:M_T K_T})\right], \tag{5.45}$$

with $\mathrm{vec}(\mathbf{H}_{d,:,(n-1)K_T+1:nK_T}) \in \mathbb{C}^{N_R K_T \times 1}$ obtained by stacking columns $(n-1)K_T+1$ to nK_T of $\mathbf{H}_d$. As the transmitter-side array response can be written as

$$\begin{aligned}
\mathbf{a}_T(\phi_{T,l}) &= \mathbf{a}_{T,M_T}(\phi_{T,l}) \otimes \mathbf{a}_{T,K_T}(\phi_{T,l}) \\
\mathbf{a}_{T,M_T}(\phi_{T,l}) &\triangleq \frac{1}{\sqrt{M_T}} \left[1, \mathrm{e}^{jK_T \frac{2\pi}{\lambda_c}\beta\cos(\phi_{T,l})}, \cdots, \mathrm{e}^{jK_T(M_T-1)\frac{2\pi}{\lambda_c}\beta\cos(\phi_{T,l})}\right]^T \\
\mathbf{a}_{T,K_T}(\phi_{T,l}) &\triangleq \frac{1}{\sqrt{K_T}} \left[1, \mathrm{e}^{j\frac{2\pi}{\lambda_c}\beta\cos(\phi_{T,l})}, \cdots, \mathrm{e}^{j(K_T-1)\frac{2\pi}{\lambda_c}\beta\cos(\phi_{T,l})}\right]^T.
\end{aligned} \tag{5.46}$$

Then

$$\overline{\mathbf{H}}_d = \sqrt{\frac{N_T N_R}{L}} \sum_{l=0}^{L-1} \alpha_l p(dT_s - \tau_l) \left(\mathbf{a}^*_{T,K_T}(\phi_{T,l}) \otimes \mathbf{a}_R(\theta_{R,l})\right) \mathbf{a}^H_{T,M_T}(\phi_{T,l}) \tag{5.47}$$

where $\left(\mathbf{a}^*_{T,K_T}(\phi_{T,l}) \otimes \mathbf{a}_R(\theta_{R,l})\right) \mathbf{a}^H_{T,M_T}(\phi_{T,l})$ is the reshaped version of $\mathbf{a}_R(\theta_{R,l})\mathbf{a}^H_T(\phi_{T,l})$ and both form rank-1 matrices. Clearly, such reshaping also preserves the rank.

The shift-invariance property can be utilized to obtain different versions of $\overline{\mathbf{H}}_d$. However, the options of $\mathbf{H}_d$ in (5.43) and (5.47) ensure that $\text{vec}(\mathbf{H}) = \text{vec}(\overline{\mathbf{H}})$. Therefore, the sensing operator $\mathcal{A}(\cdot)$ and received signal $\mathbf{Y}$ in (5.10) are not affected by reshaping. However, with the reshaped target matrix $\overline{\mathbf{H}}$ to be estimated, the problem in (5.20) is reformulated by obtaining the low-rank estimate of $\overline{\mathbf{H}} \triangleq \overline{\mathbf{U}}\overline{\mathbf{V}}^H$ as

$$\min_{\overline{\mathbf{U}},\overline{\mathbf{V}}} \frac{1}{2}\|\mathbf{K}\text{vec}(\overline{\mathbf{U}}\overline{\mathbf{V}}^H) - \text{vec}(\mathbf{Y})\|_F^2 + \frac{1}{2}\mu(\|\overline{\mathbf{U}}\|_F^2 + \|\overline{\mathbf{V}}\|_F^2). \tag{5.48}$$

Note that the dimensions of the factor matrices $\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$ depend on the adopted reshaping option, which differ from those in (5.20). After obtaining the low-rank estimate of $\widehat{\overline{\mathbf{H}}}$ using the GCG-ALTMIN algorithm similar to Section 5.3.1, it can be reshaped back to the original dimension, giving $\widehat{\mathbf{H}}$.

Note that in the above we have considered the reshaping for the LRMS estimation of $\mathbf{H}$, for which the shift-invariance property can apply. It is, however, not directly applicable to the transformed matrix $\mathbf{C}$ in the LRMC approach in Section 5.3.3. However, when switches are available, such as considered in [50], the LRMC can work directly on $\mathbf{H}$ and reshaping can be applied straightforwardly to enhance performance.

5.3.5 Performance Enhancement via Spectrum Denoising (SD)

The proposed LRMS estimator (including the special case of LRMC) exploits the low-rank nature of the channel and estimates the channel matrix from compressed observations. This reduces the training overhead in hybrid MIMO systems. The solution can be used alone as a robust wideband mmWave channel estimator which does not rely on assumptions other than low-rankness. On the other hand, LRMS does not exploit the array response knowledge. We now extend the SD approach in [108, 131] to further enhance the performance by exploiting such knowledge. With SD, the AoAs and AoDs shared by all delay taps are first estimated from $\widehat{\mathbf{H}}$. Then the path gains for each tap are estimated using the angle information and received signal.

5.3.5.1 Estimation of Path Angles

Based on the low-rank estimate of the channel matrix $\mathbf{H}$, the sparsity of the channel in the angular domain and the array response can be leveraged to estimate the AoAs $\boldsymbol{\theta}_R = \{\theta_{R,1}, \theta_{R,2}, \ldots, \theta_{R,L}\}$ and AoDs $\boldsymbol{\phi}_T = \{\phi_{T,1}, \phi_{T,2}, \ldots, \phi_{T,L}\}$ of the propagation paths. Let us first consider the AoA estimation from the row subspace of $\widehat{\mathbf{H}}$. Note that (5.3) can be rewritten in terms of transmitter/receiver array response matrices and path gains for different channel taps as

$$\begin{aligned}\mathbf{H} =& \left[\mathbf{A}_R(\boldsymbol{\theta}_R)\mathbf{G}_0\mathbf{A}_T^H(\boldsymbol{\phi}_T), \ldots, \mathbf{A}_R(\boldsymbol{\theta}_R)\mathbf{G}_{N_C-1}\mathbf{A}_T^H(\boldsymbol{\phi}_T)\right] \\ =& \mathbf{A}_R(\boldsymbol{\theta}_R)\left[\mathbf{G}_0\mathbf{A}_T^H(\boldsymbol{\phi}_T), \ldots, \mathbf{G}_{N_C-1}\mathbf{A}_T^H(\boldsymbol{\phi}_T)\right] \\ =& \mathbf{A}_R(\boldsymbol{\theta}_R)\boldsymbol{\Lambda}_R\end{aligned} \tag{5.49}$$

with $\boldsymbol{\Lambda}_R = \left[\mathbf{G}_0\mathbf{A}_T^H(\boldsymbol{\phi}_T), \ldots, \mathbf{G}_{N_C-1}\mathbf{A}_T^H(\boldsymbol{\phi}_T)\right] \in \mathbb{C}^{L\times N_C N_T}$ acts as the "source" for generating "snapshots" of the receiver ULA given by the columns of $\mathbf{H}$. Since the angles are shared by all the taps, the number of "snapshots" for estimating the

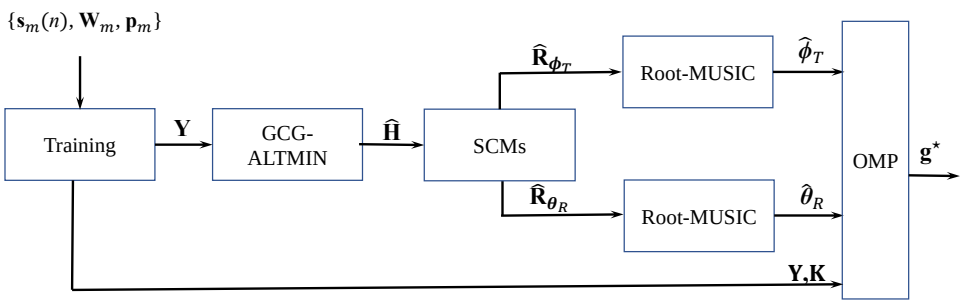


Figure 5.2: Flowchart of the proposed LRMS-SD-based scheme.

AoAs can be large, which can contribute to the reliable estimation of the angles using subspace-based super-resolution methods.

The low-rank channel matrix estimate $\widehat{\mathbf{H}}$ can be written as

$$\widehat{\mathbf{H}} = \mathbf{H} + \mathbf{E}, \tag{5.50}$$

where $\mathbf{E}$ represents the estimation error of the LRMS estimator. The SCM for the AoA estimation is written as

$$\widehat{\mathbf{R}}_{\boldsymbol{\theta}_R} = \frac{1}{N_C N_T} \widehat{\mathbf{H}} \widehat{\mathbf{H}}^H. \tag{5.51}$$

The signal and noise subspaces are estimated from the SCM in (5.51) and the AoAs can be estimated by applying root-MUSIC [93–95], exploiting its good trade-off between estimation accuracy and computational complexity. The computation complexity is kept low by utilizing polynomial rooting rather than spectrum peak search.

Similarly, to estimate the AoDs, we first note that

$$\begin{aligned}
\check{\mathbf{H}} \triangleq & \left[\mathbf{H}_0^H, \cdots, \mathbf{H}_{N_C-1}^H\right] \\
= & \left[\mathbf{A}_T(\boldsymbol{\phi}_T)\mathbf{G}_0\mathbf{A}_R^H(\boldsymbol{\theta}_R), \cdots, \mathbf{A}_T(\boldsymbol{\phi}_T)\mathbf{G}_{N_C-1}\mathbf{A}_R^H(\boldsymbol{\theta}_R)\right] \\
= & \mathbf{A}_T(\boldsymbol{\phi}_T)\left[\mathbf{G}_0\mathbf{A}_R^H(\boldsymbol{\theta}_R), \cdots, \mathbf{G}_{N_C-1}\mathbf{A}_R^H(\boldsymbol{\theta}_R)\right] \\
= & \mathbf{A}_T(\boldsymbol{\phi}_T)\boldsymbol{\Lambda}_T
\end{aligned} \tag{5.52}$$

with $\mathbf{\Lambda}_T = \left[\mathbf{G}_0\mathbf{A}_R^H(\boldsymbol{\theta}_R), \cdots, \mathbf{G}_{N_C-1}\mathbf{A}_R^H(\boldsymbol{\theta}_R)\right] \in \mathbb{C}^{L\times N_C N_R}$. The SCM for AoD estimation can be obtained by

$$\widehat{\mathbf{R}}_{\phi_T} = \frac{1}{N_C N_R}\widehat{\check{\mathbf{H}}}\widehat{\check{\mathbf{H}}}^H \tag{5.53}$$

where $\widehat{\check{\mathbf{H}}}$ denotes the estimate of $\check{\mathbf{H}}$ obtained from $\widehat{\mathbf{H}}$ in a way similar to (5.52). Then the root-MUSIC algorithm is used to estimate the AoDs.

