

Hybrid mmWave MIMO Systems under Hardware Impairments and Beam Squint: Channel Model and Dictionary Learning-aided Configuration

Hongxiang Xie, Joan Palacios and Nuria González-Prelcic *Senior Member, IEEE*

Abstract—Low overhead channel estimation based on compressive sensing (CS) has been widely investigated for hybrid wideband millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems. The channel sparsifying dictionaries used in prior work are built from ideal array response vectors evaluated on discrete angles of arrival/departure. In addition, these dictionaries are assumed to be the same for all subcarriers, without considering the impacts of hardware impairments and beam squint. In this manuscript, we derive a general channel and signal model that explicitly incorporates the impacts of hardware impairments, practical pulse shaping functions, and beam squint, overcoming the limitations of mmWave MIMO channel and signal models commonly used in previous work. Then, we propose a dictionary learning (DL) algorithm to obtain the sparsifying dictionaries embedding hardware impairments, by considering the effect of beam squint without introducing it into the learning process. We also design a novel CS channel estimation algorithm under beam squint and hardware impairments, where the channel structures at different subcarriers are exploited to enable channel parameter estimation with low complexity and high accuracy. Numerical results demonstrate the effectiveness of the proposed DL and channel estimation strategy when applied to realistic mmWave channels.

Index Terms—Dictionary learning, millimeter wave (mmWave), massive MIMO, beam squint, spatial wideband effect, hardware impairments, mutual coupling, antenna spacing error, channel estimation, sparse coding, dictionary update.

I. INTRODUCTION

The acquisition of channel state information (CSI) is crucial for mmWave link configuration, and challenging when operating with hybrid beamforming architectures. To reduce the training overhead associated with CSI acquisition, prior work has made full use of the sparse

nature of mmWave channels in the angular or delay domains [1]–[5]. Nevertheless, some relevant practical aspects have not been fully considered in previous compressive channel models and estimation algorithms: the beam squint effect, calibration errors, and hardware impairments. Specifically, the channel sparsifying dictionaries used in prior work are typically assumed to be (overcomplete) discrete Fourier transform (DFT) matrices, or constructed from the ideal array response matrices (IARM) evaluated on discrete grids of quantized angles of arrivals and departures (AoAs/AoDs) [2], [4]. These assumptions are valid, however, only when the beam squint effect is negligible and no hardware impairments or calibration errors exist. In this paper, we show that under hardware impairments such as mutual coupling or antenna separation disturbances, the array response vectors will no longer be the Vandermonde vectors, and that different array response vectors should be considered at every frequency for channel modeling under beam squint. In other words, the assumptions and modeling of wideband mmWave MIMO channels in prior work are not valid, and therefore, the prior CSI acquisition strategies are not effective when beam squint and hardware impairments are considered.

The impact of beamsquint has been analyzed in prior work. As shown in [6]–[8], the time delay of the same data symbol across the antenna array aperture is non-negligible in the large-scale MIMO configurations and/or the wideband systems. Due to the spatial delay difference of each data symbol at different antennas, the array steering vectors will have different responses at each frequency, what leads to the beam squint effect. The work presented in [6] includes the derivation of a channel model for the large-scale MIMO system under beam squint, showing that the array response vectors for channel modeling have to be frequency-dependent. Aware of this beam squint impact, [9] and [10] also considered the frequency domain channel models by using explicit frequency-dependent array response vectors for different subcarriers. Meanwhile, to avoid the impacts of beam

This material is based upon work partially supported by the National Science Foundation under Grant 2147955

Hongxiang Xie was with the Department of Electrical and Computer Engineering at the University of Texas at Austin when his work on this manuscript was performed.

Nuria González-Prelcic and Joan Palacios are with the Electrical and Computer Engineering Department, North Carolina State University, USA. The corresponding author is Nuria González-Prelcic: ngprel-cic@ncsu.edu.

squint on channel estimation at different subcarriers, [11]–[13] proposed to estimate the channel parameters from the perspective of angle and delay domains, together with user scheduling to alleviate inter-user interference. The works in [14] and [15] consider a massive MIMO setting at mmWave or THz bands with single antenna users, and propose solutions for beam tracking and channel estimation, respectively, considering beam squint, but they do not introduce the filtering effect or hardware impairments in the channel model. Generally, these prior works on channel estimation/tracking under beam squint do not consider beam squint combined with filtering effects or other hardware imperfections. In this manuscript, we will show that beam squint not only leads to the frequency-dependence on array steering vectors, but also yields additional distortions at different antennas across all subcarriers, especially on those at the band edge, when combined with the filtering operations at the transceivers. The new channel model that results from these considerations has not been derived in previous work. The pioneering work in [16] showed the need and impact of considering the pulse shaping function in the MIMO channel model, but beam squint was neglected. In summary, prior work did not consider the combined effect of beam squint and the filtering effects, and also neglected the impact of other hardware impairments such as calibration error, mutual coupling or antenna separation disturbances when developing the signal model and algorithms for a mmWave system operating with a hybrid MIMO architecture.

The inclusion of hardware impairments on MIMO channel models has been investigated in [17]–[20]. There are many different hardware impairments in the practical radio frequency (RF) chains, although three of them dominate the effects in the resulting channel. First, due to the manufacture and calibration errors, the antenna array will generate unexpected radiation patterns, including both gain and phase errors on each antenna element. Second, any perturbation on antenna locations or inter-element spacing between antenna elements will result in irregular linear arrays rather than perfect half-wavelength uniform linear arrays (ULAs). Finally, the antenna spacing disturbance also creates the mutual coupling effect between antenna elements. Taking into account all these hardware impairments, it is apparent that the aforementioned sparsifying dictionaries constructed from IARM evaluated at quantized angles are no longer the best choice for exploiting channel sparsity. Previous work has shown that dictionary learning (DL) is an effective technique to capture the underlying structure of mmWave MIMO channels associated to specific types of sites [21] or with various hardware impairments [18]–[20]. Specifically, [21] exploited the K-SVD algorithm to find a dictionary to represent a collection of observed

channel realizations. Following this idea, [18] proposed a joint uplink/downlink sparsifying DL algorithm for narrow band massive MIMO systems operating at lower frequencies. Our previous work [19], [20] further investigated the DL and channel estimation strategy for hybrid wideband mmWave MIMO systems under low SNR conditions. Nevertheless, the beam squint effect was not incorporated in previous work, and it cannot be ignored under certain relationships between the carrier frequency and the bandwidth [7].

Motivated by these limitations, we propose a DL-based channel estimation strategy for hybrid mmWave MIMO systems under the impacts of both hardware impairments and beam squint. The main contributions of this manuscript are summarized as follows:

- We derive a general wideband mmWave MIMO channel model under both hardware impairments and beam squint. The new model not only incorporates the hardware impairments of antenna spacing disturbances, gain/phase errors, and array mutual coupling, but also explicitly considers the impacts of combined pulse shaping, filtering and beam squint, showing the limitations of existing MIMO channel models with beam squint and the associated channel estimation schemes. The derivation of the combined effect of beam squint and pulse shaping/filtering is of particular interest. Previous literature does not provide an alternative channel model including all these effects.
- We propose a DL algorithm for finding sparsifying dictionaries that embed hardware impairments, which fully exploits the channel properties at different subcarriers under beam squint. This algorithm does not learn the impact of beam squint but rather considers its model. Comparing to existing DL strategies which has to learn a general dictionary for both hardware impairments and beam squint, the newly proposed DL scheme enables better adaptation to the impacts of hardware impairments and facilitates the management of the beam squint impact at different subcarriers.
- We design a novel orthogonal matching pursuit (OMP)-based algorithm for compressive channel estimation under beam squint, which exploits the simple structures of the new channel model at central subcarriers to obtain initial parameter estimates with low complexity. Then it compensates the additional distortions at side subcarriers induced by beam squint. In this way, the measurements at all subcarriers can be used to achieve higher parameter estimation accuracy with lower complexity.
- We evaluate the proposed DL and channel estimation algorithms via numerical simulations. Results show that the training overhead of channel estima-

tion with learned dictionaries can be significantly reduced compared to traditional dictionaries without considering hardware impairment or beam squint. This validates the developed channel models and corroborates the effectiveness of the proposed DL algorithms for hybrid wideband mmWave MIMO systems under both hardware impairments and beam squint effects.

Notations: Vectors and matrices are denoted by bold-face small and capital letters; the transpose, conjugate, Hermitian (conjugate transpose), inverse, and pseudo-inverse of the matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\bar{\mathbf{A}}$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$; $\mathbf{I}_M$ is an $M \times M$ identity matrix; $[\mathbf{a}]_n$ denotes the n -th element of $\mathbf{a}$ and $[\mathbf{A}]_{:,j}$ denotes the j -th column vector of $\mathbf{A}$; $[\mathbf{A}]_{m,n}$ denotes the (m, n) -th element of $\mathbf{A}$; $\triangleq$ represents new definitions; $\mathcal{I}(N) \triangleq \{0, 1, \dots, N-1\}$ denotes the index set of cardinality N ; $\mathbb{C}$ and $\mathbb{R}$ denote the sets of complex and real numbers; $\text{tr}\{\mathbf{A}\}$ is the trace of $\mathbf{A}$; $\text{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with its diagonal elements given in $\mathbf{a}$ and $\text{diag}\{\mathbf{A}\}$ formulates a vector by extracting the diagonal elements of $\mathbf{A}$; $\otimes$, $\odot$ and $\star$ denote the Kronecker, Hadamard and Khatri-Rao product between vectors/matrices; $j = \sqrt{-1}$ denotes the imaginary unit; $|\mathcal{K}|$ denotes the cardinality of a set; $\lfloor x \rfloor$ denotes the largest integer less than or equal to x and $\lceil x \rceil$ denotes the smallest integer great than or equal to x ; $\mathcal{F}(\cdot)[f]$ denotes the Fourier transform evaluated at frequency f ; and $\|\mathbf{a}\|_0$ denotes the ℓ_0 norm of vector, i.e., the number of nonzero elements of $\mathbf{a}$.

II. NEW SIGNAL AND CHANNEL MODELS INCORPORATING HARDWARE IMPAIRMENTS AND BEAM SQUINT

In this section, we derive the new general time domain signal model and frequency domain channel model for the hybrid mmWave MIMO systems under the impact of both hardware impairments and beam squint. We will prove that the channel models with beam squint assumed in previous works [6], [9]–[13] are not complete or valid in some cases.

We consider a fully connected hybrid mmWave MIMO system. The transmitter (TX) is equipped with N_T antennas and L_T RF chains, and the receiver (RX) has N_R antennas and L_R RF chains. The channel between the TX and the RX is frequency-selective. Pulse shaped orthogonal frequency division multiplexing (OFDM) with K subcarriers is considered to simultaneously transmit N_s ($\leq \min(L_T, L_R)$) data streams. The system sampling period is denoted by T_s . We also consider a pulse shaping function with a roll-off factor β and an overall system bandwidth of $(1 + \beta)/T_s$. The center carrier frequency and wavelength are denoted by f_c and $\lambda_c = c/f_c$ (with

c the speed of light). We use the index k to denote the frequency domain subcarriers, with

$$f_k = f_c + \Delta f_k = f_c - \frac{1}{2T_s} + \frac{k}{KT_s}, \quad k = 0, \dots, K-1. \quad (1)$$

the corresponding subcarrier frequency value [22].

Let us start by defining a model for practical antenna arrays. As shown in [17]–[19], [23], [24], various impairments exist in practical implementations of antenna arrays, such as the gain/phase errors on each antenna element, the errors at each antenna element location, and the coupling effects between antenna elements. Specifically, we denote $\mathbf{C}_R \in \mathbb{C}^{N_R \times N_R}$ as the symmetric mutual coupling matrix for the RX antenna array, representing the unwanted interchange of energy between elements in the arrays, and denote $\gamma_R \in \mathbb{C}^{N_R \times 1}$ with $[\gamma_R]_{n_R} = g_{R,n_R} e^{j\nu_{R,n_R}}$ as the antenna gain and phase errors, in which g_{R,n_R} and ν_{R,n_R} are the receive gain error normalized to a reference amplitude, and the additional receive phase error for the n_R -th antenna element. Moreover, let $\epsilon_R \in \mathbb{R}^{N_R \times 1}$ with $[\epsilon_R]_{n_R} = \epsilon_{R,n_R}$ be the vector of antenna location errors at all RX antenna elements, where the location error of the first antenna element is normalized to $\epsilon_{R,0} = 0$. The variables defining the hardware impairments for the TX antenna array are defined in an analogous way.

