Joint Optimization of the Channel Estimator, Transmit Precoder and Receiver in Large-Scale MIMO Systems

Tao Yu, Chengwen Xing, *Member, IEEE*, Xiaqing Miao, Shiqi Gong, and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Channel estimation and data detection constitute a pair of pivotal modules in multiple-input multiple-output (MIMO) communication systems, where achieving accurate channel estimation is particularly important for large-scale MIMO communications. However, more accurate channel estimation requires more resources. Hence, we investigate the joint optimization of the channel estimator, transmit precoder and receiver in large-scale MIMO systems. In contrast to the classic signal processing philosophy, the joint optimization aims for solving two equations in the face of realistic channel estimation and data transmission imperfections. Closed-form solutions are derived for a pair of schemes. For the first one, the joint optimization consists of the three procedures of channel estimation, data estimation and channel refinement. In this method, the estimated data symbols are also harnessed as pilots, based on which the channel estimation performance is improved. As for the second scheme, since data estimation is our final goal and channel estimation is only an intermediate step, the channel estimation procedure is substituted into the data estimation regime without deriving an explicit solution for the estimated channel. Given our objective of optimizing the data estimation performance, the channel estimator and data transceiver are jointly optimized, and the intricate linkages between these two methods are discussed. Finally, several numerical results are provided for demonstrating the performance advantages over the traditional designs.**

***Index Terms*—Joint optimization, robust transceiver design, channel estimation refinement, and MIMO communications.**

I. Introduction

The performance of pilot-aided multiple-input multiple-output (MIMO) communications critically depends on that of channel estimation and data detection [1]–[7]. In the channel estimation procedure, the pilots are transmitted from the source to destination for estimating the MIMO channel state information (CSI) harnessed for data estimation [8]–[10]. Hence, it is plausible that accurate CSI is an important premise to guarantee spectral-and power-efficient MIMO communications [11]–[16].

T. Yu and C. Xing are with School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: taoyu.bit@gmail.com, xingchengwen@gmail.com).

X. Miao and S. Gong are with School of Cyberspace Science and Technology, Beijing Institute of technology, Beijing 100081, China (e-mail: xqmiao@bit.edu.cn, gsqyx@163.com).

L. Hanzo is with the Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK (e-mail: lh@ecs.soton.ac.uk).

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1, EP/X01228X/1 and EP/Y026721/1 as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028)

Below we briefly highlight the development of pilot-aided MIMO communications. The most naive design relies on assuming that the estimated CSI is perfect and proceeds to data estimation. Therefore, the channel estimation and data estimation are decoupled into a pair of independent designs. This naive technique significantly simplifies pilot-aided MIMO designs and adequate performance can be guaranteed, provided that the estimated CSI is accurate enough. Naturally, when considering the practical limitations during transmissions, like the coherence time, training interval, etc., the CSI becomes inaccurate. Robust statistics may be adopted for analyzing the effects of channel estimation errors and for robust transceiver optimization in the face of channel errors [17]–[19]. In this category, there is an abundance of literature on robust MIMO transceiver designs conceived for different system models or error models [20]–[23]. Regarding the robust designs or the joint optimization of the pilots and the transmit precoder (TPC), the kernel idea behind them is to take the effects of channel estimation errors into account in data transmissions. The alternative concept of decision-directed channel estimation is based on exploiting the specific data symbols that are reliably estimated as pilots in turn. This idea is widely harnessed in coded communication systems [24]–[27].

In addition, some methods of joint channel estimation and data estimation were proposed for RIS-assisted communication systems to strike an attractive trade-off between the system performance and computational complexity. Specifically, the authors of [28] discussed channel estimation in RIS-assisted systems where message passing algorithms are applied to deal with a large number of unknown parameters. Then, the authors of [29] extended this scheme to solve the joint optimization of channel estimation and data recovery. By contrast, the authors of [30] proposed a low-complexity channel estimation and precoder design, aiming for maximizing the achievable rate of RIS systems.

Throughout the evolution of MIMO technologies, their operating frequencies gradually increased and the dimensionality of antenna arrays also became much higher [31]–[36]. These facts intrinsically erode the optimality of the traditional separate designs. On one hand, more and more resources might be needed to achieve accurate channel estimation for high-dimensional MIMO systems [37]. On the other hand, high-performance communications always require accurate CSI. Furthermore, the effects of limited coherence time and the hardware costs make the problem much tougher. Clearly, channel estimation in large-scale MIMO communications is usually performed under strictly limited bandwidth resources and the ensuing data transmission has to cope with non-negligible channel estimation errors, where a certain target performance has to be guaranteed. This represents a challenge. Again, channel estimation and data estimation have a close relationship with each other and this is put under the microscope in the topical context of large-scale MIMO systems in the absence of channel coding. They are also boldly contrasted to the authoritative literature in Table I.

TABLE I
KEY ATTRIBUTES OF THIS PAPER CONTRASTED TO THE EXISTING JOINT MIMO DESIGNS

	[25]	[27]	FM-JO	QMO-JO
Transmit-side correlation channel model			√	√
Joint design of channel estimator and data transceiver	√	√	√	√
Precoder design			√	√
Refinement of channel	√	√	√	
Optimization criterion MSE	√		√	√

Thus, the joint optimization of the channel estimator and data transceiver subject to a specific channel estimation error is proposed, which unveils the intricate connection of channel estimation and data transmission more accurately and it strikes a compelling trade-off between the channel estimation accuracy and the data transmission efficiency. The contributions of our work are discussed below in more depth.

