S. Noh, J. Lee, G. Lee, K. Seo, Y. Sung, and H. Yu, "Channel estimation techniques for RIS-assisted communication: Millimeter-wave and sub-THz systems," IEEE Vehicular Technology Magazine, to be published, Mar. 2022, doi: 10.1109/MVT.2022.3158765.

Channel Estimation Techniques for RIS-Assisted Communication: Millimeter-Wave and Sub-THz Systems

Song Noh, *Member, IEEE,* Junse Lee, *Member, IEEE,* Gilwon Lee, *Member, IEEE,* Kyungsik Seo, Youngchul Sung, *Senior Member, IEEE,* and Heejung Yu,‡ *Senior Member, IEEE*

***Abstract*—Millimeter-wave (mmWave) and sub-terahertz (sub-THz) communications are expected to be one of the biggest beneficiaries of the emerging reconfigurable intelligent surface (RIS) technology. RISs can compensate for large path loss and blockage inherent to the mmWave and sub-THz frequency bands to yield enhanced communication performance in these bands. To achieve high beamforming gain and realize the enhanced performance in RIS-assisted wireless communication, the acquisition of accurate channel state information is critical. In this article, we provide an overview of channel estimation for RIS-assisted mmWave/sub-THz communication to address technical challenges, trade-offs, channel estimation frameworks, and training signal design. We summarize the recent RIS-related sparse channel estimation approaches based on beam-space, sparse recovery, array signal processing, and data-driven techniques, highlighting several challenges for future research.**

I. Introduction

Millimeter-wave (mmWave) and sub-terahertz (sub-THz) communications are considered as one of the key technologies for 5G and beyond to support rate-demanding mobile applications such as extended reality encapsulating augmented/virtual/mixed realities. The large path loss in the sub-THz bands can be compensated for by beamforming based on large-scale antenna arrays. Recently, reconfigurable intelligent surfaces (RISs) have emerged as one of the most promising candidates as an evolving wireless technology [1], [2]. An RIS consists of low-cost passive elements and reflects incident signals like a scatterer in the propagation environment. Each RIS reflecting element can be controlled in a software-defined manner so that the reflected signals create a desirable multi-path effect such as signal focusing and interference cancellation, and the radio environment is controlled to improve wireless communication performance. Especially, combining mmWave/sub-THz and RISs can yield highly positive effects such as mitigating propagation blockage in the mmWave/sub-THz bands with additional paths and reducing the number of required antennas at transceivers to achieve target performance.

To harness such benefits of RIS-assisted communication, it is crucial to acquire accurate channel state information (CSI) for high-gain beamforming. An RIS-assisted communication channel is composed of two subchannels: the *direct channel* between the base station (BS) and a user equipment (UE) and the *indirect channel* generated by the RIS. The end-to-end channel between the BS and the UE can be estimated at a link end in a conventional manner. However, it is challenging to estimate individual channels between the RIS and two link ends and separate the effect of the RIS from the channel estimate due to the passive operation of RISs. In addition, the sparse scattering nature of mmWave/sub-THz channels should be exploited for channel estimation to work with fewer training signals or suppress channel estimation errors. There has been active research on channel estimation for RIS-assisted communication [1], [2]. Compared with channel estimation techniques under dense scattering propagation valid in sub-6GHz bands, those under sparse scattering propagation in the mmWave and sub-THz bands are less mature. Existing works for sparse channel estimation are specific to different system and channel models, thereby hindering their applicability. Thus, it is important to combine the categorization of the existing research contributions and results with systematizing future research directions.

This article provides an overview of channel estimation for RIS-assisted systems, emphasizing mmWave and sub-THz communication. In addition to summarizing recent development in sparse channel estimation techniques, in-depth discussions on technical challenges and trade-offs are presented to inspire future research in this field.

II. RIS-assisted Millimeter-wave and sub-THz Communication

This section explains an RIS-assisted mmWave/sub-THz system model and key features relevant to channel estimation.

A. System and Channel Model

For ease of exposition, we consider the basic RIS-assisted communication model employing an RIS to assist the transmission between a BS and a single-antenna UE, shown in Fig. 1. To establish a favorable channel environment, the RIS adjusts the phase of the reflected signal by using its reflecting elements, which are dynamically programmable through a separate control link. The main feature of mmWave and sub-THz propagation channels is *sparsity*. One of the widely-used channel models incorporating the sparse scattering nature of channels is the geometry-based channel model [3]–[9], in which a sparse channel is parameterized by a set of angle-of-arrivals (AoAs), angle-of-departures (AoDs), and path gains. For example, seven geometric parameters are used to determine the channel paths among the BS, RIS, and UE in Fig. 1. The path angles of the BS from/to the UE and RIS are denoted as $\boldsymbol{\theta}_{\text{BS-UE}}$ and $\boldsymbol{\theta}_{\text{BS-RIS}}$, and the path angles of the RIS from/to the BS and the UE are denoted as $\boldsymbol{\psi}_{\text{RIS-BS}}$ and $\boldsymbol{\psi}_{\text{RIS-UE}}$. Accordingly, the path gains for the BS-UE, BS-RIS, and RIS-UE are given by $\boldsymbol{\alpha}_{\text{BS-UE}}$, $\boldsymbol{\beta}_{\text{BS-RIS}}$, and $\boldsymbol{\alpha}_{\text{RIS-UE}}$, respectively.