We define now the signal model for the transmitted signal during training. To sound the channel, the TX sends Q symbols over a $L_T \times 1$ signal vector $\mathbf{s}(t)$ at time instant t , where the l_T -th ($l_T = 1, \dots, L_T$) element of the complex exponential representation of the transmitted signal $\mathbf{s}(t)$ is defined as

$$s_{l_T}(t) \triangleq [\mathbf{s}(t)]_{l_T} = \sum_{q=1}^Q [\mathbf{F}_{BB} \mathbf{S}]_{l_T, q} p_T(t - (q-1)T_s) e^{j2\pi f_c t}, \quad (2)$$

where $p_T(\cdot)$ is the transmit pulse, $e^{j2\pi f_c t}$ models the RF upconversion stage, $\mathbf{S} \in \mathbb{C}^{L_T \times Q}$ collects all the time domain signal symbols to be transmitted and $\mathbf{F}_{BB} \in \mathbb{C}^{L_T \times L_T}$ is the digital precoder. Moreover, $\mathbf{s}(t)$ satisfies $\mathbb{E}\{\mathbf{s}(t)\mathbf{s}(t)^*\} = \frac{P_T}{L_T} \mathbf{I}_{L_T}$, with P_T the transmit power constraint. Using an analog precoder $\mathbf{F}_{RF} \in \mathbb{C}^{N_T \times L_T}$, the ideal transmitted signal at the n_T -th antenna element, $n_T = 1, \dots, N_T$, can be computed as

$$i_{n_T}(t) = \sum_{l_T=1}^{L_T} [\mathbf{F}_{RF}]_{n_T, l_T} [\mathbf{s}(t)]_{l_T}. \quad (3)$$

Due to hardware imperfections, the signal at the n_T -th transmit antenna is multiplied by a phase and amplitude perturbation $[\gamma_T]_{n_T}^*$, and then leaks into the signal at the n'_T -th transmit antenna due to mutual coupling by a

$$[\mathbf{y}[d]]_{l_R} = \sum_{\ell, n_R, n_T, l_T} \alpha_l^P [\check{\mathbf{W}}]_{n_R, l_R}^* [\check{\mathbf{F}}]_{n_T, l_T} \sum_{q=1}^Q p((d-q)T_s - \tau_{\ell, n_R, n_T}) e^{-j2\pi f_c \tau_{\ell, n_R, n_T}} [\mathbf{s}[q]]_{l_T} + [\check{\mathbf{W}}]_{n_R, l_R}^* z_{n_R}(dT_s), \quad (10)$$

$$[\mathbf{y}[k]]_{l_R} = \sum_{\ell, n_R, n_T, l_T} \alpha_l^P [\check{\mathbf{W}}]_{n_R, l_R}^* [\check{\mathbf{F}}]_{n_T, l_T} g(k, \tau_{\ell, n_R, n_T}) e^{-j2\pi(f_c + \Delta f_k) \tau_{\ell, n_R, n_T}} [\mathbf{s}[k]]_{l_T} + [\check{\mathbf{W}}]_{n_R, l_R}^* [\mathbf{z}[k]]_{n_R}, \quad (11)$$

factor $[\mathbf{C}_T]_{n_T, n'_T}^*$, leading to the expression of the actual transmitted signal

$$\begin{aligned} x_{n_T}(t) &= \sum_{n'_T=1}^{N_T} [\mathbf{C}_T]_{n'_T, n_T}^* [\gamma_T]_{n'_T}^* i_{n'_T}(t) \\ &= \sum_{n'_T=1}^{N_T} [(\text{diag}(\gamma_T) \mathbf{C}_T)^*]_{n_T, n'_T} i_{n'_T}(t). \end{aligned} \quad (4)$$

Assuming a passband geometric channel model with L paths, the ideal received signal at the n_R -th antenna element, $n_R = 1, \dots, N_R$, can be written as

$$i_{n_R}(t) = \sum_{\ell=1}^L \sum_{n_T=1}^{N_T} \alpha_l^P \cdot x_{n_T}(t - \tau_{\ell, n_R, n_T}) + z_{n_R}(t), \quad (5)$$

where $z_{n_R}(t)$ is the noise term, $\alpha_l^P \in \mathbb{C}$ is the path gain of the passband channel including phase changes, and τ_{ℓ, n_R, n_T} is the delay of ℓ -th path between the n_R -th receive antenna and the n_T -th transmit antenna. Similarly to what happens at the transmit array, owing to hardware imperfections, the signal at each receive antenna n_R leaks into the signal at each antenna n'_R through mutual coupling by a factor $[\mathbf{C}_R]_{n'_R, n_R}$, and then receives a gain/phase weight $[\gamma_R]_{n'_R}$ to generate the actual receive signal

$$\begin{aligned} y_{n_R}(t) &= \sum_{n'_R=1}^{N_R} [\gamma_R]_{n'_R} [\mathbf{C}_R]_{n_R, n'_R} i_{n'_R}(t) \\ &= \sum_{n'_R=1}^{N_R} [\text{diag}(\gamma_R) \mathbf{C}_R]_{n_R, n'_R} i_{n'_R}(t). \end{aligned} \quad (6)$$

After an analog combining stage modeled by the matrix $\mathbf{W}_{RF} \in \mathbb{C}^{N_R \times L_R}$, we obtain the received signal vector $\mathbf{y}'(t) \in \mathbb{C}^{L_R \times 1}$, with the l_R -th element given as

$$[\mathbf{y}'(t)]_{l_R} = \sum_{n_R=1}^{N_R} [\mathbf{W}_{RF}]_{n_R, l_R}^* y_{n_R}(t). \quad (7)$$

To obtain the downconverted signal we get rid of the complex representation of the signal and multiply by $2e^{-j2\pi f_c t}$ to get $[\mathbf{y}''(t)]_{l'_R} = ([\mathbf{y}'(t)]_{l'_R} +$

$[\mathbf{y}'(t)]_{l'_R}^*) e^{-j2\pi f_c t}$. After matched filtering and sampling with sampling period T_s , we have

$$[\mathbf{y}_{BB}[d]]_{l_R} = (p_R * [\mathbf{y}''(t)]_{l'_R})(d-1)T_s, \quad d \in \mathbb{N}, \quad (8)$$

where p_R is the matched filter. Since the bandwidths of p_T and p_R , namely B_T and B_R , satisfy $B_T + B_R \leq 2f_c$ and $[\mathbf{y}']_{l'_R}^*(t) e^{-j2\pi f_c t}$ is centered at $-2f_c$, the convolution between the matched filter and the conjugate signal is zero, i.e. $p_R(t) * ([\mathbf{y}']_{l'_R}^*(t) e^{-j2\pi f_c t}) = 0$. After the digital combining stage represented by $\mathbf{W}_{BB}$, the discrete time received signal can be written as

$$[\mathbf{y}[d]]_{l_R} = \sum_{l'_R} [\mathbf{W}_{BB}]_{l'_R, l_R}^* [\mathbf{y}_{BB}[d]]_{l_R}, \quad d \in \mathbb{N}. \quad (9)$$

By developing the previous steps, we reach an equivalent expression for the received signal as shown in 10, at the top of this page, where $\check{\mathbf{W}} \triangleq (\text{diag}(\gamma_R) \mathbf{C}_R)^* \mathbf{W}$ and $\check{\mathbf{F}} \triangleq (\text{diag}(\gamma_T) \mathbf{C}_T)^* \mathbf{F}$, being $\mathbf{W} = \mathbf{W}_{RF} \mathbf{W}_{BB}$ and $\mathbf{F} = \mathbf{F}_{RF} \mathbf{F}_{BB}$ the hybrid combiner and precoder, respectively, while $p = p_R * p_T$ combines the effects of transmit and receive filters. Next, following the derivation in the Appendix I, the DFT of (10) is computed as shown in (11), where $g(k, \tau_{\ell, n_R, n_T})$ is defined in (49) in the Appendix I, representing the frequency response at the k -th subcarrier of the combined filter p for the ℓ -th path between the n_R -th RX antenna and the n_T -th TX antenna. Moreover, $[\mathbf{s}[k]]_{l_T}$ is the frequency domain version of the transmitted signal defined in (2), while $[\mathbf{z}[k]]_{n_R} \triangleq \sum_{d=0}^{K-1} z_{n_R}(dT_s) e^{-j\frac{2\pi d k}{K}}$ is the noise term in the frequency domain. Regarding the delay τ_{ℓ, n_R, n_T} , as shown in Fig. 1, it can be expressed as

$$\begin{aligned} \tau_{\ell, n_R, n_T} &= \tau_{\ell} + (n_R d_R + \epsilon_{R, n_R}) \sin(\phi_{\ell}) / c \\ &\quad - (n_T d_T + \epsilon_{T, n_T}) \sin(\theta_{\ell}) / c. \end{aligned} \quad (12)$$

Now, by defining the frequency-dependent array steering vectors

$$\check{\mathbf{a}}_{R, k}(\phi) = \text{diag}(\gamma_R) \mathbf{C}_R \check{\mathbf{a}}_{R, k}(\phi), \quad (13)$$

$$\check{\mathbf{a}}_{T, k}(\theta) = \text{diag}(\gamma_T) \mathbf{C}_T \check{\mathbf{a}}_{T, k}(\theta), \quad (14)$$

$$\mathbf{y}[k] = \mathbf{W}^* \sum_{\ell=1}^L \left(\alpha_{\ell} e^{-j2\pi \Delta f_k \tau_{\ell}} \mathbf{G}_k(\tau_{\ell}, \phi_{\ell}, \theta_{\ell}) \odot (\check{\mathbf{a}}_{R,k}(\phi_{\ell}) \check{\mathbf{a}}_{T,k}^*(\theta_{\ell})) \right) \mathbf{F} \hat{\mathbf{s}}[k] + \check{\mathbf{W}}^* \mathbf{z}[k]. \quad (18)$$

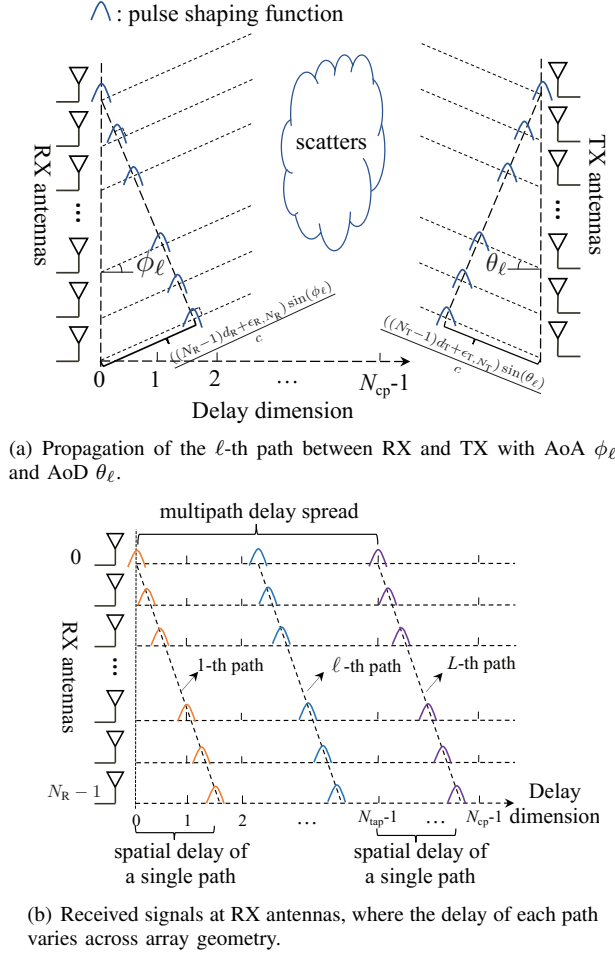


Fig. 1: Illustration of the spatial wideband effect (from the perspective of time domain) or the beam squint effect (from the perspective of frequency domain) under large-scale antenna array regimes.

with

$$[\check{\mathbf{a}}_{R,k}(\phi)]_n = \frac{1}{\sqrt{N_R}} e^{-j2\pi f_k (nd_R + \epsilon_{R,n}) \sin(\phi)/c} \quad (15)$$

$$[\check{\mathbf{a}}_{T,k}(\theta)]_n = \frac{1}{\sqrt{N_T}} e^{-j2\pi f_k (nd_T + \epsilon_{T,n}) \sin(\theta)/c}, \quad (16)$$

denoting the distortion matrix by

$$[\mathbf{G}_k(\tau_{\ell}, \phi_{\ell}, \theta_{\ell})]_{n_R, n_T} = g(k, \tau_{\ell, n_R, n_T}), \quad (17)$$

and defining the equivalent complex gain $\alpha_{\ell} = \alpha_{\ell}^P e^{-j2\pi f_c \tau_{\ell}}$, we can rewrite (11) as shown in 18, at the

top of this page. Based on this expression, we can define the general frequency domain channel matrix as

$$\mathbf{H}[k] = \sum_{\ell=1}^L \alpha_{\ell} e^{-j2\pi \Delta f_k \tau_{\ell}} \mathbf{G}_k(\tau_{\ell}, \phi_{\ell}, \theta_{\ell}) \odot (\check{\mathbf{a}}_{R,k}(\phi_{\ell}) \check{\mathbf{a}}_{T,k}^*(\theta_{\ell})), \quad (18)$$

and then, the final expression of the received signal under these definitions becomes

$$\mathbf{y}[k] = \mathbf{W}^* \mathbf{H}[k] \mathbf{F} \hat{\mathbf{s}}[k] + \check{\mathbf{W}}^* \mathbf{z}[k]. \quad (19)$$

Note that the equivalent combiner that impacts the noise term includes the hardware impairments as well, and thereby the noise covariance matrix is $\mathbf{C}_{\mathbf{z}[k]} = \sigma^2 \check{\mathbf{W}}^* \check{\mathbf{W}}$.

We define $g(k, \tau)$ as shown in (20), where $\mathcal{F}(p)[\Delta f_k]$ is the Fourier transform of the combined filter $p(t)$ evaluated for the frequency difference Δf_k . Then, the general channel matrix in (18) can be rewritten as in (21), where the second case is due to the fact that when $|\Delta f_k| \leq (1 - \beta)/2T_s$, the factor $g(k, \tau_{\ell, n_R, n_T}) = \frac{\mathcal{F}(p)[\Delta f_k]}{T_s}$ is independent of the delay τ_{ℓ, n_R, n_T} at different antenna indices, and thus enables the simplification. Note that (21) is valid for any pulse shaping function (as shown in the derivation in Appendix I), and therefore, this does not introduce any new assumption on the channel model.