- We provide novel mathematical insights for characterizing pilot-aided MIMO communications. By employing linear operators at both the transmitter and receiver, we solve a pair of linear equation arrays in the face of random noise terms in the channel estimation and data estimation procedures, respectively. Although the effect of additive noise is impossible to remove completely, the formulations of the equation arrays are quite similar, because the error models follow a similar robust design philosophy.
- Channel estimation and data estimation can be iteratively optimized by our Functional Module Based Joint Optimization (FM-JO). Based on the estimation error model of the channel estimation procedure, the accuracy of data estimation can be improved. On the other hand, with the estimation error model in our data estimation procedure and the estimated data used as pilots, the accuracy of channel estimation can be improved.
- Since our ultimate goal is to transmit data, where channel estimation is just an intermediate step for pilot-aided transmissions, the channel estimation signal model may be incorporated into the data transmission signal model. Then, Quadratic Matrix Optimization Based Joint Optimization (QMO-JO) is proposed, where the channel estimator and signal transceiver can be jointly optimized by relying on quadratic matrix optimization. Explicitly, the channel estimator, the receiver's equalizer and the TPC may be jointly optimized. Additionally, the optimized data equalizer and TPC matrices can be substituted into our channel estimation refinement scheme for improving the channel estimation accuracy as in our FM-JO scheme.
- Based on the numerical results, we will demonstrate that the joint optimization has better performance than the separate designs. Our joint optimization philosophy strikes an estimation accuracy vs. computational complexity trade-off and subsumes the traditional separate designs as special cases.

The organization of the paper is as follows. In Section II, our signal models are introduced, and both the channel estimation and data estimation problems are formulated mathematically. In Section III, the FM-JO algorithm is proposed, which reflects the mutual relationships between the channel estimation and data estimation. In Section IV, the QMO-JO algorithm is conceived, which intrinsically amalgamates the channel estimation module with the data estimator without obtaining an explicit channel estimation solution. With the objective of optimizing the data estimation performance, the channel estimator and data transceiver are jointly optimized. Then, the relationships between these two schemes are investigated, followed by our numerical results in Section V to demonstrate the performance advantages of the proposed joint optimization algorithms over the traditional separate optimization algorithms. Finally, we conclude in Section VI.

TABLE II
MAIN NOTATIONS

$\boldsymbol{F}$	$N_{\mathrm{T}} \times r$, linear TPC matrix
$\boldsymbol{G}_{\mathrm{CE}}$	$T_{\mathrm{P}} \times N_{\mathrm{T}}$, linear channel estimator
$\widetilde{\boldsymbol{G}}_{\mathrm{CE}}$	$T_{\mathrm{P}} \times N_{\mathrm{T}}$, linear channel estimator after channel refinement
$\boldsymbol{G}_{\mathrm{DE}}$	$r \times N_{\mathrm{R}}$, linear equalizer
$\widetilde{\boldsymbol{G}}_{\mathrm{DE}}$	$r \times N_{\mathrm{R}}$, linear equalizer after channel refinement
$\boldsymbol{H}$	$N_{\mathrm{R}} \times N_{\mathrm{T}}$, channel matrix
$\boldsymbol{N}_{\mathrm{D}}$	$N_{\mathrm{R}} \times T_{\mathrm{D}}$, noise matrix during data transmission
$\boldsymbol{N}_{\mathrm{P}}$	$N_{\mathrm{R}} \times T_{\mathrm{P}}$, noise matrix during pilot transmission
T_{D}	length of data signal
T_{P}	length of pilot sequence
N_{R}	number of receive antennas
N_{T}	number of transmit antennas
r	number of data stream
$\boldsymbol{X}_{\mathrm{D}}$	$r \times T_{\mathrm{P}}$, data signal matrix
$\boldsymbol{X}_{\mathrm{P}}$	$N_{\mathrm{T}} \times T_{\mathrm{P}}$, pilot signal matrix
$\boldsymbol{\Psi}$	$N_{\mathrm{T}} \times N_{\mathrm{T}}$, transmit-side correlation matrix

The main notations are illustrated in Table II. The symbol $\boldsymbol{A}^{\mathrm{H}}$ denotes the Hermitian transpose of a general matrix $\boldsymbol{A}$. $\mathrm{Tr}(\boldsymbol{A})$ represents the trace of a square matrix $\boldsymbol{A}$. For a positive semidefinite matrix $\boldsymbol{A}$, the matrix $\boldsymbol{A}^{\frac{1}{2}}$ denotes the Hermitian square root of $\boldsymbol{A}$.

II. Signal Model and Problem Formulation

We commence by outlining the signal models of channel estimation and data transmission, as the basis of jointly optimizing the channel estimator and transceiver. The most widely studied point-to-point MIMO channel model is investigated, where the transmit antenna spatial correlation is taken into account [8]. The scenario is that of the large-scale MIMO uplink transmission, where we assume that the antenna separation and the angle of spread at the base station (BS) are adequate. Thus, the antenna's spatial correlation at the BS may be ignored, and hence it is modeled by an identity matrix. The corresponding channel matrix $\boldsymbol{H}$ is given by the following equation [38], [39]

$$\boldsymbol{H} = \boldsymbol{H}_{\mathrm{W}}\boldsymbol{\Psi}^{\frac{1}{2}}, \tag{1}$$

where the elements of the matrix $\boldsymbol{H}_{\mathrm{W}}$ are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance. The positive definite matrix $\boldsymbol{\Psi} > \boldsymbol{0}$ is the transmit-side correlation matrix [38]. This is a widely used channel model in the authoritative literature [39]. Based on the channel model given in (1), the pilots and data are assumed to be transmitted in a time division duplexing manner. Specifically, the pilot sequences are sent first from the transmitter to the receiver for the estimation of the channel matrix. Then, the receiver sends the estimated channel matrix back to the transmitter through the feedback channel. Based on the estimated channel matrix, the transmitter appropriately configures the TPC matrix, and then the data signals are precoded and transmitted over the wireless channels. Finally, the signals are recovered at the receiver, for example by a linear equalizer.

Based on the above discussions, our received signal is formulated as

$$\boldsymbol{Y} = \boldsymbol{H}[\boldsymbol{X}_{\mathrm{P}}, \boldsymbol{F}\boldsymbol{X}_{\mathrm{D}}] + [\boldsymbol{N}_{\mathrm{P}}, \boldsymbol{N}_{\mathrm{D}}], \tag{2}$$

which is processed in two phases, namely channel estimation and data transmission. On the righthand side of the equation, the first signal block $\boldsymbol{X}_{\mathrm{P}}$ represents the pilot signal matrix that is known to both the source and destination [6]. The matrix $\boldsymbol{F}$ denotes the linear TPC matrix, which is designed based on the CSI [40], [41]. The matrix $\boldsymbol{X}_{\mathrm{D}}$ is the data signal matrix, which will be recovered at the receiver. Finally, the matrices $\boldsymbol{N}_{\mathrm{P}}$ and $\boldsymbol{N}_{\mathrm{D}}$ represent the corresponding random additive noise matrix during pilot transmission and data transmission, respectively.