‡ Corresponding author

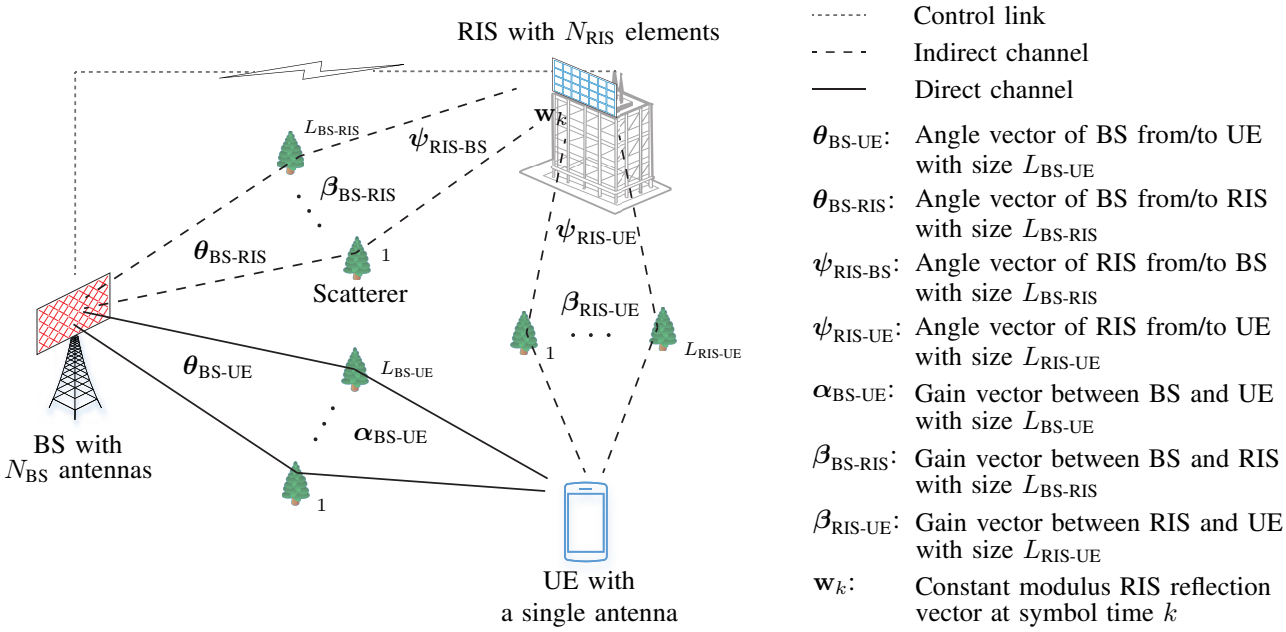

Fig. 1: Illustration of RIS-assisted mmWave and sub-THz communication system.

Here, the numbers of paths for the BS-RIS, BS-UE, and RIS-UE channels are denoted as $L_{\text{BS-RIS}}$, $L_{\text{BS-UE}}$, and $L_{\text{RIS-UE}}$, respectively. When uniform planar arrays are adopted, each path is characterized by azimuth and elevation angles, but only the angles in one dimension are considered for simplicity.

In time-division duplexing systems, a common approach to channel estimation is exploiting the channel reciprocity that the downlink CSI can be obtained from its uplink training. In general, this channel reciprocity is not applicable for frequency-division duplexing systems that utilize the downlink training and uplink feedback for downlink channel estimation. The considered model in Fig. 1 can be extended to multi-user cases. In the downlink, the BS can have common pilot symbols shared by all UEs, and each UE estimates its downlink CSI similarly to the single-user case as in [4]. In the uplink, the UEs send orthogonal pilot sequences in time slots [5], [7], frequency bands, or codes so that the BS estimates its uplink CSIs separately.

During the training period, the uplink or downlink pilot signal arrives at the receiver through the direct (solid line) and indirect (dashed line) channels, as shown in Fig. 1. Unlike the direct BS-UE channel that is similar to a conventional point-to-point mmWave channel, the indirect channel is a two-hop channel, and each of the BS-RIS and RIS-UE channels is not directly observable due to the passive operation of the RIS. Thus, it is not simple to apply conventional channel estimation schemes to indirect channel estimation. Furthermore, there exist additional degrees of freedom in designing the phase shift $\mathbf{w}_k$ of the RIS during the training period for enhanced channel estimation.

B. Individual versus Cascaded Channel Estimation

Most of the methods for channel estimation in RIS-assisted communication can be classified into two categories based on the way to handling the indirect BS-RIS-UE channel. One approach is to estimate all individual BS-RIS, RIS-UE, and BS-UE channels: we refer to this approach as *individual* channel estimation. In this case, the number of angle parameters to be estimated is $L_{\text{BS-RIS}} + L_{\text{BS-UE}}$ for the BS and $L_{\text{BS-RIS}} + L_{\text{RIS-UE}}$ for the RIS. This approach enables conventional relay-based data precoding techniques to be applied to RIS-assisted communication because all the CSI is identified. However, it is required to use additional training symbols or complicated processing for individual channel estimation because the receiver obtains only the end-to-end indirect channel measurement.