In summary, this stage estimates the AoAs and AoDs separately using the root-MUSIC algorithm from $\widehat{\mathbf{H}}$. As compared to the narrowband case of [108], we can access more "snapshots" for obtaining the SCMs for angle estimation, thanks to that the channel taps share the AoAs and AoDs. The abundance of samples also eliminates the necessity of the treatments, such as Toeplitz rectification or spatial smoothing, for low sample supports, though they can also be implemented here with low extra costs.

5.3.5.2 Estimation of Path Gains

The obtained AoA estimates $\widehat{\boldsymbol{\theta}}_R$ and AoD estimates $\widehat{\boldsymbol{\phi}}_T$ are now used to denoise the channel estimate. The key task is to obtain the path gains corresponding to each delay tap using the observations $\mathbf{Y}$ and estimated AoAs and AoDs.

The channel matrix can be written as $\mathbf{H} = \mathbf{A}_R(\boldsymbol{\theta}_R)\mathbf{G}(\mathbf{I}_{N_C} \otimes \mathbf{A}_T^H(\boldsymbol{\phi}_T))$ where $\mathbf{G} = [\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{N_C-1}] \in \mathbb{C}^{N_R\times N_T N_C}$ contains the beamspace gains of all the delay taps. Then after vectorization,

$$\mathbf{h} \triangleq \mathrm{vec}(\mathbf{H}) = ((\mathbf{I}_{N_C} \otimes \mathbf{A}_T^*(\boldsymbol{\phi})) \otimes \mathbf{A}_R(\boldsymbol{\theta}_R))\,\mathrm{vec}(\mathbf{G}) \triangleq \mathbf{\Psi}\mathbf{g}, \tag{5.54}$$

where $\mathbf{\Psi} = (\mathbf{I}_{N_C} \otimes \mathbf{A}_T^*(\boldsymbol{\phi}_T)) \otimes \mathbf{A}_R(\boldsymbol{\theta}_R) \in \mathbb{C}^{N_T N_R N_C \times L^2 N_C}$ and $\mathbf{g} \in \mathbb{C}^{L^2 N_C \times 1}$. When the AoAs $\{\theta_{R,l}\}$ and AoDs $\{\phi_{T,l}\}$ are known, within $\mathbf{G}$ the channel gain matrix of the d-th tap, $\mathbf{G}_d, \forall d$, can be estimated by using the standard LS method. However, since the correspondence between $\widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\phi}}_T$ is unknown, the LS estimate of $\mathbf{G}_d$ may contain more than $\widehat{L}$ nonzero entries in the noisy cases, where $\widehat{L}$ is the rank

of the low-rank channel estimate $\widehat{\mathbf{H}}$. This can also result in significant estimation errors when the PNR is low.

Alternatively, we estimate the channel gain vector by solving the following sparse recovery problem:

$$\mathbf{g}^\star = \arg\min_{\mathbf{g}} ||\text{vec}(\mathbf{Y}) - \mathbf{K}\widehat{\mathbf{\Psi}}\mathbf{g}||^2, \qquad ||\mathbf{g}||_0 = \widehat{L}N_C, \tag{5.55}$$

where $\widehat{\mathbf{\Psi}} = (\mathbf{I}_{N_C} \otimes \mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T)) \otimes \mathbf{A}_R(\widehat{\boldsymbol{\theta}}_R) \in \mathbb{C}^{N_T N_R \times \widehat{L}^2}$. We suggest applying OMP [132] to solve the above problem for simplicity but other sparse recovery methods can also be used. The sparsity pattern of $\mathbf{g}_d^\star$ within $\mathbf{g}^\star$ reveals the correspondence of the AoAs, AoDs, and path gains for the d-th delay tap.

If the impulsive training structure is used, the channel gains of each delay tap can also be estimated separately. The received signal corresponding to the d-th delay tap can be extracted as $\mathbf{Y}_d \triangleq \mathbf{Y}_{:,d:N_C:N_C M} \in \mathbb{C}^{Q_R \times M}$ when impulsive training is applied. Recall the channel for the d-th delay tap in (5.3). After vectorization, the channel for the d-th delay tap can be written as

$$\mathbf{h}_d \triangleq \text{vec}(\mathbf{H}_d) = (\mathbf{A}_T^*(\boldsymbol{\phi}_T) \otimes \mathbf{A}_R(\boldsymbol{\theta}_R))\,\text{vec}(\mathbf{G}_d) \triangleq \mathbf{\Psi}\mathbf{g}_d, \tag{5.56}$$

where $\mathbf{\Psi} = \mathbf{A}_T^*(\boldsymbol{\phi}_T) \otimes \mathbf{A}_R(\boldsymbol{\theta}_R) \in \mathbb{C}^{N_T N_R \times L^2}$ and $\mathbf{g}_d \in \mathbb{C}^{L^2 \times 1}$. We estimate the channel gain vector by solving the following sparse recovery problem:

$$\mathbf{g}_d^\star = \arg\min_{\mathbf{g}_d} ||\text{vec}(\mathbf{Y}_d) - \mathbf{K}_d\widehat{\mathbf{\Psi}}\mathbf{g}_d||^2, \qquad ||\mathbf{g}_d||_0 = \widehat{L}, \tag{5.57}$$

where $\widehat{\mathbf{\Psi}} = \mathbf{A}_T^*(\widehat{\boldsymbol{\phi}}_T) \otimes \mathbf{A}_R(\widehat{\boldsymbol{\theta}}_R) \in \mathbb{C}^{N_T N_R \times \widehat{L}^2}$. In the above, the sensing matrix for

the d-th delay tap is given as

$$\mathbf{K}_d = \begin{bmatrix} \mathbf{p}_1^T \otimes \mathbf{W}_1^H \\ \mathbf{p}_2^T \otimes \mathbf{W}_2^H \\ \vdots \\ \mathbf{p}_M^T \otimes \mathbf{W}_M^H \end{bmatrix} \in \mathbb{C}^{MQ_R \times N_T N_R}, \forall d. \tag{5.58}$$

Finally, the channel matrix is reconstructed as

$$\widehat{\mathbf{H}}_d = \mathrm{vec}^{-1}(\widehat{\mathbf{\Psi}}\mathbf{g}_d^\star), \forall d. \tag{5.59}$$

The flowchart of the proposed LRMS-SD-based estimator is depicted in Fig. 5.2. The SD treatments for the LRMC-based training and estimation are similar and are skipped for brevity. The solutions presented in this chapter can be applied to narrowband channels in Chapter 3 by assuming $N_C = 1$.

5.4 Performance of the Proposed Estimators

In this section, we evaluate the performance of the proposed channel estimators by comparing their complexity and estimation accuracy with several representative estimators.

5.4.1 Complexity Comparison With Alternative Estimators

The proposed estimators are compared with the LS estimator, an OMP-based [132] CS estimator exploiting only the sparsity, and an ADMM-based estimator exploiting both the sparsity and low-rankness of the channel following [92].

For the LS estimator, the received signal is modeled as $\check{\mathbf{Y}} = \mathbf{X}_R^H \mathbf{H} \overline{\mathbf{X}}_T + \mathbf{X}_R^H \mathbf{N}$, which can be viewed as a fully sampled version of the received signal used for the LRMC estimator. The channel estimate is obtained using $\widehat{\mathbf{H}} = (\mathbf{X}_R^H)^\dagger \check{\mathbf{Y}} \overline{\mathbf{X}}_T^\dagger$ where $(\cdot)^\dagger$ denotes pseudo-inverse. This LS estimator has complexity $\mathcal{O}(N_R^2 N_T N_C +$

$N_R N_T^2 N_C^2$) if $\mathbf{X}_R$ and $\overline{\mathbf{X}}_T$ are square and their inverses are pre-calculated. Exploiting neither the sparsity nor the low-rankness of the channel, the LS estimator generally requires high training overhead, but exhibits lower computational complexity.

The OMP-based estimator is based on the following sparse model of the channel matrix: $\mathbf{H} = \mathbf{D}_R \tilde{\mathbf{Z}} (\mathbf{I}_{N_C} \otimes \mathbf{D}_T^H)$, where $\mathbf{D}_R \in \mathbb{C}^{N_R \times G_R}$ and $\mathbf{D}_T \in \mathbb{C}^{N_T \times G_T}$ are DFT matrices and G_R and G_T are the grid sizes for the AoA and AoD, respectively. Here $\tilde{\mathbf{Z}} = [\tilde{\mathbf{Z}}_0, \tilde{\mathbf{Z}}_1, \ldots, \tilde{\mathbf{Z}}_{N_C-1}]$ is assumed sparse with $\tilde{\mathbf{Z}}_d$ denoting the virtual channel gains for the d-th tap. The training signal can be generated the same as the proposed method. Let $\mathbf{y} = \text{vec}(\mathbf{Y}) = \tilde{\mathbf{\Psi}}\mathbf{z} + \mathbf{n}$, with $\tilde{\mathbf{\Psi}} = \mathbf{K}\left((\mathbf{I} \otimes \mathbf{D}_T^H)^T \otimes \mathbf{D}_R\right)$ denoting the measurement matrix, $\mathbf{z} = \text{vec}(\tilde{\mathbf{Z}})$ and $\mathbf{n}$ representing the noise vector. The OMP-based solution solves heuristically the following problem:

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{s.t.} \|\mathbf{y} - \tilde{\mathbf{\Psi}}\mathbf{z}\|_F^2 \leq \rho_o \tag{5.60}$$

where $\|\cdot\|_0$ denotes the number of nonzero entries and ρ_o a threshold. The computational complexity with I_O iterations is $\mathcal{O}(I_O(G_R N_T N_C (Q_R M/N_T + G_T) + I_O N_R N_C N_T + I_O N_C Q_R M))$. The dictionary dependence makes the computational complexity of this OMP-based solution high especially for large systems. Also in order to cope with the power leakage of on-grid CS, a relatively large number of iterations is often required for the OMP to achieve good performance.

