RIS-aided Joint Channel Estimation and Localization at mmWave under Hardware Impairments: A Dictionary Learning-based Approach

Murat Bayraktar, *Graduate Student Member, IEEE*, Nuria González-Prelcic, *Senior Member, IEEE*, George C. Alexandropoulos, *Senior Member, IEEE* and Hao Chen

***Abstract*—Reconfigurable intelligent surface (RIS)-aided millimeter wave (mmWave) wireless systems offer robustness to blockage and enhanced coverage. In this paper, we develop an algorithmic solution that shows how RISs can also enhance the positioning performance in a joint localization and communication setting, even when hardware impairments are considered. We propose a realistic system architecture that considers the clock offset between the transmitter and the receiver, impairments at transmit and receive arrays, and mutual coupling between the RIS elements. We formulate the estimation of the composite channel in a RIS-aided mmWave system as a multidimensional orthogonal matching pursuit problem, which can be solved with high accuracy and low complexity, even when operating with large antenna arrays as required at mmWave. In addition, we introduce a dictionary learning stage to calibrate the hardware impairments at the user array. To complete our design, we devise a localization scheme that exploits the estimated composite channel while accounting for the clock offset between the transmitter and the receiver. Numerical results show how RIS-aided mmWave systems can significantly improve the localization accuracy in a realistic 3D indoor scenario simulated by ray tracing.**

***Index Terms*—Reconfigurable intelligent surface, joint localization and channel estimation, hardware impairments, dictionary learning.**

I. Introduction

Multiple-input multiple-output (MIMO) communication utilizing large arrays and high bandwidths, such as those employed in millimeter wave (mmWave) bands, enables high data rates for wireless links while increasing angle and delay resolvability during channel parameter estimation. Moreover, the sparsity of the channels enhances the sensing performance of the communication waveform [2]. Despite these advantages, state-of-the-art solutions for mmWave localization still face challenges in achieving the required accuracy for specific use cases that depend on precise position information. Current approaches include ultra-dense deployments for beam-based coordinated measurements from multiple base stations (BSs) [3]–[5], deep networks exploiting power delay profiles or other channel parameters [6]–[9], and two-stage methods that involve a sparse channel estimation phase and a subsequent stage that maps the estimated channel parameters to the position and orientation of the user by exploiting geometric relationships [10]–[17].

Reconfigurable intelligent surfaces (RISs) have recently emerged as a promising technology for improving the performance of wireless systems [18]–[22]. By digitally configuring the coefficients of its passive elements, the reflection angle of a RIS can be adjusted, enabling the creation of controllable propagation environments for wireless communications. This feature is especially useful in situations where there is a blockage between the BS and the mobile station (MS) [23], a challenge that is particularly acute in the mmWave band, where the line-of-sight (LoS) path is typically much stronger than the non-line-of-sight (NLoS) paths. Although the benefits of RISs in communication systems are limited when the LoS path between the BS and the MS is present [18], the accuracy of positioning solutions that exploit the communication waveform could be significantly increased if a second strong path is available [1], [24], [25]. Moreover, higher accuracy could also be obtained in NLoS scenarios, which are usually very challenging for precise localization.

A. Prior Work

One of the primary challenges in RIS-aided mmWave systems is channel estimation, as the large number of passive elements in RISs requires the estimation of large matrices [26]. The passive nature of the RIS also complicates the channel estimation process since a cascade channel comprising the BS-RIS-MS links needs to be estimated. There are several recent attempts to estimate the large dimensional channels of RIS-aided systems [27]–[39]. A compressed channel estimation method is proposed to estimate the narrowband cascade channel via a RIS where the BS-MS channel is blocked in [27]. The BS-RIS channel is determined by beam searching, and the RIS-MS channel is estimated with a sparse recovery algorithm in [28]. With the knowledge of the LoS path of the BS-RIS channel, the BS-MS and the RIS-MS channels are

A preliminary version of this work, which considers a similar system model without antenna array or RIS impairments was presented in [1]. This material is based upon work supported in part by the National Science Foundation under grant no. 2433782 and is supported in part by funds from the federal agency and industry partners as specified in the Resilient & Intelligent NextG Systems (RINGS) program and by Samsung Research America.

M. Bayraktar and N. González-Prelcic are with the Electrical and Computer Engineering Department, University of California, San Diego, USA (e-mail: {mbayraktar,ngprelcic}@ucsd.edu). G. C. Alexandropoulos is with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece (email: alexandg@di.uoa.gr). H. Chen is with Samsung Research America, El Plano, TX (email: hao.chen1@samsung.com).

jointly estimated with a compressed sensing approach in [29]. A fixed-rank manifold optimization algorithm is adopted for alternatingly estimating the BS-RIS and the RIS-MS channels in [30]. A convolutional neural network architecture is utilized to sequentially estimate the BS-MS and the BS-RIS-MS channels in [31]. Sparse multiuser channel estimation using nuclear norm is proposed in [32]. Newtonized orthogonal matching pursuit algorithm (OMP) is proposed to estimate the MS-RIS channel with the knowledge of the LoS path of the RIS-BS channel in [33]. In a multiuser scenario, [34] proposes a double-structured OMP algorithm that first estimates the common BS-RIS channel while the RIS-MS channel of each user is subsequently estimated. In [35], the direction of arrival (DoA) at the MS and the direction of departure (DoD) at the BS are estimated by atomic norm minimization followed by MUSIC/ESPRIT in a narrowband system. Then, the DoA and the DoD at the RIS are estimated using a similar atomic norm minimization approach. In [36], [37], dictionary learning methods are utilized to estimate the ill-conditioned narrowband cascade RIS channel. The work in [38] proposes a two-stage multiuser narrowband channel estimation algorithm where the common angles at the BS are estimated in the first stage by using the pilots transmitted by a single MS, while the cascade channels of MSs are estimated in the second stage. In a similar setting, [39] proposes an improved algorithm that leverages the pilots transmitted by multiple MSs in the first stage. In [40], a general framework for channel estimation in RIS-aided communication systems is presented with various channel models, array structures, and channel estimation algorithms.

Most channel estimation methods, such as the ones that utilize compressed sensing for RIS-assisted mmWave systems, exploit the sparsity of the channel and estimate the angular and delay parameters of the paths, which can also be used for localization. A RIS can enhance the localization performance with or without the presence of a LoS path between the BS and the MS since it creates another relatively strong path whose direction can be controlled [1], [25]. Theoretical lower bounds on the positioning error for RIS-aided localization were obtained in [41]–[43], whereas practical positioning algorithms exploiting RIS were proposed in [44]–[48]. Hierarchical codebooks were designed for RIS phase configurations and training beamformers to estimate the DoA, DoD, and delay parameters in [44]. A tensor-based channel estimator was derived in [45] to extract the angle and delay parameters using a two-plane twin RIS structure, followed by a location mapping. A path parameter estimation approach that exploits the MS position likelihood was designed in [46]. The received signal strength approach is utilized for localization in a RIS-aided system in [47]. Multiple RISs are employed to localize and track the users based on the Bayesian approach and message passing algorithm in [48].

The work on mmWave systems and RISs often assumes that the hardware behaves ideally. However, hardware impairments have significant effects on system modeling. For instance, mutual coupling-aware end-to-end RIS channel models are developed in [49], [50]. The work in [51] shows the degradation in channel estimation accuracy if the mutual coupling at the RIS is not considered. Although this work indicates that the degradation is more prominent if the spacing between the RIS elements is small, the degradation increases significantly when the number of elements of the RIS gets larger, even in the case of large spacing between the elements. Mutual coupling and other hardware impairments (e.g., phase/gain errors for the arrays) create a mismatch in the assumed channel model for channel estimation methods. Hence, measurement and calibration of the arrays is a necessary stage for practical applications [52]–[56]. Calibration of a RIS under mutual coupling in an experimental setup is considered in [52]. Calibration of arrays in the presence of mutual coupling and gain/phase errors is studied during the DoA estimation in [53]. Dictionary learning-based joint calibration of transmit and receive arrays of a point-to-point mmWave MIMO system is proposed in [54]. In addition to the hardware impairments, the beam squint effect is compensated with dictionary learning algorithms in [55], [56]. Furthermore, dictionary learning is used for adapting to the cell characteristics during channel estimation in [57]. In this paper, we follow a similar dictionary learning approach to calibrate the user array.

When it comes to channel estimation and localization in RIS-assisted mmWave systems, recent literature has proposed several efficient solutions [27]–[39], [44]–[48]. However, some of these solutions make unrealistic assumptions and neglect important effects that are present in real-world systems. For instance, hardware impairments such as mutual coupling and gain/phase errors are not considered in practical channel estimation and localization algorithms. Furthermore, certain studies rely on single-antenna or fully digital multiantenna architectures at MSs [27], [29]–[34], [37]–[39], [46]–[48], instead of a hybrid analog-digital architecture, which is the commonly accepted solution for mmWave MIMO if high-resolution data converters are considered. Additionally, some papers assume either a blocked or perfectly known direct link between the BS and the MS, limiting the estimation to the BS-RIS-MS channel [27], [28], [30], [32]–[35], [37]–[39], [44]–[46], [48]. Moreover, filtering effects of the pulse shaping function at the transmitter and the receiver are disregarded in all the related work except [29]. Most prior work on RIS-aided channel estimation and localization ignores filtering effects [58], which does not hold in practical systems. These assumptions lead to an artificial enhancement of delay sparsity, while realistic pulse shaping and filtering in the discrete-time equivalent channel model leads to a leakage in neighboring delay taps [59], [60]. If recovering the delays of the paths is not necessary, which is the case in many communication applications, frequency-domain channel estimation can be performed by exploiting only the angular sparsity [61]. However, delay information is crucial for localization applications; thus, sparse methods that consider pulse shaping and filtering effects are necessary. Note that the tensor decomposition-based and the ESPRIT-based channel estimation methods cannot handle delay recovery in the presence of pulse shaping and filtering [45], [62], [63] since the delay no longer appears as the exponent in a pure exponential, which is required for ESPRIT to operate. Lastly, previous work on RIS-aided localization does not account for the impact of the clock offset between the transmitter and the receiver. Although the clock offset

does not affect communication systems, it significantly affects the localization performance since the absolute delays of the paths cannot be measured. Most works in the literature assume perfect synchronization and, thus, perform the localization with angular and delay information of only the LoS path. These assumptions oversimplify the channel estimation and localization tasks, which can result in unreliable and inaccurate results in practical systems.

B. Contributions

In this paper, we propose a novel approach for joint channel estimation and localization in RIS-aided frequency-selective mmWave MIMO systems, taking into account the impact of hardware impairments, such as mutual coupling and gain/phase errors at the BS, RIS, and MS. We build our solution upon our initial work that does not consider the hardware impairments [1]. We consider a fully connected hybrid analog-digital architecture at the BS and the MS, that are not perfectly synchronized in time, which makes the system realistic but adds to the complexity of the problem. To develop our approach, we first establish a composite channel model incorporating two potential components: the BS-MS and the BS-RIS-MS channels, whose presence depend on the channel realization. We then investigate cases where one or both components are available. To simplify the channel model, we make a fundamental assumption that the DoA and the DoD associated with the LoS path of the BS-RIS channel are known, as the locations of the BS and the RIS are fixed [1], [29], [33], [64]. We consider the remaining paths of the BS-RIS channel to be interference. Thus, the BS-RIS-MS channel is simplified to a RIS-MS channel, which is similar to the BS-MS channel. In addition, we assume that the BS and the RIS hardware impairments are known and that both devices have been calibrated in previous stages. Thus, we propose a dictionary learning solution to calibrate the MS array, while noting that the same approach can be used to calibrate the BS array and the RIS. Note that the clock offset between the BS and the MS, which can be seen as a hardware impairment, is still unknown. These assumptions are reasonable, particularly for indoor scenarios, which are the focus of this paper. Considering the described system model, our contributions can be summarized as follows:

- We present a novel approach for solving the composite channel estimation problem in RIS-aided frequency-selective mmWave MIMO systems. The problem is formulated as a sparse recovery problem with independent dictionaries for the angle and delay domains while accounting for the effects of pulse shaping. To address this problem, we propose a modified version of the multidimensional orthogonal matching pursuit (MOMP) algorithm [14], [15] that jointly estimates the BS-MS and the RIS-MS channels. This approach enables low complexity sparse channel estimation for large arrays and RISs, which prevent the use of a single, large dictionary.
- To further reduce the complexity of the channel estimation problem, we propose using the MOMP algorithm by omitting the estimation of the DoA at the MS. Instead, we show that path gains and DoAs can be recovered afterward using the parameters estimated by the MOMP algorithm.
- We develop a novel dictionary learning problem based on the DoA recovery stage of our proposed channel estimation method. The dictionary learning algorithm used in this paper updates the DoA dictionary, accounting for the hardware impairments of the MS array. This approach enables joint channel estimation and user array calibration, improving the accuracy of the localization and channel estimation results.
- We introduce a versatile framework for 3D localization that can be applied to either the BS-MS or the BS-RIS-MS links, as well as to the LoS paths of both links. For the one-link case, we exploit the characteristics of the NLoS paths in indoor settings, whereas, in the two-link case, we utilize the equality satisfied by the LoS paths of both links to solve for the unknown clock offset. The location of the MS can then be determined by using the estimated clock offset and the parameters of the LoS path of either the BS-MS or the RIS-MS channel.
- We utilize ray tracing to generate a channel dataset for a RIS-aided mmWave communication system in an indoor factory environment. The simulation results obtained using the generated dataset demonstrate the improvement in the localization accuracy when the RIS-MS channel is leveraged instead of the BS-MS channel for localization. Furthermore, we show that sub-meter accuracy can be achieved when the LoS paths of both the BS-MS and the RIS-MS channels are utilized. The channel dataset and the code used to obtain the simulation results are available to the research community in [65] and [66], respectively.

Notation: The following notation is used throughout the paper: Lowercase x denotes a scalar, bold lowercase $\mathbf{x}$ denotes a column vector, bold uppercase $\mathbf{X}$ denotes a matrix (or a tensor), and $\mathcal{X}$ represents a set. Moreover, the conjugate, transpose, conjugate transpose, and pseudo-inverse of the matrix $\mathbf{X}$ are denoted by $\mathbf{X}^*$, $\mathbf{X}^{\mathrm{T}}$, $\mathbf{X}^{\mathrm{H}}$, and $\mathbf{X}^{\dagger}$, respectively. The Euclidean and Frobenius norms of the vector $\mathbf{x}$ and the matrix $\mathbf{X}$ are represented by $\|\mathbf{x}\|$ and $\|\mathbf{X}\|_{\mathrm{F}}$, respectively. The expectation operator is denoted by $\mathbb{E}[\cdot]$. The n-th element of the vector $\mathbf{x}$ is denoted by $[\mathbf{x}]_n$, and the (n,m)-th entry of the matrix $\mathbf{X}$ is represented by $[\mathbf{X}]_{n,m}$. Furthermore, $[\mathbf{X}]_{n,:}$ and $[\mathbf{X}]_{:,m}$ denote the n-th row and m-th column of the matrix $\mathbf{X}$, respectively. Multi-index notation $\mathbf{n} = (n_1, \ldots, n_N)$ is used to refer to the entries of the N-dimensional tensor $\mathbf{X}$ as $[\mathbf{X}]_{\mathbf{n}} = [\mathbf{X}]_{n_1,\ldots,n_N}$. The identity matrix is represented by $\mathbf{I}$, and the vector with all zeros is denoted by $\mathbf{0}$. The Kronecker product of the matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\mathbf{X} \otimes \mathbf{Y}$.

II. System and Channel Model

We consider a frequency-selective mmWave system that incorporates a RIS between the BS and the MS to enhance the communication and sensing performance. A visual representation of the RIS-aided joint communication and sensing system is given in Fig. 1. The BS, RIS, and MS are equipped

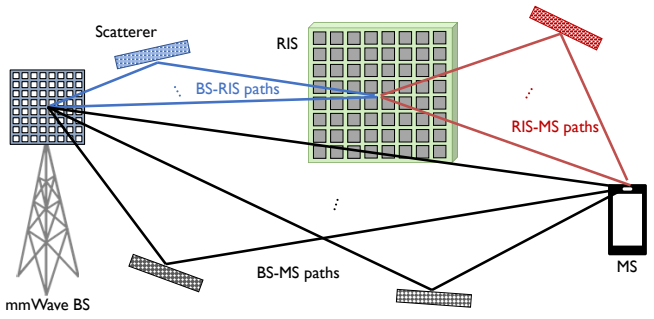


Fig. 1. Visual representation of the considered RIS-aided joint localization and communication system.

with uniform planar arrays (UPAs), where the number of antennas (passive elements for the RIS) are denoted by N_{B}, N_{R}, and N_{M}, respectively. A fully connected hybrid analog-digital architecture is employed at the BS and the MS with $N_{\mathrm{RF,B}}$ and $N_{\mathrm{RF,M}}$ radio frequency (RF) chains, respectively.