The other approach is to estimate the direct channel and the indirect cascaded channel instead of estimating the individual channels because the knowledge of the cascaded channel is sufficient for coherent decoding at the receiver and passive beamforming at the RIS. We refer to this approach as *cascaded* channel estimation. Note that there are $L_{\text{BS-RIS}}L_{\text{RIS-UE}}$ effective cascaded paths from the combination of $L_{\text{BS-RIS}}$ and $L_{\text{RIS-UE}}$ paths. With transforming physical angles to the spatial frequency domain, the cascaded spatial frequency is defined as the sum of the BS-RIS and RIS-UE spatial frequencies. Since there is a one-to-one mapping between spatial frequency and physical angle, we refer to the cascaded spatial frequency as the cascaded angle for simplicity. Compared with individual channel estimation, the number of cascaded angle parameters to estimate generally increases since the product $L_{\text{BS-RIS}}L_{\text{RIS-UE}}$ is larger than the sum in general. This is different from the dense channel assumption in which the indirect cascaded channel matrix is defined by the element-wise product of two BS-RIS and RIS-UE channel matrices, and thus it does not increase the number of cascaded channel parameters to estimate.

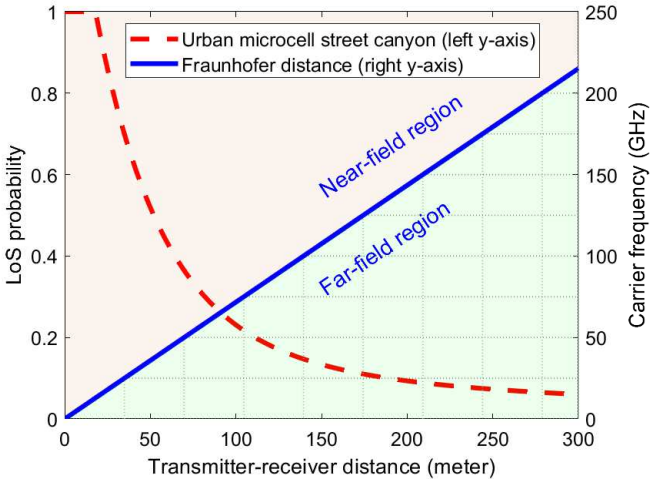

Fig. 2: LoS probability and the Fraunhofer boundary versus the transmitter-receiver distance when the size of an antenna array is 0.5m.

C. Separate versus Joint Estimation of Direct and Indirect Channels

When there exists only one of direct and indirect channels due to blockage on the other of the two, conventional methods can be applied to channel estimation with some modification. When both direct and indirect channels are considered, on the other hand, the channel estimation is not straightforward. One framework for channel estimation in this case is to treat the direct and indirect channels separately. One approach under this framework is a two-stage approach. The BS estimates the direct channel with a conventional sparse channel estimation scheme at the first stage with the RIS turned off. Then, the indirect channel is estimated by turning on the RIS at the second stage, where the direct channel's contribution to the received signal is subtracted by the estimate of the direct channel obtained at the first stage. In this approach, the periodical on-off switching of all the RIS elements can introduce a stringent requirement for time synchronization, especially for multi-user cases. The second stage may suffer from residual interference caused by imperfect direct channel estimation.

Although the separate estimation of the direct and indirect channels simplifies the overall estimation procedure, it is suboptimal due to separate use of all information. The other framework for channel estimation is to jointly exploit all channel measurements from both direct and indirect channels. This joint estimation of direct and indirect channels has not been fully investigated because it is difficult to develop an estimation algorithm based on the superimposed pilot signals received from both direct and indirect channels. However, this joint approach potentially enhances the estimation performance by joint use of all training signals. In addition, the joint approach is practically important because this enables the deployment of RISs fully transparent to wireless users by eliminating the need for switching between direct and indirect channel estimation.

D. Features of RIS-assisted Wireless Environment

The low-cost energy-efficient RIS elements enable the dense deployment of RISs for enhanced system performance. In this regard, the RIS-assisted mmWave and sub-THz communications present several propagation features of the radio channel that should be addressed in the deployment of RISs.

1) LoS versus Non-LoS: The low-cost RIS feature enables dense deployment to mitigate high pathloss and blockage encountered at mmWave frequencies. The resulting short-range propagation is potentially line-of-sight (LoS)-dominant channels. Fig. 2 shows the distance dependence of the LoS link probability in an urban micro-cellular scenario [10]. This dense RIS deployment enables UEs to access the BS through virtual LoS channels, which is beneficial, especially for single unicast and broadcast. However, in conventional multi-input multi-output (MIMO) systems, a LoS channel is approximated as a low-rank matrix [1], [2]. This implies that it is difficult to exploit the multiplexing gain of the RIS-assisted MIMO system for supporting multi-stream transmission. Therefore, the RIS deployment should be well planned to leverage the full potential of MIMO gains.

2) Near-field versus Far-field: A massive number of reflecting elements can be implemented at an RIS to achieve the RIS gain. With the extended RIS size, the phase difference of the received signal across the reflecting elements becomes non-negligible, i.e., the near-field effect arises [11]. In the near-field region, the directional characteristics of the radio channel are not simply decoupled as plane waves perpendicular to the direction of propagation due to distance-dependent angular distributions. Moreover, for the same definition of the RIS, the broader area in the radio coverage becomes the near-field region with increasing carrier frequency up to sub-THz. For instance, Fig. 2 shows the Fraunhofer distance separating near- and far-field regions with respect to carrier frequency. Therefore, it is necessary to consider the effect of spherical waves in algorithmic and signal processing aspects.