Our objective is to recover $\boldsymbol{X}_{\mathrm{D}}$ as accurately as possible based on the observation $\boldsymbol{Y}$. The traditional technique decouples (2) into a pair of successive phases [20], [42], i.e., $\boldsymbol{Y} = [\boldsymbol{Y}_{\mathrm{P}}, \boldsymbol{Y}_{\mathrm{D}}]$ formulated as

$$\begin{aligned} \boldsymbol{Y}_{\mathrm{P}} &= \boldsymbol{H}\boldsymbol{X}_{\mathrm{P}} + \boldsymbol{N}_{\mathrm{P}}, \\ \boldsymbol{Y}_{\mathrm{D}} &= \boldsymbol{H}\boldsymbol{F}\boldsymbol{X}_{\mathrm{D}} + \boldsymbol{N}_{\mathrm{D}}. \end{aligned} \tag{3}$$

In the first phase, the channel matrix $\boldsymbol{H}$ is estimated based on the first equation in (3). In the second phase, upon substituting the estimated $\boldsymbol{H}$ into the second equation in (3), $\boldsymbol{X}_{\mathrm{D}}$ will be recovered. This has been widely used as an efficient pilot-aided MIMO technique, in which the estimation error is passed on to the second step, hence degrading the accuracy of the data recovery. Here we analyze this problem more deeply. Due to the limited channel coherence time, the length (the number of columns) in the transmitted signals is also constrained. When the dimensions of $\boldsymbol{H}$ are high, as in massive MIMO systems, more resources have to be allocated to channel estimation for achieving higher channel estimation accuracy and hence high-integrity data recovery [42]. Since the maximum transmit power is usually fixed, the easiest way of allocating more resources to channel estimation is to increase the length of pilots, i.e, the number of columns in $\boldsymbol{X}_{\mathrm{P}}$ [39], which inevitably reduces the fraction of data symbols. As a result, the normalized payload of the system is reduced. Again, we note that the key objective here is to recover $\boldsymbol{X}_{\mathrm{D}}$, and the estimation of $\boldsymbol{H}$ is only an intermediate step/computation. From a pure mathematical perspective, we aim for solving the equation array given in (3). However, in contrast to classical linear algebra, there are random noise matrices in the equation array in (3). Bearing in mind the computational complexity, only linear transceivers are taken into account in this work. The received signals are processed by the receiver, and the signal model in (3) is transformed into

$$\begin{aligned} \boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}} &= \boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}} + \boldsymbol{N}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}, \\ \boldsymbol{G}_{\mathrm{DE}}\boldsymbol{Y}_{\mathrm{D}} &= \boldsymbol{G}_{\mathrm{DE}}\boldsymbol{H}\boldsymbol{F}\boldsymbol{X}_{\mathrm{D}} + \boldsymbol{G}_{\mathrm{DE}}\boldsymbol{N}_{\mathrm{D}}, \end{aligned} \tag{4}$$

where $\boldsymbol{G}_{\mathrm{CE}}$ is the linear channel estimator harnessed for estimating the channel matrix and $\boldsymbol{G}_{\mathrm{DE}}$ is the linear equalizer employed for recovering the desired data signals. The traditional logic is to find $\boldsymbol{G}_{\mathrm{CE}}$ and $\boldsymbol{G}_{\mathrm{DE}}$ in a sequential manner. The channel estimator $\boldsymbol{G}_{\mathrm{CE}}$ is designed first based on the former equation in (4), which may rely on different performance metrics. Then, the estimated channel matrix is assumed to be the perfect CSI $\boldsymbol{H}$, and the linear transceiver matrices $\boldsymbol{F}$ and $\boldsymbol{G}_{\mathrm{DE}}$ are optimized based on the second equation in (4) according to the minimum mean square error (MMSE) criterion, for example.

For practical communication systems, such as massive MIMO communication systems, where the channel matrix is of large dimension, accurate channel estimation will need a large amount of resources. However, considering the stringent resource limitations and coherence time, it is difficult to estimate the channel accurately based on the first equation in (4), and increasing the pilot overhead leads to efficiency losses for the whole system. To strike a trade-off between the channel estimation accuracy and training resources, the linear estimator $\boldsymbol{G}_{\mathrm{CE}}$, linear equalizer $\boldsymbol{G}_{\mathrm{DE}}$ and linear TPC $\boldsymbol{F}$ should be jointly optimized for improving the data integrity, which is the focus of this paper.

On the other hand, the data is also subjected to the same channel as the pilots and hence it may be exploited for improving the channel estimation accuracy. Because of the channel impairment imposed on both the channel estimation and data estimation procedures, it seems much more challenging to solve (4) in a joint optimization manner than in the absence of channel impairments, namely when linear algebra may be applied.

Our objective is to jointly optimize the channel estimator and data transceiver from a pure signal processing viewpoint. In the following, a pair of different techniques is exploited for the joint optimization of the channel estimator and data transceiver. The above-mentioned FM-JO is an extension of the existing widely used rationale. The mutual relationship between channel estimation and data estimation is established based on the corresponding estimation error models. This technique accrues from the robust transceiver optimization philosophy [1], in which the specific channel estimation error model is taken into account during the transceiver optimization for attaining performance improvements. Once the estimated data and its specific estimation error model become available, a similar robust channel estimation process can be performed in order to improve the channel estimation accuracy. As a result, the first kind of joint optimization consists of three parts, namely channel estimation, data estimation and channel refinement. The main difference among the existing algorithms [22], [23], [38], [39] arises from the channel refinement procedure. Explicitly, based on the estimated data and its corresponding estimation error model as well as the pilot, the channel matrix is estimated again for improving the estimation accuracy. It is worth noting that the last two procedures can be implemented in an iterative manner. By contrast, the QMO-JO technique is significantly different from the previous one. Inspired by the fact that the final goal of pilot-aided communications is to transmit high-integrity data, we note that channel estimation is only an intermediate step. Therefore, strictly speaking, the performance of channel estimation should be quantified in terms of the final data integrity. In the second joint optimization regime, the channel estimation procedure is directly substituted into the data estimation procedure, optionally followed by the channel refinement for further enhancing the system performance.