Remark 1. We have derived the new general model for MIMO channels under both hardware impairments and beam squint. This new model shows that the beam squint not only induces the frequency-dependence on array steering vectors, but also yields additional distortions at different antennas across all subcarriers, especially on side subcarriers. Comparing (21) to the existing MIMO channel models with beam squint in [6], [9]–[13], it is obvious that those models only considered the second case in (21), and assumed the same expression for all subcarriers. In other words, the prior work only considered the frequency-dependence impact of beam squint, but ignored the additional distortions. This is due to the fact that the model was derived in the continuous time domain, and did not take into account the impact of the extra bandwidth of the combined filter (as shown in Appendix I). For ease of subsequent exposition, we will denote the set of central subcarriers by $\mathcal{K}_{cen} = \{k | |\Delta f_k| \leq (1 - \beta)/2T_s\}$ and the set of side subcarriers by $\mathcal{K}_{side} = \{k | |\Delta f_k| > (1 - \beta)/2T_s\}$. As shown in Fig. 2, there are approximately $|\mathcal{K}_{cen}| \approx \lfloor (1 - \beta)K \rfloor$

$$g(k, \tau) = \begin{cases} \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} + \frac{\mathcal{F}(p)[\Delta f_k - 1/T_s]}{T_s} e^{j2\pi\tau/T_s} & \text{if } 2\Delta f_k > (1 - \beta)/T_s, \\ \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} & \text{if } 2|\Delta f_k| \leq (1 - \beta)/T_s, \\ \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} + \frac{\mathcal{F}(p)[\Delta f_k + 1/T_s]}{T_s} e^{-j2\pi\tau/T_s} & \text{if } 2\Delta f_k < -(1 - \beta)/T_s, \end{cases} \quad (20)$$

$$\mathbf{H}[k] = \begin{cases} \sum_{\ell=1}^L \alpha_\ell e^{-j2\pi\Delta f_k \tau_\ell} \mathbf{G}_k(\tau_\ell, \phi_\ell, \theta_\ell) \odot (\check{\mathbf{a}}_{R,k}(\phi_\ell) \check{\mathbf{a}}_{T,k}^*(\theta_\ell)), & \forall |\Delta f_k| > (1 - \beta)/2T_s \\ \sum_{\ell=1}^L \alpha_\ell \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} e^{-j2\pi\Delta f_k \tau_\ell} \check{\mathbf{a}}_{R,k}(\phi_\ell) \check{\mathbf{a}}_{T,k}^*(\theta_\ell), & \forall |\Delta f_k| \leq (1 - \beta)/2T_s \end{cases} \quad (21)$$

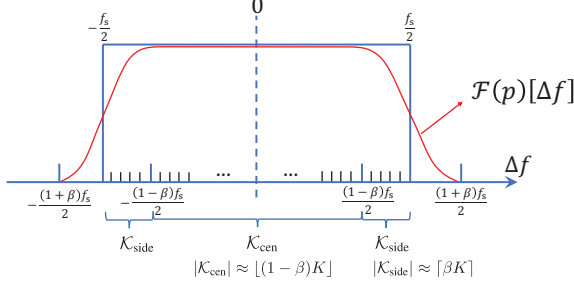


Fig. 2: Frequency response of the pulse shaping function with roll-off factor β , center frequency f_c and $f_s = 1/T_s$. For OFDM systems with K subcarriers, there are approximately $\lfloor (1 - \beta)K \rfloor$ central subcarriers inside $\mathcal{K}_{cen}$, and $\lceil \beta K \rceil$ side subcarriers inside $\mathcal{K}_{side}$.

central subcarriers inside $\mathcal{K}_{cen}$ and $|\mathcal{K}_{side}| \approx \lceil \beta K \rceil$ side subcarriers inside $\mathcal{K}_{side}$.

A. Frequency domain signal model for training

With the aforementioned system and channel model, the received signal model for training is described as follows. During the training phase, the TX will send pilot signals over several OFDM frames, so the RX can collect measurements for initial channel estimation and DL. Specifically, during the m -th ($m = 1, 2, \dots$) training OFDM symbol, the TX will send an $L_T \times 1$ pilot signal $\mathbf{s}_m[k]$ at the k -th subcarrier using a frequency-flat precoder $\mathbf{F}_m \in \mathbb{C}^{N_T \times L_T}$, while the RX employs a frequency-flat combiner $\mathbf{W}_m \in \mathbb{C}^{N_R \times L_R}$. The received signal is then given as

$$\mathbf{y}_m[k] = \mathbf{W}_m^* \mathbf{H}[k] \mathbf{F}_m \mathbf{s}_m[k] + \mathbf{n}_m[k], \quad (22)$$

where $\mathbf{n}_m[k] \in \mathbb{C}^{L_R \times 1}$ is the additive Gaussian noise vector, distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \check{\mathbf{W}}_m^* \check{\mathbf{W}}_m)$. Generally, we can decompose the transmitted signal as $\mathbf{s}_m[k] \triangleq \mathbf{q}_m s[k]$, with $\mathbf{q}_m \in \mathbb{C}^{L_T \times 1}$ a frequency-flat training vector and $s[k]$ a frequency-dependent training symbol. In doing so, we can multiply the received signal $\mathbf{y}_m[k]$

by $(s[k])^{-1}$ and get a frequency-flat observation matrix at the RX as follows

$$\begin{aligned} \tilde{\mathbf{y}}_m[k] &\triangleq \text{vec}(\mathbf{y}_m[k](s[k])^{-1}) \\ &= (\mathbf{q}_m^T \mathbf{F}_m^T \otimes \mathbf{W}_m^*) \text{vec}(\mathbf{H}[k]) + \tilde{\mathbf{n}}_m[k] \\ &= \Phi_m \text{vec}(\mathbf{H}[k]) + \tilde{\mathbf{n}}_m[k], \end{aligned} \quad (23)$$

where $\tilde{\mathbf{n}}_m[k] \triangleq \text{vec}(\mathbf{n}_m[k](s[k])^{-1})$ and Φ_m is defined accordingly, representing the sensing matrix for the m -th OFDM symbol. To get a higher effective SNR for the received measurements, we will use training spreading to average out the noise [19]. Moreover, to enable the initial channel estimate, the measurements over M OFDM symbols are stacked together such that

$$\begin{aligned} \underbrace{[\tilde{\mathbf{y}}_1[k]^T, \dots, \tilde{\mathbf{y}}_M[k]^T]^T}_{\tilde{\mathbf{y}}[k]} &= \underbrace{[\Phi_1^T, \dots, \Phi_M^T]^T}_{\Phi} \mathbf{h}[k] \quad (24) \\ &+ \underbrace{[\tilde{\mathbf{n}}_1[k]^T, \dots, \tilde{\mathbf{n}}_M[k]^T]^T}_{\tilde{\mathbf{n}}[k]}, \end{aligned} \quad (25)$$

where $\tilde{\mathbf{y}}[k] \in \mathbb{C}^{ML_R \times 1}$, $\Phi \in \mathbb{C}^{ML_R \times N_R N_T}$ and $\tilde{\mathbf{n}}[k] \in \mathbb{C}^{ML_R \times 1}$ are defined accordingly. Therefore, we can obtain the initial least squares (LS) channel estimate as $\hat{\mathbf{h}}[k] = \Phi^\dagger \tilde{\mathbf{y}}[k]$, while the corresponding channel matrix $\hat{\mathbf{H}}[k] = \text{unvec}(\hat{\mathbf{h}}[k])$.

During the training phase, it becomes necessary to collect initial channel measurements at different locations across the coverage area as the RX moves around. This not only creates a large training data set for learning the hardware imperfections, but also ensures the data set is diverse for the environment. In doing so, the learned sparsifying dictionary is not dedicated for a specific location, but adapted to the fixed hardware impairments. As in [19], this procedure of collecting measurements at multiple locations can be done at the stage of network setup, like in the case of an indoor WiFi scenario. Therefore, we assume the measurements are collected at N_{sa} locations and the initial channel estimate at the u -th location is then represented as $\hat{\mathbf{h}}^{(u)}[k]$ for $u \in \mathcal{I}(N_{sa})$. In the following, these initial channel estimates obtained at the training phases will be used for learning the hardware impairments related dictionaries.

III. DICTIONARY LEARNING FOR HARDWARE IMPAIRMENTS UNDER BEAM SQUINT (DLHWBS)

In this section, we present the problem formulation and optimization for learning the TX and RX dictionaries adapted to hardware impairments. Generally, the over-all dictionaries are partitioned into two parts, i.e., one frequency-flat part adapting to the hardware impairments and one frequency-dependent part accounting for the beam squint effect. The learning of hardware impairment related dictionaries will take into consideration of the channel properties under beam squint at different subcarriers.

A. Formulation of the dictionary learning problem

As shown in (21), the frequency domain channel model can be partitioned into two sets. We treat each one separately here. For central subcarriers inside $\mathcal{K}_{\text{cen}}$, the channel model can be simplified so that it is more convenient and straightforward to separate the hardware impairments and beam squint impact on the channel matrix. Let us first rewrite the channel matrix for $k \in \mathcal{K}_{\text{cen}}$ as follows

$$\begin{aligned} \mathbf{H}[k] &= \sum_{\ell=1}^L \alpha_{\ell} \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} e^{-j2\pi \Delta f_k \tau_{\ell}} \check{\mathbf{a}}_{\text{R},k}(\phi_{\ell}) \check{\mathbf{a}}_{\text{T},k}^*(\theta_{\ell}) \\ &= \check{\mathbf{A}}_{\text{R},k}(\phi) \mathbf{\Lambda}[k] \check{\mathbf{A}}_{\text{T},k}^*(\theta) \\ &\approx \check{\mathbf{A}}_{\text{R},k}^{\text{v}} \mathbf{\Omega}[k] \check{\mathbf{A}}_{\text{T},k}^{\text{v}*}, \end{aligned} \quad (26)$$

where $\mathbf{\Lambda}[k] \triangleq \text{diag}\left\{ \left[\alpha_1 \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} e^{-j2\pi \Delta f_k \tau_1}, \dots, \alpha_L \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} e^{-j2\pi \Delta f_k \tau_L} \right]^T \right\}$ is a diagonal matrix containing the frequency domain path gains, and $\check{\mathbf{A}}_{\text{R},k}(\phi) \triangleq [\check{\mathbf{a}}_{\text{R},k}(\phi_1), \dots, \check{\mathbf{a}}_{\text{R},k}(\phi_L)] \in \mathbb{C}^{N_{\text{R}} \times L}$ collects the receive array response vectors at all paths. Moreover, the approximation is obtained by discretizing the AoA/AoD spaces with on-grid angles, i.e., $\check{\mathbf{A}}_{\text{R},k}^{\text{v}} \triangleq [\check{\mathbf{a}}_{\text{R},k}(\phi_1^{\text{v}}), \dots, \check{\mathbf{a}}_{\text{R},k}(\phi_{K_{\text{R}}}^{\text{v}})] \in \mathbb{C}^{N_{\text{R}} \times K_{\text{R}}}$ and $\check{\mathbf{A}}_{\text{T},k}^{\text{v}} \triangleq [\check{\mathbf{a}}_{\text{T},k}(\theta_1^{\text{v}}), \dots, \check{\mathbf{a}}_{\text{T},k}(\theta_{K_{\text{T}}}^{\text{v}})] \in \mathbb{C}^{N_{\text{T}} \times K_{\text{T}}}$ collect the virtual receive and transmit array response vectors evaluated on K_{R} quantized angles $\phi^{\text{v}} = \{\phi_i^{\text{v}}\}_{i=1}^{K_{\text{R}}}$ for AoAs and K_{T} quantized angles $\theta^{\text{v}} = \{\theta_i^{\text{v}}\}_{i=1}^{K_{\text{T}}}$ for AoDs, and $\mathbf{\Omega}[k] \in \mathbb{C}^{K_{\text{R}} \times K_{\text{T}}}$ is a sparse matrix, containing the path gains of these discrete quantized AoAs/AoDs at its non-zero elements. To separate the impact of hardware impairments and beam squint on channel matrix, we can rewrite the n_{R} -th element of the array steering vector in (15) as

$$\begin{aligned} [\check{\mathbf{a}}_{\text{R},k}(\phi)]_{n_{\text{R}}} &= \frac{1}{\sqrt{N_{\text{R}}}} e^{-j2\pi f_k (n_{\text{R}} d_{\text{R}} + \epsilon_{\text{R},n_{\text{R}}}) \sin(\phi)/c} \\ &\approx e^{-j2\pi (f_c \epsilon_{\text{R},n_{\text{R}}}) \sin(\phi)/c} \\ &\quad \cdot \frac{1}{\sqrt{N_{\text{R}}}} e^{-j2\pi (f_k n_{\text{R}} d_{\text{R}}) \sin(\phi)/c}, \end{aligned} \quad (27)$$

so that

$$\check{\mathbf{a}}_{\text{R},k}(\phi) \approx \mathbf{e}_{\text{R}}(\phi) \odot \mathbf{a}_{\text{R},k}(\phi), \quad (28)$$

where $[\mathbf{e}_{\text{R}}(\phi)]_{n_{\text{R}}} = e^{-j2\pi (f_c \epsilon_{\text{R},n_{\text{R}}}) \sin(\phi)/c}$ comprises the hardware antenna element location imperfection effects. Therefore, the expression of $\check{\mathbf{A}}_{\text{R},k}^{\text{v}}$ in (26) can be rewritten as