The ADMM estimator is modified from [92] which originally targets different hybrid transceivers. It aims to jointly exploit the low-rankness of $\mathbf{H}$ and the sparsity of $\widetilde{\mathbf{Z}} \in \mathbb{C}^{G_R \times N_C G_T}$. The training scheme is designed such that $\widetilde{\mathbf{Y}} = \mathbf{\Omega} \circ \mathbf{Y}$ is observed, where $\mathbf{Y} = \mathbf{W}^H \mathbf{H} \mathbf{\Xi} + \mathbf{W}^H \mathbf{N} \in \mathbb{C}^{N_R \times N_C M}$, $\mathbf{\Omega} \in \{0,1\}^{N_R \times N_C M}$ denotes the sampling matrix with Q_R ones and $N_R - Q_R$ zeros in each column, $\circ$ the Hadamard product, and $\mathbf{N}$ is the noise matrix. This is implemented by employing a hybrid combiner $\mathbf{W} = \mathbf{X}_R$ and selecting Q_R columns of $\mathbf{X}_R$ at each time instance and generating $\mathbf{\Xi}$ according to (5.9) using random 4-QAM symbols and random precoders. The low-rankness of the noiseless version of $\mathbf{Y}$ (arising from that of $\mathbf{H}$) and the sparsity of $\tilde{\mathbf{Z}}$

are then jointly exploited by the iterative ADMM algorithm. The estimate can be further refined by exploiting *a priori* angle information as suggested in [92], which can be obtained by applying the ADMM estimator for another time. This is referred to as "ADMM with angle". The complexity of the ADMM estimator is dominated by the SVD of a $N_R \times N_C M$ matrix in line 2 of [92, Alogrithm 1] and the pseudo-inverse of a $MN_CN_T \times N_CN_RN_T$ matrix in line 5 of [92, Algorithm 1]. Utilizing a gradient descent (GD) algorithm and incomplete SVD the overall complexity is $\mathcal{O}(I_A(N_R^3N_C^3N_T^2M + N_R^2N_C^2N_TM + N_R^2N_C^2N_T^2 + N_RN_CM))$. The complexity may be reduced by exploiting further iterative algorithms but it is generally high in large systems as the number of iterations I_A required tends to be high.

By contrast, the proposed estimators can exploit the low rankness and sparsity at separate stages. The complexity of the LRMS-based approach using the GCG-ALTMIN algorithm is dominated by the computation of (5.21) and (5.23). By utilizing the PCG algorithm and the factored forms, their complexity is kept low. Additionally, similar factored forms can be used to reduce the complexity of calculating (5.18) and (5.19). Then the complexity of the GCG-ALTMIN-based LRMS is $\mathcal{O}(I_M(I_P(M\widehat{L}^2(Q_RN_R + N_TN_C + Q_RN_C))) + N_RN_C\widehat{L}(MN_T + MQ_R + N_T))$ with I_P denoting the number of the PCG iterations and I_M denoting number of the ALTMIN iterations. Both I_M and I_P are low due to the fast convergence of the ALTMIN algorithm and preconditioning, respectively. The complexity for obtaining the LRMC-based estimate using the GCG-ALTMIN algorithm is significantly lower since $\mathbf{K}_\Omega$ is a sparse matrix, which is $\mathcal{O}(I_M\widehat{L}^2((N_CN_T+N_R)(\widehat{L}^2+\widehat{L}+1)+\widehat{L}Q_RMN_C)+\widehat{L}N_RN_TN_C)$ [50]. This complexity assumes direct computation of (5.21) and (5.23), and it may be reduced further by using PCG.

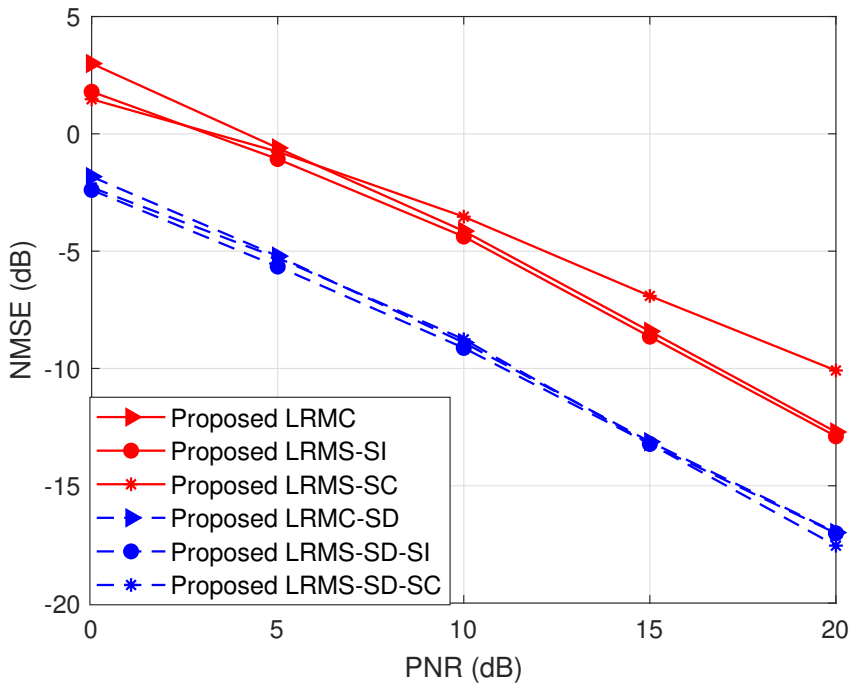
The matrix reshaping strategy in Section 5.3.4 does not increase the computational complexity of the LRMS estimators as the ALTMIN steps for solving (5.48) require fewer computations than those for (5.20) due to the changed matrix dimensions. The SD method in Section 5.3.5 requires additional computational efforts. The complexity of AoA and AoD estimation is dominated by the root-MUSIC algo-

rithm and given by $\mathcal{O}(N_R^3 + N_T^3)$. For the d-th delay tap, the OMP-based channel path gain estimation has a complexity of $\mathcal{O}(\widehat{L}(\widehat{L}N_T(Q_R M/N_T + \widehat{L}) + \widehat{L}N_T N_R + \widehat{L}MN_C Q_R))$, which is low as $\widehat{L}$ is typically small for mmWave and THz bands.

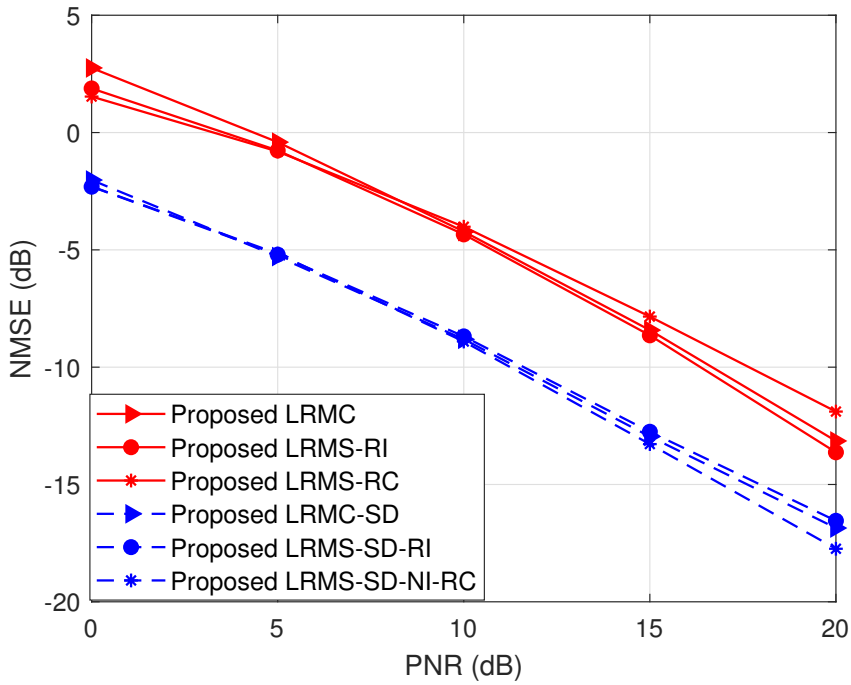
5.4.2 Simulation Results

This subsection presents the simulation results. We employ (5.1) to model the small-scale fading of the mmWave and THz channels, following [36, 38, 52, 87, 117–120]. The pilot-to-noise-ratio is defined as $\text{PNR} \triangleq 10\log_{10}(\frac{P}{\sigma_n^2})$, with $P = \|\mathbf{p}_m\mathbf{s}_m\|_F^2 = 1$ denoting the total transmitted power during a subframe. The path gains $\{\alpha_l\}$ in (5.1) follow $\mathcal{CN}(0,1)$, the delays $\{\tau_l\}$ are uniformly distributed in $[0,(N_C-1)T_s]$, while all the AoAs and AoDs are uniformly distributed in $[0,2\pi)$. The roll-off factor for the raised-cosine filter is assumed to be 0.8 [36]. For the GCG-ALTMIN algorithm, the regularization parameter $\mu = \sigma_n^2$ and the stopping thresholds for the ALTMIN algorithm and the GCG algorithm are set as $\epsilon_a = 0.1$ and $\epsilon = 0.01$, respectively, following [50]. To compensate for the power leakage for the OMP estimator, the sparsity level of $\mathbf{Z}$ is assumed to be 100. The grid sizes for the OMP and ADMM estimators are assumed as $G_R = N_R$ and $G_T = N_T$. The normalized mean squared error (NMSE) for estimating the concatenated channel matrix is obtained as the mean of $\frac{\|\mathbf{H}-\widehat{\mathbf{H}}\|_F^2}{\|\mathbf{H}\|_F^2}$.