III. Joint Transceiver and Channel Estimator Optimization

In this section, the joint optimization of the transceiver and channel estimator is investigated by relying on alternating optimization iterating between these two functional modules based on matrix-monotonic optimization [17]. This can be viewed as an extension based on the classic separate designs. Specifically, in traditional designs, only the estimated channel and its estimation error model are incorporated in the transceiver designs. However, there is a paucity of literature on jointly exploiting the estimated data and its estimation error model in the channel estimation procedure. Hence, in this paper the iterations between these two procedures are taken into account. In the channel estimation procedure, both the estimated channel matrix and the corresponding channel estimation error model are derived. Following that, in the data estimation procedure, the estimated data matrix and the corresponding data estimation error model can be obtained. It is worth noting that the data estimation error model is relevant to the channel estimation error model in the precoding procedure. Then, the estimated data are exploited as pilots and the channel estimation is refined based on the estimated data, the data estimation error model and the pilots. Finally, given the refined channel matrix and the updated channel error model, the transceiver may be further optimized to re-estimate the data signals.

A. Channel Estimation Procedure

According to the popular MMSE criterion and based on the signal model discussed in (4), the estimated channel matrix equals to [17]:

$$\widehat{\boldsymbol{H}} = \boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE}, \tag{5}$$

where the channel estimator $\boldsymbol{G}_{\rm CE}$ is chosen to minimize the following channel estimation MSE matrix

$$\begin{aligned} &\mathbb{E}\{\Delta\boldsymbol{H}^{\rm H}\Delta\boldsymbol{H}\} \\ =&(\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE}-\boldsymbol{I})^{\rm H}\boldsymbol{R}_{\rm H}(\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE}-\boldsymbol{I})+\boldsymbol{G}_{\rm CE}^{\rm H}\boldsymbol{R}_{\rm N}\boldsymbol{G}_{\rm CE}, \end{aligned} \tag{6}$$

where $\Delta\boldsymbol{H}=\boldsymbol{H}-\widehat{\boldsymbol{H}}$ is the channel estimation error, and the noise covariance matrix $\boldsymbol{R}_{\rm N}$ as well as the channel covariance matrix $\boldsymbol{R}_{\rm H}$ are defined as follows

$$\begin{aligned} \boldsymbol{R}_{\rm N} &= \mathbb{E}\{\boldsymbol{N}^{\rm H}\boldsymbol{N}\} = \sigma_{\rm n}^2 N_{\rm R}\boldsymbol{I}, \\ \boldsymbol{R}_{\rm H} &= \mathbb{E}\{\boldsymbol{H}^{\rm H}\boldsymbol{H}\} = N_{\rm R}\boldsymbol{\Psi}. \end{aligned} \tag{7}$$

The optimal LMMSE channel estimator $\boldsymbol{G}_{\rm CE}$ minimizing the MSE matrix in (6) is [39]

$$\boldsymbol{G}_{\rm CE} = (\boldsymbol{X}_{\rm P}^{\rm H}\boldsymbol{R}_{\rm H}\boldsymbol{X}_{\rm P}+\sigma_{\rm n}^2 N_{\rm R}\boldsymbol{I})^{-1}\boldsymbol{X}_{\rm P}^{\rm H}\boldsymbol{R}_{\rm H}. \tag{8}$$

The corresponding channel estimation error model is [38], [39]

$$\boldsymbol{H} = \widehat{\boldsymbol{H}} + \Delta\boldsymbol{H}, \quad \Delta\boldsymbol{H} = \boldsymbol{E}_{\rm C,W}\boldsymbol{\Phi}_{\rm C}^{1/2}, \tag{9}$$

where $\boldsymbol{E}_{\rm C,W}$ has i.i.d. Gaussian random elements, and $\boldsymbol{\Phi}_{\rm C}$ equals to

$$\boldsymbol{\Phi}_{\rm C} = \left(\boldsymbol{\Psi}^{-1} + \frac{1}{\sigma_n^2}\boldsymbol{X}_{\rm P}\boldsymbol{X}_{\rm P}^{\rm H}\right)^{-1}. \tag{10}$$

Given the estimated MIMO channel matrix, the key task is now to recover the desired data signals. Then the question arises whether the channel estimation accuracy is satisfactory for the data estimation. Here, we would like to point out an important fact that when the channel has been estimated, there exists a simple method of verifying the channel estimation accuracy, which is formulated in the following.

Based on the estimated channel matrix $\widehat{\boldsymbol{H}}$ and the channel estimation error model in (9), the signal model in the channel estimation procedure can be further rewritten as

$$\boldsymbol{Y}_{\rm P} = (\widehat{\boldsymbol{H}} + \Delta\boldsymbol{H})\boldsymbol{X}_{\rm P} + \boldsymbol{N}_{\rm P}. \tag{11}$$

Based on the reformulated signal model, there is a very simple strategy to evaluate whether the channel estimation performance is satisfactory. If the performance target is not met, further actions are needed to improve the estimation performance. Based on (11), if taking $\boldsymbol{X}_{\rm P}$ as the unknown data, the corresponding estimated pilot signals using the classic Wiener filter $\boldsymbol{G}$ are given by:

$$\widehat{\boldsymbol{X}}_{\rm P} = \boldsymbol{G}\boldsymbol{Y}_{\rm P}, \tag{12}$$

where $\boldsymbol{G} = \widehat{\boldsymbol{H}}^{\rm H}\big[\widehat{\boldsymbol{H}}\widehat{\boldsymbol{H}}^{\rm H} + (\sigma_n^2 + {\rm Tr}(\boldsymbol{\Phi}_{\rm C}))\boldsymbol{I}\big]^{-1}$ is closely related to the error-dependent estimated channel. Based on the estimated $\widehat{\boldsymbol{X}}_{\rm P}$, if the following inequality is satisfied

$$\|\widehat{\boldsymbol{X}}_{\rm P} - \boldsymbol{X}_{\rm P}\|_{\rm F}^2 > \tau, \tag{13}$$

where $\|\cdot\|_{\rm F}$ denotes the Frobenius norm and the positive scalar τ is a predefined estimation accuracy target, this means that the channel estimation accuracy is not good enough. Therefore, more resources or other methods are required for improving the estimation accuracy. In other words, the estimated channel should further be refined in order to achieve higher estimation accuracy.