We are interested in the initial access stage, where the BS transmits the training frames and the MS sounds the channel with the aid of a RIS using a set of M_{B} training transmit configurations and M_{M} training combiners. A transmit configuration includes a particular choice of the pilot signal, the training precoder, and the phase shifts for the RIS. To measure the channel response for all possible combinations of transmit configurations and combiners, the BS transmits $M_{\mathrm{B}} \times M_{\mathrm{M}}$ training frames, each containing N_{s} streams of length-N training sequences. From the received signals corresponding to the training frames, directly from the BS and/or through the RIS, the MS estimates the downlink channel and its own location. The proposed training architecture can be easily extended to the multiuser case. Since we focus on downlink channel estimation, and the transmit configurations are independent of the MS locations and channels, the same channel estimation and localization procedure can be concurrently performed at different MSs. Furthermore, uplink channel estimation at the BS is also possible, even in the multiuser case, if orthogonal pilot sequences are assigned to MSs.

We set the number of streams to $N_{\mathrm{s}} = N_{\mathrm{RF,B}}$ during the training. The training precoder for the m_{B}-th training transmit configuration, $m_{\mathrm{B}} = 1, \ldots, M_{\mathrm{B}}$, is represented by $\mathrm{F}_{m_{\mathrm{B}}} = \mathrm{F}_{\mathrm{RF},m_{\mathrm{B}}}\mathrm{F}_{\mathrm{BB},m_{\mathrm{B}}} \in \mathbb{C}^{N_{\mathrm{B}} \times N_{\mathrm{RF,B}}}$, with analog precoder $\mathrm{F}_{\mathrm{RF},m_{\mathrm{B}}} \in \mathbb{C}^{N_{\mathrm{B}} \times N_{\mathrm{RF,B}}}$ and baseband precoder $\mathrm{F}_{\mathrm{BB},m_{\mathrm{B}}} \in \mathbb{C}^{N_{\mathrm{RF,B}} \times N_{\mathrm{RF,B}}}$. Analogously, the m_{M}-th training combiner, $m_{\mathrm{M}} = 1, \ldots, M_{\mathrm{M}}$, is denoted by $\mathbf{W}_{m_{\mathrm{M}}} = \mathbf{W}_{\mathrm{RF},m_{\mathrm{M}}}\mathbf{W}_{\mathrm{BB},m_{\mathrm{M}}} \in \mathbb{C}^{N_{\mathrm{M}} \times N_{\mathrm{RF,M}}}$ with analog combiner $\mathbf{W}_{\mathrm{RF},m_{\mathrm{M}}} \in \mathbb{C}^{N_{\mathrm{M}} \times N_{\mathrm{RF,M}}}$ and baseband combiner $\mathbf{W}_{\mathrm{BB},m_{\mathrm{M}}} \in \mathbb{C}^{N_{\mathrm{RF,M}} \times N_{\mathrm{RF,M}}}$. The training symbol vector at the n-th time instance transmitted with the m_{B}-th training configuration is denoted by $\mathbf{s}_{m_{\mathrm{B}}}[n] \in \mathbb{C}^{N_{\mathrm{RF,B}}}$, satisfying $\mathbb{E}[\mathbf{s}_{m_{\mathrm{B}}}[n]\mathbf{s}_{m_{\mathrm{B}}}[n]^*] = \frac{1}{N_{\mathrm{RF,B}}}\mathbf{I}_{N_{\mathrm{RF,B}}}$. To mitigate the intersymbol interference, a zero-prefix with length $D-1$ is added at the beginning of each frame, where D is the delay tap length.

The training signal generated by the BS is sent through the channel and can reach the MS directly or via the RIS. Thus, the channel model can be written in terms of two components: the BS-MS channel and the BS-RIS-MS channel. Note that the RIS employs a set of different phase configurations during training, which leads to different BS-RIS-MS channel matrices for different transmissions. The overall channel matrix observed while transmitting with the m_{B}-th configuration for the d-th delay tap, $d = 0, \ldots, D-1$, can be written as

$$\mathbf{H}_d^{(m_{\mathrm{B}})} = \mathbf{H}_{\mathrm{BM},d} + \mathbf{H}_{\mathrm{BRM},d}^{(m_{\mathrm{B}})}, \tag{1}$$

where $\mathbf{H}_{\mathrm{BM},d} \in \mathbb{C}^{N_{\mathrm{M}} \times N_{\mathrm{B}}}$ is the channel matrix of the BS-MS link and $\mathbf{H}_{\mathrm{BRM},d}^{(m_{\mathrm{B}})} \in \mathbb{C}^{N_{\mathrm{M}} \times N_{\mathrm{B}}}$ is the cascade channel matrix of the BS-RIS-MS link corresponding to the m_{B}-th training transmit configuration. Leveraging the geometric channel model with L paths, the BS-MS channel is written as

$$\mathbf{H}_{\mathrm{BM},d} = \sum_{l=1}^{L} \alpha_{\mathrm{BM},l}\mathbf{M}_{\mathrm{M}}\mathbf{\Gamma}_{\mathrm{M}}\mathbf{a}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{BM},l}) \times \left(\mathbf{M}_{\mathrm{B}}\mathbf{\Gamma}_{\mathrm{B}}\mathbf{a}_{\mathrm{B}}(\boldsymbol{\phi}_{\mathrm{BM},l})\right)^{\mathrm{H}} p(dT_{\mathrm{s}} + t_0 - \tau_{\mathrm{BM},l}), \tag{2}$$

where each path l is described by complex gain $\alpha_{\mathrm{BM},l}$, DoA $\boldsymbol{\theta}_{\mathrm{BM},l}$, DoD $\boldsymbol{\phi}_{\mathrm{BM},l}$, and delay $\tau_{\mathrm{BM},l}$; the sampling period is denoted by T_{s}; the time between the beginning of the transmission and the beginning of the reception, i.e., clock offset, is t_0; the time response of the pulse shaping function, which includes the effects of the transmitted signal and the filtering operations at the BS and the MS, is represented by $p(t)$; the array response vectors of the BS and the MS are denoted by $\mathbf{a}_{\mathrm{B}}(\boldsymbol{\phi}) \in \mathbb{C}^{N_{\mathrm{B}}}$ and $\mathbf{a}_{\mathrm{M}}(\boldsymbol{\theta}) \in \mathbb{C}^{N_{\mathrm{M}}}$ for DoD $\boldsymbol{\phi}$ and DoA $\boldsymbol{\theta}$, respectively. We consider 3D direction vectors for the DoD and the DoA instead of polar coordinates, which will be useful in the proceeding operations. Any direction vector $\boldsymbol{\theta}$ can be expressed as $\boldsymbol{\theta} = [\theta_{\mathrm{x}}, \theta_{\mathrm{y}}, \theta_{\mathrm{z}}]^{\mathrm{T}}$ with $\|\boldsymbol{\theta}\| = 1$. If a UPA that has $N_{\mathrm{x}} \times N_{\mathrm{y}}$ antennas with half wavelength spacing is placed on the xy-plane, its array response vector can be written as $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{a}_{\mathrm{x}}(\theta_{\mathrm{x}}) \otimes \mathbf{a}_{\mathrm{y}}(\theta_{\mathrm{y}})$ with expressions $[\mathbf{a}_{\mathrm{x}}(\theta_{\mathrm{x}})]_{n_{\mathrm{x}}} = e^{-j\pi(n_{\mathrm{x}}-1)\theta_{\mathrm{x}}}$ and $[\mathbf{a}_{\mathrm{y}}(\theta_{\mathrm{y}})]_{n_{\mathrm{y}}} = e^{-j\pi(n_{\mathrm{y}}-1)\theta_{\mathrm{y}}}$ for $n_{\mathrm{x}} = 1, \ldots, N_{\mathrm{x}}$ and $n_{\mathrm{y}} = 1, \ldots, N_{\mathrm{y}}$. Moreover, the gain and phase errors at the array elements are modeled by diagonal matrices $\mathbf{\Gamma}_{\mathrm{B}} \in \mathbb{C}^{N_{\mathrm{B}} \times N_{\mathrm{B}}}$ for the BS and $\mathbf{\Gamma}_{\mathrm{M}} \in \mathbb{C}^{N_{\mathrm{M}} \times N_{\mathrm{M}}}$ for the MS. The mutual coupling matrices for the arrays at the BS and the MS are denoted by $\mathbf{M}_{\mathrm{B}} \in \mathbb{C}^{N_{\mathrm{B}} \times N_{\mathrm{B}}}$ and $\mathbf{M}_{\mathrm{M}} \in \mathbb{C}^{N_{\mathrm{M}} \times N_{\mathrm{M}}}$ in (2), respectively [54], [56]. The off-diagonal entries of the mutual coupling matrices model the coupling between different antenna elements, which is usually omitted in parametric channel estimation methods. This leads to a mismatch in the model used for channel estimation, resulting in erroneous parameter estimates. We assume the perfect knowledge of the hardware impairments at the BS, i.e., $\mathbf{M}_{\mathrm{B}}$ and $\mathbf{\Gamma}_{\mathrm{B}}$, while they can be estimated by using the array calibration solution proposed for the MS in Section IV. In this paper, we are interested in estimating the channel and the impairments at the MS. Hence, we rewrite (2) as $\mathbf{H}_{\mathrm{BM},d} = \tilde{\mathbf{H}}_{\mathrm{BM},d}\mathbf{\Gamma}_{\mathrm{B}}^{\mathrm{H}}\mathbf{M}_{\mathrm{B}}^{\mathrm{H}}$ where $\tilde{\mathbf{H}}_{\mathrm{BM},d} \in \mathbb{C}^{N_{\mathrm{M}} \times N_{\mathrm{B}}}$ is the matrix that includes the parameters to be estimated defined as

$$\tilde{\mathbf{H}}_{\mathrm{BM},d} = \sum_{l=1}^{L} \alpha_{\mathrm{BM},l}\tilde{\mathbf{a}}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{BM},l})\mathbf{a}_{\mathrm{B}}(\boldsymbol{\phi}_{\mathrm{BM},l}^{\mathrm{H}})p(dT_{\mathrm{s}} + t_0 - \tau_{\mathrm{BM},l}), \tag{3}$$

where the array response vector at the MS that includes

$$\mathbf{H}_{\mathrm{BRM},d}^{(m_\mathrm{B})} = \sum_{q=1}^{Q}\sum_{p=1}^{P} \alpha_{\mathrm{RM},q}\alpha_{\mathrm{BR},p}\tilde{\mathbf{a}}_\mathrm{M}(\boldsymbol{\theta}_{\mathrm{RM},q}) \left(\mathbf{M}_\mathrm{R}\boldsymbol{\Gamma}_\mathrm{R}\mathbf{a}_\mathrm{R}(\boldsymbol{\phi}_{\mathrm{RM},q})\right)^\mathrm{H} \times \boldsymbol{\Omega}^{(m_\mathrm{B})} \times \mathbf{M}_\mathrm{R}\boldsymbol{\Gamma}_\mathrm{R}\mathbf{a}_\mathrm{R}(\boldsymbol{\theta}_{\mathrm{BR},p}) \left(\mathbf{M}_\mathrm{B}\boldsymbol{\Gamma}_\mathrm{B}\mathbf{a}_\mathrm{B}(\boldsymbol{\phi}_{\mathrm{BR},p})\right)^\mathrm{H} p(dT_\mathrm{s} + t_0 - \tau_{\mathrm{BR},p} + \tau_{\mathrm{RM},q}), \quad (4)$$

$$\mathbf{H}_{\mathrm{BRM},d}^{(m_\mathrm{B})} \approx \underbrace{\sum_{q=1}^{Q} \tilde{\alpha}_{\mathrm{RM},q}\tilde{\mathbf{a}}_\mathrm{M}(\boldsymbol{\theta}_{\mathrm{RM},q})\mathbf{a}_\mathrm{R}^\mathrm{H}(\boldsymbol{\phi}_{\mathrm{RM},q})p(dT_\mathrm{s} + t_0 - \tau_{\mathrm{BR},1} - \tau_{\mathrm{RM},q})}_{\triangleq \mathbf{H}_{\mathrm{RM},d}} \times \underbrace{\boldsymbol{\Gamma}_\mathrm{R}^\mathrm{H}\mathbf{M}_\mathrm{R}^\mathrm{H}\boldsymbol{\Omega}^{(m_\mathrm{B})}}_{\triangleq \bar{\boldsymbol{\Omega}}^{(m_\mathrm{B})}} \times \underbrace{\mathbf{M}_\mathrm{R}\boldsymbol{\Gamma}_\mathrm{R}\mathbf{a}_\mathrm{R}(\boldsymbol{\theta}_{\mathrm{BR},1})\mathbf{a}_\mathrm{B}^\mathrm{H}(\boldsymbol{\phi}_{\mathrm{BR},1})\boldsymbol{\Gamma}_\mathrm{B}^\mathrm{H}\mathbf{M}_\mathrm{B}^\mathrm{H}}_{\triangleq \mathbf{H}_{\mathrm{BR}}}, \quad (5)$$

the hardware impairments is denoted by $\tilde{\mathbf{a}}_\mathrm{M}(\boldsymbol{\theta}_{\mathrm{BM},l}) = \mathbf{M}_\mathrm{M}\boldsymbol{\Gamma}_\mathrm{M}\mathbf{a}_\mathrm{M}(\boldsymbol{\theta}_{\mathrm{BM},l}) \in \mathbb{C}^{N_\mathrm{M}}$. On the other hand, the second contribution to the channel in (1) that represents the BS-RIS-MS link is given in (4) at the top of this page, where the diagonal phase reflection matrix at the RIS for the m_B-th training transmit configuration is represented by $\boldsymbol{\Omega}^{(m_\mathrm{B})} = \mathrm{diag}(\boldsymbol{\omega}^{(m_\mathrm{B})}) \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{R}}$, with a phase reflection vector $\boldsymbol{\omega}^{(m_\mathrm{B})} \in \mathbb{C}^{N_\mathrm{R}}$ which has unit modulus entries [18]–[21]. In (4), the number of paths of the BS-RIS and the RIS-MS channels are denoted by P and Q, respectively. The channel parameters (i.e., complex gain, DoA, DoD, and delay) for each path are defined similarly to those of the BS-MS channel. The array response vector for the RIS is denoted by $\mathbf{a}_\mathrm{R}(\cdot) \in \mathbb{C}^{N_\mathrm{R}}$. The mutual coupling and the gain/phase error matrices for the RIS are represented by $\mathbf{M}_\mathrm{R} \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{R}}$ and $\boldsymbol{\Gamma}_\mathrm{R} \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{R}}$, respectively. Note that we model the mutual coupling as a linear effect [67], instead of including it in an end-to-end channel model which accounts for the coupling between the RIS and the transmit/receive arrays [49]–[51], since we consider a far-field mmWave system that presents sparse channels. Although the RIS-MS link has Q paths, the MS receives P signals with different delays at the DoA of the q-th path, as shown in Fig. 1. Estimating the BS-RIS and the RIS-MS channel parameters from (4) is not straightforward since it is hard to decouple the two channels as the delays are embedded in the pulse shaping function.

Several options exist to simplify the cascade channel expression given in (4). In mmWave systems, it is known that the LoS path is much stronger than the NLoS paths. In this work, it is assumed that the locations of the BS and the RIS are known. Therefore, it is plausible to arrange the locations such that a strong LoS path exists for the BS-RIS link [29], [33]. Relying on this assumption, let us only consider $p = 1$, which is the LoS path of the BS-RIS link, and treat others (i.e., NLoS paths) as interference. In addition, hardware impairments at the RIS, i.e., $\mathbf{M}_\mathrm{R}$ and $\boldsymbol{\Gamma}_\mathrm{R}$, are assumed to be known similar to those of the BS. If the NLoS paths of the BS-RIS link are assumed to be negligible, the cascade channel in (4) can be simplified to the expression given in (5) at the top of this page, where the parameters related to the LoS path of the BS-RIS link (i.e., $\boldsymbol{\theta}_{\mathrm{BR},1}$, $\boldsymbol{\phi}_{\mathrm{BR},1}$ and $\tau_{\mathrm{BR},1}$) are assumed to be perfectly known as the locations of the BS and the RIS are known a priori. The parameters related to the paths of the RIS-MS link that need to be estimated are encompassed in the equivalent channel matrix represented by $\mathbf{H}_{\mathrm{RM},d} \in \mathbb{C}^{N_\mathrm{M}\times N_\mathrm{R}}$ in (5), which has the same form as $\tilde{\mathbf{H}}_{\mathrm{BM},d}$. Furthermore, the equivalent LoS channel for the BS-RIS link is denoted by $\mathbf{H}_{\mathrm{BR}} \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{B}}$, which is known and independent of the delay taps. The unknown parameter of the LoS path of the BS-RIS link (i.e., complex gain $\alpha_{\mathrm{BR},1}$) is included in the equivalent channel of the RIS-MS link by $\tilde{\alpha}_{\mathrm{RM},q} = \alpha_{\mathrm{RM},q}\alpha_{\mathrm{BR},1}$.