We first compare the influence of the training schemes on the proposed methods. We consider two types of precoders/combiners: the structured precoder/combiner as used for the LRMC estimator in Section 5.3.3, and the random precoder/combiner having phase shifts randomly selected from the set of feasible phase shifts. Note that LRMC needs to use the structured precoder/combiner and impulsive training while LRMS can accommodate different precoder/combiner and training structures. Fig.5.3a shows the performance with structured precoders/combiners. It is seen that the impulsive training (LRMS-SI) shows slightly better performance. The LRMS estimator with structured precoder/combiner and continuous symbol transmission



(a) Structured precoder and combiner (SI: structured precoder/combiner with impulsive training signal; SC: structured precoder/combiner with continuous training signal.)



(b) Random precoder and combiner (RI: random precoder/combiner with impulsive training signal; RC: random precoder/combiner with continuous training signal.)

Figure 5.3: Channel estimation performance versus PNR with different training schemes for a wideband hybrid MIMO system with $N_T = 8, N_R = 32, Q_R = 4, Q_T = 2, L = 3, N_C = 4, M = 40$.

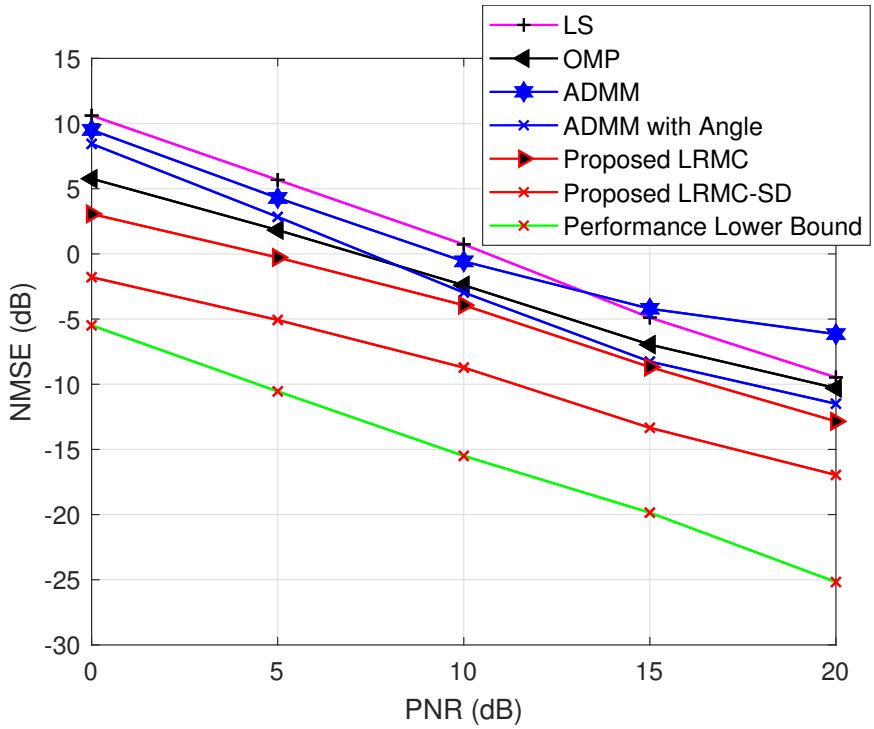


Figure 5.4: Channel estimation performance versus PNR for $N_T = 8, N_R = 32, Q_R = 4, Q_T = 2, L = 3, N_C = 4, M = 40, M_{\text{LS}} = 64$.

(LRMS-SC) suffers performance degradation of around 2 dB, but this loss diminishes when SD is used by gleaning the angle information. Fig.5.3b uses random precoding/combining, where it can be seen that the different training schemes show similar performance, with and without SD with marginal degradation in performance shown by LRMS-RC. Recall that impulsive training leads to lower computational complexity of the LRMS estimator and the LRMC estimator can further reduce the complexity. However, note also that continuous training has a lower dynamic range of the transmitted training signal. For its simplicity, impulsive training will be used for the proposed estimators in the following simulations.

Comparisons of the proposed estimators with several existing solutions is presented in Fig. 5.4. The training overhead is set the same for the different solutions except that the LS-based solution uses an extended training overhead of $M_{\text{LS}} = N_T N_R / Q_R$ subframes. The proposed LRMC-SD estimator performs the best and improves the proposed LRMC estimator by around 4 dB for the different PNR values considered. However, it is worth noting that the LRMC estimator (and also

the LRMS estimator) do not require knowledge of the array responses, in contrast to the LRMC-SD, OMP, and ADMM estimators. The ADMM estimator attempts to reconstruct the $N_R \times N_C M = 32 \times 160$ received signal matrix (see Section 5.4.1). This is in contrast to the proposed LRMC and LRMS approaches which aim to recover directly the $N_R \times N_C N_T = 32 \times 32$ channel matrix which has a smaller size but the same rank. The performance lower bound is also added where perfect AoA and AoD information is used to estimate the channel path gains using (5.57) for reconstructing channel for all delay taps. It can be seen the proposed LRMC-SD is the closest to the performance lower bound.

The comparison of channel estimation performance versus the training overhead is shown in Fig. 5.5. The proposed LRMC-SD approach performs the best even at very low training overhead. The ADMM approach with angle information performs better than the proposed LRMC and OMP approaches at low training overhead. After 40 subframes, the proposed LRMC approach outperforms the ADMM approach with angle information. The performance gains for the LRMC and LRMC-SD estimators increase further at higher training. At 64 subframes, the LRMC and LRMC-SD approaches show improvements of 2 dB and 5 dB, respectively, over the ADMM approach with angle information. The influence of the number of channel paths on the channel estimation performance is demonstrated in Fig. 5.6. It is seen that the proposed estimators are most suitable for low-rank channels and their estimation accuracy reduces when the channel rank increases. However, the proposed LRMC-SD approach provides around 1.5 dB gain for $L = 5$ over the ADMM approach exploiting the angle information.

Next, we report the effectiveness of matrix reshaping for the proposed LRMS estimator. The comparison with the ADMM solution is skipped due to its increased computational requirements for the channel dimensions considered. In Fig. 5.7, the 8×256 channel matrix $\mathbf{H}$ is reshaped into a matrix $\overline{\mathbf{H}}$ of size 32×64. The results in Fig. 5.7a suggest that the proposed LRMS estimators suffer performance loss when $\mathbf{H}$ is far from square. However, the reshaped LRMS (R-LRMS) estimator,

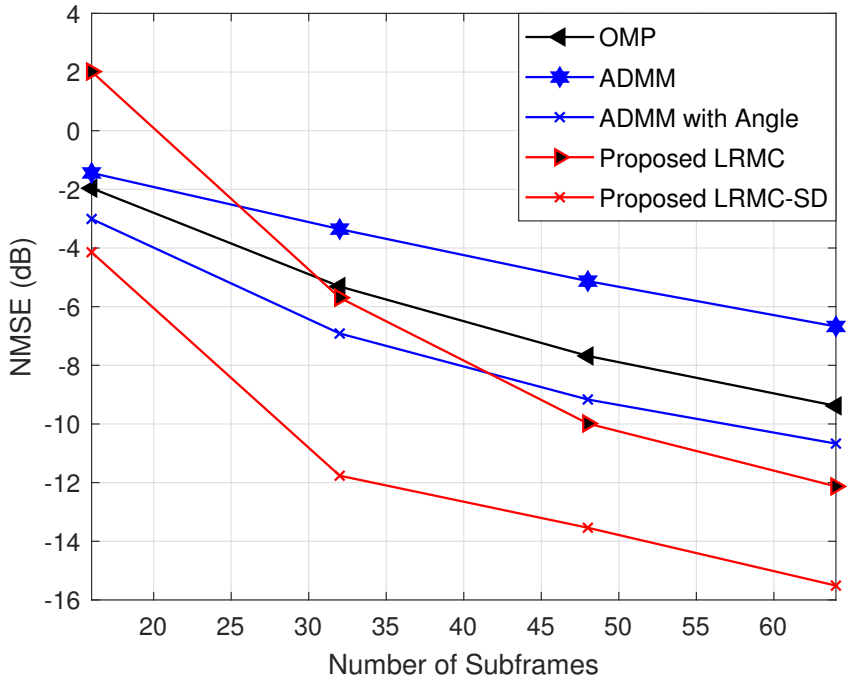


Figure 5.5: Channel estimation performance versus the number of training subframes M for a wideband hybrid MIMO system with $N_T = 8, N_R = 32, Q_T = 2, Q_R = 4, L = 3, N_C = 4, \text{PNR} = 15$ dB.

which works on $\overline{\mathbf{H}}$ instead of $\mathbf{H}$, can significantly improve the performance. The improvement is also translated into the R-LRMS-SD estimator that applies SD to refine the R-LRMS estimate. The gap between the performance of the R-LRMS-SD estimator and the LRMC-SD ranges from 1 dB to 10 dB when the PNR is increased from 0 dB to 20 dB. Fig. 5.8a shows the rank estimates obtained for the LRMC, LRMS, and R-LRMS estimators using the GCG-ALTMIN algorithm. The results suggest that reshaping improves the accuracy of the rank estimation. The condition number of the reshaped channel matrix, which can generally indicate the difficulty in solving LRMS problems [133], is also examined. The empirical cumulative distribution function (CDF) of the condition number is demonstrated in Fig. 5.8b for the original and reshaped channel matrix. It is seen that on average reshaping improves the conditioning of the matrix to be estimated, which is helpful.