B. Data Estimation Procedure

Based on the estimated channel and the corresponding channel estimation error model, in the data estimation procedure, the signal model is given as follows

$$\boldsymbol{Y}_{\mathrm{D}} = (\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}} + \Delta\boldsymbol{H})\boldsymbol{F}\boldsymbol{X}_{\mathrm{D}} + \boldsymbol{N}_{\mathrm{D}}. \tag{14}$$

It is plausible that when the estimation error is high, robust designs are of great importance to guarantee the overall system performance. When the channel estimation model is taken into account, the resultant regime is referred as robust transceiver optimization. In the following, some further results are given, which constitute the necessary basis for the ensuing analysis and design. Based on the signal model in (4), the estimated data is represented as

$$\begin{aligned}\widehat{\boldsymbol{X}}_{\mathrm{D}} =& \boldsymbol{G}_{\mathrm{DE}}\boldsymbol{Y}_{\mathrm{D}} \\ =& \boldsymbol{G}_{\mathrm{DE}}(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}} + \Delta\boldsymbol{H})\boldsymbol{F}\boldsymbol{X}_{\mathrm{D}} + \boldsymbol{G}_{\mathrm{DE}}\boldsymbol{N}_{\mathrm{D}}. \end{aligned} \tag{15}$$

When employing the MMSE criterion, the data transceiver $\boldsymbol{G}_{\mathrm{DE}}$ and $\boldsymbol{F}$ aim for minimizing the following data estimation MSE

$$\begin{aligned}&\mathbb{E}\{\mathrm{Tr}(\Delta\boldsymbol{X}_{\mathrm{D}}\Delta\boldsymbol{X}_{\mathrm{D}}^{\mathrm{H}})\} \\ =&\mathbb{E}\left\{\mathrm{Tr}\left((\widehat{\boldsymbol{X}}_{\mathrm{D}} - \boldsymbol{X}_{\mathrm{D}})(\widehat{\boldsymbol{X}}_{\mathrm{D}} - \boldsymbol{X}_{\mathrm{D}})^{\mathrm{H}}\right)\right\} \\ =&T_{\mathrm{D}}\mathrm{Tr}\big\{\boldsymbol{I} - \boldsymbol{G}_{\mathrm{DE}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}\boldsymbol{F} - (\boldsymbol{G}_{\mathrm{DE}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}\boldsymbol{F})^{\mathrm{H}} \\ &+ (\boldsymbol{G}_{\mathrm{DE}}[\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}} + \sigma_n^2\boldsymbol{I} \\ &+ \boldsymbol{I}\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{C}})]\boldsymbol{G}_{\mathrm{DE}}^{\mathrm{H}})\big\}. \end{aligned} \tag{16}$$

The detailed derivations are given in Appendix A. The LMMSE equalizer $\boldsymbol{G}_{\mathrm{DE}}$ minimizing the sum MSE in (16) may be formulated as:

$$\begin{aligned}\boldsymbol{G}_{\mathrm{DE}} =& (\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}\boldsymbol{F})^{\mathrm{H}}[\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}} + \sigma_n^2\boldsymbol{I} \\ &+ \boldsymbol{I}\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{C}})]^{-1}. \end{aligned} \tag{17}$$

Then the corresponding data estimation error model is represented as follows

$$\boldsymbol{X}_{\mathrm{D}} = \widehat{\boldsymbol{X}}_{\mathrm{D}} + \Delta\boldsymbol{X}_{\mathrm{D}}, \tag{18}$$

with the estimation error term $\Delta\boldsymbol{X}_{\mathrm{D}}$ given by:

$$\begin{aligned}\Delta\boldsymbol{X}_{\mathrm{D}} &= \boldsymbol{\Phi}_{\mathrm{D}}^{1/2}\boldsymbol{E}_{\mathrm{D,W}}, \\ \boldsymbol{\Phi}_{\mathrm{D}} &= \left(\frac{\boldsymbol{F}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}\boldsymbol{F}}{\sigma_n^2 + \mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{C}})} + \boldsymbol{I}\right)^{-1}. \end{aligned} \tag{19}$$

The elements of $\boldsymbol{E}_{\mathrm{D,W}}$ are i.i.d. distributed random variables with zero mean and unit variance.

The TPC optimization aims for solving the following equivalent optimization problem subject to the power constraints P_{D}

$$\begin{aligned}\min_{\boldsymbol{F}} \quad & \mathrm{Tr}\left[\left(\frac{\boldsymbol{F}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}\boldsymbol{F}}{\sigma_n^2 + \mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{C}})} + \boldsymbol{I}\right)^{-1}\right] \\ \text{s.t.} \quad & \mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}) \leq P_{\mathrm{D}}, \end{aligned} \tag{20}$$

which can be efficiently solved by using the matrix-monotonic optimization framework of [17]. Due to space limitation, the detailed derivation procedure of the optimal TPC is omitted here. The optimal linear TPC is given by [17]

$$\boldsymbol{F} = \sqrt{\frac{P_{\mathrm{D}}}{\mathrm{Tr}[(\sigma_{\mathrm{n}}^2\boldsymbol{I} + P_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{C}})^{-1}\boldsymbol{F}_{\mathrm{eq}}\boldsymbol{F}_{\mathrm{eq}}^{\mathrm{H}}]}}(\sigma_{\mathrm{n}}^2\boldsymbol{I} + P_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{C}})^{-\frac{1}{2}}\boldsymbol{F}_{\mathrm{eq}}, \tag{21}$$