Given this definition of the channel, the received signal at the MS for the n-th time instance and the $(m_\mathrm{B}, m_\mathrm{M})$-th training configuration can be written as

$$\mathbf{y}_{m_\mathrm{B},m_\mathrm{M}}[n] = \sqrt{P_\mathrm{t}}\mathbf{W}_{m_\mathrm{M}}^\mathrm{H} \sum_{d=0}^{D-1} \mathbf{H}_d^{(m_\mathrm{B})}\mathrm{F}_{m_\mathrm{B}}\mathbf{s}_{m_\mathrm{B}}[n-d] + \mathbf{W}_{m_\mathrm{M}}^\mathrm{H}\mathbf{v}_{m_\mathrm{B},m_\mathrm{M}}[n], \quad (6)$$

for $n = 1, \ldots, N$, where P_t is the transmit power, and $\mathbf{v}_{m_\mathrm{B},m_\mathrm{M}}[n] \in \mathbb{C}^{N_\mathrm{M}}$ is the noise vector for the $(m_\mathrm{B}, m_\mathrm{M})$-th training configuration at the n-th time instance, with independent and identically distributed entries obeying $\mathcal{NC}(0, \sigma^2)$. Note that the noise at the output of the combiner in (6) becomes correlated if the combiner is not orthogonal.

III. MOMP-based Compressed Channel Estimation

In this section, we formulate the channel estimation problem as a sparse recovery problem that can be solved with the MOMP algorithm exploiting independent dictionaries in the angular and delay domain [15]. This algorithm enables a low complexity sparse recovery solution in a scenario where the introduction of the RIS increases the dimensions of the sensing matrix, and the final localization application requires high-resolution dictionaries. The MOMP algorithm can recover a multidimensional sparse signal based on a set of observations provided that it can be represented by a product of projections on a given set of sparsifying dictionaries. Exploiting the results in [15] is not straightforward, however, since the received signal is a combination of paths coming from two sources, namely, the BS-MS link and the BS-RIS-MS link, for the RIS-aided scenario. To overcome this difficulty, we present a procedure that exploits two sets of dictionaries to model the two parts in the composite channel.

A. Sparse Channel Representations

The first step to solving the channel estimation problem is to find a sparse representation for the channels. We need preliminary definitions and explanations before introducing the sparsifying dictionaries used for the sparse representations. First, the delay domain in the discrete equivalent channel

model can be represented by a vector $\mathbf{a}_{\mathrm{T}}(\tau) \in \mathbb{C}^D$ such that $[\mathbf{a}_{\mathrm{T}}(\tau)]_d = p(dT_{\mathrm{s}} - \tau)$ for a given delay τ, where the function $p(t)$ includes the effect of pulse shaping and the different filtering stages at the RF front end. To represent the angular domain, we can define two dictionaries for the array responses at the BS, RIS, and MS since, as explained before, a UPA has an array response vector that can be written as the Kronecker product of two array responses. Consider the UPAs at the BS, RIS, and MS with $N_{\mathrm{B}} = N_{\mathrm{B,x}} \times N_{\mathrm{B,y}}$, $N_{\mathrm{R}} = N_{\mathrm{R,x}} \times N_{\mathrm{R,y}}$ and $N_{\mathrm{M}} = N_{\mathrm{M,x}} \times N_{\mathrm{M,y}}$ elements, respectively. Then, the array response at the BS for the incident direction $\boldsymbol{\phi}$ can be expressed as $\mathbf{a}_{\mathrm{B}}(\boldsymbol{\phi}) = \mathbf{a}_{\mathrm{B,x}}(\phi_{\mathrm{x}}) \otimes \mathbf{a}_{\mathrm{B,y}}(\phi_{\mathrm{y}})$, with $\mathbf{a}_{\mathrm{B,x}}(\phi_{\mathrm{x}}) \in \mathbb{C}^{N_{\mathrm{B,x}}}$ and $\mathbf{a}_{\mathrm{B,y}}(\phi_{\mathrm{y}}) \in \mathbb{C}^{N_{\mathrm{B,y}}}$. The array responses at the RIS and the MS can be expressed similarly. First, let us define the dictionaries for the BS-MS channel given in (3). In this case, we can define $N_{\mathrm{D}} = 4$ dictionaries $\boldsymbol{\Psi}_{\mathrm{BM},k} \in \mathbb{C}^{N^{\mathrm{s}}_{\mathrm{BM},k} \times N^{\mathrm{a}}_{\mathrm{BM},k}}$, $k = 1, \ldots, 4$ that sparsify the BS-MS channel as

$$\begin{aligned}
\boldsymbol{\Psi}_{\mathrm{BM},1} &= [\mathbf{a}_{\mathrm{T}}(\overline{\tau}_1), \cdots, \mathbf{a}_{\mathrm{T}}(\overline{\tau}_{N^{\mathrm{a}}_{\mathrm{BM},1}})], \\
\boldsymbol{\Psi}_{\mathrm{BM},2} &= \left[\mathbf{a}^*_{\mathrm{B,x}}(\overline{\phi}^{\mathrm{x}}_{\mathrm{BM},1}), \cdots, \mathbf{a}^*_{\mathrm{B,x}}(\overline{\phi}^{\mathrm{x}}_{\mathrm{BM},N^{\mathrm{a}}_{\mathrm{BM},2}})\right], \\
\boldsymbol{\Psi}_{\mathrm{BM},3} &= \left[\mathbf{a}^*_{\mathrm{B,y}}(\overline{\phi}^{\mathrm{y}}_{\mathrm{BM},1}), \cdots, \mathbf{a}^*_{\mathrm{B,y}}(\overline{\phi}^{\mathrm{y}}_{\mathrm{BM},N^{\mathrm{a}}_{\mathrm{BM},3}})\right], \\
\boldsymbol{\Psi}_{\mathrm{BM},4} &= \left[\tilde{\mathbf{a}}_{\mathrm{M}}(\overline{\boldsymbol{\theta}}_{\mathrm{BM},1}), \cdots, \tilde{\mathbf{a}}_{\mathrm{M}}(\overline{\boldsymbol{\theta}}_{\mathrm{BM},N^{\mathrm{a}}_{\mathrm{BM},4}})\right],
\end{aligned} \tag{7}$$

where the directions and the delays are discretized using some predefined resolutions, i.e., a given number of atoms for each dictionary, $N^{\mathrm{a}}_{\mathrm{BM},k}$, for $k = 1, \ldots, N_{\mathrm{D}}$, such that the dictionaries $\left\{\overline{\tau}_1, \ldots, \overline{\tau}_{N^{\mathrm{a}}_{\mathrm{BM},1}}\right\}$, $\left\{\overline{\phi}^{\mathrm{x}}_{\mathrm{BM},1}, \ldots, \overline{\phi}^{\mathrm{x}}_{\mathrm{BM},N^{\mathrm{a}}_{\mathrm{BM},2}}\right\}$, $\left\{\overline{\phi}^{\mathrm{y}}_{\mathrm{BM},1}, \ldots, \overline{\phi}^{\mathrm{y}}_{\mathrm{BM},N^{\mathrm{a}}_{\mathrm{BM},3}}\right\}$ and $\left\{\overline{\boldsymbol{\theta}}_{\mathrm{BM},1}, \ldots, \overline{\boldsymbol{\theta}}_{\mathrm{BM},N^{\mathrm{a}}_{\mathrm{BM},4}}\right\}$ correspond to the discrete forms of $\tau - t_0$, $\phi_{\mathrm{BM,x}}$, $\phi_{\mathrm{BM,y}}$ and $\boldsymbol{\theta}_{\mathrm{BM}}$, respectively. Furthermore, the sizes of the atoms in each dictionary are given as $N^{\mathrm{s}}_{\mathrm{BM},1} = D$, $N^{\mathrm{s}}_{\mathrm{BM},2} = N_{\mathrm{B,x}}$, $N^{\mathrm{s}}_{\mathrm{BM},3} = N_{\mathrm{B,y}}$ and $N^{\mathrm{s}}_{\mathrm{BM},4} = N_{\mathrm{M}}$. For ease of notation, we use the indices $i_k = 1, \ldots, N^{\mathrm{s}}_{\mathrm{BM},k}$ and $j_k = 1, \ldots, N^{\mathrm{a}}_{\mathrm{BM},k}$ to point to the entries in the k-th dictionary for $k = 1, \ldots, N_{\mathrm{D}}$. In addition, we utilize the multi-index notation $\mathbf{i} \in \mathcal{I}_{\mathrm{BM}} = \{(i_1, \ldots, i_{N_{\mathrm{D}}}), i_k = 1, \ldots, N^{\mathrm{s}}_{\mathrm{BM},k}\}$ and $\mathbf{j} \in \mathcal{J}_{\mathrm{BM}} = \{(j_1, \ldots, j_{N_{\mathrm{D}}}), j_k = 1, \ldots, N^{\mathrm{a}}_{\mathrm{BM},k}\}$ to concurrently point to the entries in multiple dictionaries. Note that the Kronecker product form of the array response of the MS is not exploited in $\boldsymbol{\Psi}_{\mathrm{BM},4}$ since the overall hardware impairment matrix $\mathbf{M}_{\mathrm{M}}\boldsymbol{\Gamma}_{\mathrm{M}}$ does not necessarily have a Kronecker product form. However, this is not a limitation, as the MSs typically contain fewer elements than the BS and the RIS.

The dictionaries defined in (7) quantize the directions and the delays, which results in a mismatch with the actual values of the channel paths. Since the MOMP algorithm can operate with very high resolution, the quantization error can be neglected [15]. Therefore, we can define the sparse tensor $\mathbf{C}_{\mathrm{BM}} \in \mathbb{C}^{N^{\mathrm{a}}_{\mathrm{BM},1} \times \cdots \times N^{\mathrm{a}}_{\mathrm{BM},N_{\mathrm{D}}}}$ that contains the channel gains corresponding to specific atoms with multi-index $\mathbf{j} \in \mathcal{J}_{\mathrm{BM}}$ as

$$[\mathbf{C}_{\mathrm{BM}}]_{\mathbf{j}} = \begin{cases} \alpha_{\mathrm{BM},l} & \text{if } \begin{array}{rcl} \tau_{\mathrm{BM},l} &=& \overline{\tau}_{\mathrm{BM},j_1}, \\ \phi^{\mathrm{x}}_{\mathrm{BM},l} &=& \overline{\phi}^{\mathrm{x}}_{\mathrm{BM},j_2}, \\ \phi^{\mathrm{y}}_{\mathrm{BM},l} &=& \overline{\phi}^{\mathrm{y}}_{\mathrm{BM},j_3}, \\ \boldsymbol{\theta}_{\mathrm{BM},l} &=& \overline{\boldsymbol{\theta}}_{\mathrm{BM},j_4}, \end{array} \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The dictionaries and the sparse channel gain tensor are used to rewrite the entries of the BS-MS channel expression as

$$[\tilde{\mathbf{H}}_{\mathrm{BM},i_1}]_{i_4, i_2 N_{\mathrm{B,y}} + i_3} = \sum_{\mathbf{j} \in \mathcal{J}_{\mathrm{BM}}} \prod_{k=1}^{N_{\mathrm{D}}=4} [\boldsymbol{\Psi}_{\mathrm{BM},k}]_{i_k, j_k} [\mathbf{C}_{\mathrm{BM}}]_{\mathbf{j}}. \tag{9}$$

Analogously, for the RIS-MS channel $\mathbf{H}_{\mathrm{RM},d}$ introduced in (5), the dictionaries $\boldsymbol{\Psi}_{\mathrm{RM},k} \in \mathbb{C}^{N^{\mathrm{s}}_{\mathrm{RM},k} \times N^{\mathrm{a}}_{\mathrm{RM},k}}$ are defined as

$$\begin{aligned}
\boldsymbol{\Psi}_{\mathrm{RM},1} &= [\mathbf{a}_{\mathrm{T}}(\overline{\tau}_1), \cdots, \mathbf{a}_{\mathrm{T}}(\overline{\tau}_{N^{\mathrm{a}}_{\mathrm{RM},1}})], \\
\boldsymbol{\Psi}_{\mathrm{RM},2} &= \left[\mathbf{a}^*_{\mathrm{R,x}}(\overline{\phi}^{\mathrm{x}}_{\mathrm{RM},1}), \cdots, \mathbf{a}^*_{\mathrm{R,x}}(\overline{\phi}^{\mathrm{x}}_{\mathrm{RM},N^{\mathrm{a}}_{\mathrm{RM},2}})\right], \\
\boldsymbol{\Psi}_{\mathrm{RM},3} &= \left[\mathbf{a}^*_{\mathrm{R,y}}(\overline{\phi}^{\mathrm{y}}_{\mathrm{RM},1}), \cdots, \mathbf{a}^*_{\mathrm{R,y}}(\overline{\phi}^{\mathrm{y}}_{\mathrm{RM},N^{\mathrm{a}}_{\mathrm{RM},3}})\right], \\
\boldsymbol{\Psi}_{\mathrm{RM},4} &= \left[\tilde{\mathbf{a}}_{\mathrm{M}}(\overline{\boldsymbol{\theta}}_{\mathrm{RM},1}), \cdots, \tilde{\mathbf{a}}_{\mathrm{M}}(\overline{\boldsymbol{\theta}}_{\mathrm{RM},N^{\mathrm{a}}_{\mathrm{RM},4}})\right],
\end{aligned} \tag{10}$$

where the directions and the delays fall on a grid of possible values similar to the BS-MS channel. The number of atoms of each dictionary are denoted by $N^{\mathrm{a}}_{\mathrm{RM},k}$, $k = 1, \ldots, N_{\mathrm{D}}$, such that the dictionaries $\left\{\overline{\tau}_1, \ldots, \overline{\tau}_{N^{\mathrm{a}}_{\mathrm{RM},1}}\right\}$, $\left\{\overline{\phi}^{\mathrm{x}}_{\mathrm{RM},1}, \ldots, \overline{\phi}^{\mathrm{x}}_{\mathrm{RM},N^{\mathrm{a}}_{\mathrm{RM},2}}\right\}$, $\left\{\overline{\phi}^{\mathrm{y}}_{\mathrm{RM},1}, \ldots, \overline{\phi}^{\mathrm{y}}_{\mathrm{RM},N^{\mathrm{a}}_{\mathrm{RM},3}}\right\}$ and $\left\{\overline{\boldsymbol{\theta}}_{\mathrm{BM},1}, \ldots, \overline{\boldsymbol{\theta}}_{\mathrm{BM},N^{\mathrm{a}}_{\mathrm{BM},4}}\right\}$ correspond to the discrete forms of $\tau + \tau_{\mathrm{BR},1} - t_0$, $\phi_{\mathrm{RM,x}}$, $\phi_{\mathrm{RM,y}}$, $\theta_{\mathrm{RM,x}}$ and $\boldsymbol{\theta}_{\mathrm{BM}}$, respectively. The sizes of the atoms in each dictionary are given as $N^{\mathrm{s}}_{\mathrm{RM},1} = D$, $N^{\mathrm{s}}_{\mathrm{RM},2} = N_{\mathrm{R,x}}$, $N^{\mathrm{s}}_{\mathrm{RM},3} = N_{\mathrm{R,y}}$ and $N^{\mathrm{s}}_{\mathrm{RM},4} = N_{\mathrm{M}}$. The multi-index notations that are used for the RIS-MS channel are given as $\mathbf{i} \in \mathcal{I}_{\mathrm{RM}} = \{(i_1, \ldots, i_{N_{\mathrm{D}}}), i_k = 1, \ldots, N^{\mathrm{s}}_{\mathrm{RM},k}\}$ and $\mathbf{j} \in \mathcal{J}_{\mathrm{RM}} = \{(j_1, \ldots, j_{N_{\mathrm{D}}}), j_k = 1, \ldots, N^{\mathrm{a}}_{\mathrm{RM},k}\}$. Finally, we define the sparse channel gain tensor for the RIS-MS channel $\mathbf{C}_{\mathrm{RM}} \in \mathbb{C}^{N^{\mathrm{a}}_{\mathrm{RM},1} \times \cdots \times N^{\mathrm{a}}_{\mathrm{RM},N_{\mathrm{D}}}}$ as

$$[\mathbf{C}_{\mathrm{RM}}]_{\mathbf{j}} = \begin{cases} \tilde{\alpha}_{\mathrm{RM},q} & \text{if } \begin{array}{rcl} \tau_{\mathrm{RM},q} &=& \overline{\tau}_{\mathrm{RM},j_1}, \\ \phi^{\mathrm{x}}_{\mathrm{RM},q} &=& \overline{\phi}^{\mathrm{x}}_{\mathrm{RM},j_2}, \\ \phi^{\mathrm{y}}_{\mathrm{RM},q} &=& \overline{\phi}^{\mathrm{y}}_{\mathrm{RM},j_3}, \\ \boldsymbol{\theta}_{\mathrm{RM},l} &=& \overline{\boldsymbol{\theta}}_{\mathrm{RM},j_4}, \end{array} \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

The resultant sparse representation for the entries of the RIS-MS channel expression in (5) can be written as

$$[\mathbf{H}_{\mathrm{RM},i_1}]_{i_4, i_2 N_{\mathrm{R,y}} + i_3} = \sum_{\mathbf{j} \in \mathcal{J}_{\mathrm{RM}}} \prod_{k=1}^{N_{\mathrm{D}}=4} [\boldsymbol{\Psi}_{\mathrm{RM},k}]_{i_k, j_k} [\mathbf{C}_{\mathrm{RM}}]_{\mathbf{j}}. \tag{12}$$

The sparse channel representations in (9) and (12) will be used to express the observations of the received signals during training, which will be exploited by the sparse recovery algorithm developed in the following subsection.