3) Wideband Communication: In wideband communication systems, the delay difference of the received signal on a large-scale array of antennas can be comparable to a sampling period in time domain. Such delay difference makes the system see different spatial directions for the same physical path over sub-bands, referred to as *beam squint*. Since the effect of beam squint increases with the number of antennas or the bandwidth, ignoring beam squint can degrade the performance of conventional channel estimation or beamforming methods. From a hardware perspective, the RIS reflecting elements exhibit frequency-dependent responses [12], which is different from the simple assumption for analysis that the RIS's reflection response is frequency-independent. The aforementioned issues need to be investigated thoroughly to enable the RIS-assisted communication systems to be practically feasible.

III. Channel Estimation Methodologies

This section discusses major channel estimation strategies and frameworks for RIS-assisted mmWave and sub-THz communication and their pros and cons. Fig. 3 shows the representative methodologies, leveraging the sparse nature of mmWave and sub-THz channels to enhance performance.

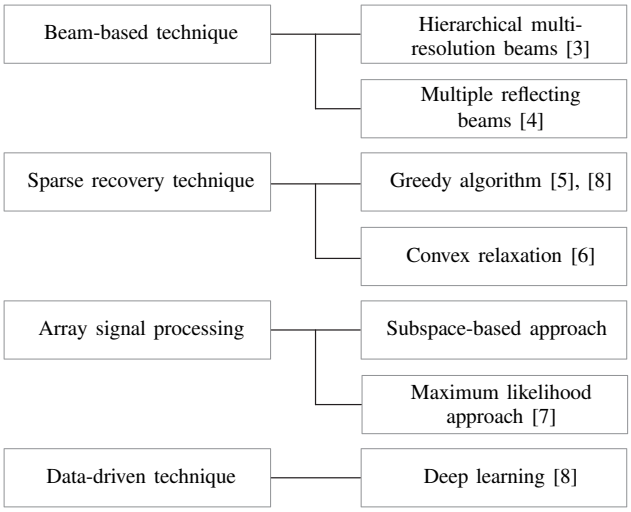

Fig. 3: Classification of sparse channel estimation methodologies.

A. Beam-based Technique

In 5G new radio (NR)'s initial access procedure, a BS periodically transmits multiple beamformed NR synchronization signals, and a UE selects the best beam to establish the connection. This beam-based technique can be applied to RIS-assisted wireless systems in which the BS, RIS, and UE should find the best beam directions. The beam-based technique can be viewed as an approximate version of channel estimation in a sense that a selected beam index reflects the rough information of the angle parameter.

A hierarchical beam search method was proposed to reduce beam search latency [3]. The idea of this approach is that the BS and UE perform an exhaustive search with wide beams at the initial stage and subsequently refine the selected wide beam with narrow beams. In [4], the RIS is treated as an array-of-subarrays by dividing the N_{RIS} RIS elements into N_{subarray} subarrays, each consisting of $N_{\text{RIS}}/N_{\text{subarray}}$ elements. This RIS structure enables multi-beam training that multiple reflecting directions are simultaneously probed by steering each subarray's reflecting pattern to a distinct direction. Since such beam-based techniques are usually based on a finite set of pre-designed beams, it is feasible to feed back the best beam index to the transmitter through a finite-bit feedback channel. However, using the wide beam at each subarray or the initial stage can constrain the service coverage due to its low directivity gain. In addition, an early-stage beam misalignment can cause beam mismatch in subsequent beam refinement stages.

B. Sparse Recovery Technique

By considering the sparsity of mmWave/sub-THz channels, a two-phase uplink channel estimation method was proposed in the multi-user case [5]. In the first phase, the uplink channel of one typical UE is estimated in three steps: (i) AoA estimation at the BS, (ii) cascaded channel estimation associated with a particular BS-RIS path, and (iii) estimation of the remaining $L_{\text{BS-RIS}} - 1$ BS-RIS paths. The devised algorithms at Steps (i) and (iii) are correlation-based methods exploiting *a priori* knowledge of the BS and RIS array manifold. At Step (ii), the compressed sensing (CS)-based reconstruction algorithm is applied by approximating the system model as a function of the cascaded channel only, assuming that the spatial BS-RIS paths are asymptotically orthogonal with infinitely-many BS antennas. Step (ii) determines the training overhead for the first phase to guarantee to recover the modeled sparse channel vector. In the second phase, the cascaded channels of the remaining users are estimated by exploiting the estimated BS-RIS CSI in the first phase. The employed algorithm at Step (ii) is an approximation method because the off-grid mismatch exists between actual continuous angles and discrete angles in the dictionary matrix. In [8], a similar CS-based algorithm was proposed based on a semi-passive RIS configuration that allows the RIS to embed a few active sensors capable of sensing. The added sensing mode at the RIS enables estimating individual indirect channels at the RIS and simplifies the channel estimation protocol.