In the above, we have considered the channel model (5.1). In Fig. 5.9, we further

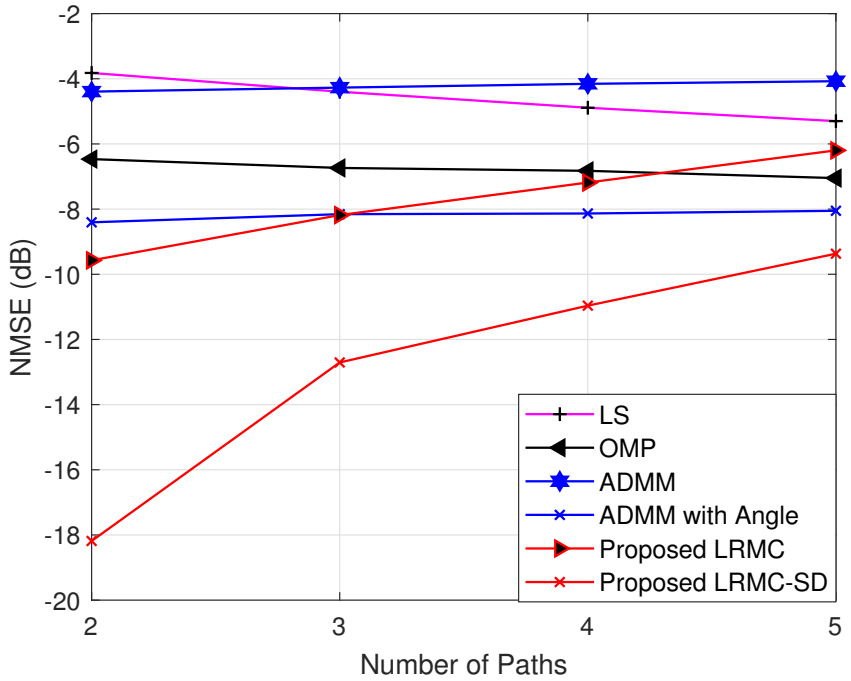


Figure 5.6: Channel estimation performance versus number of paths L for a wideband hybrid MIMO system with $N_T = 8, N_R = 32, Q_T = 2, Q_R = 4, N_C = 4, M = 40, M_{\text{LS}} = 64, \text{PNR} = 15$ dB.

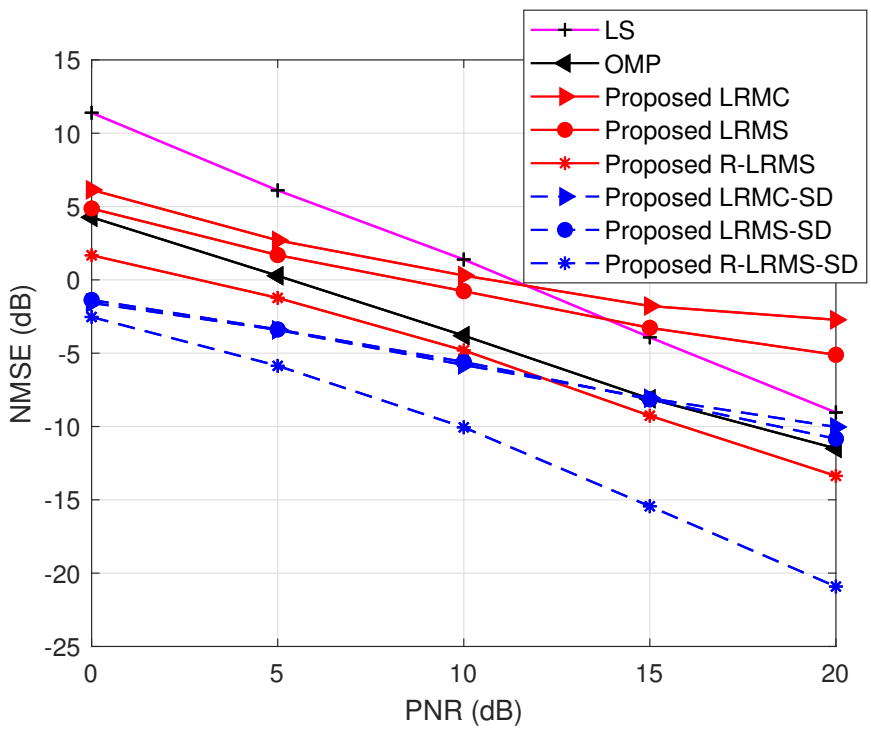
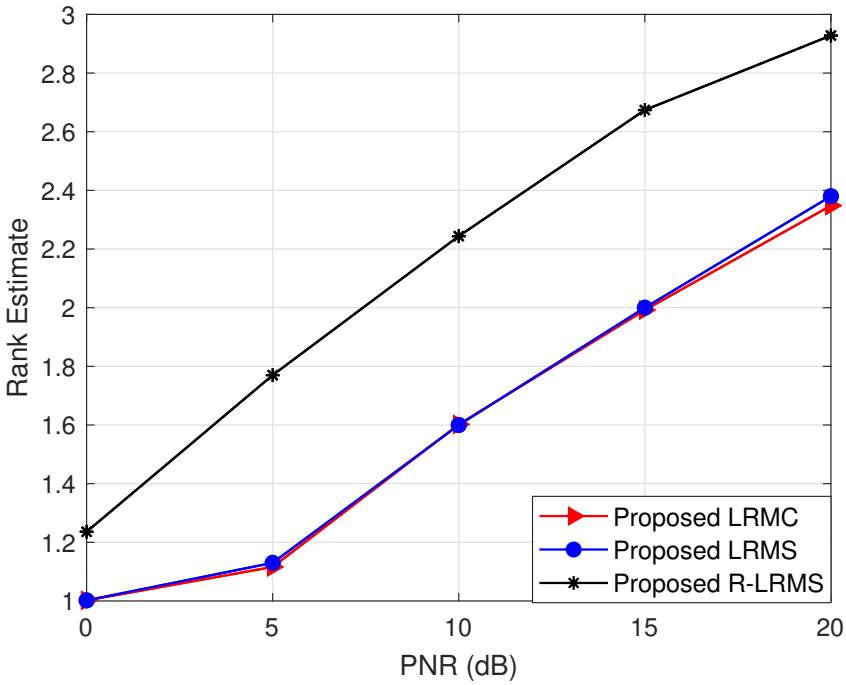
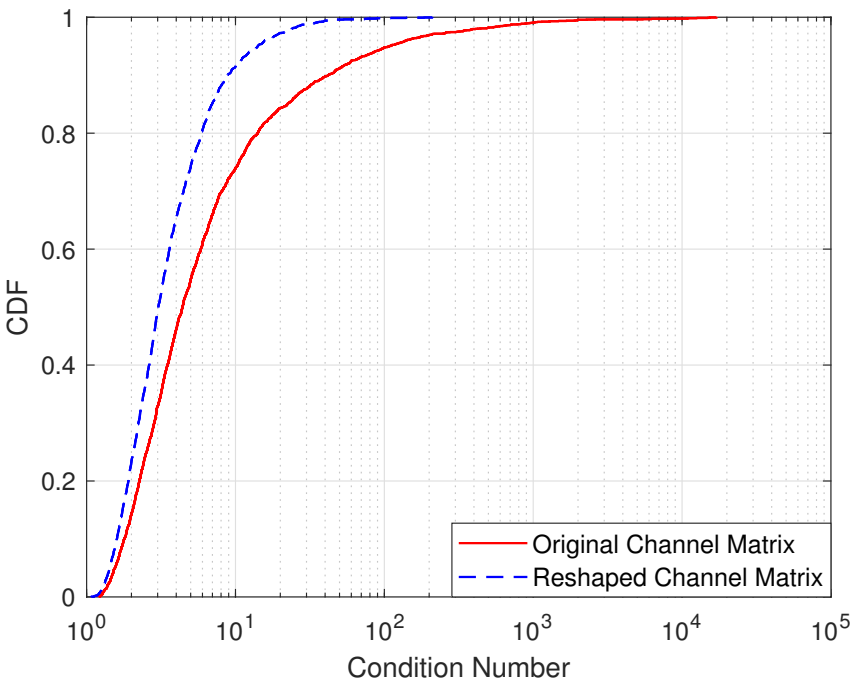


Figure 5.7: Channel estimation performance versus PNR for a wideband hybrid MIMO system with $N_T = 64, N_R = 8, N_C = 4, Q_T = 4, Q_R = 2, L = 3, M = 128, M_{\text{LS}} = 256, \mathbf{H} \in \mathbb{C}^{8\times 256}, \overline{\mathbf{H}} \in \mathbb{C}^{32\times 64}, K_T = 4, M_T = 16$.