B. Channel Estimation Exploiting Multidimensional Sparsifying Dictionaries

In this subsection, we will first derive the sparse recovery problem to be solved with MOMP to estimate the composite channel. To this end, we need to express the set of observations by using the sparse representation of the channels developed in the previous section. However, the MOMP algorithm requires the observation noise to be white [15]. Therefore, we will include a whitening stage before constructing the sparse recovery problem. To whiten the received signal in (6), we left multiply it by $\mathbf{L}^{-1}_{m_{\mathrm{M}}}$, which can be found from

the Cholesky decomposition of the noise correlation matrix, i.e., $\mathbf{L}_{m_\mathrm{M}}\mathbf{L}_{m_\mathrm{M}}^\mathrm{H} = \mathbf{W}_{m_\mathrm{M}}^\mathrm{H}\mathbf{W}_{m_\mathrm{M}}$. The whitened received signal $\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}[n] \in \mathbb{C}^{N_{\mathrm{RF,M}}}$, which is defined as $\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}[n] = \mathbf{L}_{m_\mathrm{M}}^{-1}\mathbf{y}_{m_\mathrm{B},m_\mathrm{M}}[n]$, can be written as

$$\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}[n] = \sqrt{P_\mathrm{t}}\bar{\mathbf{W}}_{m_\mathrm{M}}^\mathrm{H}\sum_{d=0}^{D-1}\mathbf{H}_d^{(m_\mathrm{B})}\mathrm{F}_{m_\mathrm{B}}\mathbf{s}_{m_\mathrm{B}}[n-d] + \bar{\mathbf{v}}_{m_\mathrm{B},m_\mathrm{M}}[n], \quad (13)$$

where $\bar{\mathbf{W}}_{m_\mathrm{M}} = \mathbf{W}_{m_\mathrm{M}}[\mathbf{L}_{m_\mathrm{M}}^{-1}]^\mathrm{H} \in \mathbb{C}^{N_{\mathrm{RF,M}}\times N_\mathrm{M}}$ is the whitened combiner. Moreover, $\bar{\mathbf{v}}_{m_\mathrm{B},m_\mathrm{M}}[n] = \bar{\mathbf{W}}_{m_\mathrm{M}}^\mathrm{H}\mathbf{v}_{m_\mathrm{B},m_\mathrm{M}}[n] \in \mathbb{C}^{N_{\mathrm{RF,M}}}$ is the noise after whitening. Using the overall channel expression in (1) and the simplified channel expression for the BS-RIS-MS link in (5), the whitened received signal in (13) can be rewritten as

$$\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}[n] = \sqrt{P_\mathrm{t}}\bar{\mathbf{W}}_{m_\mathrm{M}}^\mathrm{H}\sum_{d=0}^{D-1}\Big(\tilde{\mathbf{H}}_{\mathrm{BM},d}\mathbf{\Gamma}_\mathrm{B}^\mathrm{H}\mathbf{M}_\mathrm{B}^\mathrm{H} + \mathbf{H}_{\mathrm{RM},d}\bar{\mathbf{\Omega}}^{(m_\mathrm{B})}\mathbf{H}_\mathrm{BR}\Big)\mathrm{F}_{m_\mathrm{B}}\mathbf{s}_{m_\mathrm{B}}[n-d] + \bar{\mathbf{v}}_{m_\mathrm{B},m_\mathrm{M}}[n], \quad (14)$$

where the received signal for the BS-MS and the BS-RIS-MS links can be separated as $\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}[n] = \bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{BM}}[n] + \bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{RM}}[n] + \bar{\mathbf{v}}_{m_\mathrm{B},m_\mathrm{M}}[n]$. The noiseless received signals over the BS-MS and the BS-RIS-MS channels are denoted by $\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{BM}}[n] \in \mathbb{C}^{N_{\mathrm{RF,M}}}$ and $\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{RM}}[n] \in \mathbb{C}^{N_{\mathrm{RF,M}}}$, respectively. By defining the equivalent precoder for the BS-MS link as $\mathrm{F}_{m_\mathrm{B}}^{'} = \mathbf{\Gamma}_\mathrm{B}^\mathrm{H}\mathbf{M}_\mathrm{B}^\mathrm{H}\mathrm{F}_{m_\mathrm{B}} \in \mathbb{C}^{N_\mathrm{B}\times N_{\mathrm{RF,B}}}$ and for the BS-RIS-MS link as $\mathrm{F}_{m_\mathrm{B}}^{''} = \bar{\mathbf{\Omega}}^{(m_\mathrm{B})}\mathbf{H}_\mathrm{BR}\mathrm{F}_{m_\mathrm{B}} \in \mathbb{C}^{N_\mathrm{R}\times N_{\mathrm{RF,B}}}$, the noiseless received signals over the two links can be written as

$$\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{BM}}[n] = \sqrt{P_\mathrm{t}}\bar{\mathbf{W}}_{m_\mathrm{M}}^\mathrm{H}\sum_{d=0}^{D-1}\tilde{\mathbf{H}}_{\mathrm{BM},d}\mathrm{F}_{m_\mathrm{B}}^{'}\mathbf{s}_{m_\mathrm{B}}[n-d], \quad (15)$$

$$\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{RM}}[n] = \sqrt{P_\mathrm{t}}\bar{\mathbf{W}}_{m_\mathrm{M}}^\mathrm{H}\sum_{d=0}^{D-1}\mathbf{H}_{\mathrm{RM},d}\mathrm{F}_{m_\mathrm{B}}^{''}\mathbf{s}_{m_\mathrm{B}}[n-d]. \quad (16)$$

If we substitute the sparse representations of the BS-MS and the RIS-MS channels in (9) and (12) into the expressions above, we can rewrite them as

$$\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{BM}}[n] = \sqrt{P_\mathrm{t}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathrm{BM}}}[\bar{\mathbf{W}}_{m_\mathrm{M}}]_{i_4,:}^\mathrm{H}[\mathrm{F}_{m_\mathrm{B}}^{'}]_{i_2N_{\mathrm{B},y}+i_3,:}\mathbf{s}_{m_\mathrm{B}}[n-i_1] \times \sum_{\mathbf{j}\in\mathcal{J}_{\mathrm{BM}}}\prod_{k=1}^{N_\mathrm{D}=4}[\mathbf{\Psi}_{\mathrm{BM},k}]_{i_k,j_k}[\mathbf{C}_{\mathrm{BM}}]_\mathbf{j}, \quad (17)$$

$$\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{RM}}[n] = \sqrt{P_\mathrm{t}}\sum_{\mathbf{i}\in\mathcal{I}_{\mathrm{RM}}}[\bar{\mathbf{W}}_{m_\mathrm{M}}]_{i_4,:}^\mathrm{H}[\mathrm{F}_{m_\mathrm{B}}^{''}]_{i_2N_{\mathrm{R},y}+i_3,:}\mathbf{s}_{m_\mathrm{B}}[n-i_1] \times \sum_{\mathbf{j}\in\mathcal{J}_{\mathrm{RM}}}\prod_{k=1}^{N_\mathrm{D}=4}[\mathbf{\Psi}_{\mathrm{RM},k}]_{i_k,j_k}[\mathbf{C}_{\mathrm{RM}}]_\mathbf{j}. \quad (18)$$

To build the overall observation matrix, we need to consider all the time instances during training. Hence, we define the observation for a given training configuration as $\mathbf{Y}_{m_\mathrm{B},m_\mathrm{M}} = [\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}[1],\cdots,\bar{\mathbf{y}}_{m_\mathrm{B},m_\mathrm{M}}[N]]^\mathrm{T} \in \mathbb{C}^{N\times N_{\mathrm{RF,M}}}$. The matrices $\mathbf{Y}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{BM}}, \mathbf{Y}_{m_\mathrm{B},m_\mathrm{M}}^{\mathrm{RM}}, \mathbf{V}_{m_\mathrm{B},m_\mathrm{M}} \in \mathbb{C}^{N\times N_{\mathrm{RF,M}}}$ can be defined by using a similar form. Furthermore, we can group all the measurements in a single observation matrix $\mathbf{Y} \in \mathbb{C}^{M_\mathrm{B}N\times M_\mathrm{M}N_{\mathrm{RF,M}}}$ as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{1,1} & \cdots & \mathbf{Y}_{1,M_\mathrm{M}} \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_{M_\mathrm{B},1} & \cdots & \mathbf{Y}_{M_\mathrm{B},M_\mathrm{M}} \end{bmatrix}. \quad (19)$$

We can construct $\mathbf{Y}_{\mathrm{BM}}, \mathbf{Y}_{\mathrm{RM}}, \mathbf{V} \in \mathbb{C}^{M_\mathrm{B}N\times M_\mathrm{M}N_{\mathrm{RF,M}}}$ by following the same steps. Finally, we vectorize all matrices as $\mathbf{y} = \mathrm{vec}(\mathbf{Y}), \mathbf{y}_{\mathrm{BM}} = \mathrm{vec}(\mathbf{Y}_{\mathrm{BM}}), \mathbf{y}_{\mathrm{RM}} = \mathrm{vec}(\mathbf{Y}_{\mathrm{RM}}), \mathbf{v} = \mathrm{vec}(\mathbf{V}) \in \mathbb{C}^{M_\mathrm{M}N_{\mathrm{RF,M}}M_\mathrm{B}N}$. By using the representations in (17) and (18), $\mathbf{y}_{\mathrm{BM}}$ and $\mathbf{y}_{\mathrm{RM}}$ can be written as

$$\mathbf{y}_{\mathrm{BM}} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathrm{BM}}}\sum_{\mathbf{j}\in\mathcal{J}_{\mathrm{BM}}}[\mathbf{\Phi}_{\mathrm{BM}}]_{:,\mathbf{i}}\left(\prod_{k=1}^{N_\mathrm{D}=4}[\mathbf{\Psi}_{\mathrm{BM},k}]_{i_k,j_k}\right)[\mathbf{C}_{\mathrm{BM}}]_\mathbf{j}, \quad (20)$$

$$\mathbf{y}_{\mathrm{RM}} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathrm{RM}}}\sum_{\mathbf{j}\in\mathcal{J}_{\mathrm{RM}}}[\mathbf{\Phi}_{\mathrm{RM}}]_{:,\mathbf{i}}\left(\prod_{k=1}^{N_\mathrm{D}=4}[\mathbf{\Psi}_{\mathrm{RM},k}]_{i_k,j_k}\right)[\mathbf{C}_{\mathrm{RM}}]_\mathbf{j}, \quad (21)$$

where the entries of the BS-MS and the RIS-MS measurement tensors, i.e., $\mathbf{\Phi}_{\mathrm{BM}} \in \mathbb{C}^{M_\mathrm{M}N_{\mathrm{RF,M}}M_\mathrm{B}N\times N_{\mathrm{BM},1}^\mathrm{s}\times\cdots\times N_{\mathrm{BM},N_\mathrm{D}}^\mathrm{s}}$ and $\mathbf{\Phi}_{\mathrm{RM}} \in \mathbb{C}^{M_\mathrm{M}N_{\mathrm{RF,M}}M_\mathrm{B}N\times N_{\mathrm{RM},1}^\mathrm{s}\times\cdots\times N_{\mathrm{RM},N_\mathrm{D}}^\mathrm{s}}$, are defined as

$$[\mathbf{\Phi}_{\mathrm{BM}}]_{m_\mathrm{M}N_{\mathrm{RF,M}}M_\mathrm{B}N+n_{\mathrm{RF,M}}M_\mathrm{B}N+m_\mathrm{B}N+n,\mathbf{i}} = \sqrt{P_\mathrm{t}}[\bar{\mathbf{W}}_{m_\mathrm{M}}]_{i_4,n_{\mathrm{RF,M}}}^*[\mathrm{F}_{m_\mathrm{B}}^{'}]_{i_2N_{\mathrm{B},y}+i_3,:}\mathbf{s}_{m_\mathrm{B}}[n-i_1], \quad (22)$$

$$[\mathbf{\Phi}_{\mathrm{RM}}]_{m_\mathrm{M}N_{\mathrm{RF,M}}M_\mathrm{B}N+n_{\mathrm{RF,M}}M_\mathrm{B}N+m_\mathrm{B}N+n,\mathbf{i}} = \sqrt{P_\mathrm{t}}[\bar{\mathbf{W}}_{m_\mathrm{M}}]_{i_4,n_{\mathrm{RF,M}}}^*[\mathrm{F}_{m_\mathrm{B}}^{''}]_{i_2N_{\mathrm{R},y}+i_3,:}\mathbf{s}_{m_\mathrm{B}}[n-i_1] \quad (23)$$

for $n = 1,\ldots,N$, $n_{\mathrm{RF,M}} = 1,\ldots,N_{\mathrm{RF,M}}$, $m_\mathrm{M} = 1,\ldots,M_\mathrm{M}$ and $m_\mathrm{B} = 1,\ldots,M_\mathrm{B}$. The goal is to simultaneously estimate the BS-MS and the RIS-MS channels by solving the minimization problem that is defined as

$$\min_{\mathbf{C}_{\mathrm{BM}},\mathbf{C}_{\mathrm{RM}}}\left\|\mathbf{y} - \left(\sum_{\substack{\mathbf{i}\in\mathcal{I}_{\mathrm{BM}},\\ \mathbf{j}\in\mathcal{J}_{\mathrm{BM}}}}[\mathbf{\Phi}_{\mathrm{BM}}]_{:,\mathbf{i}}\prod_{k=1}^{N_\mathrm{D}=4}[\mathbf{\Psi}_{\mathrm{BM},k}]_{i_k,j_k}[\mathbf{C}_{\mathrm{BM}}]_\mathbf{j} + \sum_{\substack{\mathbf{i}\in\mathcal{I}_{\mathrm{RM}},\\ \mathbf{j}\in\mathcal{J}_{\mathrm{RM}}}}[\mathbf{\Phi}_{\mathrm{RM}}]_{:,\mathbf{i}}\prod_{k=1}^{N_\mathrm{D}=4}[\mathbf{\Psi}_{\mathrm{RM},k}]_{i_k,j_k}[\mathbf{C}_{\mathrm{RM}}]_\mathbf{j}\right)\right\|^2. \quad (24)$$

The MOMP algorithm proposed in [15] aims to solve the conventional sparse channel estimation problem for a given BS-MS link with much lower complexity than other greedy solutions. This algorithm is similar to the classical OMP in [68], in the sense that both algorithms follow a greedy approach such that the channel paths are found one by one in descending channel gain order by using matching projections. At each iteration, the residual vector $\mathbf{y}_{\mathrm{res}}$, which is initialized by $\mathbf{y}$, is updated by subtracting the contribution of the estimated path. The reason for preferring MOMP over OMP is that it can operate with large arrays and high-resolution dictionaries at a lower complexity since it exploits multiple dictionaries. In contrast, OMP leverages a single larger dictionary constructed by taking the Kronecker product of all dictionaries [60]. The

readers are referred to [15] for more details about the MOMP algorithm.

To estimate the BS-MS and the RIS-MS channels concurrently, we need to modify the MOMP algorithm, which is tailored to estimate a single channel such as the BS-MS channel. Since the BS-MS and the RIS-MS channels have different transmit configurations and dictionaries, it is possible to distinguish the paths of different channels and estimate them simultaneously. To this end, we propose operating with two sources by independently repeating the projection step at each iteration for both sources and then selecting the best matching projection. The proposed approach is summarized in Algorithm 1. The algorithm can be terminated when a convergence condition is satisfied or a target number of paths $N_{\max}$ are estimated. the BS-MS and the RIS-MS channels can be reconstructed by substituting the estimated $\mathbf{C}_{\mathrm{BM}}$ and $\mathbf{C}_{\mathrm{RM}}$ tensors at the output of Algorithm 1 into (9) and (12), respectively. Furthermore, the sets of multi-indices that refer to the atoms in the dictionaries for the estimated paths are denoted by $\mathcal{C}_{\mathrm{BM}}$ and $\mathcal{C}_{\mathrm{RM}}$. The directions and the delays of the estimated paths are found using the dictionaries defined in (7) and (10). In other words, for the l-th multi-index $\mathbf{j}_{\mathrm{BM},l} = (j^1_{\mathrm{BM},l}, j^2_{\mathrm{BM},l}, j^3_{\mathrm{BM},l}, j^4_{\mathrm{BM},l}) \in \mathcal{C}_{\mathrm{BM}}$, we can recover the path parameters as $\tau_{\mathrm{BM},l} - t_0 = \overline{\tau}_{\mathrm{BM},j^1_{\mathrm{BM},l}}$, $\phi^{\mathrm{x}}_{\mathrm{BM},l} = \overline{\phi}^{\mathrm{x}}_{\mathrm{BM},j^2_{\mathrm{BM},l}}$, $\phi^{\mathrm{y}}_{\mathrm{BM},l} = \overline{\phi}^{\mathrm{y}}_{\mathrm{BM},j^3_{\mathrm{BM},l}}$ and $\boldsymbol{\theta}_{\mathrm{BM},l} = \overline{\boldsymbol{\theta}}_{\mathrm{BM},j^4_{\mathrm{BM},l}}$. Then, we can recover the z-axis parameters of the DoDs by $\phi^{\mathrm{z}}_{\mathrm{BM},l} = \sqrt{1 - (\phi^{\mathrm{x}}_{\mathrm{BM},l})^2 - (\phi^{\mathrm{y}}_{\mathrm{BM},l})^2}$. The path parameters of the RIS-MS channel can be recovered by following the same procedure. Although this algorithm has low computational complexity, the memory requirements to store the measurement tensors $\boldsymbol{\Phi}_{\mathrm{BM}}$ and $\boldsymbol{\Phi}_{\mathrm{RM}}$ become the bottleneck when we have large arrays, which is the case for RIS-aided mmWave systems.