To alleviate the off-grid mismatch, a discretization-free technique based on atomic norm minimization was introduced for downlink channel estimation [6]. The atomic norm-based approach addresses the grid mismatch problem well because an atomic set has infinitely many atoms with the same dimension of the signals to be recovered. At the first stage, two decoupled angle parameters (i.e., AoDs for a BS and AoAs for a UE) are estimated by formulating the atomic norm minimization as dual semidefinite programming (SDP) and root-finding methods in addition. Given the estimated angles, the transmit/receive precoders at the second stage are designed to approximately decorrelate the array manifold for the BS and UE in the received training signal so that the observation model is dependent only on the cascaded angle and gain for an RIS. This approximation is similar to that in [5] in which the array manifold becomes asymptotically orthogonal with a sufficiently large number of antennas. After decorrelation, the cascaded angles for the RIS are estimated from the SDP solution as in the first stage.

C. Array Signal Processing

During the training period, the RIS reflection can be used to manipulate the spatio-temporal signature (or subspace) of the reflected signal. The RIS reflection coefficients can be integrated into the reflected signal as a multiplicative factor that enables spatial encoding with the phase shifts. Similarly, the temporal signature of the reflected signal can be modified by changing the RIS reflection coefficient over multiple training symbols. This spatio-temporal signature of the received signal can be used to construct signal subspace to which conventional subspace-based methods can be applied for channel estimation. Furthermore, the problem of sparse channel estimation can be approached with a general maximum likelihood (ML) criterion based on the prior knowledge of the array manifold and the training reflection pattern [7]. This approach aims to maximize the likelihood iteratively to reduce the computational complexity involved with joint ML estimation. To be specific, each maximization step corresponds

Framework		Direct channel	Channel estimation approach		Minimum training overhead
Cascaded channel estimation	Separate estimation of direct and indirect channels	Unblocked and unknown	[3]	Three-phase beam search by combining partial search at RIS and hierarchical search at BS and UE	$6N_{\text{RIS}} + 4\log_3(\max\{N_{\text{BS}}, N_{\text{UE}}\}) - 1$
		Blocked or known	[4]	Multiple beam sweeping with multi-directional reflecting beams based on an array-of-subarrays RIS architecture	$\frac{N_{\text{RIS}}}{N_{\text{subarray}}}\left(1 + \frac{\log_2 N_{\text{subarray}}}{2}\right)$
			[5]	Two-phase estimation based on correlation-based filtering and orthogonal matching pursuit (OMP)-based sparse recovery algorithm	$8L_{\text{RIS-UE}} - 2$
			[6]	Two-stage estimation based on dual SDP of the atomic norm problem and least square estimator	$N_{1\text{st}} + N_{2\text{nd}}L_{\text{BS-RIS}}$
Individual channel estimation			[8]	Two OMP- and deep learning-based algorithms using a few active RIS sensors	2
Cascaded channel estimation	Joint estimation	Unblocked and unknown / Blocked or known	[7]	Two-stage estimation based on root-finding methods and iterative ML estimation	$\max\{L_{\text{RIS-UE}}, L_{\text{BS-UE}}\} + 2$

Table 1. A summary of sparse channel estimation approaches for RIS-assisted communication where N_{UE} denotes the number of antenna elements of the UE [3]; N_{subarray} denotes the number of subarrays at the RIS [4]; $N_{1\text{st}}$ and $N_{2\text{nd}}$ denote the number of training symbols at the first and second stages [6]. The remaining notations in the last column follow those in Fig. 1.

to the basis update for the noise subspace estimation, which is formulated as a quadratic optimization problem. Once the iteration converges, the path angles can be estimated from the polynomial roots characterizing the noise subspace.

D. Data-driven Technique

The aforementioned channel estimation approaches are in the form of analytical methods optimized for specific system models. Devising an algorithm for general systems is challenging, but it is possible to obtain approximate solutions based on deep learning techniques. A deep learning approach can learn correlations in a set of data representing an environment of interest. Therefore, this approach works on channel estimation problems by composing training datasets with the received pilot signal and estimated CSI. For example, the design of RIS reflection pattern with implicit channel estimation was proposed under the assumption that a semi-passive RIS receives pilot symbols from a BS and a UE using a few active sensors [8]. Then, a deep neural network on the RIS uses the collected data as input to choose a good RIS reflection pattern as output. In general, conventional deep-learning methods can be applied to uplink training because the BS can collect all training datasets from associated UEs and run model training in a centralized manner. On the other hand, a UE has only its own training datasets with a limited computing capability in downlink training. This constraint can be alleviated by allowing multiple UEs to collaborate on model training through federated learning.

It is interesting to see the training overhead for all the considered algorithms in the single user case, as summarized in Table 1. The methods in [3], [4] require training overhead in the order of the number of BS/RIS/UE antenna (or reflecting) elements. The methods in [5]–[7] require relatively small training overhead related to the number of channel paths under the assumption that the number of channel paths is known or pre-estimated. The lowest overhead is achieved with [8] due to the semi-passive RIS assumption. The comparison of these