(a) Rank estimation accuracy versus PNR



(b) Condition number comparison

Figure 5.8: The influence of matrix reshaping on the rank estimation and condition number for $N_T = 64, N_R = 8, N_C = 4, Q_T = 4, Q_R = 2, L = 3, M = 128, \mathbf{H} \in \mathbb{C}^{8\times256}, \overline{\mathbf{H}} \in \mathbb{C}^{32\times64}, K_T = 4, M_T = 16$.

consider the ray-cluster channel model [134]

$$\mathbf{H}_d = \sqrt{\frac{N_T N_R}{L_c L_r}} \sum_{c=0}^{L_c-1} \sum_{r=0}^{L_r-1} \alpha_{c,r} p(dT_s - \tau_c) \mathbf{a}_R(\theta_{r,c}) \mathbf{a}_T^H(\phi_{r,c}), \tag{5.61}$$

where the total number of paths is given as $L = L_c L_r$, L_c denotes the number of clusters, L_r denotes the number of rays and τ_c denotes the path delay for the c-th cluster. The path delays for different rays within a cluster are assumed to be the same [135]. Similarly to [50], L_c follows $\max(\text{Poisson}(1.8), 1)$, L_r is uniformly distributed in $[1, 20]$ and the other path parameters are generated following [50]. Although the total number of paths $L = L_c L_r$ can be high, the rank of $\mathbf{H}_d, \forall d$, can still be low [50] due to the high correlation of the rays within each cluster. Therefore, the proposed LRMS solutions are still effective for channels modelled by (5.61), outperforming the LS and OMP based estimators. Furthermore, compared to OMP, the LRMS-based solutions do not require knowledge of the array responses and hence they can be more robust against the phase and gain errors of the antennas [50]. When L_r is large, the angles of the rays within each cluster can be very close and the SD treatment becomes less effective due to the limitation of root-MUSIC in estimating very close path angles.

5.5 Summary

In summary, this chapter presents LRMS-based solutions for estimating the wideband channel of mmWave and THz MIMO systems equipped with hybrid precoders and combiners. The proposed estimators leverage the low rankness of the channel matrix. The computational complexity of the LRMS estimator is lower when adopting the preconditioned conjugate gradient implementations or the LRMC formulation. Such LRMS-based solutions do not rely on the knowledge of array responses and thus can be used as stand-alone, robust estimators of low-rank wideband MIMO channels, which differ from many existing estimators. Their performance can be

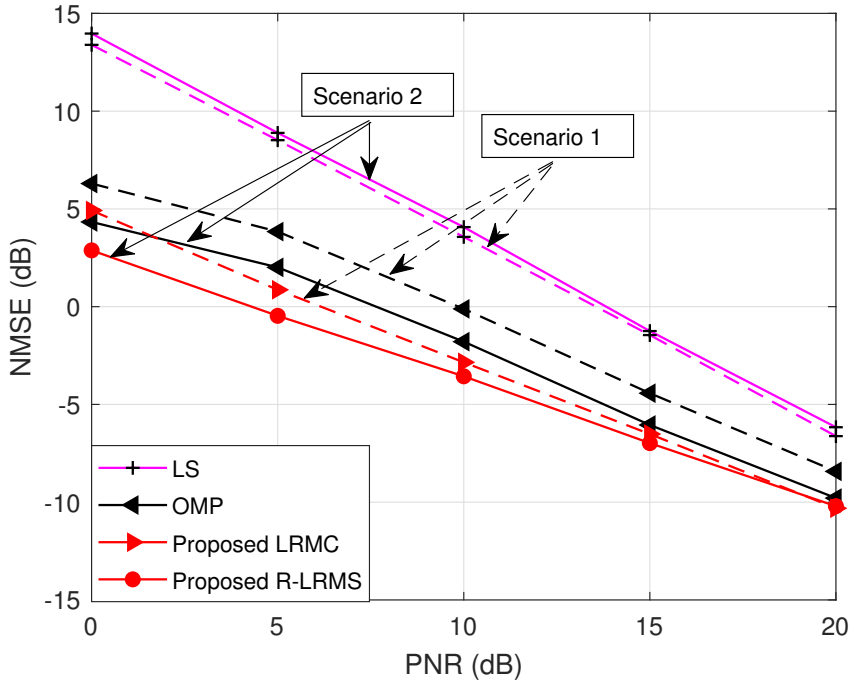


Figure 5.9: Channel estimation performance versus PNR for the ray-cluster channel model with $L = L_c L_r, L_c \sim \max(\text{Poisson}(1.8), 1), L_r \sim \mathcal{U}[1, 20]$, $N_C = 4$. Scenario 1: $N_T = 8, N_R = 32, Q_T = 2, Q_R = 4, M = 40, M_{\text{LS}} = 64$. Scenario 2: $N_T = 64, N_R = 8, Q_T = 4, Q_R = 2, M = 128, M_{\text{LS}} = 256, \mathbf{H} \in \mathbb{C}^{8\times 256}, \overline{\mathbf{H}} \in \mathbb{C}^{32\times 64}, K_T = 4, M_T = 16$.

further improved if knowledge of the array response is available. In particular, the shift-invariance property of the transmitter or receiver arrays can be exploited by the matrix reshaping approach to enhance performance. If the array responses are known, low-complexity spectrum denoising steps can be included to obtain the path directions and reconstruct the channel. The different precoder/combiner structures and training schemes that are suitable for the proposed estimators are also presented. As demonstrated by the simulation results, the proposed estimators can noticeably improve the channel estimation performance and strike an attractive balance between performance and complexity.

Chapter 6

Conclusions and Future Works

6.1 Conclusions

In this thesis, channel estimation solutions for future wireless communication systems are proposed. The proposed solutions target MIMO systems operating at mmWave and THZ bands. All the provided solutions require low training overhead and are computationally efficient. Hybrid analog-digital transceivers are considered throughout the thesis and passive IRS is also considered. The findings show that by first utilizing the low-rank nature of the channel and then exploiting the knowledge of array response significant improvement in channel estimation accuracy can be achieved. More specifically,

- In Chapter 3, a three-stage approach is presented for narrowband mmWave channels. The proposed solution is able to estimate channel parameters like AoAs, AoDs, and channel path gains in different stages with each stage benefiting from the previous stage. The simulation results show that the proposed narrowband channel estimator outperforms various existing solutions like the OMP estimator, ADMM-based estimator, and IR-SR-based estimator. Additionally, it is shown that the proposed estimator allows low training overhead and its computational effort is less than the existing solutions.

- In Chapter 4, a channel estimation solution is proposed for IRS-assisted hybrid MIMO systems. The proposed solution progressively estimates the channel parameters including transmitter AoDs, receiver AoAs, the angle difference between AoAs and AoDs at the IRS, and the composite channel gains for the Tx-IRS and IRS-Rx channels. The proposed solution consists of two stages, Stage 1 provides the estimates for transmitter AoDs and receiver AoAs,based on which Stage 2 estimates the IRS angle difference and composite path gains. The simulation results show significant improvements as compared to existing solutions like LS-estimator, ANM-based solution, and TRICE estimator. The training overhead of the proposed solution is kept low by adopting a two-stage training in which Stage 2 benefits from already estimated angles and activates only a subarray from IRS elements. The computational effort for the proposed solution is significantly reduced as compared to existing solutions. This reduction in computational complexity is achieved by estimating the channel parameters in two stages and also by adopting low-complexity solutions while estimating each channel parameter.

- In Chapter 5, channel estimation solutions for single-carrier wideband MIMO channels are presented. First, the channel estimation problem is formulated as low-rank matrix recovery and solved using LRMS. Then two complexity reduction techniques are presented, first by utilizing an iterative matrix inversion within the LRMS and then by using LRMC as a special case of LRMS. For performance enhancement, a rank-preserving matrix reshaping strategy is proposed for channel matrices far from square. Then the performance is further enhanced by applying SD on the LRMS estimate to estimate the channel parameters like AoAs, AoDs that are shared by all delay taps, and effective path gains for each delay tap. The simulation results suggest that the proposed solution with SD provides better estimation accuracy as compared to LS-estimator, OMP-based estimator, and ADMM-based solutions. The pro-

posed solution is suitable for continues as well as impulsive training making it flexible for practical scenarios.

Overall the solutions provided in Chapters 3, 4, and 5 exploit the low-rankness of the channel matrix and the knowledge of array responses. All chapters provide parametric channel estimation with super-resolution angle estimation. Having said that, the problems considered in each chapter have their own challenges.

- Chapter 3 aims to provide a low-complexity solution for the narrowband channel estimation. The low-rank property of the $N_R \times N_T$ channel matrix is exploited and the using low-rank channel estimate AoAs, AoDs, and finally the path gains are estimated to further improve the overall channel estimate.

- Chapter 4 considers the narrowband IRS-assisted systems. The training and estimation process is divided into two stages to reduce the training overhand as well as computational cost. As compared to Chapter 3 the low-rankness of the effective channel matrix $\boldsymbol{\mathcal{H}}_0 = \mathbf{G}\boldsymbol{\Omega}_0\mathbf{F} \in \mathbb{C}^{N_R \times N_T}$ is first exploited to only provide the AoDs and AoAs at the transmitter and receiver. The low-rank estimate of $\boldsymbol{\mathcal{H}}_0$ is not a standalone channel estimation solution in contrast to Chapter 3. Additionally, the IRS angles are also estimated and the collective path gains are estimated as compared to path gains of a single channel in Chapter 3.

- Finally, Chapter 5 considers wideband systems and first provides the LRMS-based solution that is computational expensive, then complexity reduction techniques are introduced. In contrast to Chapters 3 and 4, the channel model also accounts for the path delays and have different delay taps. Different from LRMC a generic framework based on the LRMS is first introduced that can also be applicable to Chapter 3. Also, rank-preserving matrix reshaping is introduced that was not considered in Chapters 3 and 4.

The trade-off between computational complexity and estimation accuracy is of the proposed solutions is better than the existing solutions. This can be seen from

the computational costs of the proposed solutions and the existing solutions. The proposed solutions are less complex than the CS-based, ADMM-based (or IRSR and ANM-based) solutions. In addition to low complexity, the proposed solutions also outperform the above mentioned existing solutions as shown in the simulation results.

6.2 Future Work

The work in this thesis can be extended to various promising directions:

- Extension to OFDM: The solutions in Chapter 3, Chapter 5, and Chapter 4 can be extended to OFDM systems with similar settings [36]. Effective LRMS schemes may be developed, followed by the spectrum denoising techniques to recover the AoAs, AoDs and delays shared by different subcarirers.

- Extension to time-varying channels: The solutions presented in Chapter 3, Chapter 4, and Chapter 5 can be extended to time-varying channels where Doppler shifts capture the time variation of the channel [81, 136, 137].