C. Reduced Complexity Channel Estimation

To further reduce the complexity of Algorithm 1, instead of defining separate dictionaries for the DoA, DoD, and delays for each channel component, we only consider dictionaries for the DoD and the delay so that the DoA information is embedded into the equivalent complex gain for each path. Note that the DoA of each path can be recovered from the resultant equivalent complex gains. For the compressed sensing problem, we will use the first $N_{\mathrm{D}} = 3$ dictionaries in (7) and (10) for the BS-MS and the RIS-MS channels, respectively. To obtain a sparse representation for the received signal, we will use the observation matrix in (19). The observation matrix for the BS-MS channel can be expressed as

$$\mathbf{Y}_{\mathrm{BM}} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathrm{BM}}} \sum_{\mathbf{j}\in\mathcal{J}_{\mathrm{BM}}} [\bar{\boldsymbol{\Phi}}_{\mathrm{BM}}]_{:,\mathbf{i}} \left(\prod_{k=1}^{N_{\mathrm{D}}=3} [\boldsymbol{\Psi}_{\mathrm{BM},k}]_{i_k,j_k} \right) [\bar{\mathbf{C}}_{\mathrm{BM}}]_{\mathbf{j},:}, \tag{25}$$

where $\bar{\boldsymbol{\Phi}}_{\mathrm{BM}} \in \mathbb{C}^{M_{\mathrm{B}}N \times N^{\mathrm{s}}_{\mathrm{BM},1} \times \cdots \times N^{\mathrm{s}}_{\mathrm{BM},N_{\mathrm{D}}}}$ is the measurement tensor. Moreover, the tensor that captures the channel gain and the DoA information of the BS-MS channel is denoted by $\bar{\mathbf{C}}_{\mathrm{BM}} \in \mathbb{C}^{N^{\mathrm{a}}_{\mathrm{BM},1} \times \ldots N^{\mathrm{a}}_{\mathrm{BM},N_{\mathrm{D}}} \times M_{\mathrm{M}} N_{\mathrm{RF,M}}}$. For the three dictionaries, the multi-index variables are defined as $\mathbf{i} \in \mathcal{I}_{\mathrm{BM}} = \{(i_1, i_2, i_3), i_k = 1, \ldots, N^{\mathrm{s}}_{\mathrm{BM},k}\}$ and $\mathbf{j} \in \mathcal{J}_{\mathrm{BM}} = \{(j_1, j_2, j_3), j_k = 1, \ldots, N^{\mathrm{a}}_{\mathrm{BM},k}\}$. On the other hand, the observation matrix for the RIS-MS channel is written as

$$\mathbf{Y}_{\mathrm{RM}} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathrm{RM}}} \sum_{\mathbf{j}\in\mathcal{J}_{\mathrm{RM}}} [\bar{\boldsymbol{\Phi}}_{\mathrm{RM}}]_{:,\mathbf{i}} \left(\prod_{k=1}^{3} [\boldsymbol{\Psi}_{\mathrm{RM},k}]_{i_k,j_k} \right) [\bar{\mathbf{C}}_{\mathrm{RM}}]_{\mathbf{j},:}, \tag{26}$$

where $\bar{\boldsymbol{\Phi}}_{\mathrm{RM}} \in \mathbb{C}^{M_{\mathrm{B}}N \times N^{\mathrm{s}}_{\mathrm{RM},1} \times \cdots \times N^{\mathrm{s}}_{\mathrm{RM},N_{\mathrm{D}}}}$ is the sensing matrix corresponding to the RIS-MS link and $\bar{\mathbf{C}}_{\mathrm{RM}} \in \mathbb{C}^{N^{\mathrm{a}}_{\mathrm{RM},1} \times \ldots N^{\mathrm{a}}_{\mathrm{RM},N_{\mathrm{D}}} \times M_{\mathrm{M}} N_{\mathrm{RF,M}}}$ contains the RIS-MS channel coefficients. The multi-index variables are defined as $\mathbf{i} \in \mathcal{I}_{\mathrm{RM}} = \{(i_1, i_2, i_3), i_k = 1, \ldots, N^{\mathrm{s}}_{\mathrm{RM},k}\}$ and $\mathbf{j} \in \mathcal{J}_{\mathrm{RM}} = \{(j_1, j_2, j_3), j_k = 1, \ldots, N^{\mathrm{a}}_{\mathrm{RM},k}\}$. With the first three dictionaries given in (7) and (10), we can express $\bar{\mathbf{C}}_{\mathrm{BM}}$ and $\bar{\mathbf{C}}_{\mathrm{RM}}$ with the BS-MS and the RIS-MS channel coefficients including information about the path complex gains and the DoA responses as

$$[\bar{\mathbf{C}}_{\mathrm{BM}}]_{\mathbf{j},:} = \begin{cases} \boldsymbol{\beta}^{\mathrm{T}}_{\mathrm{BM},l} & \text{if } \begin{array}{rcl} \tau_{\mathrm{BM},l} & = & \overline{\tau}_{\mathrm{BM},j_1}, \\ \phi^{\mathrm{x}}_{\mathrm{BM},l} & = & \overline{\phi}^{\mathrm{x}}_{\mathrm{BM},j_2}, \\ \phi^{\mathrm{y}}_{\mathrm{BM},l} & = & \overline{\phi}^{\mathrm{y}}_{\mathrm{BM},j_3}, \end{array} \\ \mathbf{0}^{\mathrm{T}} & \text{otherwise,} \end{cases} \tag{27}$$

$$[\bar{\mathbf{C}}_{\mathrm{RM}}]_{\mathbf{j},:} = \begin{cases} \boldsymbol{\beta}^{\mathrm{T}}_{\mathrm{RM},q} & \text{if } \begin{array}{rcl} \tau_{\mathrm{RM},q} & = & \overline{\tau}_{\mathrm{RM},j_1}, \\ \phi^{\mathrm{x}}_{\mathrm{RM},q} & = & \overline{\phi}^{\mathrm{x}}_{\mathrm{RM},j_2}, \\ \phi^{\mathrm{y}}_{\mathrm{RM},q} & = & \overline{\phi}^{\mathrm{y}}_{\mathrm{RM},j_3}, \end{array} \\ \mathbf{0}^{\mathrm{T}} & \text{otherwise.} \end{cases} \tag{28}$$

for the channel coefficients denoted by $\boldsymbol{\beta}_{\mathrm{BM},l}, \boldsymbol{\beta}_{\mathrm{RM},q} \in \mathbb{C}^{M_{\mathrm{M}} N_{\mathrm{RF},M}}$ whose entries are given as

$$[\boldsymbol{\beta}_{\mathrm{BM},l}]_{m_{\mathrm{M}} N_{\mathrm{RF,M}} + n_{\mathrm{RF},M}} = \alpha_{\mathrm{BM},l} [\bar{\mathbf{W}}_{m_{\mathrm{M}}}]^{\mathrm{H}}_{:,n_{\mathrm{RF},M}} \tilde{\mathbf{a}}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{BM},l}), \tag{29}$$

$$[\boldsymbol{\beta}_{\mathrm{RM},q}]_{m_{\mathrm{M}} N_{\mathrm{RF,M}} + n_{\mathrm{RF},M}} = \tilde{\alpha}_{\mathrm{RM},q} [\bar{\mathbf{W}}_{m_{\mathrm{M}}}]^{\mathrm{H}}_{:,n_{\mathrm{RF},M}} \tilde{\mathbf{a}}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{RM},q}) \tag{30}$$

Algorithm 1 MOMP-based Channel Estimation

Input: $\mathbf{y}$, $\boldsymbol{\Phi}_{\mathrm{BM}}$, $\boldsymbol{\Phi}_{\mathrm{RM}}$, $\boldsymbol{\Psi}_{\mathrm{BM},k}$, $\boldsymbol{\Psi}_{\mathrm{RM},k}$ for $k = 1, \ldots, N_{\mathrm{D}}$
Initialize: $\mathbf{y}_{\mathrm{res}} = \mathbf{y}$, $\mathcal{C}_{\mathrm{BM}} = \varnothing$, $\mathcal{C}_{\mathrm{RM}} = \varnothing$, iteration $t = 0$

```
while not converged do
    Estimate j_BM and α_BM by using only the BS-MS projections for the MOMP problem
    Estimate j_RM and α_RM by using only the RIS-MS projections for the MOMP problem
    if |α_BM|^2 > |α_RM^2| then
        C_BM ← C_BM ∪ j_BM
        [C_BM]_{j_BM} ← α_BM
        Update y_res by subtracting the path {j_BM, α_BM}
    else
        C_RM ← C_RM ∪ j_RM
        [C_RM]_{j_RM} ← α_RM
        Update y_res by subtracting the path {j_RM, α_RM}
    end if
    t ← t + 1
    if t = N_max then
        Terminate the algorithm
    end if
end while
```

Output: $\mathbf{C}_{\mathrm{BM}}$, $\mathbf{C}_{\mathrm{RM}}$, $\mathcal{C}_{\mathrm{BM}}$, $\mathcal{C}_{\mathrm{RM}}$

for $n_{\mathrm{RF,M}} = 1, \ldots, N_{\mathrm{RF,M}}$ and $m_{\mathrm{M}} = 1, \ldots, M_{\mathrm{M}}$. The entries of the tensors $\bar{\mathbf{\Phi}}_{\mathrm{BM}}$ and $\bar{\mathbf{\Phi}}_{\mathrm{RM}}$ satisfying the models in (25) and (26) to reconstruct the received signal are given as

$$[\bar{\mathbf{\Phi}}_{\mathrm{BM}}]_{m_{\mathrm{B}}N+n,\mathbf{i}} = \sqrt{P_{\mathrm{t}}}[\mathrm{F}'_{m_{\mathrm{B}}}]_{i_2 N_{\mathrm{B}}^{\mathrm{y}}+i_3,:}\mathbf{s}_{m_{\mathrm{B}}}[n-i_1], \quad (31)$$

$$[\bar{\mathbf{\Phi}}_{\mathrm{RM}}]_{m_{\mathrm{B}}N+n,\mathbf{i}} = \sqrt{P_{\mathrm{t}}}[\mathrm{F}''_{m_{\mathrm{B}}}]_{i_2 N_{\mathrm{R}}^{\mathrm{y}}+i_3,:}\mathbf{s}_{m_{\mathrm{B}}}[n-i_1] \quad (32)$$

for $n = 1, \ldots, N$ and $m_{\mathrm{B}} = 1, \ldots, M_{\mathrm{B}}$. With all these definitions, we can write the overall received signal as $\mathbf{Y} = \mathbf{Y}_{\mathrm{BM}} + \mathbf{Y}_{\mathrm{RM}} + \mathbf{V}$. Then, the channel estimation problem is defined as

$$\min_{\bar{\mathbf{C}}_{\mathrm{BM}},\bar{\mathbf{C}}_{\mathrm{RM}}} \Bigg\| \mathbf{Y} - \Bigg(\sum_{\substack{\mathbf{i}\in\mathcal{I}_{\mathrm{BM}},\\ \mathbf{j}\in\mathcal{J}_{\mathrm{BM}}}} [\bar{\mathbf{\Phi}}_{\mathrm{BM}}]_{:,\mathbf{i}} \prod_{k=1}^{N_{\mathrm{D}}=3} [\mathbf{\Psi}_{\mathrm{BM},k}]_{i_k,j_k} [\bar{\mathbf{C}}_{\mathrm{BM}}]_{\mathbf{j},:} + \sum_{\substack{\mathbf{i}\in\mathcal{I}_{\mathrm{RM}},\\ \mathbf{j}\in\mathcal{J}_{\mathrm{RM}}}} [\bar{\mathbf{\Phi}}_{\mathrm{RM}}]_{:,\mathbf{i}} \prod_{k=1}^{N_{\mathrm{D}}=3} [\mathbf{\Psi}_{\mathrm{RM},k}]_{i_k,j_k} [\bar{\mathbf{C}}_{\mathrm{RM}}]_{\mathbf{j},:} \Bigg) \Bigg\|_{\mathrm{F}}^2, \quad (33)$$

similar to the previously defined problem in (24). The MOMP algorithm can solve this problem by following the approach introduced in Algorithm 1. The output of the algorithm provides estimates of the DoDs and the delays of the paths. Furthermore, the MOMP algorithm estimates the coefficient vectors defined in (27) and (28). Using the expressions given in (29) and (30), the coefficient vectors can be written as

$$\boldsymbol{\beta}_{\mathrm{BM},l} = \alpha_{\mathrm{BM},l}\bar{\mathbf{W}}^{\mathrm{H}}\tilde{\mathbf{a}}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{BM},l}), \quad (34)$$

$$\boldsymbol{\beta}_{\mathrm{RM},q} = \tilde{\alpha}_{\mathrm{RM},q}\bar{\mathbf{W}}^{\mathrm{H}}\tilde{\mathbf{a}}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{RM},q}), \quad (35)$$

where $\bar{\mathbf{W}} = [\bar{\mathbf{W}}_1, \ldots, \bar{\mathbf{W}}_{M_{\mathrm{M}}}] \in \mathbb{C}^{N_{\mathrm{M}} \times M_{\mathrm{M}} N_{\mathrm{RF,M}}}$. In the rest of the section, we introduce the steps to estimate DoAs and the complex gains of the BS-MS channel paths, while the same steps apply to the RIS-MS channel paths. The estimated coefficient vector for the l-th path of the BS-MS channel $\mathbf{z}_{\mathrm{BM},l} \in \mathbb{C}^{M_{\mathrm{M}} N_{\mathrm{RF,M}}}$ can be expressed as

$$\mathbf{z}_{\mathrm{BM},l} = \alpha_{\mathrm{BM},l}\bar{\mathbf{W}}^{\mathrm{H}}\tilde{\mathbf{a}}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{BM},l}) + \mathbf{n}_{\mathrm{BM},l}, \quad (36)$$

where $\mathbf{n}_{\mathrm{BM},l} \in \mathbb{C}^{M_{\mathrm{M}} N_{\mathrm{RF,M}}}$ is the estimation noise. Therefore, the estimate of the DoA can be found by using the estimate of $\boldsymbol{\beta}_{\mathrm{BM},l}$ as

$$\hat{\boldsymbol{\theta}}_{\mathrm{BM},l} = \arg\max_{\boldsymbol{\theta}} \mathbf{z}_{\mathrm{BM},l}^{\mathrm{H}} \bar{\mathbf{W}}^{\mathrm{H}} (\bar{\mathbf{W}}\bar{\mathbf{W}}^{\mathrm{H}})^{-1} \tilde{\mathbf{a}}_{\mathrm{M}}(\boldsymbol{\theta}), \quad (37)$$

which is a structured search problem that can be solved using a discrete dictionary for the DoA, i.e., $\mathbf{\Psi}_{\mathrm{BM},4}$ as in the previous section. Then, the channel gain can be recovered by using $\mathbf{z}_{\mathrm{BM},l}$, $\bar{\mathbf{W}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{BM},l}$ as

$$\hat{\alpha}_{\mathrm{BM},l} = \frac{\tilde{\mathbf{a}}_{\mathrm{M}}^{\mathrm{H}}(\hat{\boldsymbol{\theta}}_{\mathrm{BM},l})\bar{\mathbf{W}}\mathbf{z}_{\mathrm{BM},l}}{\|\bar{\mathbf{W}}^{\mathrm{H}}\tilde{\mathbf{a}}_{\mathrm{M}}(\hat{\boldsymbol{\theta}}_{\mathrm{BM},l})\|^2}. \quad (38)$$

All the channel parameters in (2) are recovered by following the described two-stage procedure. The DoAs and the channel gains of the RIS-MS channel can be estimated by following the same steps for constructing $\mathbf{z}_{\mathrm{RM},q} \in \mathbb{C}^{M_{\mathrm{M}} N_{\mathrm{RF,M}}}$ and following the provided solution. The described channel estimation algorithm works with given hardware impairments at the MS, which are not known at the beginning of the initial access stage. That is, the exact expressions for the atoms of the dictionaries $\mathbf{\Psi}_{\mathrm{BM},4}$ and $\mathbf{\Psi}_{\mathrm{RM},4}$ are not known. The following section will introduce a dictionary learning approach to estimate the channel with the described algorithm while learning the hardware impairments.