$$\begin{aligned} \check{\mathbf{A}}_{\text{R},k}^{\text{v}} &\triangleq [\check{\mathbf{a}}_{\text{R},k}(\phi_1^{\text{v}}), \dots, \check{\mathbf{a}}_{\text{R},k}(\phi_{K_{\text{R}}}^{\text{v}})] \\ &= \underbrace{\text{diag}(\gamma_{\text{R}}) \mathbf{C}_{\text{R}}}_{\mathbf{D}_{\text{R},1}} \left(\underbrace{[\mathbf{e}_{\text{R}}(\phi_1^{\text{v}}), \dots, \mathbf{e}_{\text{R}}(\phi_{K_{\text{R}}}^{\text{v}})]}_{\mathbf{D}_{\text{R},2}} \odot \underbrace{[\mathbf{a}_{\text{R},k}(\phi_1^{\text{v}}), \dots, \mathbf{a}_{\text{R},k}(\phi_{K_{\text{R}}}^{\text{v}})]}_{\mathbf{A}_{\text{R},k}^{\text{v}}} \right) \\ &= \mathbf{D}_{\text{R},1} (\mathbf{D}_{\text{R},2} \odot \mathbf{A}_{\text{R},k}^{\text{v}}), \end{aligned} \quad (29)$$

where the general RX dictionary $\check{\mathbf{A}}_{\text{R},k}^{\text{v}}$ is partitioned into three parts, i.e., $\mathbf{D}_{\text{R},1}$ includes the impacts of antenna coupling and gain/phase errors, $\mathbf{D}_{\text{R},2}$ accounts for the antenna location perturbations, and $\mathbf{A}_{\text{R},k}^{\text{v}} \in \mathbb{C}^{N_{\text{R}} \times K_{\text{R}}}$ handles the beam squint effect at each subcarriers. Note that $\mathbf{D}_{\text{R},1}$ and $\mathbf{D}_{\text{R},2}$ are hardware impairments related dictionary components to be learned while $\mathbf{A}_{\text{R},k}^{\text{v}}$ is known for all subcarriers. Moreover, recalling the notation of antenna location error vector ϵ_{R} and the definition of $\mathbf{e}_{\text{R}}(\phi)$ in (27), we can further express $\mathbf{D}_{\text{R},2}$ as a function of the antenna location error vector ϵ_{R} , i.e., $\mathbf{D}_{\text{R},2} = f(\epsilon_{\text{R}}) = e^{-j2\pi f_c \epsilon_{\text{R}} \sin(\phi^{\text{v}})^T/c}$. Similarly, the general TX dictionary $\check{\mathbf{A}}_{\text{T},k}^{\text{v}}$ can be expressed as $\check{\mathbf{A}}_{\text{T},k}^{\text{v}} = \mathbf{D}_{\text{T},1} (\mathbf{D}_{\text{T},2} \odot \mathbf{A}_{\text{T},k}^{\text{v}})$ and $\mathbf{D}_{\text{T},2}$ is a function of the antenna location error vector ϵ_{T} , i.e., $\mathbf{D}_{\text{T},2} = f(\epsilon_{\text{T}}) = e^{-j2\pi f_c \epsilon_{\text{T}} \sin(\theta^{\text{v}})^T/c}$. Then the vectorization of $\mathbf{H}[k]$, $k \in \mathcal{K}_{\text{cen}}$ in (26) can be expressed as

$$\begin{aligned} \mathbf{h}[k] &= \text{vec}(\mathbf{H}[k]) \approx \left(\overline{\check{\mathbf{A}}_{\text{T},k}^{\text{v}}} \otimes \check{\mathbf{A}}_{\text{R},k}^{\text{v}} \right) \text{vec}(\mathbf{\Omega}[k]) \\ &= (\overline{\mathbf{D}_{\text{T},1}} \otimes \mathbf{D}_{\text{R},1}) \left[(\overline{\mathbf{D}_{\text{T},2}} \otimes \mathbf{D}_{\text{R},2}) \odot \left(\overline{\mathbf{A}_{\text{T},k}^{\text{v}}} \otimes \mathbf{A}_{\text{R},k}^{\text{v}} \right) \right] \text{vec}(\mathbf{\Omega}[k]). \end{aligned} \quad (30)$$

The channel model for side subcarriers $k \in \mathcal{K}_{\text{side}}$ in (21) can be represented and approximated as follows

$$\begin{aligned} \mathbf{H}[k] &= \sum_{\ell=1}^L \alpha_{\ell} e^{-j2\pi \Delta f_k \tau_{\ell}} \mathbf{G}_{\text{K}}(\tau_{\ell}, \phi_{\ell}, \theta_{\ell}) \\ &\quad \odot (\check{\mathbf{a}}_{\text{R},k}(\phi_{\ell}) \check{\mathbf{a}}_{\text{T},k}^*(\theta_{\ell})) \\ &\approx \sum_{l=1}^{L_{\text{v}}} \sum_{i=1}^{K_{\text{R}}} \sum_{j=1}^{K_{\text{T}}} b_{l,i,j} e^{-j2\pi \Delta f_k \tau_l^{\text{v}}} \mathbf{G}_{\text{K}}(\tau_l^{\text{v}}, \phi_i^{\text{v}}, \theta_j^{\text{v}}) \\ &\quad \odot (\check{\mathbf{a}}_{\text{R},k}(\phi_i^{\text{v}}) \check{\mathbf{a}}_{\text{T},k}^*(\theta_j^{\text{v}})), \end{aligned} \quad (35)$$

where the approximation is obtained by first discretizing the AoA/AoD spaces as did in (26) and then discretizing

$$\begin{aligned}
& \min_{\mathbf{D}_{T,1}, \mathbf{D}_{T,2}, \mathbf{D}_{R,1}, \mathbf{D}_{R,2}, \mathbf{\Omega}^{(u)}[k], \mathbf{b}_l^{(u)}} \sum_{u \in \mathcal{I}(N_{sa})} \sum_{k \in \mathcal{K}_{cen}} \left\| \hat{\mathbf{h}}^{(u)}[k] - \left(\overline{\check{\mathbf{A}}_{T,k}^v} \otimes \check{\mathbf{A}}_{R,k}^v \right) \text{vec}(\mathbf{\Omega}^{(u)}[k]) \right\|_F^2 \\
& + \sum_{u \in \mathcal{I}(N_{sa})} \sum_{k \in \mathcal{K}_{side}} \left\| \hat{\mathbf{h}}^{(u)}[k] - \sum_{l=1}^{L_v} \left[\Psi_k(\tau_l^v, \phi^v, \theta^v) \odot \left(\overline{\check{\mathbf{A}}_{T,k}^v} \otimes \check{\mathbf{A}}_{R,k}^v \right) \right] \mathbf{b}_l^{(u)} \right\|_F^2, \\
& \text{subject to} \quad \left\| \text{vec}(\mathbf{\Omega}^{(u)}[k]) \right\|_0 \leq L^{(u)}; \quad \left\| \mathbf{b}_l^{(u)} \right\|_0 \leq 1, \quad \sum_{l=1}^{L_v} \left\| \mathbf{b}_l^{(u)} \right\|_0 \leq L^{(u)}. \tag{33}
\end{aligned}$$

$$\begin{aligned}
& \min_{\mathbf{\Omega}[k], \mathbf{b}_l} \sum_{k \in \mathcal{K}_{cen}} \left\| \hat{\mathbf{h}}[k] - \left(\overline{\check{\mathbf{A}}_{T,k}^v} \otimes \check{\mathbf{A}}_{R,k}^v \right) \text{vec}(\mathbf{\Omega}[k]) \right\|_F^2 + \sum_{k \in \mathcal{K}_{side}} \left\| \hat{\mathbf{h}}[k] - \sum_{l=1}^{L_v} \left[\Psi_k(\tau_l^v, \phi^v, \theta^v) \odot \left(\overline{\check{\mathbf{A}}_{T,k}^v} \otimes \check{\mathbf{A}}_{R,k}^v \right) \right] \mathbf{b}_l \right\|_F^2, \\
& \text{subject to} \quad \left\| \text{vec}(\mathbf{\Omega}[k]) \right\|_0 \leq L; \quad \left\| \mathbf{b}_l \right\|_0 \leq 1, \quad \sum_{l=1}^{L_v} \left\| \mathbf{b}_l \right\|_0 \leq L. \tag{34}
\end{aligned}$$

the delay space with L_v on-grid delays $\{\tau_l^v\}_{l=1}^{L_v}$. Moreover, $b_{l,i,j}$ is the corresponding path gain on each pair of discretized AoA/AoD/delay grids, which theoretically is nonzero only at the discrete AoA/AoD/delay bin $(\phi_i^v, \theta_j^v, \tau_l^v)$ corresponding to the true AoA/AoD/delay $(\phi_\ell, \theta_\ell, \tau_\ell)$. Then the vectorization of $\mathbf{H}[k], k \in \mathcal{K}_{side}$ can be given as

$$\begin{aligned}
\mathbf{h}[k] &= \text{vec}(\mathbf{H}[k]) = \sum_{\ell=1}^L \alpha_\ell e^{-j2\pi \Delta f_k \tau_\ell} \text{vec}(\mathbf{G}_k(\tau_\ell, \phi_\ell, \theta_\ell)) \\
& \odot \left(\overline{\check{\mathbf{A}}_{T,k}^v}(\theta_\ell) \otimes \check{\mathbf{A}}_{R,k}^v(\phi_\ell) \right) \\
& \approx \sum_{l=1}^{L_v} \left[\Psi_k(\tau_l^v, \phi^v, \theta^v) \odot \left(\overline{\check{\mathbf{A}}_{T,k}^v} \otimes \check{\mathbf{A}}_{R,k}^v \right) \right] \mathbf{b}_l, \tag{32}
\end{aligned}$$

where $\Psi_k(\tau_l^v, \phi^v, \theta^v) \in \mathbb{C}^{N_R N_T \times K_R K_T}$ and its columns are defined as $[\Psi_k(\tau_l^v, \phi^v, \theta^v)]_{:, (j-1)K_T + i} = e^{-j2\pi \Delta f_k \tau_l^v} \text{vec}(\mathbf{G}_k(\tau_l^v, \phi_i^v, \theta_j^v)), \forall i = 1, \dots, K_R, j = 1, \dots, K_T$, and $\mathbf{b}_l \in \mathbb{C}^{K_R K_T \times 1}$ collects $b_{l,i,j}$, satisfying $\|\mathbf{b}_l\|_0 \leq 1$ and $\sum_{l=1}^{L_v} \|\mathbf{b}_l\|_0 \leq L$.

Remark 2. Comparing the channel approximation expressions for central subcarriers in (26) and for side subcarriers in (35), it is clear that the additional distortion matrix $\mathbf{G}_k(\tau_\ell, \phi_\ell, \theta_\ell)$ for side subcarriers entangles the three parameters of delay, AoA and AoD, such that a combination of discretized delay, AoA and AoD grids is needed. This will increase the sparsifying dictionary dimension significantly and induce overwhelming computational complexity during the CS recovery of these parameters. Instead, for the central subcarriers, the discretization of delay, AoA and AoD is decoupled without the additional distortion matrix and thus even one-dimensional search can be done sequentially for each

parameter, which will help reduce the computational complexity to a large extent.

Next, as per the approximate channel models in (26)-(30) as well as (35)-(32), we can formulate the final DL problem for hardware impairments. Specifically, stacking the initial channel estimates at all subcarriers from all locations $u \in \mathcal{I}(N_{sa})$, the problem formulation of DLHW can be expressed as shown in (33). Note that the two sum terms in (33) are format-consistent as $[\Psi_k(\tau_l^v, \phi^v, \theta^v)]_{:, (j-1)K_T + i} = \frac{\mathcal{F}(p)[f_k]}{T_s} e^{-j2\pi f_k \tau_l^v} \mathbf{1}_{N_R N_T}$ for central subcarriers $k \in \mathcal{K}_{cen}$. We will see that the simplified expression for central subcarriers can help facilitate the derivations of DL algorithms.

B. Speeding up the dictionary learning with a new sparse coding algorithm

The optimization problem in (33) is not jointly convex with respect to the variables $\mathbf{D}_{T,1}, \mathbf{D}_{T,2}, \mathbf{D}_{R,1}, \mathbf{D}_{R,2}, \mathbf{\Omega}^{(u)}[k]$, and $\mathbf{b}_l^{(u)}$, but it can still be solved by the alternating optimization techniques. As in the typical DL problems [25], the optimization of (33) is split into two stages: sparse coding and dictionary update.