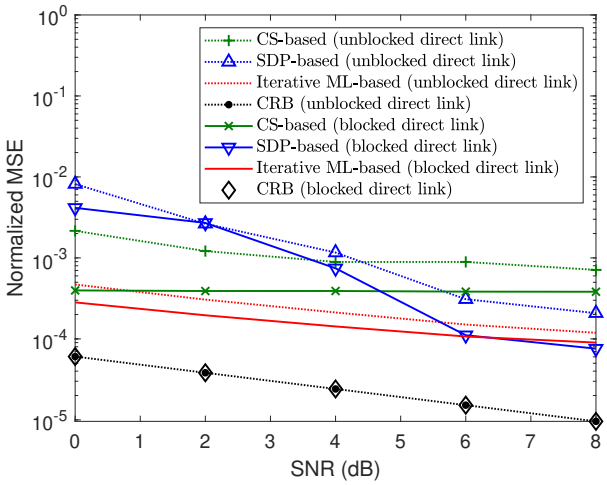

Fig. 4: Performance comparison of several methods, including CS-based [5], SDP-based [6], iterative ML-based [7], CRB [9], where the number of BS antennas $N_{\text{BS}} = 8$, the number of RIS reflecting elements $N_{\text{RIS}} = 32$, the number of pilot symbols $K = 22$, and SNR denotes the signal-to-noise ratio.

algorithms in terms of the minimum training overhead only is not complete because they are optimized for different system and channel models. For example, in the presence of direct link, direct comparison of [5]–[7] may not be fair because the methods in [5], [6] require additional pilot symbols for direct channel estimation.

E. Performance Evaluation

Fig. 4 shows the normalized mean square error (MSE) performance of several methods. With the same number of pilot symbols for all methods, the ratio of pilot symbols over two-stage estimation or separate direct and indirect channel estimation in [5], [6] is fine-tuned offline for a fair comparison. The performance difference of the joint estimation framework [7] with or without the direct link is relatively small compared with the separate estimation framework [5], [6]. Because

Approach	Pilot sequence design		RIS reflection pattern design
Algorithm-specific design	[3]	Hierarchical multi-resolution beams	DFT-based (narrow) beam pattern
	[4]	Rank-one beamforming with perfect CSI	DFT-based (wide) beam pattern
	[5]	Constant pilot symbol	MC-based pattern
	[6]	Random training matrix and array steering matrix with estimated CSI	Pseudo-noise (PN) pattern
	[7]	Constant pilot symbol	Partial on-off reflection pattern
Systematic design	[9]	Any pilot symbol satisfying the constant power constraint	Bayesian and hybrid CRB-based pattern

Table 2. A summary of training signal design approaches.

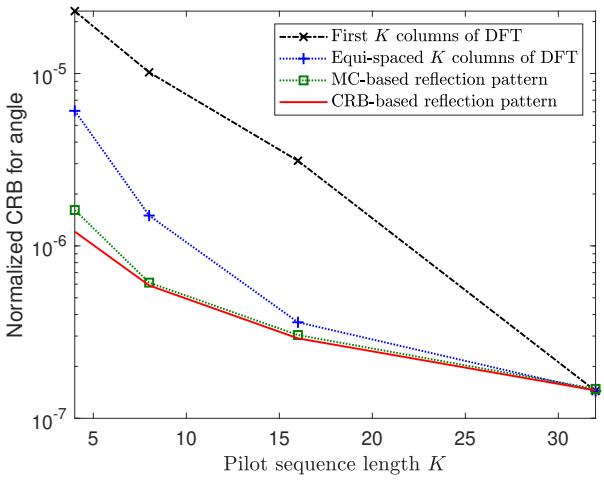

Fig. 5: Performance comparison of RIS reflection patterns where the number of RIS reflecting elements $N_{\text{RIS}} = 32$ and SNR $= 5$ dB.

the separate estimation framework can be subject to error propagation from imperfect direct channel estimation. This numerical study also clarifies that there is room for further improvement compared with the CRB.

IV. Training Signal Design Approaches

Up to now, we went over the channel estimation frameworks and algorithms. To enhance channel estimation performance, it is essential to properly design training signals, including the pilot sequence and the RIS reflection pattern. Unlike dense propagation environments, the system model is related to a nonlinear observation model containing angle parameters as the exponent term of the array manifold. Thus, a common approach based on MSE is inapplicable to training signal design because the MSE for sparse channel estimation is not derived analytically in general [13]. In this section, we briefly introduce recent developments in training signal design for sparse channel estimation.

A. Algorithm-specific Design

One approach to training signal design is to specify the channel estimation algorithm and then optimize the training signal under specific design criteria, summarized in Table 2. The authors in [3], [4] proposed several beam broadening approaches so that passive/active beam sweeping over the entire spatial horizon is done with low beam training overhead. In [4], a wide beamwidth reflection codebook for RIS is generated by using an array-of-subarrays. In [4], variable beamwidth codebooks are designed using a subset of antenna elements or synthesizing multiple beams, depending on the number of available radio frequency chains at the transceiver. In [5], the RIS reflection pattern is designed by minimizing the mutual coherence (MC) that measures extreme correlations in the equivalent dictionary defined as the product of the RIS reflection pattern and a dictionary matrix. The MC is an analytically tractable metric that is used to access the CS recovery capability. This design problem is non-convex due to the constant modulus constraint for the RIS reflection pattern, and thus a sub-optimal solution is obtained by relaxing the constraint. The authors in [6], [7] considered specific training signals to facilitate their algorithms. In [6], the transmit/receive precoders are generated from random matrices for diversified channel sounding and designed as an array steering matrix with the previously estimated angle. In [7], channel estimation is carried out by switching the on-off state of the RIS elements one by one so that *a priori* knowledge of RIS geometry is exploited for channel estimation. Such an on-off reflection pattern is modeled as some columns of the identity matrix, and thus the reflected signal structure becomes equivalent to the array manifold of the RIS. In [8], the codebook-based reflection pattern was proposed to reflect the practical requirement that the RIS phase shifters are controlled with a set of quantized angles. In this approach, the RIS reflection pattern is used only for data transmission because the RIS uses the integrated active sensors for receiving the training signal without passive reflection. Note that the joint design of the pilot sequence and RIS reflection pattern along with algorithms is non-trivial [2]. To tackle this challenge, most of existing methods, including the methods in this subsection, consider the separate design of the training signal by means of the known behavior of their chosen algorithms.