with $\boldsymbol{F}_{\mathrm{eq}}$ being equivalent to

$$\boldsymbol{F}_{\mathrm{eq}} = \boldsymbol{U}_{\mathrm{C}}\left[\left(\frac{1}{\sqrt{\mu}}\boldsymbol{\Lambda}_{\mathrm{C}}^{-\frac{1}{2}} - \boldsymbol{\Lambda}_{\mathrm{C}}^{-1}\right)^{+}\right]^{\frac{1}{2}}, \tag{22}$$

where μ is the Lagrange multiplier for the sum power constraint, which is found by bisection search and ensures that the Pareto optimal solution $\mathrm{Tr}(\boldsymbol{F}_{\mathrm{eq}}\boldsymbol{F}_{\mathrm{eq}}^{\mathrm{H}}) = P_{\mathrm{D}}$ achieved. The matrices $\boldsymbol{U}_{\mathrm{C}}$ and $\boldsymbol{\Lambda}_{\mathrm{C}}$ are defined based on the following eigenvalue decomposition (EVD)

$$\begin{aligned}&(\sigma_{\mathrm{n}}^2\boldsymbol{I} + P_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{C}})^{-\frac{1}{2}}\boldsymbol{G}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE}}(\sigma_{\mathrm{n}}^2\boldsymbol{I} + P_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{C}})^{-\frac{1}{2}} \\ =&\boldsymbol{U}_{\mathrm{C}}\boldsymbol{\Lambda}_{\mathrm{C}}\boldsymbol{U}_{\mathrm{C}}^{\mathrm{H}}, \end{aligned} \tag{23}$$

where $\boldsymbol{\Lambda}_{\mathrm{C}}$ is a diagonal matrix with nonzero diagonal elements in a non-increasing order and $\boldsymbol{U}_{\mathrm{C}}^{\mathrm{H}}\boldsymbol{U}_{\mathrm{C}} = \boldsymbol{I}$. As a result, we could optimize $\boldsymbol{F}$ in (21), and then calculate $\boldsymbol{G}_{\mathrm{DE}}$ in (17) without iterations.

C. Channel Refinement Procedure

The above discussions represent the traditional way of dealing with channel estimation and data estimation. However, as discussed in [5], provided that the length of the pilot sequences is higher than the number of antennas, the optimal overall performance may be achieved. It might happen in large-scale MIMO systems that the channel block length is not long enough, or the energy to be allocated is insufficient. Here, we propose an extension of the traditional scheme, to deal with this potential deficiency. First, a few pilots are sent for channel estimation. The number of the pilots may be set equal to or even lower than the number of antennas. The corresponding performance will be discussed in Section V. Then, the data signals are transmitted, and estimated at the receiver. The recovered data symbols may then be used as pilots given their known statistical characteristics $\boldsymbol{\Phi}_{\mathrm{D}}$, in order to re-estimate the channel. Based on the estimated data signals and the associated estimation error model, there is more information that can be exploited for improving the channel estimation performance without increasing the pilot overhead. Similar to the above-mentioned robust transceiver optimization, this kind of design may be termed as robust channel estimation. Specifically, upon reusing the estimated data symbols as pilots and together with the corresponding data estimation error model, the resultant channel estimation model is now rewritten as

$$\boldsymbol{Y} = \boldsymbol{H}[\boldsymbol{X}_{\mathrm{P}}, \boldsymbol{F}(\widehat{\boldsymbol{X}}_{\mathrm{D}} + \Delta\boldsymbol{X}_{\mathrm{D}})] + [\boldsymbol{N}_{\mathrm{P}}, \boldsymbol{N}_{\mathrm{D}}]. \tag{24}$$

In contrast to the first equation in (3), the term $\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\mathrm{D}} + \Delta\boldsymbol{X}_{\mathrm{D}})$ in (24) is also harnessed as pilot signals. After performing linear channel estimation, the channel is re-estimated as

$$\widetilde{\boldsymbol{H}} = \boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}. \tag{25}$$

The channel estimator $\widetilde{\boldsymbol{G}}_{\mathrm{CE}}$ aims for minimizing the following MSE matrix

$$\begin{aligned}&\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\mathrm{H}}\Delta\widetilde{\boldsymbol{H}}\}\\=&\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}-\boldsymbol{R}_{\mathrm{H}}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}}]\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\\&-\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}+\boldsymbol{R}_{\mathrm{H}}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}},\end{aligned}\tag{26}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}&=\begin{bmatrix}\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}}\\ \widehat{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \widehat{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}}+\sigma_1\boldsymbol{I}\end{bmatrix},\\ \sigma_1&=\mathrm{Tr}(\boldsymbol{\Phi}_{\mathrm{D}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}).\end{aligned}\tag{27}$$

The detailed derivations can be found in Appendix B. Based on (26) and the definition of $\boldsymbol{\Sigma}$, the optimal channel estimator $\widetilde{\boldsymbol{G}}_{\mathrm{CE}}$ can be derived to be [43]

$$\widetilde{\boldsymbol{G}}_{\mathrm{CE}}=(\boldsymbol{\Sigma}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\boldsymbol{I})^{-1}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}},\tag{28}$$

and the corresponding data estimation MSE matrix of (26), given the optimal channel estimator in (28), is further rewritten in the following form:

$$\begin{aligned}&\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\mathrm{H}}\Delta\widetilde{\boldsymbol{H}}\}\\=&\boldsymbol{R}_{\mathrm{H}}-\boldsymbol{R}_{\mathrm{H}}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}}](\boldsymbol{\Sigma}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\boldsymbol{I})^{-1}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\\\triangleq&N_{\mathrm{R}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}}.\end{aligned}\tag{29}$$

According to Appendix C, the value of MSE matrix of channel refinement is reduced:

$$\begin{aligned}&\mathrm{Tr}(\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\mathrm{H}}\Delta\widetilde{\boldsymbol{H}}\}-\mathbb{E}\{\Delta\boldsymbol{H}^{\mathrm{H}}\Delta\boldsymbol{H}\})\\=&-\mathrm{Tr}[N_{\mathrm{R}}(\boldsymbol{\Phi}_{\mathrm{C}}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}})(\frac{\sigma_1}{N_{\mathrm{R}}}\boldsymbol{I}+\sigma_{\mathrm{n}}^2\boldsymbol{I}+\widehat{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{C}}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}})^{-1}\\&\times(\boldsymbol{\Phi}_{\mathrm{C}}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\mathrm{D}})^{\mathrm{H}}]<0.\end{aligned}\tag{30}$$

Since MSE involves expectation operation, the knowledge of channel statistics is what matters. In our work, we assume perfect knowledge on channel statistics such as noise variance and covariance matrices. Hence, the error propagation could be reduced. Additionally, the scenario of imperfect channel statistics is left for the future work.