1) *Sparse coding stage:* In this stage, we fix all dictionary parts and update the channel coefficients $\mathbf{\Omega}^{(u)}[k], \forall k \in \mathcal{K}_{cen}, u \in \mathcal{I}(N_{sa})$ and $\mathbf{b}_l^{(u)}, \forall k \in \mathcal{K}_{side}, u \in \mathcal{I}(N_{sa})$. Specifically, for each $u \in \mathcal{I}(N_{sa})$, the optimization problem of (33) is reduced to the expression in (34) (omitting the superscript $^{(u)}$ for simplicity). Generally, this problem can be optimized by various CS techniques, such as orthogonal matching pursuit (OMP) [26], simultaneous OMP (SOMP) [27], or simultaneous weighted OMP (SWOMP) [2], to name a few. In this paper, we propose a new sparse coding algorithm to solve (34), named *Dictionary Adaptive OMP under beam*

$$\begin{aligned} \min_{\mathbf{D}_{R,1}} \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{cen}}} & \left\| \hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{R,1} \underbrace{(\mathbf{D}_{R,2} \odot \mathbf{A}_{R,k}^{\text{v}}) \boldsymbol{\Omega}^{(u)}[k] (\mathbf{D}_{T,2} \odot \mathbf{A}_{T,k}^{\text{v}})^* \mathbf{D}_{T,1}^*}_{\mathbf{X}_{R,1}^{(u)}[k]} \right\|_F^2 \\ + \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{side}}} & \left\| \hat{\mathbf{H}}^{(u)}[k] - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}^{(u)}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \odot (\check{\mathbf{a}}_{R,k}(\hat{\phi}_{\hat{l}}^{(u)}) \check{\mathbf{a}}_{T,k}^*(\hat{\theta}_{\hat{l}}^{(u)})) \right\|_F^2. \end{aligned} \quad (35)$$

$$\begin{aligned} \min_{\mathbf{D}_{R,1}} \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{cen}}} & \left\| \hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{R,1} \underbrace{(\mathbf{D}_{R,2} \odot \mathbf{A}_{R,k}^{\text{v}}) \boldsymbol{\Omega}^{(u)}[k] (\mathbf{D}_{T,2} \odot \mathbf{A}_{T,k}^{\text{v}})^* \mathbf{D}_{T,1}^*}_{\mathbf{X}_{R,1}^{(u)}[k]} \right\|_F^2 \\ + \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{side}}} & \left\| \hat{\mathbf{H}}^{(u)}[k] - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}^{(u)}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \right. \\ & \left. \odot \underbrace{(\mathbf{D}_{R,1} (\mathbf{e}_R(\hat{\phi}_{\hat{l}}^{(u)}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}}^{(u)})) (\mathbf{e}_T(\hat{\theta}_{\hat{l}}^{(u)}) \odot \mathbf{a}_{T,k}(\hat{\theta}_{\hat{l}}^{(u)}))^* \mathbf{D}_{T,1}^*)}_{\mathbf{X}_{R,1,\hat{l}}^{(u)}[k]} \right\|_F^2, \end{aligned} \quad (36)$$

squint (DA-OMP-BS) as shown in Algorithm 1. This algorithm exploits the following two important properties of the channel models under beam squint, which will enable low-complexity and high-accuracy recovery of the channel coefficients and parameters:

- First, the common sparsity support property of channel vectors across subcarriers will still be considered but with frequency-dependent dictionaries accounting for beam squint impacts. For SOMP and OMP in previous work, the common sparsity support of channel coefficients between subcarriers is assumed under the same dictionary for all subcarriers. This is an approximate result for the case without considering beam squint effect. When the steering vector at the center frequency is used for all subcarriers, there exist approximation errors no matter how significant the beam squint effect is. Under beam squint circumstances, we can exploit the true property of common sparsity support between subcarriers but with different frequency-dependent dictionaries accounting for the beam squint impacts. The principle behind this argument is that the physical AoAs and AoDs associated with propagation paths are constant and independent of subcarriers, and thereby when the channel vector at each subcarrier is projected to the corresponding sparsifying dictionary, only a few bins out of the K_R or K_T virtual angular bins corresponding to the physical AoAs and AoDs are nonzero.
- Second, as mentioned above, the additional distortion matrix at side subcarriers tangles the three parameters of delay, AoA and AoD. Then if tradi-

tional SOMP or OMP is directly applied for (34), the overall sparsifying dictionary will be a three-dimensional (3D) dictionary [4] and the dimension of this 3D dictionary $[\Psi_k(\tau_1^{\text{v}}, \phi^{\text{v}}, \theta^{\text{v}}) \odot (\check{\mathbf{A}}_{T,k}^{\text{v}} \otimes \check{\mathbf{A}}_{R,k}^{\text{v}}), \dots, \Psi_k(\tau_{L_v}^{\text{v}}, \phi^{\text{v}}, \theta^{\text{v}}) \odot (\check{\mathbf{A}}_{T,k}^{\text{v}} \otimes \check{\mathbf{A}}_{R,k}^{\text{v}})]$ will be $N_R N_T \times L_v K_R K_T$, proportional to the product of the numbers of delay/AoA/AoD grids. This will induce overwhelming complexity. As for the newly proposed DA-OMP-BS, it will first exploit the channel models at central subcarriers to obtain initial estimates of the delay/AoA/AoD parameters. Without the impact of additional distortion matrix at central subcarriers, this can be done even by one-dimensional search over each parameter space iteratively. These initial estimates of parameters are then used to reduce the effective dictionary dimension at side subcarriers and compensate the distortions induced by beam squint, so that all the subcarriers can be collected to improve the estimates of delays/AoAs/AoDs again. In doing so, the sparse coding of (34) can be solved with much lower complexity and higher accuracy.

In line of these ideas, the DA-OMP-BS is expected to outperform SOMP and OMP without considering beam squint. We summarize the procedure of DA-OMP-BS in Algorithm 1. Once the sparse coding stage of (34) is done, the coefficients $\boldsymbol{\Omega}^{(u)}[k], k \in \mathcal{K}_{\text{cen}}$ and $\mathbf{b}_{\hat{l}}^{(u)}, k \in \mathcal{K}_{\text{side}}$ as well as the estimates of path gains, delays, AoAs, and AoDs $\hat{\varphi}_{\hat{l}}^{(u)} \triangleq \{\hat{\alpha}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}, \hat{\tau}_{\hat{l}}^{(u)}\}$ for $\hat{l} = 1, \dots, \hat{L}^{(u)}, \forall u \in \mathcal{I}(N_{\text{sa}})$ can be obtained, which

will be used to update the hardware impairments related dictionaries in the next subsections.

Algorithm 1 : DA-OMP-BS algorithm for sparse coding

- 1: **procedure** DA-OMP-BS ($\mathbf{y}^{(u)}[k]$, Φ , $\mathbf{D}_{T,1}$, $\mathbf{D}_{T,2}$, $\mathbf{D}_{R,1}$, $\mathbf{D}_{R,2}$, K_R , K_T , L_v)
 - 2: **Construct** $\check{\mathbf{A}}_{R,k}^v$ and $\check{\mathbf{A}}_{T,k}^v$ **at each subcarrier with input dictionaries**
 - 3: **Compute the whitened equivalent sensing matrix and received signals**
 - 4: $\mathbf{L}_w = \text{blkdiag} \left\{ (\check{\mathbf{W}}_1^* \check{\mathbf{W}}_1)^{-1/2}, \dots, (\check{\mathbf{W}}_M^* \check{\mathbf{W}}_M)^{-1/2} \right\}$
and $\Phi_w = \mathbf{L}_w \Phi$
 - 5: $\mathbf{y}_w^{(u)}[k] = \mathbf{L}_w \mathbf{y}^{(u)}[k]$, for $k = 0, \dots, K - 1$, $u \in \mathcal{I}(N_{sa})$
 - 6: **for** $u = 0, \dots, N_{sa} - 1$ **do**
 - 7: **Initialize the residual vectors and set of estimated parameters**
 - 8: $\mathbf{r}[k] = \Phi_w^* \mathbf{y}_w^{(u)}[k]$, $\forall k = 0, \dots, K - 1$ and $\hat{L}^{(u)} = 0$, $\hat{\varphi}_l^{(u)} = \emptyset$
 - 9: **while** $\text{MSE} > \epsilon$ **do**:
 - 10: **Define:** $O(\tau, \phi, \theta, \mathcal{K}_{set}) = \frac{\left\| \sum_{k \in \mathcal{K}_{set}} (\check{\mathbf{a}}_{T,k}^v(\theta) \otimes \check{\mathbf{a}}_{R,k}^v(\phi))^* (\mathbf{r}[k] e^{j2\pi \Delta f_k \tau}) \right\|_F^2}{\left\| \sum_{k \in \mathcal{K}_{set}} (\check{\mathbf{a}}_{T,k}^v(\theta) \otimes \check{\mathbf{a}}_{R,k}^v(\phi))^* (\Phi_w^* e^{j2\pi \Delta f_k \tau}) \right\|_F^2}$ =
 - 11: **Initial parameter estimation using central subcarriers**
 - 12: **delay:** $\hat{\tau} = \underset{\tau_l^v}{\text{argmax}} \left\| \sum_{k \in \mathcal{K}_{cen}} \mathbf{r}[k] e^{j2\pi \Delta f_k \tau_l^v} \right\|_F^2$ =
 - 13: **AoA:** $\hat{\phi} = \underset{\phi_i^v}{\text{argmax}} \left\| \sum_{k \in \mathcal{K}_{cen}} (\mathbf{I}_{N_T} \otimes \check{\mathbf{a}}_{R,k}^v(\phi_i^v))^* (\mathbf{r}[k] e^{j2\pi \Delta f_k \hat{\tau}}) \right\|_F^2$ =
 $\left\| \sum_{k \in \mathcal{K}_{cen}} (\mathbf{I}_{N_T} \otimes \check{\mathbf{a}}_{R,k}^v(\phi_i^v))^* (\Phi_w^* e^{j2\pi \Delta f_k \hat{\tau}}) \right\|_F^2$
 - 14: **AoD:** $\hat{\theta} = \underset{\theta_j^v}{\text{argmax}} O(\hat{\tau}, \hat{\phi}, \theta_j^v, \mathcal{K}_{cen})$
 - 15: **Refine the initial parameter estimates**
 - 16: **delay:** $\hat{\tau} = \underset{\tau_l^v}{\text{argmax}} O(\tau_l^v, \hat{\phi}, \hat{\theta}, \mathcal{K}_{cen})$
 - 17: **AoA:** $\hat{\phi} = \underset{\phi_i^v}{\text{argmax}} O(\hat{\tau}, \phi_i^v, \hat{\theta}, \mathcal{K}_{cen})$
 - 18: **AoD:** $\hat{\theta} = \underset{\theta_j^v}{\text{argmax}} O(\hat{\tau}, \hat{\phi}, \theta_j^v, \mathcal{K}_{cen})$
 - 19: **Distortion compensation at side subcarriers**
 - 20: $\mathbf{r}[k] = \text{diag} \{ \text{vec}(\mathbf{G}_k(\hat{\tau}, \hat{\phi}, \hat{\theta})) \}^{-1} \mathbf{r}[k]$, $\forall k \in \mathcal{K}_{side}$
 - 21: **Update parameter estimates using all subcarriers**
 - 22: **delay:** $\hat{\tau} = \underset{\tau_l^v}{\text{argmax}} O(\tau_l^v, \hat{\phi}, \hat{\theta}, \mathcal{K}_{cen} \cup \mathcal{K}_{side})$
 - 23: **AoA:** $\hat{\phi} = \underset{\phi_i^v}{\text{argmax}} O(\hat{\tau}, \phi_i^v, \hat{\theta}, \mathcal{K}_{cen} \cup \mathcal{K}_{side})$
 - 24: **AoD:** $\hat{\theta} = \underset{\theta_j^v}{\text{argmax}} O(\hat{\tau}, \hat{\phi}, \theta_j^v, \mathcal{K}_{cen} \cup \mathcal{K}_{side})$
 - 25: **Update the set of estimated parameters:**
 $\hat{L}^{(u)} = \hat{L}^{(u)} + 1$, $\hat{\varphi}_l^{(u)} = \{\hat{\tau}, \hat{\phi}, \hat{\theta}\}$
-

2) *Dictionary update stage:* Next, we fix the channel coefficients and path parameters in (33) in preparation of

Algorithm 1 : DA-OMP-BS algorithm for sparse coding (continued)

- 26: **Update the path gains $\hat{\alpha}_l$ and coefficient vectors by minimizing**
 - 27: $\sum_{k=0}^{K-1} \left\| \mathbf{y}_w^{(u)}[k] - \Phi_w \sum_{l=1}^{\hat{L}^{(u)}} \alpha_l e^{-j2\pi \Delta f_k \hat{\tau}_l} \text{vec}(\mathbf{G}_k(\hat{\tau}_l, \hat{\phi}_l, \hat{\theta}_l)) \odot (\check{\mathbf{a}}_{T,k}^v(\hat{\theta}_l) \otimes \check{\mathbf{a}}_{R,k}^v(\hat{\phi}_l)) \right\|_2^2$
 - 28: **Update the residual for each subcarrier**
 - 29: $\mathbf{r}[k] = \Phi_w^* (\mathbf{y}_w^{(u)}[k] - \Phi_w \sum_{l=1}^{\hat{L}^{(u)}} \hat{\alpha}_l e^{-j2\pi \Delta f_k \hat{\tau}_l} \text{vec}(\mathbf{G}_k(\hat{\tau}_l, \hat{\phi}_l, \hat{\theta}_l)) \odot (\check{\mathbf{a}}_{T,k}^v(\hat{\theta}_l) \otimes \check{\mathbf{a}}_{R,k}^v(\hat{\phi}_l)))$
 - 30: **Update the current MSE:**
 - 31: $\text{MSE} = \frac{1}{ML_R K} \sum_{k=0}^{K-1} \|(\Phi_w^*)^\dagger \mathbf{r}[k]\|_2^2$
 - 32: **end while**
 - 33: **end for**
 - 34: **Output:** $\hat{\varphi}_l^{(u)} = \{\hat{\alpha}_l, \hat{\tau}_l, \hat{\phi}_l, \hat{\theta}_l\}$, $\text{vec}(\hat{\Omega}^{(u)}[k])$ and $b_l^{(u)}$, for $u \in \mathcal{I}(N_{sa})$
-

updating the hardware impairments related dictionaries $\mathbf{D}_{T,1}$, $\mathbf{D}_{T,2}$, $\mathbf{D}_{R,1}$, $\mathbf{D}_{R,2}$. Note that due to the special structure in this problem that the TX and RX dictionaries are entangled with the beam squint effect at different subcarriers, the typical dictionary update algorithms, like the method of optimal directions (MOD) [28] or K-SVD [25], cannot be directly used. Therefore, we apply alternating optimization in this sub-stage as well to subsequently update the four dictionaries.