D. Complexity Analysis

The complexity order of the MOMP algorithm is computed in [15]. However, since we modify the MOMP algorithm in this work, we calculate the complexity of the proposed algorithm and also compute the complexity of the proposed approach if OMP is used instead of MOMP. Assuming that $N_{\mathrm{BM,p}}$ and $N_{\mathrm{RM,p}}$ are the number of estimated paths for the BS-MS and the RIS-MS channels, the computational complexity of the reduced complexity version of the MOMP algorithm is given as

$$O\Bigg((N_{\mathrm{iter}}+1)(M_{\mathrm{B}}N + M_{\mathrm{M}}N_{\mathrm{RF,M}})(N_{\mathrm{BM,p}} + N_{\mathrm{RM,p}}) \times \Bigg[\Bigg(\sum_{k=1}^{N_{\mathrm{D}}=3} N_{\mathrm{BM},k}^{\mathrm{a}} \Bigg) \prod_{k=1}^{N_{\mathrm{D}}=3} N_{\mathrm{BM},k}^{\mathrm{s}} + \Bigg(\sum_{k=1}^{N_{\mathrm{D}}=3} N_{\mathrm{RM},k}^{\mathrm{a}} \Bigg) \prod_{k=1}^{N_{\mathrm{D}}=3} N_{\mathrm{RM},k}^{\mathrm{s}} \Bigg] \Bigg), \quad (39)$$

where N_{iter} is the number of inner iterations [15]. The sizes of the atoms for the BS-MS channel are given as $N_{\mathrm{BM},1}^{\mathrm{s}} = D$, $N_{\mathrm{BM},2}^{\mathrm{s}} = N_{\mathrm{B,x}}$, and $N_{\mathrm{BM},3}^{\mathrm{s}} = N_{\mathrm{B,y}}$ with corresponding number of atoms $N_{\mathrm{BM},1}^{\mathrm{a}}$, $N_{\mathrm{BM},2}^{\mathrm{a}}$, and $N_{\mathrm{BM},3}^{\mathrm{a}}$, respectively. Similarly, the sizes of the atoms for the RIS-MS channel are given as $N_{\mathrm{RM},1}^{\mathrm{s}} = D$, $N_{\mathrm{RM},2}^{\mathrm{s}} = N_{\mathrm{R,x}}$, and $N_{\mathrm{RM},3}^{\mathrm{s}} = N_{\mathrm{R,y}}$ with corresponding number of atoms $N_{\mathrm{RM},1}^{\mathrm{a}}$, $N_{\mathrm{RM},2}^{\mathrm{a}}$, and $N_{\mathrm{RM},3}^{\mathrm{a}}$, respectively. If OMP algorithm is used instead of MOMP, the terms $\sum_{k=1}^{N_{\mathrm{D}}=3} N_{\mathrm{BM},k}^{\mathrm{a}}$ and $\sum_{k=1}^{N_{\mathrm{D}}=3} N_{\mathrm{RM},k}^{\mathrm{a}}$ in (39) would be replaced by $\prod_{k=1}^{N_{\mathrm{D}}=3} N_{\mathrm{BM},k}^{\mathrm{a}}$ and $\prod_{k=1}^{N_{\mathrm{D}}=3} N_{\mathrm{RM},k}^{\mathrm{a}}$, and the number of iterations N_{iter} would be zero. Considering that we would like to utilize very high-resolution dictionaries, i.e., large $N_{\mathrm{BM},k}^{\mathrm{a}}$ and $N_{\mathrm{RM},k}^{\mathrm{a}}$, the complexity of the OMP algorithm is extremely prohibitive, and large arrays with high-resolution dictionaries cannot be supported.

The computational complexity of the DoA recovery stage given in (37) is low compared to the channel estimation stage. The pseudo-inverse of the matrix $\bar{\mathbf{W}}$ can be calculated once with a computational complexity of $O((M_{\mathrm{M}}N_{\mathrm{RF,M}})^2 N_{\mathrm{M}})$, and the search over the angular grid would result in a complexity of $O(M_{\mathrm{M}}N_{\mathrm{RF,M}}(N_{\mathrm{BM,p}}N_{\mathrm{BM},4}^{\mathrm{a}} + N_{\mathrm{RM,p}}N_{\mathrm{RM},4}^{\mathrm{a}}))$. Thus, the overall complexity of the DoA recovery stage can be expressed as $O(M_{\mathrm{M}}N_{\mathrm{RF,M}}(M_{\mathrm{M}}N_{\mathrm{RF,M}}N_{\mathrm{M}} + N_{\mathrm{BM,p}}N_{\mathrm{BM},4}^{\mathrm{a}} + N_{\mathrm{RM,p}}N_{\mathrm{RM},4}^{\mathrm{a}}))$. Since this complexity order is negligible, the overall computational complexity order of the proposed algorithm can be simply expressed by (39).

IV. Dictionary Learning

The channel parameters are estimated using the procedure described in Section III exploiting given dictionaries. However,

the dictionaries related to the DoAs, i.e., $\mathbf{\Psi}_{\mathrm{BM},4}$ and $\mathbf{\Psi}_{\mathrm{RM},4}$, include the impairments of the array at the MS. If these impairments are not known, the channel estimation stage will provide inaccurate results. To overcome this limitation, we leverage a dictionary learning approach to update the dictionary related to the DoAs. Without loss of generality, we use the same dictionary for the BS-MS and the RIS-MS channels, denoted by $\mathbf{\Psi}_{\mathrm{M}} = \mathbf{\Psi}_{\mathrm{BM},4} = \mathbf{\Psi}_{\mathrm{RM},4} \in \mathbb{C}^{N_{\mathrm{M}} \times G_{\mathrm{M}}}$, where G_{M} is the number of atoms. Let us express the estimated coefficient vectors of the BS-MS and the RIS-MS channel paths in terms of the dictionary $\mathbf{\Psi}_{\mathrm{M}}$ as

$$\mathbf{z}_{\mathrm{BM},l} = \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \mathbf{x}_{\mathrm{BM},l} + \mathbf{n}_{\mathrm{BM},l}, \tag{40}$$

$$\mathbf{z}_{\mathrm{RM},q} = \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \mathbf{x}_{\mathrm{RM},q} + \mathbf{n}_{\mathrm{RM},q}, \tag{41}$$

where $\mathbf{x}_{\mathrm{BM},l} \in \mathbb{C}^{G_{\mathrm{M}}}$ and $\mathbf{x}_{\mathrm{RM},q} \in \mathbb{C}^{G_{\mathrm{M}}}$ are 1-sparse vectors with the only non-zero entries $\alpha_{\mathrm{BM},l}$ and $\tilde{\alpha}_{\mathrm{RM},q}$, respectively. The channel estimation problem for the l-th path of the BS-MS channel given a sparsifying dictionary is

$$\begin{aligned} \min_{\mathbf{\Psi}_{\mathrm{M}}, \mathbf{x}_{\mathrm{BM},l}} \quad & \left\| \mathbf{z}_{\mathrm{BM},l} - \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \mathbf{x}_{\mathrm{BM},l} \right\|_2^2 \\ \text{subject to} \quad & \left\| \mathbf{x}_{\mathrm{BM},l} \right\|_0 = 1, \\ & \left\| [\mathbf{\Psi}_{\mathrm{M}}]_{:,j} \right\|_2 = 1, \; j = 1, \ldots, G_{\mathrm{M}}, \end{aligned} \tag{42}$$

where the l_0-norm constraint ensures that the vector $\mathbf{x}_{\mathrm{BM},l}$ has only one non-zero entry. The same problem can be constructed for the paths of the RIS-MS channel. To formulate the joint sparse recovery and dictionary learning problem when the dictionary is unknown, we collect the observations corresponding to all paths of the BS-MS and the RIS-MS channels as $\mathbf{Z} = \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \mathbf{X} + \mathbf{N}$ where

$$\begin{aligned} \mathbf{Z} &= [\mathbf{z}_{\mathrm{BM},1}, \cdots, \mathbf{z}_{\mathrm{BM},N_{\mathrm{BM,p}}}, \mathbf{z}_{\mathrm{RM},1}, \cdots, \mathbf{z}_{\mathrm{BM},N_{\mathrm{RM,p}}}], \\ \mathbf{X} &= [\mathbf{x}_{\mathrm{BM},1}, \cdots, \mathbf{x}_{\mathrm{BM},N_{\mathrm{BM,p}}}, \mathbf{x}_{\mathrm{RM},1}, \cdots, \mathbf{x}_{\mathrm{BM},N_{\mathrm{RM,p}}}], \\ \mathbf{N} &= [\mathbf{n}_{\mathrm{BM},1}, \cdots, \mathbf{n}_{\mathrm{BM},N_{\mathrm{BM,p}}}, \mathbf{n}_{\mathrm{RM},1}, \cdots, \mathbf{n}_{\mathrm{BM},N_{\mathrm{RM,p}}}], \end{aligned} \tag{43}$$

with dimensions $\mathbf{Z} \in \mathbb{C}^{M_{\mathrm{M}} N_{\mathrm{RF,M}} \times (N_{\mathrm{BM,p}} + N_{\mathrm{RM,p}})}$, $\mathbf{X} \in \mathbb{C}^{G_{\mathrm{M}} \times (N_{\mathrm{BM,p}} + N_{\mathrm{RM,p}})}$ and $\mathbf{N} \in \mathbb{C}^{M_{\mathrm{M}} N_{\mathrm{RF,M}} \times (N_{\mathrm{BM,p}} + N_{\mathrm{RM,p}})}$, respectively. In addition to exploiting the observations from different paths, we can also use the measurements obtained at different MS locations, as proposed in [54]. The goal is to increase the number of different sparse patterns, a common consideration in the dictionary learning literature. Let the number of different MS locations be denoted by N_{loc}. The estimated coefficient matrix for the i-th MS location can be represented by $\mathbf{Z}_i$, $i = 1, \ldots, N_{\mathrm{loc}}$. Then, the observations for all the MS locations can be collected as $\bar{\mathbf{Z}} = \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \bar{\mathbf{X}} + \bar{\mathbf{N}}$ where

$$\begin{aligned} \bar{\mathbf{Z}} &= [\mathbf{Z}_i, \cdots, \mathbf{Z}_{N_{\mathrm{loc}}}] \in \mathbb{C}^{M_{\mathrm{M}} N_{\mathrm{RF,M}} \times N_{\mathrm{loc}} (N_{\mathrm{BM,p}} + N_{\mathrm{RM,p}})}, \\ \bar{\mathbf{X}} &= [\mathbf{X}_i, \cdots, \mathbf{X}_{N_{\mathrm{loc}}}] \in \mathbb{C}^{G_{\mathrm{M}} \times N_{\mathrm{loc}} (N_{\mathrm{BM,p}} + N_{\mathrm{RM,p}})}, \\ \bar{\mathbf{N}} &= [\mathbf{N}_i, \cdots, \mathbf{N}_{N_{\mathrm{loc}}}] \in \mathbb{C}^{M_{\mathrm{M}} N_{\mathrm{RF,M}} \times N_{\mathrm{loc}} (N_{\mathrm{BM,p}} + N_{\mathrm{RM,p}})}. \end{aligned} \tag{44}$$

Finally, we can incorporate a denoising stage to combat the low signal-to-noise ratio (SNR) conditions of the initial access stage at mmWave bands. Let us express the denoised version of $\bar{\mathbf{Z}}$ by $\mathbf{Q}$, such that $\bar{\mathbf{Z}} = \mathbf{Q} + \bar{\mathbf{N}}' = \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \bar{\mathbf{X}} + \bar{\mathbf{N}}$, where $\bar{\mathbf{N}}'$ shows the mismatch between the denoised matrix and $\bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \bar{\mathbf{X}}$. The denoising operation can be embedded into the problem by including a regularization term $\left\| \bar{\mathbf{Z}} - \mathbf{Q} \right\|_{\mathrm{F}}^2$ as

$$\begin{aligned} \min_{\mathbf{\Psi}_{\mathrm{M}}, \bar{\mathbf{X}}, \mathbf{Q}} \quad & \left\| \mathbf{Q} - \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \bar{\mathbf{X}} \right\|_{\mathrm{F}}^2 + w \left\| \bar{\mathbf{Z}} - \mathbf{Q} \right\|_{\mathrm{F}}^2 \\ \text{subject to} \quad & \left\| \mathbf{x}_{\mathrm{BM},l}^i \right\|_0 = 1, \forall l, i, \; \left\| \mathbf{x}_{\mathrm{RM},q}^i \right\|_0 = 1, \forall q, i, \\ & \left\| [\mathbf{\Psi}_{\mathrm{M}}]_{:,j} \right\|_2 = 1, \; j = 1, \ldots, G_{\mathrm{M}}, \end{aligned} \tag{45}$$

where w is the regularization constant. We follow the combined dictionary learning solution given in [54]. The objective function of the given problem is convex for the individual optimization variables. However, it is not jointly convex with all the variables. Therefore, the problem can be solved with an alternating minimization approach with three sub-problems, where each sub-problem minimizes the objective function for one variable while the others are fixed. Since we assume all the paths are independent, the sparse recovery problem can be solved, as explained in the previous section. The dictionary update is achieved with the method of optimal directions as $\mathbf{\Psi}_{\mathrm{M}} = [\bar{\mathbf{W}}^{\mathrm{H}}]^{\dagger} \mathbf{Q} \bar{\mathbf{X}}^{\mathrm{H}} (\bar{\mathbf{X}} \bar{\mathbf{X}}^{\mathrm{H}})^{-1}$ [69]. The denoising stage is implemented with the least-squares (LS) solution $\mathbf{Q} = (1+w)^{-1} (w \bar{\mathbf{Z}} + \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \bar{\mathbf{X}})$. The resulting procedure is given in Algorithm 2.

Algorithm 2 Dictionary Learning

Input: $\bar{\mathbf{Z}}$, $\bar{\mathbf{W}}$, w
Initialize: $\mathbf{\Phi}_{\mathrm{M}}$ with $\mathbf{M}_{\mathrm{M}} \mathbf{\Gamma}_{\mathrm{M}} = \mathbf{I}_{N_{\mathrm{M}}}$
1: **while** not converged **do**
2: Find each $\mathbf{x}_{\mathrm{BM},l}$ and $\mathbf{x}_{\mathrm{RM},q}$ by solving (37) and (38)
3: Update the dictionary $\mathbf{\Psi}_{\mathrm{M}} = [\bar{\mathbf{W}}^{\mathrm{H}}]^{\dagger} \mathbf{Q} \bar{\mathbf{X}}^{\mathrm{H}} (\bar{\mathbf{X}} \bar{\mathbf{X}}^{\mathrm{H}})^{-1}$
4: Denoise observation $\mathbf{Q} = (1+w)^{-1} (w \bar{\mathbf{Z}} + \bar{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}_{\mathrm{M}} \bar{\mathbf{X}})$
5: **end while**
Output: $\bar{\mathbf{X}}$, $\mathbf{\Phi}_{\mathrm{M}}$

V. Localization Algorithms

The direction and the delay parameters of the paths belonging to the BS-MS and the RIS-MS channels can be estimated using the sparse channel estimation method explained in previous sections. Any 3D localization algorithm that utilizes these parameters can be adopted to estimate the locations of the MSs as a byproduct of channel estimation. However, unlike most mmWave localization schemes, we have additional paths thanks to the aid of the RIS. We consider two different geometric approaches for localization depending on whether the LoS paths of the BS-MS and the RIS-MS channels exist. The first approach applies to the classical BS-MS setting without a RIS and to the case with a RIS when only the RIS-MS link is available, i.e., the BS-MS link is obstructed. As we consider indoor localization, our derivations exploit beneficial properties of indoor propagation environments to jointly solve for the unknown clock offset and the localization problem in an original manner. The second approach applies to the case where the LoS paths of both the BS-MS and the RIS-MS channels are available.

A. Localization with a Single LoS Path

This approach applies to the case where one LoS path is available, i.e., one of the BS-MS or RIS-MS channels is obstructed while the LoS path for the other channel is available. We begin by describing the localization algorithm by using the BS-MS link. Let $\mathbf{b} \in \mathbb{R}^3$ and $\mathbf{m} \in \mathbb{R}^3$ be the locations of the BS and the MS, respectively, and c denote the speed of light. Then, the MS location can be expressed as

$$\mathbf{m} = \mathbf{b} + c\tau_{\mathrm{BM},1}\boldsymbol{\phi}_{\mathrm{BM},1}. \tag{46}$$

The same equation can be written with the help of the DoA $\boldsymbol{\theta}_{\mathrm{BM},1}$. Unfortunately, the channel estimation algorithm provides only relative delays or time difference of arrival (TDoA) for the paths (i.e., $\Delta\tau_{\mathrm{BM},l} = \tau_{\mathrm{BM},l} - t_0$) due to the unknown clock offset t_0. We will use the relative delays and a fundamental assumption of the considered indoor localization scenario, where the reflection surfaces are either horizontal or vertical, to solve this problem [14], [15]. This assumption lets us classify the NLoS reflections as wall or floor/ceiling reflections. As a preliminary for the classification algorithm, we need to recover the azimuth and elevation angles from the direction vectors, which can be achieved by $\theta^{\mathrm{az}} = \arg(\theta_{\mathrm{x}} + j\theta_{\mathrm{y}})$ and $\theta^{\mathrm{el}} = \arcsin(\theta_{\mathrm{z}})$ for a direction vector $\boldsymbol{\theta}$.