After the channel refinement, the updated channel estimation error model becomes:

$$\boldsymbol{H}=\widetilde{\boldsymbol{H}}+\Delta\widetilde{\boldsymbol{H}},\quad \Delta\widetilde{\boldsymbol{H}}=\widetilde{\boldsymbol{E}}_{\mathrm{C,W}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}}^{1/2}.\tag{31}$$

Based on the refined estimated CSI and its estimation error model, the signal transmission model can be reformulated as follows

$$\widetilde{\boldsymbol{X}}_{\mathrm{D}}=\widetilde{\boldsymbol{G}}_{\mathrm{DE}}(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}+\Delta\widetilde{\boldsymbol{H}})\boldsymbol{F}\boldsymbol{X}_{\mathrm{D}}+\widetilde{\boldsymbol{G}}_{\mathrm{DE}}\boldsymbol{N}_{\mathrm{D}}.\tag{32}$$

In data transmission, the transceiver optimization typically aims for minimizing the following sum MSE

$$\begin{aligned}&\mathbb{E}\left\{\mathrm{Tr}(\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}})\right\}\\=&\mathrm{Tr}\Big(T_{\mathrm{D}}\boldsymbol{I}-T_{\mathrm{D}}\widetilde{\boldsymbol{G}}_{\mathrm{DE}}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F}-T_{\mathrm{D}}(\widetilde{\boldsymbol{G}}_{\mathrm{DE}}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F})^{\mathrm{H}}\\&+T_{\mathrm{D}}(\widetilde{\boldsymbol{G}}_{\mathrm{DE}}[\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}+\sigma_n^2\boldsymbol{I}\\&+\boldsymbol{I}\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}})]\widetilde{\boldsymbol{G}}_{\mathrm{DE}}^{\mathrm{H}})\Big).\end{aligned}\tag{33}$$

The detailed derivations for the expectation of $\mathbb{E}\{\mathrm{Tr}(\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}})\}$ are given in Appendix D. Based on the sum objective function in (33), it is readily seen that the optimal equalizer is the LMMSE equalizer of the following form:

$$\begin{aligned}\widetilde{\boldsymbol{G}}_{\mathrm{DE}}=&(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F})^{\mathrm{H}}[\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}\\&+\sigma_n^2\boldsymbol{I}+\boldsymbol{I}\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}})]^{-1},\end{aligned}\tag{34}$$

and substituting (34) into (33) the data estimation error can be rewritten as

$$\begin{aligned}&\mathbb{E}\left\{\mathrm{Tr}(\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}})\right\}\\=&T_{\mathrm{D}}\mathrm{Tr}\left[\left(\frac{\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F}}{\sigma_n^2+\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}})}+\boldsymbol{I}\right)^{-1}\right]\\\triangleq&T_{\mathrm{D}}\mathrm{Tr}\left(\widetilde{\boldsymbol{\Phi}}_{\mathrm{D}}\right).\end{aligned}\tag{35}$$

At this stage, a single round of channel refinement was completed, but the iterative channel refinement procedures may be continued for further improving the accuracy of the hitherto received data or the future transmitted and received data. However, to strike a beneficial performance vs. complexity trade-off, we only consider a single round of improvements in the simulation part and the simulation results show that a satisfactory performance can be achieved.

D. Channel Refinement Relying on Current Received Data

At the end of Section III-C, we acquired the refined channel estimation and re-estimated data signals after a single round. Then, we could utilize the re-estimated data for improving the channel estimation again, and channel refinement as well as data re-estimation may be processed in an alternating manner until convergence, where the estimators $\widetilde{\boldsymbol{G}}_{\mathrm{CE}}$, $\widetilde{\boldsymbol{G}}_{\mathrm{DE}}$ and the corresponding estimation error models would be updated in each iteration. They are similar to the ones given in Section III-C.

In the channel refinement procedures, the updated channel estimator is

$$\widetilde{\boldsymbol{G}}_{\mathrm{CE}}=(\boldsymbol{\Sigma}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\boldsymbol{I})^{-1}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}},\tag{36}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}&=\begin{bmatrix}\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}}\\ \widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}}+\sigma_2\boldsymbol{I}\end{bmatrix},\\ \sigma_2&=\mathrm{Tr}(\widetilde{\boldsymbol{\Phi}}_{\mathrm{D}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}),\end{aligned}\tag{37}$$

and the MSE matrix of channel estimation error is written as

$$\begin{aligned}\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\mathrm{H}}\Delta\widetilde{\boldsymbol{H}}\}=&\boldsymbol{R}_{\mathrm{H}}-\boldsymbol{R}_{\mathrm{H}}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}}](\boldsymbol{\Sigma}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\boldsymbol{I})^{-1}\\&\times[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\\\triangleq&N_{\mathrm{R}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}}.\end{aligned}\tag{38}$$

Upon reverting to the data estimation, the data estimator is updated as

$$\begin{aligned}\widetilde{\boldsymbol{G}}_{\mathrm{DE}}=&(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F})^{\mathrm{H}}[\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}+\sigma_n^2\boldsymbol{I}\\&+\boldsymbol{I}\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}})]^{-1},\end{aligned}\tag{39}$$

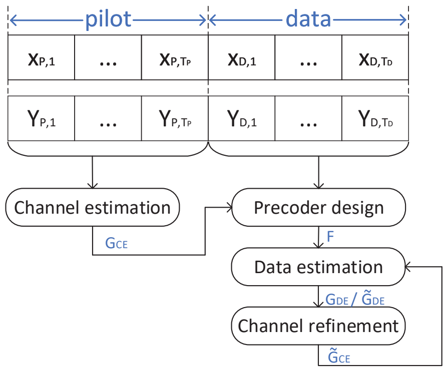

Fig. 1. Iterative optimization associated with current data.