B. Systematic Design

Two Cramér-Rao lower bound (CRB)-based design approaches were proposed to establish algorithm-independent training signal design under the assumption that both direct and indirect channels exist [9]. The CRB is associated with the signal model itself, not with any particular estimation

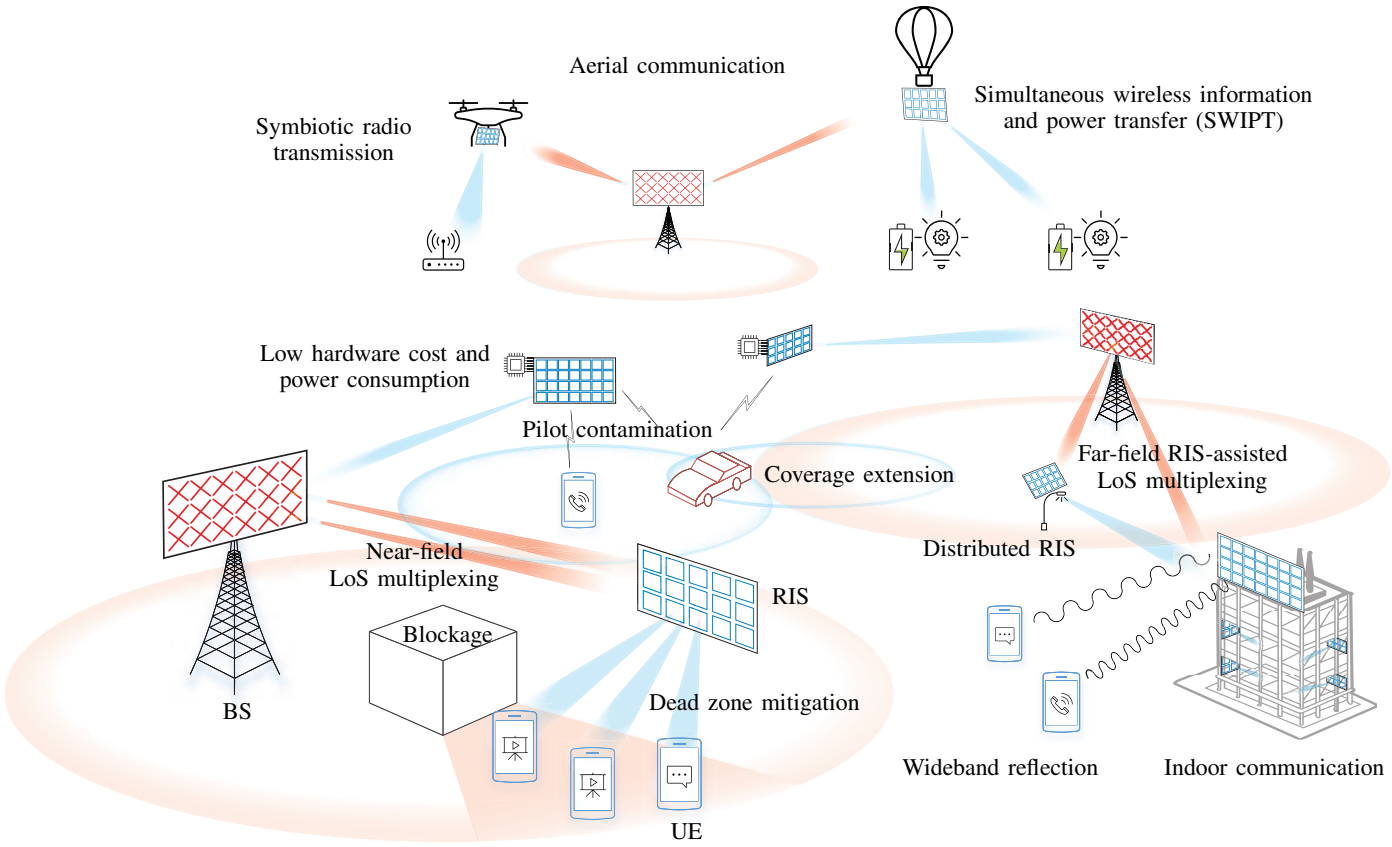

Fig. 6: RIS-assisted communication use cases and scenarios in multi-cell wireless network.