For the update of $\mathbf{D}_{R,1}$, the problem of (33) can be reduced to the expression in (35), which is equivalent to (36), where $\mathbf{X}_{R,1}^{(u)}[k]$ and $\mathbf{X}_{R,1,\hat{l}}^{(u)}[k]$ are defined accordingly for ease of expression. To update $\mathbf{D}_{R,1}$, we need to calculate the derivative of the objective function with respect to $\mathbf{D}_{R,1}$, i.e., $\frac{\partial J}{\partial \mathbf{D}_{R,1}}$, which is expressed as in (37) (a proof is provided in Appendix II). Therefore, the update of $\mathbf{D}_{R,1}$ can be obtained by (stochastic) gradient descent as

$$[\mathbf{D}_{R,1}]^{\text{new}} = [\mathbf{D}_{R,1}]^{\text{old}} - \eta \frac{\partial J}{\partial \mathbf{D}_{R,1}}, \quad (38)$$

where η is the step-size of gradient descent and can be determined by backtracking line search. Similarly, the update of $\mathbf{D}_{T,1}$ can be obtained.

For the update of $\mathbf{D}_{R,2} = f(\epsilon_R) \triangleq e^{-j2\pi f_c \epsilon_R \cdot \sin(\phi^v)^T / c}$, it is equivalent to updating ϵ_R . By stacking all subcarriers, we have the objective of updating ϵ_R as in (39), where $\mathbf{X}_{R,2}^{(u)}[k]$ and $\mathbf{X}_{R,2,\hat{l}}^{(u)}[k]$ are defined accordingly for ease of expression. To update ϵ_R , we need to calculate the derivative of the objective

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{D}_{R,1}} = & - \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{cen}}} \overline{(\hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{R,1} \mathbf{X}_{R,1}^{(u)}[k])(\mathbf{X}_{R,1}^{(u)}[k])^T} - \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{side}}} \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}^{(u)}} \\
& \cdot \left[\overline{(\hat{\mathbf{H}}^{(u)}[k] - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}^{(u)}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \odot \mathbf{D}_{R,1} \mathbf{X}_{R,1}^{(u)}[k])} \right. \\
& \left. \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \right] \cdot (\mathbf{X}_{R,1,\hat{l}}^{(u)}[k])^T.
\end{aligned} \tag{37}$$

$$\begin{aligned}
\min_{\epsilon_R} \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{cen}}} & \left\| \underbrace{\mathbf{D}_{R,1}^\dagger \hat{\mathbf{H}}^{(u)}[k] \mathbf{D}_{T,1}^{*\dagger}}_{\mathbf{Y}_{R,2}^{(u)}[k]} - (\mathbf{D}_{R,2} \odot \mathbf{A}_{R,k}^v) \underbrace{\Omega^{(u)}[k] (\mathbf{D}_{T,2} \odot \mathbf{A}_{T,k}^v)^*}_{\mathbf{X}_{R,2}^{(u)}[k]} \right\|_F^2 \\
& + \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{side}}} \left\| \hat{\mathbf{H}}^{(u)}[k] - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}^{(u)}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \right. \\
& \left. \odot \left(\mathbf{D}_{R,1} (\mathbf{e}_R(\hat{\phi}_{\hat{l}}^{(u)}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}}^{(u)})) \underbrace{(\mathbf{e}_T(\hat{\theta}_{\hat{l}}^{(u)}) \odot \mathbf{a}_{T,k}(\hat{\theta}_{\hat{l}}^{(u)}))^*}_{\mathbf{X}_{R,2,\hat{l}}^{(u)}[k]} \mathbf{D}_{T,1}^* \right) \right\|_F^2,
\end{aligned} \tag{39}$$

function with respect to ϵ_R , which can be expressed as in (40) (A proof is given in Appendix III). Therefore, the update of $\mathbf{D}_{R,2}$ is given as

$$[\mathbf{D}_{R,2}]^{\text{new}} = f([\epsilon_R]^{\text{new}}) = f\left([\epsilon_R]^{\text{old}} - \eta \frac{\partial J}{\partial \epsilon_R}\right). \tag{41}$$

Similarly, the update of $\mathbf{D}_{T,2}$ can be obtained.

The overall procedure of the proposed DL scheme is summarized in Algorithm 2, and the flow diagram is represented in Fig. 3.

Once the hardware impairments related dictionaries $\mathbf{D}_{R,1}$, $\mathbf{D}_{R,2}$, $\mathbf{D}_{T,1}$, $\mathbf{D}_{T,2}$ are obtained, the overall RX and TX sparsifying dictionaries $\mathbf{A}_{R,k}^v$ and $\mathbf{A}_{T,k}^v$ can be constructed at each subcarrier to incorporate the beam squint impact as in (29), and then used for subsequent online compressive channel estimation, which is expected to help reduce the training overhead significantly.

C. Convergence and complexity analysis

The iterative refinement between the sparse coding and the dictionary update stages is a decreasing process of the objective function in (33), which takes positive values. In other words, we are minimizing a function bounded by zero and its domain is closed. Therefore, the convergence is guaranteed, and the system will be able to learn the hardware impairments.

As for the complexity, since the DL phase can be implemented offline, the complexity involved in the dictionary update stage does not increase the overall

Algorithm 2 : Dictionary learning for hardware impairments under beam squint (DLHWBS)

- **Input:** Initial channel estimates $\hat{\mathbf{h}}^{(u)}[k], \forall k = 0, \dots, K-1, u \in \mathcal{I}(N_{\text{sa}})$.
- **Initialization:** Set the dictionary matrices $\mathbf{D}_{R,1} \in \mathbb{C}^{N_R \times N_R}$ and $\mathbf{D}_{T,1} \in \mathbb{C}^{N_T \times N_T}$ using measurement data based on DIA algorithm [29], and set the dictionary matrices $\mathbf{D}_{T,2}$ and $\mathbf{D}_{R,2}$ as all-one matrices.
- **While not converge do**
 1. *Sparse coding stage:* Fixing all dictionaries, solve (34) using Algorithm 1 to update channel coefficients $\Omega^{(u)}[k]$ and $\mathbf{b}_l^{(u)}$, as well as the path parameters $\hat{\varphi}_l^{(u)}$.
 2. *Dictionary update stage:* Fixing coefficients, update dictionaries as follows
 - While not converge do**
 - Update $\mathbf{D}_{R,1}$ using (38) and update $\mathbf{D}_{T,1}$ similarly,
 - Update $\mathbf{D}_{R,2}$ using (41) and update $\mathbf{D}_{T,2}$ similarly.
 - end while**
- **Output:** The optimal dictionaries $\mathbf{D}_{T,1}, \mathbf{D}_{T,2}, \mathbf{D}_{R,1}, \mathbf{D}_{R,2}$.

complexity of the online sparse coding stage. Therefore, we compare the computational overhead of the proposed DA-OMP-BS with those of TD-OMP [4], in

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{e}_R} = & \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{cen}}} 2\mathcal{R} \left\{ \left\{ \frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{R,2}} \odot \mathbf{D}_{R,2} \right\} \cdot \frac{-j2\pi f_c \sin(\phi^v)}{c} \right\} \\
& - \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{side}}} 2\mathcal{R} \left\{ \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}^{(u)}} \cdot \frac{-j2\pi f_c \sin(\hat{\phi}_{\hat{l}}^{(u)})}{c} \cdot \left[\mathbf{D}_{R,1}^T \right. \right. \\
& \cdot \left(\left(\widehat{\mathbf{H}}^{(u)}[k] - \sum_{\hat{l}=1}^{\hat{L}^{(u)}} \hat{\alpha}_{\hat{l}}^{(u)} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}^{(u)}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \odot (\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}^{(u)}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}}^{(u)})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k]) \right) \right. \\
& \left. \left. \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}^{(u)}, \hat{\phi}_{\hat{l}}^{(u)}, \hat{\theta}_{\hat{l}}^{(u)}) \right) (\mathbf{X}_{R,2,\hat{l}}^{(u)}[k])^T \odot (\mathbf{e}_R(\hat{\phi}_{\hat{l}}^{(u)}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}}^{(u)})) \right\}. \quad (40)
\end{aligned}$$

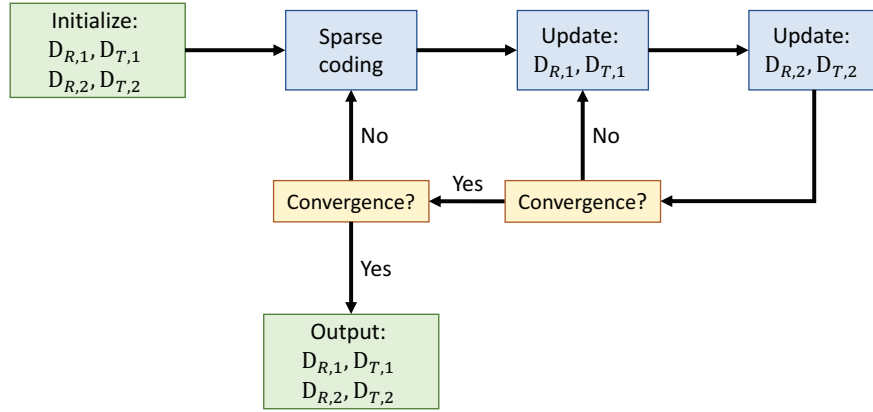


Fig. 3: Flow diagram of the dictionary learning stage.

terms of complex multiplication operations for each iteration. For ease of comparison, we set the numbers of AoA/AoD/delay grids to be the same $K_R = K_T = L_v = S$. For sparse coding with DA-OMP-BS, the parameters can be estimated by iterative one-dimensional search over each parameter space, and thus the complexity order per iteration is $\mathcal{O}((N_R + N_T + N_c)SN_{\text{sa}})$. For TD-OMP, the joint search over the 3D parameter space induces a much higher complexity order of $\mathcal{O}(N_R N_T + S^3 N_c N_{\text{sa}})$.

IV. NUMERICAL RESULTS

In this section, we present numerical results to demonstrate the effectiveness of the proposed dictionary learning and channel estimation algorithms for a realistic hybrid wideband mmWave MIMO system under both hardware impairments and beam squint.

The TX and RX are equipped with ULAs with half-wavelength spacing, i.e., $d_R = d_T = \lambda_c/2$, and $N_T = 32$, $N_R = 8$. Regarding the RF chains, $L_T = L_R = N_s = 2$. The number of OFDM subcarriers is $N_c = 64$ and $T_s = 1$ ns.

The channels are generated based on (21) with $L = 6$ multipath components, as in typical indoor scenarios.

The pulse shaping function $p(t)$ is assumed to be a raised-cosine filter with roll-off factor of $\beta = 0.25$. The angle of arrival ϕ_l and angle of departure θ_l are uniformly distributed in $[-\pi, \pi]$, while the delays are uniformly distributed in $[0, 16T_s]$. The gains are generated following a complex Gaussian distribution with the variance adjusted to achieve a specific SNR for the evaluation of the channel estimation strategies. For the parameters of hardware impairments, as in [17]–[19], the maximum gain and phase error variances for each antenna element are set as 5% and $20^\circ\pi/180^\circ$ respectively. Finally, the mutual coupling coefficients among antennas are within $[0.01, 0.4]$ and the antenna location errors are assumed to be uniformly distributed between $[-0.1\lambda_c, 0.1\lambda_c]$.

We evaluate the performance of our proposed wideband channel estimation algorithm, DA-OMP-BS, in addition to two baseline algorithms described in prior work, denoted as TD-OMP [4] and WB-ADMM [1]. The sizes for the discrete angle and delay grids are set as $K_T = 2N_T$, $K_R = 2N_R$ and $L_v = 2N_c$ for all the different algorithms evaluated in the simulations. Note, however, that the proposed Algorithm 1 is of much lower

complexity, so that finer angle and delay grids could have been used to improve parameter estimation accuracy. The performance of the different channel estimation strategies is evaluated in conjunction with different strategies for building the sparsifying dictionaries. In particular, we will consider: 1) the dictionaries obtained with our proposed approach DLHWBS; 2) a dictionary constructed from overcomplete ideal array response matrices (IARM) $\mathbf{A}_{R,k}^v$ and $\mathbf{A}_{T,k}^v$ considering beam squint, but without hardware impairments; 3) the general SeDL algorithm proposed in [19], which learns a frequency-flat combined dictionary for hardware impairments; and 4) the ideal dictionary built from the known impairments and the corresponding array response vectors.

During the DL phase, the number of OFDM symbols for training is set as 500, and a spreading factor 10 is used to increase the effective SNR by 10 dBs. For the optimization of DLHWBS, a revised version of the dictionary initialization algorithm (DIA) in [29] is used for initialization of the hardware impairments. Specifically, the original DIA algorithm [29] is first applied to initialize the combined dictionary $\mathbf{D}_R^{\text{init}} \triangleq \mathbf{D}_{R,1}(\mathbf{D}_{R,2} \odot \mathbf{A}_{R,k}^v)$. Then we set the initial values of the antenna spacing errors as zeros, i.e., $\epsilon_R = \mathbf{0}$ and $\mathbf{D}_{R,2} = \mathbf{1}_{N_R \times K_R}$, and thus the initialization of gain, phase and coupling matrix is obtained by $\mathbf{D}_{R,1} = \mathbf{D}_R^{\text{init}}(\mathbf{A}_{R,k}^v)^\dagger$.

First we evaluate the average computational complexity time for each channel estimation algorithm considering 100 channel realizations. The results are shown in Fig. 4, where it can be seen that with the selected grid sizes, fixed for all the algorithms, our method is around 500 times faster than TD-OMP, and around 100 times faster than WD-ADMM. In the next simulations we will show that despite this significant reduction in complexity, our method always outperforms WD-ADMM for any selection of the dictionary, and, depending on the system parameters, it slightly outperforms or performs similarly to TD-OMP.