Let us classify the paths as LoS, first-order wall, or floor/ceiling reflections. Firstly, if the path is LoS, the elevation and azimuth angles should satisfy $\boldsymbol{\theta}^{\mathrm{el}}_{\mathrm{BM},l} = -\boldsymbol{\phi}^{\mathrm{el}}_{\mathrm{BM},l}$ and $\boldsymbol{\theta}^{\mathrm{az}}_{\mathrm{BM},l} = \pi - \boldsymbol{\phi}^{\mathrm{az}}_{\mathrm{BM},l}$, respectively. Now, we can classify the NLoS paths by using the properties of the indoor environment considered in this paper. If an NLoS path is reflected by a wall, the vertical distance traveled by the LoS and the NLoS path is the same, satisfying

$$\phi^{\mathrm{z}}_{\mathrm{BM},l}(\Delta\tau_{\mathrm{BM},l} + t_0) = \phi^{\mathrm{z}}_{\mathrm{BM},1}(\Delta\tau_{\mathrm{BM},1} + t_0). \tag{47}$$

On the other hand, if it is a floor/ceiling reflection, the azimuth angle of the path should be equal to the azimuth angle of the LoS path. For such paths, a floor/ceiling path travels the same distance in the horizontal plane as the LoS path, satisfying

$$\begin{aligned}\sqrt{(\phi^{\mathrm{x}}_{\mathrm{BM},l})^2 + (\phi^{\mathrm{y}}_{\mathrm{BM},l})^2}(\Delta\tau_{\mathrm{BM},l} + t_0) \\ = \sqrt{(\phi^{\mathrm{x}}_{\mathrm{BM},1})^2 + (\phi^{y}_{\mathrm{BM},1})^2}(\Delta\tau_{\mathrm{BM},1} + t_0).\end{aligned} \tag{48}$$

Paths that do not satisfy the floor/ceiling or wall conditions are discarded from the path pool. Additional constraints exist for the path classification if both the DoA and the DoD are available. Floor/ceiling reflections satisfy $\theta^{\mathrm{el}}_{\mathrm{BM},l} = \phi^{\mathrm{el}}_{\mathrm{BM},l}$, whereas wall reflections satisfy $\theta^{\mathrm{el}}_{\mathrm{BM},l} = -\phi^{\mathrm{el}}_{\mathrm{BM},l}$. After classifying the NLoS paths, the estimate $\hat{t}_0$ is found by using the set of equations given by (47) and (48) [15]. Then the true $\tau_{\mathrm{BM},1}$ is computed, and the MS position is estimated using (46).

While we described the localization for the BS-MS link, the same approach also works for the case where the BS-MS link is obstructed, and a LoS path exists for the RIS-MS link. Let the position of the RIS be denoted by $\mathbf{r} \in \mathbb{R}^3$. Then, we can obtain an estimate of the MS position with the aid of the RIS, if we replace $\mathbf{b}$, $\tau_{\mathrm{BM},l}$, $\boldsymbol{\phi}_{\mathrm{BM},l}$ and $\boldsymbol{\theta}_{\mathrm{BM},l}$ with $\mathbf{r}$, $\tau_{\mathrm{RM},q}$, $\boldsymbol{\phi}_{\mathrm{RM},q}$ and $\boldsymbol{\theta}_{\mathrm{RM},q}$, respectively, in the equations above.

Algorithm 3 Proposed Localization Scheme

Input: $\mathcal{C}_{\mathrm{BM}}$, $\mathcal{C}_{\mathrm{RM}}$

```
Find the angular and delay parameters from C_BM and C_RM
Classify the paths in C_BM and C_RM
if LoS paths exist for both BS-MS and RIS-MS then
    Solve (49) to find t̂_0
else
    if LoS path exists for only BS-MS then
        Find t̂_0 using the paths C_BM in (47) and (48)
    else
        Find t̂_0 using the paths C_RM in (47) and (48)
    end if
end if
Find m̂ by using t̂_0 in the suitable equation in (49)
```

Output: $\hat{\mathbf{m}}$

B. Localization with Two LoS Paths

Now, we assume that LoS paths exist for both the BS-MS and the RIS-MS links. In this case, it is possible to find the location of the MS by using just the two LoS paths; the NLoS paths are not required, which means there is no dependency on the properties of the indoor environment. The key idea is that the MS position computed from (46) or the RIS-MS is the same, which leads to an equality expressed as

$$\begin{aligned}\mathbf{m} &= \mathbf{b} + c(\Delta\tau_{\mathrm{BM},1} + t_0)\boldsymbol{\phi}_{\mathrm{BM},1} \\ &= \mathbf{r} + c(\Delta\tau_{\mathrm{RM},1} + t_0)\boldsymbol{\phi}_{\mathrm{RM},1}.\end{aligned} \tag{49}$$

Using the estimated TDoAs, (49) is a simple linear equation with the unknown t_0, which can be solved accordingly. Then, the estimate of t_0 can be substituted in either of the equations to find the MS location. It is meaningful to use the LoS path that has a higher gain since it could possibly provide more accurate estimates. Note that (49) is constructed using only the DoDs and the delays, while it can also be constructed with DoAs and delays. The selection depends on the size of the array and the number of training sequences devoted to the estimation of the DoAs and the DoDs. The overall localization procedure is summarized in Algorithm 3.

VI. Numerical Results

We consider a ray tracing simulation implemented with Wireless InSite software for channel generation in an indoor factory environment, shown in Fig. 2. The generated channel dataset is available in [65]. The BS is located on the ceiling while the RIS is mounted on the wall at a height of $r_z = 5.5\mathrm{m}$. The location of the MS is randomly generated 100 times at a height of $m_z = 1.5\mathrm{m}$. Considering the communication system, the center frequency is set to $60\mathrm{GHz}$ while the bandwidth is $1\mathrm{GHz}$. The transmit power is set to $P_{\mathrm{t}} = 20\mathrm{dBm}$, whereas the noise variance is $\sigma^2 = -84\mathrm{dBm}$, which is the thermal noise at $15^{\circ}\mathrm{C}$ with the given bandwidth. The pulse shaping function is selected as $p(t) = \mathrm{sinc}(t)$. The delay tap length and the number of training symbols are set to $D = 64$ and $N = 64$, respectively. Training symbols are selected as the rows of the 64 element Hadamard matrix. the BS and the MS are equipped with 8×8 UPA with 8 RF chains and

Fig. 2. Indoor factory environment in Wireless-Insite. The BS is located on the ceiling, while the RIS is mounted on the wall. The RIS is represented by the red hemisphere on the left wall.

4×4 UPA with 4 RF chains, respectively. For the RIS, two different architectures with 16×16 and 32×32 elements are considered. The mutual coupling matrix and the gain/phase errors are generated by using the parameters in [53], [54], such that the off-diagonals of the mutual coupling matrix are randomly selected from a uniform distribution in the range $[0.01, 0.4]$. Furthermore, the gains are modeled as normal variables with distribution $\mathcal{N}(1, (0.05)^2)$, and the phase errors are modeled as $\mathcal{N}(0, (\pi/9)^2)$. Mutual coupling between two antennas is highly dependent on the circuit. Therefore, using random entries for the mutual coupling matrix covers a wide range of scenarios, including challenging ones. Consequently, we use random mutual coupling for the results unless specified otherwise.

Regarding the channel estimation algorithm, the dictionaries have a high resolution such that the ratio of the number of atoms to the size of the atoms in each dictionary is set to 128. Furthermore, the number of training frames, i.e., pilot overhead, for each setting is given in Table I. As the number of elements at the RIS increases, more training frames are required. Therefore, the number of training transmit configurations M_{B} is adjusted according to the number of antennas at the BS and the number of elements at the RIS. Note that the proposed approach is not optimized for reducing the pilot overhead, which would require the optimization of the training precoders, RIS phases, and combiners. We consider three different cases, namely, only the BS-MS channel is available, only the RIS-MS channel is available (i.e., the BS-MS link is blocked), and both channels are available. We assume that the link availability (or blockage) information is provided by an oracle. Although determining link availability is an interesting subject, it is beyond the scope of this work. Similar assumptions are used in many channel estimation works where the BS-MS link is assumed to be blocked [27], [28], [30], [32]–[37]. Moreover, the training combiners, which are the same for all cases, are generated using the Kronecker product of the DFT matrices for the x and y directions of the UPA at the MS. The entries of the precoders are selected with random phases for the case with only the BS-MS channel. For the case with only the RIS-MS channel, the precoders are chosen such

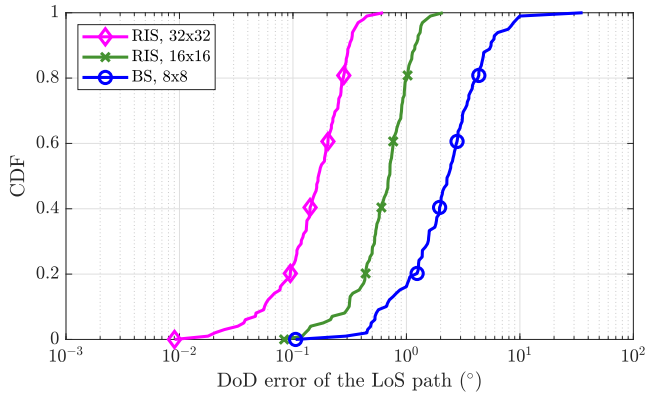
TABLE I
ARRAY ARCHITECTURES AND TRAINING FRAMES

BS	MS	RIS	M_{B}	M_{M}	Number of frames
8x8	4x4	-	8	4	32
8x8	4x4	16x16	32	4	128
8x8	4x4	32x32	128	4	512

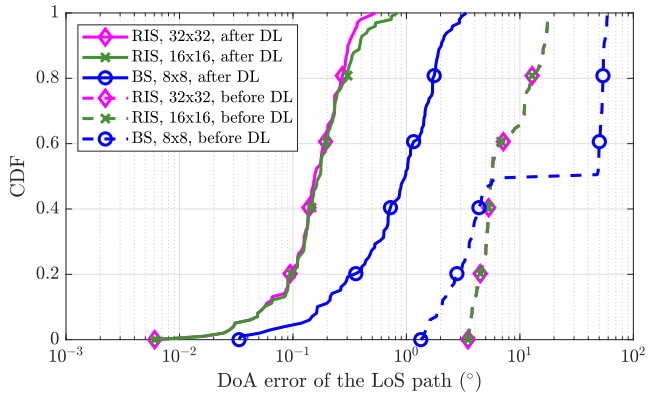
that the columns of the precoders are matched to the DoD of the known LoS path of the BS-RIS channel, i.e., $\phi_{\mathrm{BR},1}$. Finally, half of the precoders are selected with random phases, and the other half are matched to the LoS path of the BS-RIS channel for the case where both the BS-MS and the RIS-MS channels are available.

First, we evaluate the DoD and the DoA estimation accuracy of the LoS paths of the BS-MS and the RIS-MS channels. The cumulative distribution functions (CDFs) of the angle estimation errors are shown in Fig. 3. It is observed that the DoD estimation error of the RIS-MS channel is much lower than that of the BS-MS channel. Moreover, the accuracy increases with the increasing number of RIS elements. On the other hand, we observe that the DoA estimation error is high for all cases before the dictionary learning algorithm is applied. Since we consider ray tracing-based channels and randomly generated mutual coupling matrices for evaluating our proposed approach, the errors are more prominent for specific sets of angles before dictionary learning. Thus, we observe DoA estimation outage in some cases depending on the locations of the BS and the MS. After applying the dictionary learning algorithm, the DoA estimation error decreases significantly. Moreover, we investigate the delay estimation error. Since the estimated delays include the unknown clock offset, we evaluate the TDoA estimation error for the path with the lowest TDoA, i.e., the reflection that travels the shortest distance. The results, which are given in Fig. 4, show that the TDoA error is below $6\mathrm{ns}$ for all cases. The maximum ranging error for the reflected path is $1.8\mathrm{m}$. It should be noted that the TDoA error for the BS-MS channel is, in general, smaller than the TDoA error for the RIS-MS channel for both 16×16 and 32×32 RISs. This result is reasonable since BS-MS channel paths are stronger than other RIS-MS paths.

Furthermore, we investigate the CDFs of the channel estimation error and the localization error in Fig. 5. We consider the normalized mean square error (NMSE) to observe the channel estimation accuracy. The NMSE of the BS-MS channel estimate can be expressed as $\sum_d \|\hat{\mathbf{H}}_{\mathrm{BM},d} - \mathbf{H}_{\mathrm{BM},d}\|^2 / \sum_d \|\mathbf{H}_{\mathrm{BM},d}\|^2$, where $\hat{\mathbf{H}}_{\mathrm{BM},d}$ is the estimated channel. Similarly, the NMSE of the RIS-MS channel estimate can be computed as $\sum_d \|\hat{\mathbf{H}}_{\mathrm{RM},d} - \mathbf{H}_{\mathrm{RM},d}\|^2 / \sum_d \|\mathbf{H}_{\mathrm{RM},d}\|^2$, where $\hat{\mathbf{H}}_{\mathrm{RM},d}$ is the estimated channel. Since the BS-RIS-MS channel depends on the RIS configuration and we assume that the LoS path of the BS-RIS channel is strong and known, showing the NMSE of the RIS-MS channel is sufficient to assess the channel estimation performance of the channel through RIS. We observe that the NMSE of the BS-MS channel is higher than the NMSE of the RIS-MS channel, especially before the dictionary learning algorithm is applied. A similar observation can be made for the localization accuracy.



(a) CDF of the DoD estimation error with only BS-MS or RIS-MS channel.



(b) CDF of the DoA estimation error with only BS-MS or RIS-MS channel.

Fig. 3. CDF of the angle estimation error of the LoS path of either the BS-MS or the RIS-MS channel at $P_{\mathrm{t}} = 20\mathrm{dBm}$.

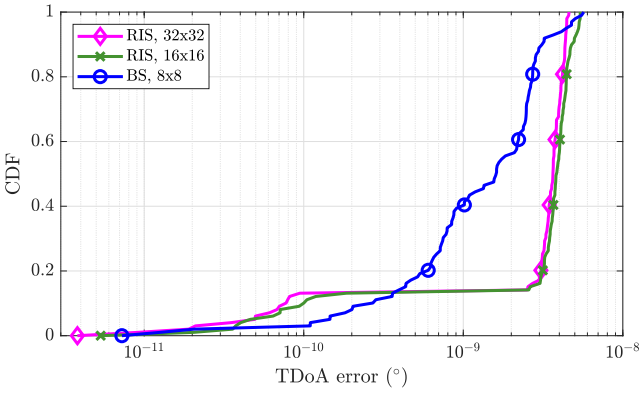
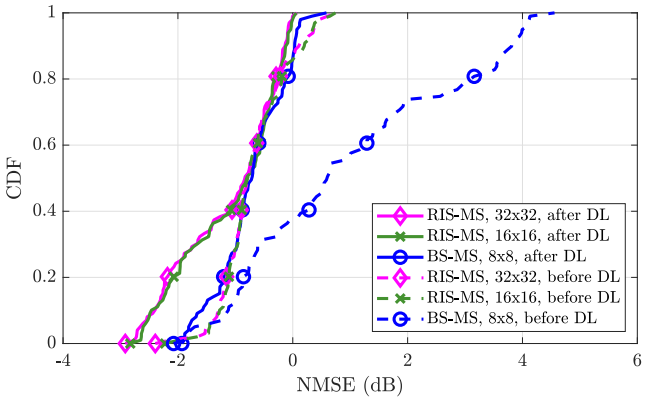
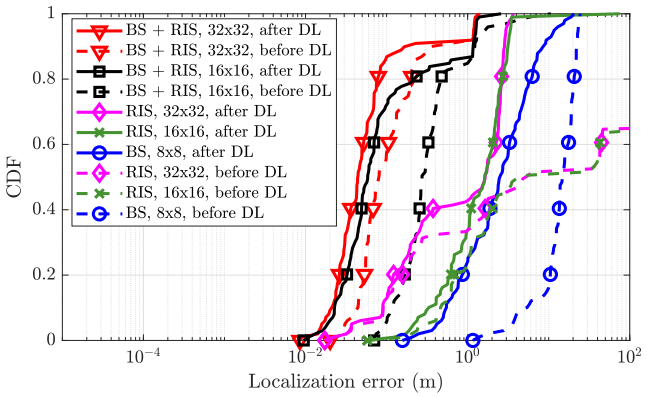


Fig. 4. CDF of the TDoA estimation error for the path with the lowest TDoA of either the BS-MS or the RIS-MS channel at $P_{\mathrm{t}} = 20\mathrm{dBm}$.

Note that the localization accuracy before dictionary learning for the cases where only the RIS-MS channel exists suffer from outage. The NLoS paths of the RIS-MS channel are weak due to the double path loss; thus, the lack of knowledge of the MS hardware impairments significantly affects the DoA estimation error. Since the localization method relies on a system of equations built using the information of the NLoS paths, the erroneous information results in low accuracy or no solution in some cases. One crucial observation is that localization error is significantly lower when the LoS paths are available for both channels. The main reason is that this method does not rely on NLoS paths. Consequently, the benefit of the RISs is



(a) CDF of the channel estimation NMSE.

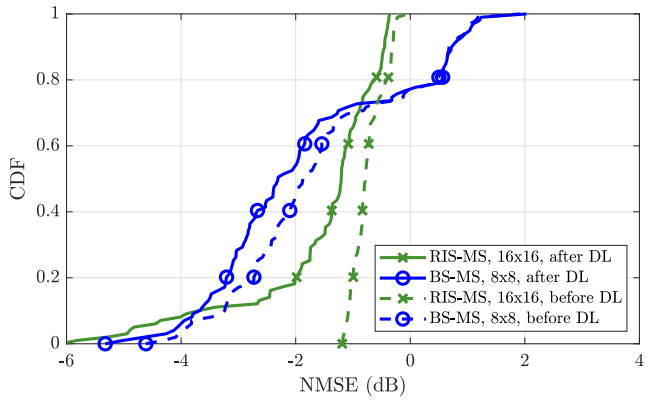


(b) CDF of the localization error.