- Extension to UPA: The solutions in the thesis considered that the transmitter and receiver are equipped with ULA. The extension to the case where the transmitter and receiver are equipped with UPAs is also feasible. The problem of 2D direction finding can be decomposed into two separate 1D problems and solved using root-MUSIC as considered for the IRS angle estimation in Chapter 4. Then the pairing can be done while estimating the path gains using CS. Alternatively, 2D direction finding methods such as 2D-ESPRIT [116] can be used if the computational cost can be afforded.

- Subarray Sampling: The transmitter AoDs and receiver AoAs in this thesis are estimated by utilizing the full arrays. For the cases when N_T (N_R) is very high, the AoDs (AoDs) estimation will require high training and computational

cost. To reduce the computational cost, subarray sampling can be utilized for estimating the AoDs (AoDs).

- Array Inherent Impairments: The solutions presented in the thesis assume that arrays do not suffer from phase or gain errors. The GCG-ALTMIN algorithm has shown to be able to provide an accurate low-rank estimate in the presence of array inherent impairments [50, 138]. However, the AoDs (AoAs) estimation will suffer when arrays are not calibrated. For such cases, direction estimation tools should take into account the array inherent impairments [139, 140].

- In addition to root-MUSIC algorithm that already provides viable angle estimation solution, angle estimation based on the sparse Bayesian learning may be explored.

Bibliography

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[2] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, I. Chih-Lin, *et al.*, "Millimeter wave communications for future mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, 2017.

[3] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[4] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Communications Surveys & Tutorials*, 2022.

[5] I. FG-NET, "2030,"network 2030: A blueprint of technology, applications and market drivers towards the year 2030 and beyond,"," 2019.

[6] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap

to 6G: Ai empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[7] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2019.

[8] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[9] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[10] A. Faisal, H. Sarieddeen, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Ultramassive MIMO systems at terahertz bands: Prospects and challenges," *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 33–42, 2020.

[11] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[12] C. Han, Y. Wang, Y. Li, Y. Chen, N. A. Abbasi, T. Kürner, and A. F. Molisch, "Terahertz wireless channels: A holistic survey on measurement, modeling, and analysis," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1670–1707, 2022.

[13] A. Shafie, G. N. Yang, C. Han, J. M. Jornet, M. Juntti, and T. Kurner, "Terahertz communications for 6G and beyond wireless networks: Challenges, key advancements, and opportunities," *IEEE Network*, 2022.

[14] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[15] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[16] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE network*, vol. 33, no. 4, pp. 70–75, 2019.

[17] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[18] S. Zeng, H. Zhang, B. Di, Y. Tan, Z. Han, H. V. Poor, and L. Song, "Reconfigurable intelligent surfaces in 6G: Reflective, transmissive, or both?," *IEEE Communications Letters*, vol. 25, no. 6, pp. 2063–2067, 2021.

[19] Y.-C. Liang, R. Long, Q. Zhang, J. Chen, H. V. Cheng, and H. Guo, "Large intelligent surface/antennas (lisa): Making reflective radios smart," *Journal of Communications and Information Networks*, vol. 4, no. 2, pp. 40–50, 2019.

[20] T. Ma, Y. Xiao, X. Lei, P. Yang, X. Lei, and O. A. Dobre, "Large intelligent surface assisted wireless communications with spatial modulation and antenna selection," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2562–2574, 2020.

[21] C. Xing, N. Wang, J. Ni, Z. Fei, and J. Kuang, "MIMO beamforming designs with partial CSI under energy harvesting constraints," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 363–366, 2013.

[22] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-

scale MIMOs," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2015.

[23] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO x channels: Interference alignment, decomposition, and performance analysis," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3457–3470, 2008.

[24] O. El Ayach, R. W. Heath, S. Rajagopal, and Z. Pi, "Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," in *2013 IEEE Global Communications Conference (GLOBECOM)*, pp. 3476–3480, IEEE, 2013.

[25] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016.

[26] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[27] I. F. Akyildiz, C. Han, and S. Nie, "Combating the distance problem in the millimeter wave and terahertz frequency bands," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 102–108, 2018.

[28] B. Ji, Y. Han, S. Liu, F. Tao, G. Zhang, Z. Fu, and C. Li, "Several key technologies for 6G: challenges and opportunities," *IEEE Communications Standards Magazine*, vol. 5, no. 2, pp. 44–51, 2021.

[29] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.

[30] B. Ning, Z. Tian, Z. Chen, C. Han, J. Yuan, and S. Li, "Prospective beamforming technologies for ultra-massive MIMO in terahertz communications: A tutorial," *arXiv preprint arXiv:2107.03032*, 2021.

[31] C. Kourogiorgas, S. Sagkriotis, and A. D. Panagopoulos, "Coverage and outage capacity evaluation in 5G millimeter wave cellular systems: Impact of rain attenuation," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, IEEE, 2015.

[32] K. Haneda, J. Zhang, L. Tan, G. Liu, Y. Zheng, H. Asplund, J. Li, Y. Wang, D. Steer, C. Li, *et al.*, "5G 3gpp-like channel models for outdoor urban microcellular and macrocellular environments," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pp. 1–7, IEEE, 2016.

[33] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 8954–8958, Jun. 2018.

[34] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "Mmwave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 13–21, 2015.

[35] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[36] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, 2017.

[37] J. Sung, J. Choi, and B. L. Evans, "Wideband channel estimation for hy-

brid beamforming millimeter wave communication systems with low-resolution ADCs," *arXiv preprint arXiv:1710.10669*, 2017.

[38] R. Zhang, J. Zhang, T. Zhao, and H. Zhao, "Block sparse recovery for wideband channel estimation in hybrid mmWave MIMO systems," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2018.

[39] C. Han, A. O. Bicen, and I. F. Akyildiz, "Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2402–2412, 2014.

[40] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. De Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[41] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2019.

[42] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Communications*, vol. 21, pp. 136–143, Dec. 2014.

[43] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2370–2386, 2016.

[44] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?," in *2015*

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2909–2913, IEEE, Apr. 2015.

[45] C. Huang, L. Liu, and C. Yuen, "Asymptotically optimal estimation algorithm for the sparse signal with arbitrary distributions," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 10070–10075, Jul. 2018.

[46] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using lse and sparse message passing for mmWave MIMO systems," *IEEE Transactions on Signal Processing*, vol. 67, pp. 245–259, Nov. 2018.

[47] D. Zhang, A. Li, M. Shirvanimoghaddam, P. Cheng, Y. Li, and B. Vucetic, "Codebook-based training beam sequence design for millimeter-wave tracking systems," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 5333–5349, Aug. 2019.

[48] D. Zhang, A. Li, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Exploring aoa/aod dynamics in beam alignment of mobile millimeter wave MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 68, pp. 6172–6176, Apr. 2019.

[49] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint CSIT acquisition based on low-rank matrix completion for fdd massive MIMO systems," *IEEE Communications Letters*, vol. 19, pp. 2178–2181, Oct. 2015.

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

[51] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Transactions on Wireless Communications*, vol. 17, pp. 1123–1133, Nov. 2017.

[52] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Massive MIMO channel estimation for millimeter wave systems via matrix completion," *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1675–1679, 2018.

[53] S. Liang, X. Wang, and L. Ping, "Semi-blind detection in hybrid massive MIMO systems via low-rank matrix completion," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 5242–5254, Aug. 2019.

[54] M. Wang, F. Gao, S. Jin, and H. Lin, "An overview of enhanced massive MIMO with array signal processing techniques," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, pp. 886–901, Aug. 2019.

[55] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based on 2-D beamspace MUSIC method," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 5384–5394, Jun. 2017.

[56] J. Zhang and M. Haardt, "Channel estimation and training design for hybrid multi-carrier mmWave massive MIMO systems: The beamspace ESPRIT approach," in *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 385–389, IEEE, Aug. 2017.

[57] J. Zhang and M. Haardt, "Channel estimation for hybrid multi-carrier mmwave MIMO systems using three-dimensional unitary ESPRIT in DFT beamspace," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 1–5, IEEE, 2017.

[58] W. Ma and C. Qi, "Beamspace channel estimation for millimeter wave massive MIMO system with hybrid precoding and combining," *IEEE Transactions on Signal Processing*, vol. 66, pp. 4839–4853, Aug. 2018.

[59] W. Ma, C. Qi, and G. Y. Li, "High-resolution channel estimation for frequency-

selective mmWave massive MIMO system," *IEEE Transactions on Wireless Communications*, Feb. 2020.

[60] G. C. Alexandropoulos and E. Vlachos, "A hardware architecture for reconfigurable intelligent surfaces with minimal active elements for explicit channel estimation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 9175–9179, IEEE, 2020.

[61] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based algebraic channel estimation for hybrid IRS-assisted MIMO-OFDM," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3770–3784, 2021.

[62] S. Liu, Z. Gao, J. Zhang, M. Di Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9223–9228, 2020.

[63] M. Jian and Y. Zhao, "A modified off-grid SBL channel estimation and transmission strategy for RIS-assisted wireless communication systems," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 1848–1853, IEEE, 2020.

[64] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 210–214, 2019.

[65] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7736–7747, 2021.

[66] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and

analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6607–6620, 2020.

[67] X. Guan, Q. Wu, and R. Zhang, "Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication," *IEEE Transactions on Wireless Communications*, 2021.

[68] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 898–912, 2021.

[69] N. K. Kundu and M. R. McKay, "Channel estimation for reconfigurable intelligent surface aided miso communications: From lmmse to deep learning solutions," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 471–487, 2021.

[70] Y. Wang, H. Lu, and H. Sun, "Channel estimation in IRS-enhanced mmWave system with super-resolution network," *IEEE Communications Letters*, vol. 25, no. 8, pp. 2599–2603, 2021.

[71] M. Ye, H. Zhang, and J.-B. Wang, "Channel estimation for intelligent reflecting surface aided wireless communications using conditional gan," *IEEE Communications Letters*, 2022.