Next we evaluate the normalized mean squared error (NMSE) for the different combinations of channel estimation algorithms and sparsifying dictionaries, as a function of the number of training symbols M and the SNR considering 100 channel realizations. Fig. 5(a) shows the NMSE when the SNR=0dB and the number of training symbols varies from 20 to 120. It can be observed that the NMSE reduction provided by DA-OMP-BS when exploiting the ideal dictionary based on known impairments varies from 1 to 1.5 dB. DA-OMP-BS outperforms any other channel estimation algorithm independently of the considered dictionary. Fig. 5(b) shows the NMSE results as function of the SNR when the number of training symbols M is set to 60. For SNR=0dB, both DA-OMP-BS leads to an NMSE value of -10 dB as TD-OMP but with a cost in complexity

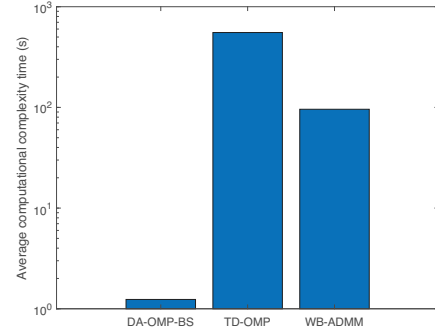


Fig. 4: Average computation complexity time of each evaluated channel estimation method.

500 times lower. Leaving aside the ideal dictionary built from known impairments, the best performing dictionary is DLHWBS, as it can be seen in both Fig. 5(a) and Fig. 5(b). DLHWBS outperforms SeDL because the latter attempts to learn a frequency-flat dictionary for all subcarriers under both hardware impairments and beam squint effect, and assumes all the channels follow the second case of (21). In other words, the distortions on the channel models at side subcarriers are ignored, and thus there already exist some modeling errors when SeDL is applied. These results confirm the effectiveness of our proposed approaches both for channel estimation and dictionary learning.

In Fig. 6, we also compare the spectral efficiency (SE) performances corresponding to the various sparsifying dictionaries and sparse coding algorithms. As in previous work [2], [19], the SE is computed by assuming fully-digital precoding and combining using estimates for the N_s dominant left and right singular vectors of the channel estimates. To be clear, SE is defined as

$$\text{SE} = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{n=1}^{N_s} \log \left(1 + \frac{\text{SNR}}{N_s} \lambda_n(\mathbf{H}_{\text{eff}}[k])^2 \right),$$

where $\mathbf{H}_{\text{eff}}[k]$ is the effective channel after precoding/combining and $\lambda_n(\mathbf{H}_{\text{eff}}[k])$ takes the singular values of $\mathbf{H}_{\text{eff}}[k]$. For both DA-OMP-BS and conventional TD-OMP, the SE can be significantly increased when exploiting the dictionaries learned with DLHWBS instead of IARM dictionaries. Moreover, the performance gap between the proposed DLHWBS algorithm and the case of ideal hardware impairment knowledge is small.

Finally, we compute the BER when considering the quadrature phase shift keying (QPSK) modulation and minimum mean squared error (MMSE) detection. Fig. 7 shows how better channel estimates translate into lower BER for DA-OMP-BS combined with DLHWBS. The average BER performance gap between the proposed

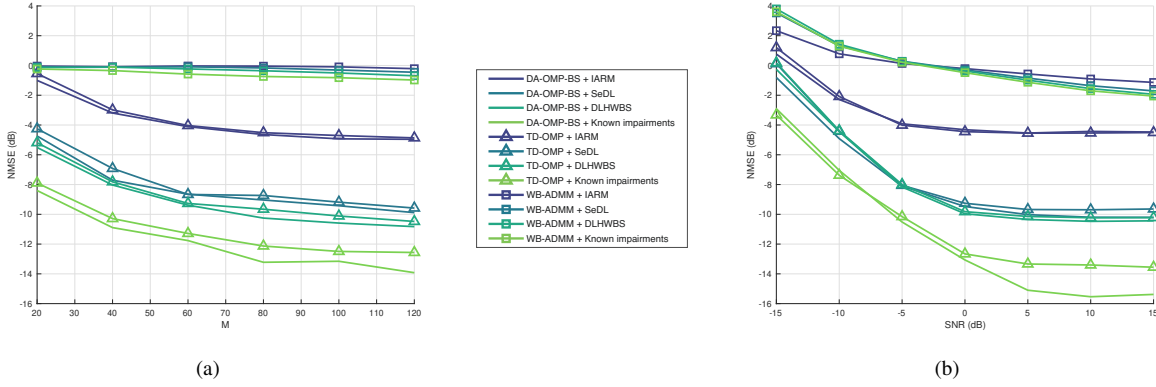


Fig. 5: Comparisons of NMSE performance: (a) as a function of the number of training OFDM symbols and SNR set to 0dB; (b) as a function of the SNR for 60 training OFDM symbols.

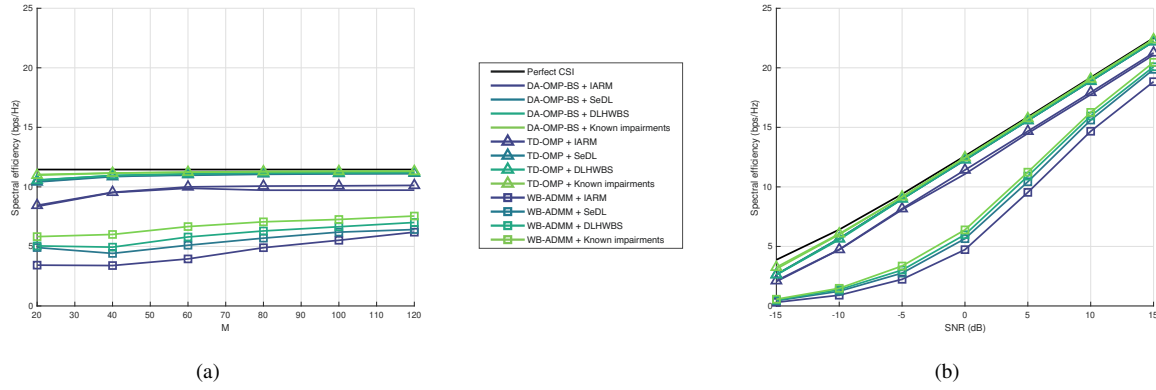


Fig. 6: Comparison of SE performance: (a) as a function of the number of training OFDM symbols and SNR set to 0dB; (b) as a function of the SNR for 60 training OFDM symbols.

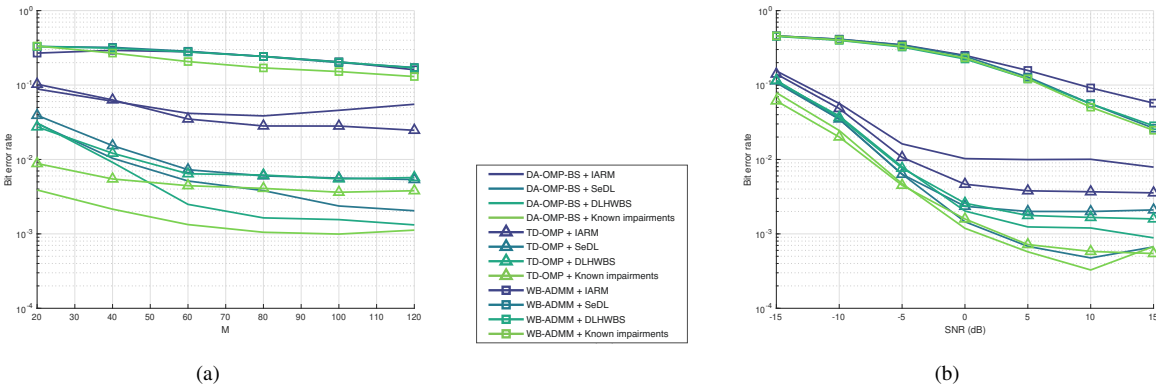


Fig. 7: Comparisons of BER performance: (a) as a function of the number of training OFDM symbols and SNR set to 0dB; (b) as a function of the SNR for 60 training OFDM symbols.

DLHWBS and the ideal case of known impairments is less than 0.5 dB. This reaffirms the effectiveness of our proposed algorithm for hybrid wideband channel under hardware impairments and beam squint.

V. CONCLUSIONS

In this manuscript, we derived a general channel model for MIMO systems by explicitly considering the combined impact of hardware impairments, pulse shaping/filtering effects and beam squint. The resulting model is an extension of existing MIMO channel models with beam squint. Based on this general channel model, we formulated a DL problem to obtain the sparsifying dictionaries for channel representation that account for hardware impairments. The effect of beam squint is considered, but it is not part of the learning process, since it can be mathematically modeled. We also proposed a novel compressive channel estimation algorithm under beam squint, which exploits the structure of the beam squint at different subcarriers to facilitate the estimation of the channel parameters at a much lower complexity. Numerical results have demonstrated the effectiveness of the proposed DL and channel estimation strategies and the significant gains with respect to the strategies proposed in prior work.

APPENDIX I PROOF OF EQUATION (11)

Let the delayed filter be $r(t) = p(t - \tau)$. If we have a set of measurements $\mathbf{r} = [r(0), r(T_s), \dots, r((K-1)T_s)]^T$, the k -th entry after DFT can be computed as

$$\hat{r}[k] = \sum_{d=0}^{K-1} [\mathbf{r}]_d \cdot e^{-j\frac{2\pi kd}{K}}. \quad (42)$$

Equivalently, we can use the continuous interpretation of the DFT to represent this as

$$\hat{r}[k] = \mathcal{F}\left(\sum_d \delta_{dT_s}(p * \delta_\tau)\right)[\Delta f_k], \quad (43)$$

where $\sum_d \delta_{dT_s} = \sum_{d=-\infty}^{\infty} \delta(t - dT_s)$ is the sampling function with period T_s and $\delta_\tau \triangleq \delta(t - \tau)$. Moreover, $*$ denotes convolution operation and $\mathcal{F}(\cdot)[\Delta f_k]$ is the Fourier transform evaluated for the frequency difference Δf_k . Next we can make use of Fourier product and convolution identities to get to

$$\hat{r}[k] = \left(\mathcal{F}\left(\sum_d \delta_{dT_s}\right) * (\mathcal{F}(p)\mathcal{F}(\delta_\tau))\right)[\Delta f_k]. \quad (44)$$

Using the Dirac comb Fourier identity formula we reach

$$\hat{r}[k] = \left(\left(\frac{1}{T_s} \sum_i \delta_{i/T_s}\right) * (\mathcal{F}(p)\mathcal{F}(\delta_\tau))\right)[\Delta f_k]. \quad (45)$$

Now we can express the convolution in terms of an integral

$$\hat{r}[k] = \frac{1}{T_s} \int \mathcal{F}(p)[f] e^{-j2\pi f\tau} \sum_i \delta_{i/T_s}(\Delta f_k - f) df, \quad (46)$$

and this can be solved by evaluating the Dirac comb as follows

$$\hat{r}[k] = \sum_i \frac{\mathcal{F}(p)[\Delta f_k + i/T_s]}{T_s} e^{-j2\pi(\Delta f_k + i/T_s)\tau} \quad (47)$$

$$= \left(\sum_i \frac{\mathcal{F}(p)[\Delta f_k + i/T_s]}{T_s} e^{-j2\pi i\tau/T_s}\right) e^{-j2\pi \Delta f_k \tau}. \quad (48)$$

Note that since $\mathcal{F}(p)$ is bounded, this sum only includes a few terms. Let us define the delay-frequency distortion as

$$g(k, \tau) = \sum_{k'} \frac{\mathcal{F}(p)[\Delta f_k + k'/T_s]}{T_s} e^{-j2\pi k'\tau/T_s}, \quad (49)$$

and then (47) can be simplified to

$$\hat{r}[k] = g(k, \tau) e^{-j2\pi \Delta f_k \tau}. \quad (50)$$

If we assume the domain of $\mathcal{F}(p)$ to be in $[-\frac{1+\beta}{2T_s}, +\frac{1+\beta}{2T_s}]$ with $\beta \in [0, 1]$, then we have the expression for $g(k, \tau)$ as in (51). In the case of a raised cosine filter with parameters $T = T_s$ and $\beta \in [0, 1]$, it is straightforward to prove the result in (52).

APPENDIX II DERIVATIVE OF THE OBJECTIVE FUNCTION IN (36) WITH RESPECT TO $\mathbf{D}_{R,1}$

First, we calculate the derivative of the first sum term in (36) with respect to $\mathbf{D}_{R,1}$. For $u \in \mathcal{I}(N_{sa}), k \in \mathcal{K}_{cen}$, we have

$$\frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{R,1}} = -(\hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{R,1} \mathbf{X}_{R,1}^{(u)}[k])(\mathbf{X}_{R,1}^{(u)}[k])^T. \quad (53)$$

Next, we calculate the derivative of the second sum term in (36) with respect to $\mathbf{D}_{R,1}$. Recalling the chain rule, we can express the Jacobian matrix of the second sum term with respect to $\mathbf{D}_{R,1}$ as in (54), $\forall u \in \mathcal{I}(N_{sa}), k \in \mathcal{K}_{side}$. From (54), we can easily obtain. Then, combining the derivatives in (53) for $k \in \mathcal{K}_{cen}$ and the derivatives in (55) for $k \in \mathcal{K}_{side}$, we can obtain the final derivative of the objective function in (36) with respect to $\mathbf{D}_{R,1}$ as (56).