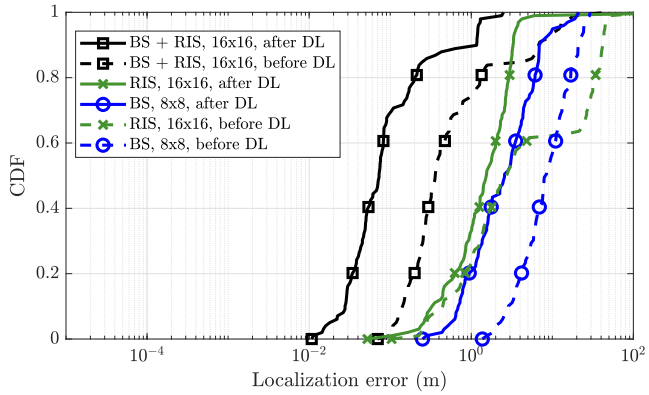
Fig. 5. CDF of the channel estimation NMSE and the localization error at $P_{\mathrm{t}} = 20\mathrm{dBm}$.

more significant if both LoS paths are available. Specifically, localization error below $10\mathrm{cm}$ is achieved for 80% of the cases. Additionally, we investigate the channel estimation and the localization error under a realistic mutual coupling effect. To that end, we generate the mutual impedances between the antennas using the analytical derivations and parameters in [51], where the antennas and the RIS elements are assumed to be cylindrical thin wires of perfectly conducting material. The obtained results with a 16×16 RIS are given in Fig. 6. It can be seen that the channel estimation NMSE for the BS-MS channel is improved compared to the random case, especially before the dictionary learning. For localization, we observe similar accuracy levels attained for the random case. That is, the localization error with the RIS after dictionary learning is lower than that with the BS. These results confirm that the proposed algorithm can operate with arbitrary mutual coupling matrices.

Finally, we analyze the dependency of the channel estimation and the localization accuracy on the transmit power in Fig. 7. We show the mean NMSE and the localization error for different cases and percentiles. The x-th percentile indicates that the first $x\%$ of the data points, sorted from lowest to highest, are used for plotting that curve. The aim of showing the results with different percentiles is to investigate the variation of the errors. The results are shown after the dictionary learning with a RIS of size 16×16. We observe that the channel estimation NMSE and the localization error with

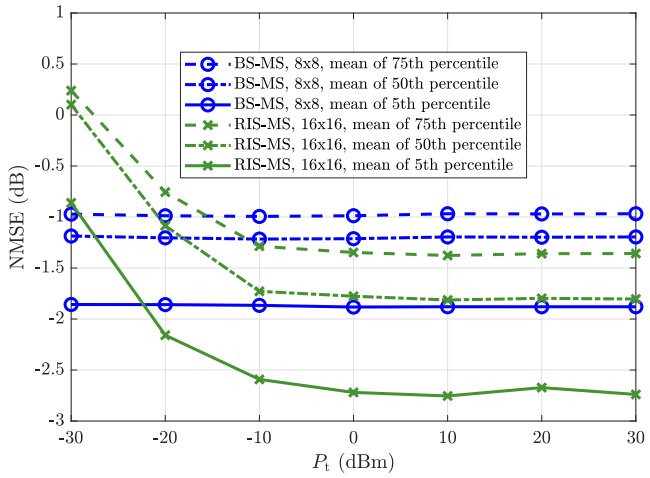


(a) CDF of the channel estimation NMSE.

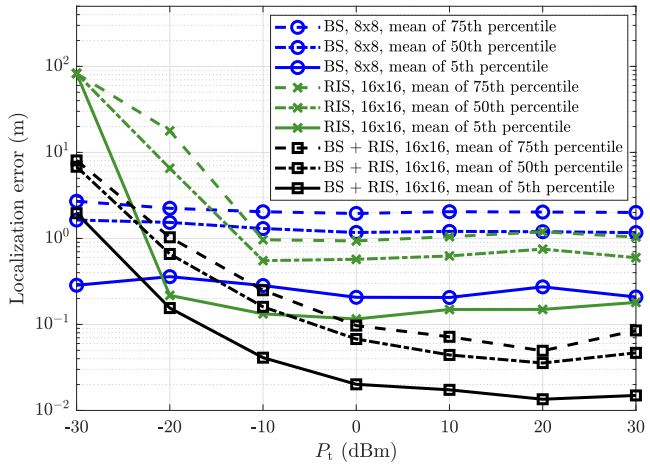


(b) CDF of the localization error.

Fig. 6. CDF of the channel estimation NMSE and the localization error at $P_\mathrm{t} = 20\mathrm{dBm}$ with realistic mutual coupling matrices.



(a) Channel estimation NMSE vs transmit power.



(b) Localization error vs transmit power.

Fig. 7. Channel estimation NMSE and the localization error with respect to transmit power.

only the BS-MS channel are relatively constant even at low transmit power. In contrast, localization errors of the cases that utilize the RIS-MS channel are much higher at low transmit power. The path through the BS-RIS-MS link suffers from double path loss, which is why the effective SNR of the signals received through this link is low. Nevertheless, it is observed that cm-level accuracy can be achieved with the availability of both the BS-MS and the RIS-MS channels at reasonable transmit power levels. The localization error can be reduced even further with larger RISs.

VII. Conclusion

We developed a low-complexity compressive channel estimation and dictionary learning strategy for a RIS-aided mmWave system, leveraging the MOMP algorithm. The proposed approach can jointly estimate the BS-MS and the RIS-MS channels while calibrating the user array with the assumption of the knowledge of the LoS path of the BS-RIS channel. We integrated this approach with an indoor localization strategy that can operate with or without the RIS. In the case where either the BS-MS or the RIS-MS channel is available, we made use of the NLoS paths and the properties of the indoor environment. We generated a set of realistic indoor channels using ray tracing, and we showed the significant improvement in localization accuracy that the RIS deployment can provide, even without perfect synchronization assumptions. The scenario where the LoS paths of both the BS-MS and the RIS-MS channels are available enables extremely high localization accuracy. These results illustrate the interest in exploiting RISs for sensing purposes, extending the benefits of this technology to situations where the LoS link between the BS and the MS is available.

References

[1] M. Bayraktar, J. Palacios, N. González-Prelcic, and C. J. Zhang, "Multidimensional orthogonal matching pursuit-based RIS-aided joint localization and channel estimation at mmWave," in *Proc. IEEE 23rd Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, 2022, pp. 1–5.

[2] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Towards dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, pp. 1–1, 2022.

[3] M. Koivisto, M. Costa, J. Werner, K. Heiska, J. Talvitie, K. Leppänen, V. Koivunen, and M. Valkama, "Joint device positioning and clock synchronization in 5G ultra-dense networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2866–2881, 2017.

[4] E. Y. Menta, N. Malm, R. Jäntti, K. Ruttik, M. Costa, and K. Leppänen, "On the performance of AoA–based localization in 5G ultra–dense networks," *IEEE Access*, vol. 7, pp. 33 870–33 880, 2019.

[5] Y. Lu, O. Kaltiokallio, M. Koivisto, J. Talvitie, E. S. Lohan, H. Wymeersch, and M. Valkama, "Bayesian filtering for joint multi-user positioning, synchronization and anchor state calibration," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10 949–10 964, 2023.

[6] J. Gante, G. Falcao, and L. Sousa, "Deep learning architectures for accurate millimeter wave positioning in 5G," *Neural Process. Lett.*, vol. 51, no. 1, pp. 487–514, 2020.

[7] C. Wu, X. Yi, W. Wang, L. You, Q. Huang, X. Gao, and Q. Liu, "Learning to localize: A 3D CNN approach to user positioning in massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4556–4570, 2021.

[8] G. Kia, L. Ruotsalainen, and J. Talvitie, "A CNN approach for 5G mm Wave positioning using beamformed CSI measurements," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, 2022, pp. 01–07.

[9] Z. Chen, Z. Zhang, Z. Xiao, C. Zhang, and Z. Yang, "CSI of each subcarrier is a fingerprint: Multi-carrier cumulative learning based positioning in massive MIMO systems," in *Proc. IEEE 34th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, 2023, pp. 1–7.

[10] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1822–1835, 2018.

[11] F. Jiang, Y. Ge, M. Zhu, and H. Wymeersch, "High-dimensional channel estimation for simultaneous localization and communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2021, pp. 1–6.

[12] F. Wen, J. Kulmer, K. Witrisal, and H. Wymeersch, "5G positioning and mapping with diffuse multipath," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1164–1174, 2021.

[13] J. Li, M. F. Da Costa, and U. Mitra, "Joint localization and orientation estimation in millimeter-wave MIMO OFDM systems via atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 70, pp. 4252–4264, 2022.

[14] J. Palacios, N. González-Prelcic, and C. Rusu, "Low complexity joint position and channel estimation at millimeter wave based on multidimensional orthogonal matching pursuit," in *Proc. 30th Eur. Signal Process. Conf. (EUSIPCO)*, 2022, pp. 1002–1006.

[15] J. Palacios, N. González-Prelcic, and C. Rusu, "Multidimensional orthogonal matching pursuit: Theory and application to high accuracy joint localization and communication at mmWave," *arXiv preprint arXiv:2208.11600*, 2022.

[16] M. A. Nazari, G. Seco-Granados, P. Johannisson, and H. Wymeersch, "mmWave 6D radio localization with a snapshot observation from a single BS," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 8914–8928, 2023.

[17] S. Weng, F. Jiang, and H. Wymeersch, "Wideband mmWave massive MIMO channel estimation and localization," *IEEE Wireless Commun. Lett.*, vol. 12, no. 8, pp. 1314–1318, 2023.

[18] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[19] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.

[20] 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 J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, 2020.

[21] 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, 2021.

[22] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, 2022.

[23] M. Rahal, B. Denis, K. Keykhosravi, M. F. Keskin, B. Uguen, G. C. Alexandropoulos, and H. Wymeersch, "Performance of RIS-aided near-field localization under beams approximation from real hardware characterization," *EURASIP J. Wireless Commun. Netw.*, vol. 2023, no. 1, p. 86, 2023.

[24] R. Liu, M. Li, H. Luo, Q. Liu, and A. L. Swindlehurst, "Integrated sensing and communication with reconfigurable intelligent surfaces: Opportunities, applications, and future directions," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 50–57, 2023.

[25] K. Keykhosravi, B. Denis, G. C. Alexandropoulos, Z. S. He, A. Albanese, V. Sciancalepore, and H. Wymeersch, "Leveraging RIS-enabled smart signal propagation for solving infeasible localization problems: Scenarios, key research directions, and open challenges," *IEEE Veh. Technol. Mag.*, vol. 18, no. 2, pp. 20–28, 2023.

[26] M. Jian, G. C. Alexandropoulos, E. Basar, C. Huang, R. Liu, Y. Liu, and C. Yuen, "Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques," *Intell. Converged Netw.*, vol. 3, no. 1, pp. 1–32, 2022.

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

[28] S. E. Zegrar, L. Afeef, and H. Arslan, "A general framework for RIS-aided mmWave communication networks: Channel estimation and mobile user tracking," *arXiv preprint arXiv:2009.01180*, 2020.

[29] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–6.

[30] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for intelligent reflecting surface-assisted millimeter wave MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.

[31] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, 2020.

[32] H. Liu, J. Zhang, Q. Wu, H. Xiao, and B. Ai, "ADMM based channel estimation for RISs aided millimeter wave communications," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 2894–2898, 2021.

[33] Y. Liu, S. Zhang, F. Gao, J. Tang, and O. A. Dobre, "Cascaded channel estimation for RIS assisted mmWave MIMO transmissions," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 2065–2069, 2021.

[34] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications—Part II: An improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, 2021.

[35] 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, 2021.

[36] J. Mirza and B. Ali, "Channel estimation method and phase shift design for reconfigurable intelligent surface assisted MIMO networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 441–451, 2021.

[37] Z. Zhou, B. Cai, J. Chen, and Y.-C. Liang, "Dictionary learning-based channel estimation for RIS-aided MISO communications," *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2125–2129, 2022.

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

[39] Z. Peng, C. Pan, G. Zhou, H. Ren, S. Jin, P. Popovski, R. Schober, and X. You, "Two-stage channel estimation for RIS-aided multiuser mmWave systems with reduced error propagation and pilot overhead," *IEEE Trans. Signal Process.*, vol. 71, pp. 3607–3622, 2023.

[40] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan, and M. Li, "Channel estimation with reconfigurable intelligent surfaces—A general framework," *Proc. IEEE*, vol. 110, no. 9, pp. 1312–1338, 2022.

[41] J. He, H. Wymeersch, L. Kong, O. Silvén, and M. Juntti, "Large intelligent surface for positioning in millimeter wave MIMO systems," in *Proc. IEEE Veh. Tech. Conf. (VTC2020-Spring)*, 2020, pp. 1–5.

[42] A. Elzanaty, A. Guerra, F. Guidi, and M.-S. Alouini, "Reconfigurable intelligent surfaces for localization: Position and orientation error bounds," *IEEE Trans. Signal Process.*, vol. 69, pp. 5386–5402, 2021.

[43] Y. Liu, E. Liu, R. Wang, and Y. Geng, "Reconfigurable intelligent surface aided wireless localization," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2021, pp. 1–6.

[44] J. He, H. Wymeersch, T. Sanguanpuak, O. Silven, and M. Juntti, "Adaptive beamforming design for mmWave RIS-aided joint localization and communication," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, 2020, pp. 1–6.

[45] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Channel estimation and user localization for IRS-assisted MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

[46] A. Albanese, P. Mursia, V. Sciancalepore, and X. Costa-Pérez, "PAPIR: Practical RIS-aided localization via statistical user information," in *Proc. IEEE 22nd Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, 2021, pp. 531–535.

[47] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Metalocalization: Reconfigurable intelligent surface aided multi-user wireless indoor localization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7743–7757, 2021.

[48] B. Teng, X. Yuan, R. Wang, and S. Jin, "Bayesian user localization and tracking for reconfigurable intelligent surface aided MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 1040–1054, 2022.

[49] G. Gradoni and M. Di Renzo, "End-to-end mutual coupling aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 938–942, 2021.

[50] R. Faqiri, C. Saigre-Tardif, G. C. Alexandropoulos, N. Shlezinger, M. F. Imani, and P. del Hougne, "PhysFad: Physics-based end-to-end channel modeling of RIS-parametrized environments with adjustable fading," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 580–595, 2023.

[51] P. Zheng, X. Ma, and T. Y. Al-Naffouri, "On the impact of mutual coupling on RIS-assisted channel estimation," *IEEE Wireless Commun. Lett.*, vol. 13, no. 5, pp. 1275–1279, 2024.

[52] P. Zheng, R. Wang, A. Shamim, and T. Y. Al-Naffouri, "Mutual coupling in RIS-aided communication: Experimental validation and performance evaluation," *arXiv preprint arXiv:2311.10544*, 2023.

[53] M. Eberhardt, P. Eschlwech, and E. Biebl, "Investigations on antenna array calibration algorithms for direction-of-arrival estimation," *Adv. Radio Sci.*, vol. 14, pp. 181–190, 2016.

[54] H. Xie and N. González-Prelcic, "Dictionary learning for channel estimation in hybrid frequency-selective mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7407–7422, 2020.

[55] P. Maity, S. Srivastava, S. Khatri, and A. K. Jagannatham, "Dictionary-learning (DL)-based sparse CSI estimation in multiuser terahertz (THz) hybrid MIMO systems under hardware impairments and beam-squint effect," *IEEE Access*, vol. 10, pp. 113 699–113 714, 2022.

[56] H. Xie, J. Palacios, and N. González-Prelcic, "Hybrid mmWave MIMO systems under hardware impairments and beam squint: Channel model and dictionary learning-aided configuration," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[57] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, 2018.

[58] A. Liao, Z. Gao, H. Wang, S. Chen, M.-S. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329–8345, 2019.

[59] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. 48th Asilomar Conf. Signals, Syst., Comput.*, 2014, pp. 273–277.

[60] K. Venugopal, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, 2017.

[61] 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 Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, 2018.

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

[63] F. Jiang, Y. Ge, M. Zhu, H. Wymeersch, and F. Tufvesson, "Low-complexity channel estimation and localization with random beamspace observations," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 5985–5990.

[64] J. He, A. Fakhreddine, H. Wymeersch, and G. C. Alexandropoulos, "Compressed-sensing-based 3D localization with distributed passive reconfigurable intelligent surfaces," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2023, pp. 1–5.

[65] "Ray tracing channel dataset for RIS-aided mmWave system," https://github.com/WiSeCom-Lab/RIS-ray-tracing.

[66] "RIS-aided joint channel estimation and localization at mmWave ," https://github.com/WiSeCom-Lab/RIS-localization.

[67] G. C. Alexandropoulos, K. Stylianopoulos, C. Huang, C. Yuen, M. Bennis, and M. Debbah, "Pervasive machine learning for smart radio environments enabled by reconfigurable intelligent surfaces," *Proc. IEEE*, vol. 110, no. 9, pp. 1494–1525, 2022.

[68] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[69] K. Engan, S. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 5, 1999, pp. 2443–2446 vol.5.