[72] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Processing Letters*, vol. 27, pp. 905–909, 2020.

[73] K. Ardah, S. Gherekhloo, A. L. de Almeida, and M. Haardt, "Trice: A channel estimation framework for RIS-aided millimeter-wave MIMO systems," *IEEE Signal Processing Letters*, vol. 28, pp. 513–517, 2021.

[74] H. Liu, X. Yuan, and Y.-J. A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2621–2636, 2020.

[75] J. He, M. Leinonen, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.

[76] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems via atomic norm minimization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5786–5797, 2021.

[77] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1141–1154, 2017.

[78] K. Dovelos, M. Matthaiou, H. Q. Ngo, and B. Bellalta, "Channel estimation and hybrid combining for wideband terahertz massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1604–1620, 2021.

[79] H. Kim, G.-T. Gil, and Y. H. Lee, "Two-step approach to time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5139–5152, 2019.

[80] S. Srivastava, C. S. K. Patro, A. K. Jagannatham, and L. Hanzo, "Sparse, group-sparse, and online bayesian learning aided channel estimation for doubly-selective mmWave hybrid MIMO OFDM systems," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 5843–5858, 2021.

[81] S. Gao, X. Cheng, and L. Yang, "Estimating doubly-selective channels for

hybrid mmWave massive MIMO systems: A doubly-sparse approach," *IEEE Transactions on Wireless Communications*, 2020.

[82] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, 2016.

[83] J. Rodríguez-Fernández, K. Venugopal, N. González-Prelcic, and R. W. Heath, "A frequency-domain approach to wideband channel estimation in millimeter wave systems," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–7, IEEE, 2017.

[84] J. Rodríguez-Fernández, N. González-Prelcic, and R. W. Heath, "A compressive sensing-maximum likelihood approach for off-grid wideband channel estimation at mmWave," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 1–5, IEEE, 2017.

[85] A. Abdallah, A. Celik, M. M. Mansour, and A. M. Eltawil, "Deep learning based frequency-selective channel estimation for hybrid mmwave MIMO systems," *IEEE Transactions on Wireless Communications*, 2021.

[86] K. Li, M. El-Hajjar, and L.-l. Yang, "Millimeter-wave based localization using a two-stage channel estimation relying on few-bit ADCs," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1736–1752, 2021.

[87] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1524–1538, 2017.

[88] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based channel estima-

tion for millimeter wave MIMO-OFDM with dual-wideband effects," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4218–4232, 2020.

[89] D. C. Araújo, A. L. De Almeida, J. P. Da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42133–42147, 2019.

[90] D. Rakhimov, J. Zhang, A. de Almeida, A. Nadeev, and M. Haardt, "Channel estimation for hybrid multi-carrier mmWave MIMO systems using 3-d unitary tensor-ESPRIT in DFT beamspace," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pp. 447–451, IEEE, 2019.

[91] J. Zhang, D. Rakhimov, and M. Haardt, "Gridless channel estimation for hybrid mmWave MIMO systems via tensor-ESPRIT algorithms in DFT beamspace," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 816–831, 2021.

[92] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Wideband MIMO channel estimation for hybrid beamforming millimeter wave systems via random spatial sampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 1136–1150, 2019.

[93] B. D. Rao and K. S. Hari, "Performance analysis of root-MUSIC," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, 1989.

[94] B. Friedlander, "The root-MUSIC algorithm for direction finding with interpolated arrays," *Signal Processing*, vol. 30, no. 1, pp. 15–29, 1993.

[95] A. Vesa, "Direction of arrival estimation using MUSIC and root–MUSIC algorithm," in *18th Telecommunications Forum, Pg*, pp. 582–585, 2010.

[96] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recov-

ery with noise," *IEEE Transactions on Information Theory*, vol. 57, pp. 4680–4688, Jul. 2011.

[97] Z. Weng and X. Wang, "Low-rank matrix completion for array signal processing," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2697–2700, IEEE, 2012.

[98] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, 2016.

[99] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[100] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1132–1143, 1990.

[101] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1081–1089, 1986.

[102] P. Stoica, R. L. Moses, *et al.*, *Spectral Analysis of Signals.* Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[103] M. Shaghaghi and S. A. Vorobyov, "Subspace leakage analysis and improved DOA estimation with small sample size," *IEEE Transactions on Signal Processing*, vol. 63, pp. 3251–3265, Apr. 2015.

[104] P. Vallet and P. Loubaton, "Toeplitz rectification and DoA estimation with MUSIC," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2237–2241, IEEE, May 2014.

[105] P. Forster, "Generalized rectification of cross spectral matrices for arrays of arbitrary geometry," *IEEE Transactions on Signal Processing*, vol. 49, pp. 972–978, May 2001.

[106] P. Vallet and P. Loubaton, "On the performance of MUSIC with toeplitz rectification in the context of large arrays," *IEEE Transactions on Signal Processing*, vol. 65, pp. 5848–5859, Aug. 2017.

[107] J. Cadzow, "Signal enhancement using canonical projection operators," in *ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 12, pp. 673–676, IEEE, Apr. 1987.

[108] K. F. Masood, R. Hu, J. Tong, J. Xi, Q. Guo, and Y. Yu, "A low-complexity three-stage estimator for low-rank mmWave channels," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5920–5931, 2021.

[109] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 806–811, 1985.

[110] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 8–15, 1989.

[111] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory.* John Wiley & Sons, 2004.

[112] H. Chung and S. Kim, "Atomic norm minimization-based low-overhead channel estimation for RIS-aided MIMO systems," *arXiv preprint arXiv:2107.09216*, 2021.

[113] N. Xi and L. Liping, "A computationally efficient subspace algorithm for 2-D DOA estimation with l-shaped array," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 971–974, 2014.

[114] Y. Wei and X. Guo, "Pair-matching method by signal covariance matrices for 2D-DOA estimation," *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp. 1199–1202, 2014.

[115] J.-F. Gu, W.-P. Zhu, and M. Swamy, "Joint 2-D DOA estimation via sparse l-shaped array," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1171–1182, 2015.

[116] M. Haardt, M. D. Zoltowski, C. P. Mathews, and J. Nossek, "2D unitary ESPRIT for efficient 2D parameter estimation," in *1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 2096–2099, IEEE, 1995.

[117] L. Yan, C. Han, and J. Yuan, "A dynamic array-of-subarrays architecture and hybrid precoding algorithms for terahertz wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2041–2056, 2020.

[118] H. Yuan, N. Yang, K. Yang, C. Han, and J. An, "Hybrid beamforming for terahertz multi-carrier systems over frequency selective fading," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6186–6199, 2020.

[119] S. Fan, Y. Wu, C. Han, and X. Wang, "Siabr: A structured intra-attention bidirectional recurrent deep learning method for ultra-accurate terahertz indoor localization," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 2226–2240, 2021.

[120] Z. Wan, Z. Gao, F. Gao, M. Di Renzo, and M.-S. Alouini, "Terahertz massive MIMO with holographic reconfigurable intelligent surfaces," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4732–4750, 2021.

[121] Y. Wang, W. Xu, H. Zhang, and X. You, "Wideband mmWave channel esti-

mation for hybrid massive MIMO with low-precision ADCs," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 285–288, 2018.

[122] A. M. Elbir, K. V. Mishra, and S. Chatzinotas, "Terahertz-band joint ultra-massive MIMO radar-communications: Model-based and model-free hybrid beamforming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1468–1483, 2021.

[123] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 23, pp. 5893–5908, 2019.

[124] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2946–2960, 2018.

[125] A. W. Yu, W. Ma, Y. Yu, J. Carbonell, and S. Sra, "Efficient structured matrix rank minimization," in *Advances in Neural Information Processing Systems*, pp. 1350–1358, 2014.

[126] S. C. Eisenstat, "Efficient implementation of a class of preconditioned conjugate gradient methods," *SIAM Journal on Scientific and Statistical Computing*, vol. 2, no. 1, pp. 1–4, 1981.

[127] O. Axelsson and G. Lindskog, "On the rate of convergence of the preconditioned conjugate gradient method," *Numerische Mathematik*, vol. 48, no. 5, pp. 499–523, 1986.

[128] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," tech. rep., School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, CMU-CS-94-125, 1994.

[129] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[130] K. Zhong, P. Jain, and I. S. Dhillon, "Efficient matrix sensing using rank-1 gaussian measurements," in *International Conference on Algorithmic Learning Theory*, pp. 3–18, Springer, 2015.

[131] K. F. Masood, J. Tong, J. Xi, J. Yuan, and Y. Yu, "Inductive matrix completion and root-music-based channel estimation for intelligent reflecting surface (IRS)-aided hybrid mimo systems," *IEEE Transactions on Wireless Communications*, 2023.

[132] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[133] T. Tong, C. Ma, and Y. Chi, "Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2396–2409, 2021.

[134] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE journal on selected areas in communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[135] G. Kwon, H. Park, and M. Z. Win, "Joint beamforming and power splitting for wideband millimeter wave swipt systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 5, pp. 1211–1227, 2021.

[136] Q. Qin, L. Gui, P. Cheng, and B. Gong, "Time-varying channel estimation for millimeter wave multiuser MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9435–9448, 2018.

[137] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36579–36589, 2019.

[138] R. Hu, J. Tong, J. Xi, Q. Guo, and Y. Yu, "Channel covariance matrix estimation via dimension reduction for hybrid MIMO mmwave communication systems," *Sensors*, vol. 19, no. 15, p. 3368, 2019.

[139] A. Liu, G. Liao, C. Zeng, Z. Yang, and Q. Xu, "An eigenstructure method for estimating DOA and sensor gain-phase errors," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5944–5956, 2011.

[140] P. Chen, Z. Chen, Z. Cao, and X. Wang, "A new atomic norm for DOA estimation with gain-phase errors," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4293–4306, 2020.