algorithm. Thus, the resulting training signal design can be considered universal. In the Bayesian framework, the channel path gains and angles are considered as random parameters. The Bayesian CRB provides a lower bound on the MSE of any Bayesian channel estimator exploiting *a priori* knowledge of the distribution for the random channel parameter. Since the Bayesian Fisher information is averaged over the parameter distribution, the Bayesian CRB is independent of parameter realizations, and the optimal training signal minimizing the Bayesian CRB is derived analytically. In [9], another RIS pattern design approach is proposed based on a hybrid parameter assumption to incorporate the case that the prior knowledge is not available, especially for the path angle [9]. In this case, the path angles are considered as deterministic parameters, whereas the path gains are treated as random nuisance parameters. Accordingly, the hybrid CRB is dependent on the realization of the path angle, and it is possible to formulate the training signal design problem relevant to a set of target angles. The resulting optimization is numerically solved by the projected gradient method. Under the joint optimization of the pilot sequence and RIS reflection pattern, transmission of a constant modulus pilot symbol at a single-antenna transmitter is optimal in terms of minimizing the CRB. However, the joint design of the training signal for a multiple-antenna transmitter has not been fully addressed, which is an interesting problem for future work.

C. Performance Evaluation

Direct comparison of training signal designs is not straightforward because the design algorithms are based on different system/channel assumptions. The CRB is used as a performance metric without relying on specific channel estimation algorithms because the CRB is associated with the system model itself. Fig. 5 shows the CRB on the estimation of path angles for the methods in [3]–[5], [9]. The design of the equispaced K columns of the discrete Fourier transform (DFT) matrix shows better performance than the consecutive first K columns of the DFT matrix. When the pilot sequence length is identical to the number of RIS elements, the DFT-based reflection pattern achieves the best performance, as proven in [14]. However, the CRB-based design in [9] is shown to be effective as the pilot sequence length K decreases. Interestingly, the design based on MC optimization performs close to the CRB-based design.

V. Conclusion and Future Directions

In this article, we have presented a comprehensive overview of channel estimation for RIS-assisted mmWave/sub-THz communication, including topics from representative channel estimation frameworks to RIS training signal design. As illustrated in Fig. 6, several challenges and open problems remain to be addressed to accommodate diverse use cases of RISs.

First, inventing a CRB-achieving estimator is essential for achieving near-optimal performance along with the joint de-

sign of the pilot sequence and RIS reflection pattern for a general setup, including a multiple-antenna transmitter, wideband transmission, and hybrid beamforming. Second, in the multi-user uplink channel estimation, orthogonal pilot sequences can be assigned across UEs in a cell to avoid inter-user interference. When RISs are deployed in the cell edge for coverage extension, the inter-user interference may not vanish due to the reuse of orthogonal pilot sequences in multiple cells, known as *pilot contamination* in multi-cell massive MIMO. Ways to handle the pilot contamination need to be investigated for the enhanced channel estimation in the multi-cell case. Next, the ideal assumption that the reflecting coefficient per RIS element is controlled independently in the continuous domain can facilitate the analysis of performance limits and reduce the complexity of algorithm design. However, implementing fine-grained reflecting elements raises the cost and design complexity. Thus, it is necessary to consider these hardware aspects for realizing efficient channel estimation algorithms in practice. Lastly, the effect of spherical waves in the near-field region can enable spatial multiplexing gain in LoS channels [15]. This opens up new possibilities for RISs configuring the LoS MIMO channel rank. The comprehension of the involved channel estimation algorithm and signal processing for transmission/reception has not been addressed and remains future research.

REFERENCES

[1] M. D. Renzo, A. Zappone, M. Debbah, M. Alouini, C. Yuen, J. D. Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450 – 2525, Nov. 2020.

[2] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313 – 3351, May 2021.

[3] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Terahertz multi-user massive MIMO with intelligent reflecting surface: Beam training and hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1376 – 1393, Feb. 2021.

[4] C. You, B. Zheng, and R. Zhang, "Fast beam training for IRS-assisted multiuser communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1845 – 1849, Aug. 2020.

[5] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, "Channel estimation for RIS-aided multiuser millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 1478 – 1492, Mar. 2022.

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

[7] S. Noh *et al.*, "Joint direct and indirect channel estimation for RIS-assisted mmWave systems based on array signal processing," in preparation for submission, 2022.

[8] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44 304 – 44 321, Mar. 2021.

[9] S. Noh, H. Yu, and Y. Sung, "Training signal design for sparse channel estimation in intelligent reflecting surface-assisted millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2399 – 2413, Apr. 2022.

[10] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," *TS 38.901 V16.1.0*, Jan. 2020.

[11] W. Tang, M. Z. Zhen, X. Chen, J. Y. Dai, Y. Han, M. D. Renzo, Y. Zong, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421 – 439, Jan. 2021.

[12] H. Li, W. Cai, Y. Liu, M. Li, Q. Liu, and Q. Wu, "Intelligent reflecting surface enhanced wideband MIMO-OFDM communications: From practical model to reflection optimization," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4807 – 4820, Mar. 2021.

[13] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684 – 698, Feb. 2015.

[14] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Process. (ICASSP)*, Barcelona, Spain, May 2020.

[15] H. Do, S. Cho, J. Park, H. Song, N. Lee, and A. Lozano, "Terahertz line-of-sight MIMO communication: Theory and practical challenges," *IEEE Wireless Commun. Mag.*, vol. 59, no. 3, pp. 104 – 109, Mar. 2021.