APPENDIX III DERIVATIVE OF THE OBJECTIVE FUNCTION IN (39) WITH RESPECT TO ϵ_R

First, we calculate the derivative of the first sum term in (39) with respect to ϵ_R . Note that the gradient of any element of $\mathbf{D}_{R,2}$ with respect to the antenna location error $\epsilon_{R,m}$ can be expressed as

$$\begin{aligned} \frac{\partial [\mathbf{D}_{R,2}]_{m,n}}{\partial \epsilon_{R,m}} &= \frac{\partial e^{-j2\pi f_c \epsilon_{R,m} \sin(\phi_n^v)/c}}{\partial \epsilon_{R,m}} \\ &= [\mathbf{D}_{R,2}]_{m,n} \cdot \frac{-j2\pi f_c \sin(\phi_n^v)}{c}. \end{aligned} \quad (57)$$

$$g(k, \tau) = \begin{cases} \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} + \frac{\mathcal{F}(p)[\Delta f_k - 1/T_s]}{T_s} e^{j2\pi\tau/T_s} & \text{if } 2\Delta f_k > (1 - \beta)/T_s \\ \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} & \text{if } 2|\Delta f_k| \leq (1 - \beta)/T_s \\ \frac{\mathcal{F}(p)[\Delta f_k]}{T_s} + \frac{\mathcal{F}(p)[\Delta f_k + 1/T_s]}{T_s} e^{-j2\pi\tau/T_s} & \text{if } 2\Delta f_k < -(1 - \beta)/T_s \end{cases} \quad (51)$$

$$g(k, \tau) = \begin{cases} \frac{1}{2}(1 + e^{j\frac{2\pi\tau}{T_s}} + (1 - e^{j\frac{2\pi\tau}{T_s}}) \cos(\frac{\pi T_s}{\beta}(|\Delta f_k| - \frac{1-\beta}{2T_s}))) & \text{if } 2\Delta f_k > (1 - \beta)/T_s \\ 1 & \text{if } 2|\Delta f_k| \leq (1 - \beta)/T_s \\ \frac{1}{2}(1 + e^{-j\frac{2\pi\tau}{T_s}} + (1 - e^{-j\frac{2\pi\tau}{T_s}}) \cos(\frac{\pi T_s}{\beta}(|\Delta f_k| - \frac{1-\beta}{2T_s}))) & \text{if } 2\Delta f_k < -(1 - \beta)/T_s \end{cases} \quad (52)$$

$$\begin{aligned} \frac{\partial J_k^{(u)}}{\partial \text{vec}(\mathbf{D}_{R,1})^T} &= \frac{\partial J_k^{(u)}}{\partial \text{vec}(\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{R,1} \mathbf{X}_{R,1,\hat{l}}^{(u)}[k])^T} \\ &\cdot \frac{\partial \text{vec}(\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{R,1} \mathbf{X}_{R,1,\hat{l}}^{(u)}[k])}{\partial \text{vec}(\mathbf{D}_{R,1} \mathbf{X}_{R,1,\hat{l}}^{(u)}[k])^T} \cdot \frac{\partial \text{vec}(\mathbf{D}_{R,1} \mathbf{X}_{R,1,\hat{l}}^{(u)}[k])}{\partial \text{vec}(\mathbf{D}_{R,1})^T} \\ &= -\text{vec}\left(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{R,1} \mathbf{X}_{R,1,\hat{l}}^{(u)}[k]\right)^T \\ &\cdot \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \text{diag}\{\text{vec}(\mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}))\} ((\mathbf{X}_{R,1,\hat{l}}^{(u)}[k])^T \otimes \mathbf{I}_{N_R}). \end{aligned} \quad (54)$$

$$\begin{aligned} \frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{R,1}} &= \text{unvec}\left\{ \frac{\partial J_k^{(u)}}{\partial \text{vec}(\mathbf{D}_{R,1})} \right\} = \text{unvec}\left\{ -\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} (\mathbf{X}_{R,1,\hat{l}}^{(u)}[k] \otimes \mathbf{I}_{N_R}) \right. \\ &\cdot \text{diag}\{\text{vec}(\mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}))\} \cdot \text{vec}\left(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{R,1} \mathbf{X}_{R,1,\hat{l}}^{(u)}[k]\right) \Big\} \\ &= -\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \left[\left(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{R,1} \mathbf{X}_{R,1,\hat{l}}^{(u)}[k] \right) \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \right] \\ &\cdot (\mathbf{X}_{R,1,\hat{l}}^{(u)}[k])^T. \end{aligned} \quad (55)$$

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{D}_{R,1}} &= -\sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{cen}}} \overline{(\hat{\mathbf{H}}^{(u)}[k] - \mathbf{D}_{R,1} \mathbf{X}_{R,1}^{(u)}[k])} (\mathbf{X}_{R,1}^{(u)}[k])^T - \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{side}}} \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \\ &\cdot \left[\left(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi\Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot \mathbf{D}_{R,1} \mathbf{X}_{R,1,\hat{l}}^{(u)}[k] \right) \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \right] (\mathbf{X}_{R,1,\hat{l}}^{(u)}[k])^T. \end{aligned} \quad (56)$$

Therefore, we have the gradient of $J_k^{(u)}$ with respect to ϵ_R as

$$\frac{\partial J_k^{(u)}}{\partial \epsilon_R} = 2\mathcal{R} \left\{ \left\{ \frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{R,2}} \odot \mathbf{D}_{R,2} \right\} \cdot \frac{-j2\pi f_c \sin(\phi^v)}{c} \right\}, \quad (58)$$

We next calculate the derivative of the second sum term in (39) with respect to ϵ_R . Similar to Appendix II, using the chain rule of Jacobian matrix, $\forall u \in \mathcal{I}(N_{\text{sa}}), k \in \mathcal{K}_{\text{side}}$, we have the expression (59), which

leads to (60). Then, combining the derivatives in (58) for $k \in \mathcal{K}_{\text{cen}}$ and the derivatives in (60) for $k \in \mathcal{K}_{\text{side}}$, we can obtain the final derivative of the objective function in (39) with respect to ϵ_R , shown in (61).

REFERENCES

- [1] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Wideband MIMO Channel Estimation for Hybrid Beamforming Millimeter Wave Systems via Random Spatial Sampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 1136–1150, 2019.
- [2] 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, May 2018.
- [3] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. of IEEE ISIT'16*, July 2016, pp. 860–864.
- [4] 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, Sept. 2017.
- [5] J. Lee, G. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *Proc. of IEEE GLOBECOM'14*, Dec. 2014, pp. 3326–3331.
- [6] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, July 2018.
- [7] J. H. Brady and A. M. Sayeed, "Wideband communication with high-dimensional arrays: New results and transceiver architectures," in *Proc. of IEEE ICCW'15*, June 2015, pp. 1042–1047.
- [8] T. Chu and H. Hashemi, "True-time-delay-based multi-beam arrays," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 8, pp. 3072–3082, Aug. 2013.
- [9] J. Rodríguez-Fernández and N. González-Prelcic, "Channel estimation for frequency-selective mmWave MIMO systems with beam-squint," in *Proc. of IEEE GLOBECOM'18*, 2018, pp. 1–6.
- [10] J. P. González-Coma, W. Utschick, and L. Castedo, "Hybrid LISA for wideband multiuser millimeter wave communication systems under beam squint," *CoRR*, vol. abs/1804.09223, 2018. [Online]. Available: <http://arxiv.org/abs/1804.09223>
- [11] M. Wang, F. Gao, N. Shlezinger, M. F. Flanagan, and Y. C. Eldar, "A block sparsity based estimator for mmWave massive MIMO channels with beam squint," *IEEE Trans. Signal Process.*, vol. 68, pp. 49–64, 2019.
- [12] M. Jian, F. Gao, Z. Tian, S. Jin, and S. Ma, "Angle-domain aided UL/DL channel estimation for wideband mmWave massive MIMO systems with beam squint," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3515–3527, 2019.
- [13] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, "Beam squint and channel estimation for wideband mmwave massive MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 5893–5908, 2019.
- [14] J. Tan and L. Dai, "Wideband Beam Tracking in THz Massive MIMO Systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1693–1710, 2021.
- [15] —, "Wideband channel estimation for THz massive MIMO," *China Communications*, vol. 18, no. 5, pp. 66–80, 2021.
- [16] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, 2014, pp. 273–277.
- [17] M. Eberhardt, P. Eschlwech, and E. Biebl, "Investigations on antenna array calibration algorithms for direction-of-arrival estimation," *Advances in Radio Science: ARS*, vol. 14, pp. 181–190, 2016.
- [18] 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.
- [19] 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.
- [20] J. Palacios, N. González-Prelcic, and J. Widmer, "Managing hardware impairments in hybrid millimeter wave MIMO systems: A dictionary learning-based approach," in *Proc. of IEEE Asilomar'19*, 2019, pp. 168–172.
- [21] T. Wiese, L. Weiland, and W. Utschick, "Low-rank approximations for spatial channel models," in *Proc. of IEEE WSA'16*, Mar. 2016, pp. 1–5.
- [22] P. Tan and N. C. Beaulieu, "Reduced ICI in OFDM systems using the 'better than' raised-cosine pulse," *IEEE Communications Letters*, vol. 8, no. 3, pp. 135–137, 2004.
- [23] B. Friedlander and A. J. Weiss, "Direction finding in the presence of mutual coupling," *IEEE Trans. Antennas Propag.*, vol. 39, no. 3, pp. 273–284, 1991.
- [24] K. V. Stavropoulos and A. Manikas, "Array calibration in the presence of unknown sensor characteristics and mutual coupling," in *Proc. of IEEE EUSIPCO'00*, 2000, pp. 1–4.
- [25] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [26] 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, Dec. 2007.
- [27] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. of IEEE ICASSP'05*, vol. 5, March 2005, pp. 721–724.
- [28] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. of IEEE ICASSP'99*, vol. 5, 1999, pp. 2443–2446.
- [29] C. Rusu and B. Dumitrescu, "An initialization strategy for the dictionary learning problem," in *Proc. of IEEE ICASSP'14*, May 2014, pp. 6731–6735.

$$\begin{aligned}
\frac{\partial J_k^{(u)}}{\partial \epsilon_R^T} &= 2\mathcal{R} \left\{ \frac{\partial J_k^{(u)}}{\partial \text{vec} \left(\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot (\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k]) \right)^T} \right. \\
&\quad \cdot \frac{\partial \text{vec} \left(\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot (\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k]) \right)}{\partial \text{vec}(\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k])^T} \\
&\quad \cdot \frac{\partial \text{vec}(\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k])}{\partial \text{vec}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}}))^T} \cdot \frac{\partial \text{vec}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}))}{\partial \epsilon_R^T} \left. \right\} \\
&= 2\mathcal{R} \left\{ -\text{vec} \left(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot (\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k]) \right)^T \right. \\
&\quad \cdot \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \text{diag} \{ \text{vec}(\mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}})) \} \\
&\quad \cdot ((\mathbf{X}_{R,2,\hat{l}}^{(u)}[k])^T \otimes \mathbf{D}_{R,1}) \cdot \text{diag} \{ \text{vec}(\mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \} \cdot \text{diag} \{ \mathbf{e}_R(\hat{\phi}_{\hat{l}}) \} \cdot \frac{-j2\pi f_c \sin(\hat{\phi}_{\hat{l}})}{c} \left. \right\}. \tag{59}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial J_k^{(u)}}{\partial \epsilon_R} &= 2\mathcal{R} \left\{ -\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \cdot \frac{-j2\pi f_c \sin(\hat{\phi}_{\hat{l}})}{c} \cdot \text{diag} \{ \mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}}) \} \right. \\
&\quad \cdot ((\mathbf{X}_{R,2,\hat{l}}^{(u)}[k]) \otimes \mathbf{D}_{R,1}^T) \cdot \text{diag} \{ \text{vec}(\mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}})) \} \\
&\quad \cdot \text{vec} \left(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot (\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k]) \right) \left. \right\} \\
&= 2\mathcal{R} \left\{ -\sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \cdot \frac{-j2\pi f_c \sin(\hat{\phi}_{\hat{l}})}{c} \cdot \left[\mathbf{D}_{R,1}^T \right. \right. \\
&\quad \cdot \left(\left(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot (\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k]) \right) \right. \\
&\quad \left. \left. \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \right) (\mathbf{X}_{R,2,\hat{l}}^{(u)}[k])^T \right] \odot (\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \left. \right\}. \tag{60}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial J}{\partial \epsilon_R} &= \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{cen}}} 2\mathcal{R} \left\{ \left\{ \frac{\partial J_k^{(u)}}{\partial \mathbf{D}_{R,2}} \odot \mathbf{D}_{R,2} \right\} \cdot \frac{-j2\pi f_c \sin(\phi^v)}{c} \right\} \\
&\quad - \sum_{u \in \mathcal{I}(N_{\text{sa}})} \sum_{k \in \mathcal{K}_{\text{side}}} 2\mathcal{R} \left\{ \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \cdot \frac{-j2\pi f_c \sin(\hat{\phi}_{\hat{l}})}{c} \cdot \left[\mathbf{D}_{R,1}^T \right. \right. \\
&\quad \cdot \left(\left(\overline{\hat{\mathbf{H}}^{(u)}[k]} - \sum_{\hat{l}=1}^{\hat{L}} \hat{\alpha}_{\hat{l}} e^{-j2\pi \Delta f_k \hat{\tau}_{\hat{l}}} \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \odot (\mathbf{D}_{R,1}(\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \mathbf{X}_{R,2,\hat{l}}^{(u)}[k]) \right) \right. \\
&\quad \left. \left. \odot \mathbf{G}_k(\hat{\tau}_{\hat{l}}, \hat{\phi}_{\hat{l}}, \hat{\theta}_{\hat{l}}) \right) (\mathbf{X}_{R,2,\hat{l}}^{(u)}[k])^T \right] \odot (\mathbf{e}_R(\hat{\phi}_{\hat{l}}) \odot \mathbf{a}_{R,k}(\hat{\phi}_{\hat{l}})) \left. \right\}. \tag{61}
\end{aligned}$$