and the data estimation error becomes:

$$\begin{aligned}&\mathbb{E}\left\{\mathrm{Tr}(\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}})\right\}\\=&T_{\mathrm{D}}\mathrm{Tr}\left[\left(\frac{\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\boldsymbol{F}}{\sigma_n^2+\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}})}+\boldsymbol{I}\right)^{-1}\right]\\\triangleq&T_{\mathrm{D}}\mathrm{Tr}\left(\widetilde{\boldsymbol{\Phi}}_{\mathrm{D}}\right).\end{aligned}\tag{40}$$

It may also be interpreted as an iterative optimization of a single data segment, which is illustrated in Fig. 1. Given the improvement of channel estimation accuracy, the sum MSE of data estimation error would tend to converge to a certain value, but several iterations might be harnessed at the cost of substantial computational overheads.

E. Channel Refinement Relying on Successive Received Data

Here, we consider iterative optimization across several data segments. If the channel's coherence time is quite long or the updated channel estimation and updated recovered data could be calculated in a timely manner, we could optimize the TPC $\boldsymbol{F}$ for the next data transmission with the results in Section III-C. The corresponding TPC optimization problem is

$$\begin{aligned}\min_{\widetilde{\boldsymbol{F}}}\ &\mathrm{Tr}\left[\left(\frac{\widetilde{\boldsymbol{F}}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\widetilde{\boldsymbol{F}}}{\sigma_n^2+\mathrm{Tr}(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}})}+\boldsymbol{I}\right)^{-1}\right]\\\mathrm{s.t.}\ &\mathrm{Tr}(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}})\leq P_{\mathrm{D}},\end{aligned}\tag{41}$$

which is the same as the optimization problem (20), and can be efficiently solved based on matrix-monotonic optimization [17].

When new data is precoded with $\widetilde{\boldsymbol{F}}$ and transmitted, we have:

$$\begin{aligned}\boldsymbol{Y}_{\mathrm{D,new}}&=\boldsymbol{H}\widetilde{\boldsymbol{F}}\boldsymbol{X}_{\mathrm{D,new}}+\boldsymbol{N}_{\mathrm{D,new}}\\&=(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}+\Delta\widetilde{\boldsymbol{H}})\widetilde{\boldsymbol{F}}\boldsymbol{X}_{\mathrm{D,new}}+\boldsymbol{N}_{\mathrm{D,new}},\end{aligned}\tag{42}$$

and then we can use the updated estimator to recover the new received data $\widehat{\boldsymbol{X}}_{\mathrm{D,new}}=\widetilde{\boldsymbol{G}}_{\mathrm{DE}}\boldsymbol{Y}_{\mathrm{D,new}}$

$$\begin{aligned}\widetilde{\boldsymbol{G}}_{\mathrm{DE}}=&(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\widetilde{\boldsymbol{F}})^{\mathrm{H}}[\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}+\sigma_n^2\boldsymbol{I}\\&+\boldsymbol{I}\mathrm{Tr}(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}})]^{-1}.\end{aligned}\tag{43}$$

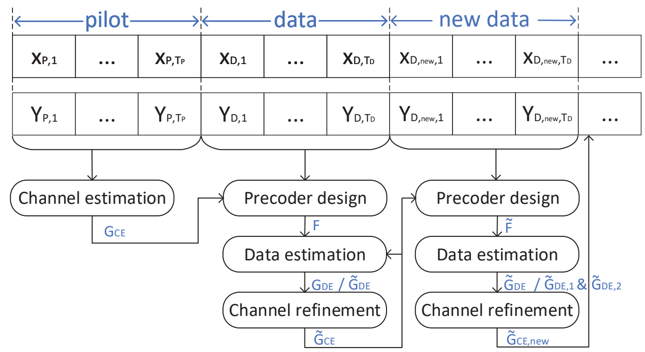

Fig. 2. Iterative optimization associated with successive data.

The corresponding data estimation error matrix becomes:

$$\begin{aligned}&\mathbb{E}\left\{\mathrm{Tr}(\Delta\widehat{\boldsymbol{X}}_{\mathrm{D,new}}\Delta\widehat{\boldsymbol{X}}_{\mathrm{D,new}}^{\mathrm{H}})\right\}\\=&T_{\mathrm{D,new}}\mathrm{Tr}\left[\left(\frac{\widetilde{\boldsymbol{F}}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE}}\widetilde{\boldsymbol{F}}}{\sigma_n^2+\mathrm{Tr}(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C}})}+\boldsymbol{I}\right)^{-1}\right]\\\triangleq&T_{\mathrm{D,new}}\mathrm{Tr}\left(\widetilde{\boldsymbol{\Phi}}_{\mathrm{D,new}}\right).\end{aligned}\tag{44}$$

Then we may exploit the new received data as pilots again for refining the channel estimation as discussed before, e.g. with $[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}(\widetilde{\boldsymbol{X}}_{\mathrm{D}}+\Delta\widetilde{\boldsymbol{X}}_{\mathrm{D}}),\widetilde{\boldsymbol{F}}(\widehat{\boldsymbol{X}}_{\mathrm{D,new}}+\Delta\widehat{\boldsymbol{X}}_{\mathrm{D,new}})]$. The optimal channel estimator is updated as

$$\widetilde{\boldsymbol{G}}_{\mathrm{CE,new}}=(\boldsymbol{\Sigma}_{\mathrm{new}}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\boldsymbol{I})^{-1}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}},\widetilde{\boldsymbol{F}}\widehat{\boldsymbol{X}}_{\mathrm{D,new}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}},\tag{45}$$

where $\boldsymbol{\Sigma}_{\mathrm{new}}$ is defined in (46).

The modified channel estimation is $\widetilde{\boldsymbol{H}}_{\mathrm{new}}=\boldsymbol{Y}_{\mathrm{new}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,new}}$, where $\boldsymbol{Y}_{\mathrm{new}}=[\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{Y}_{\mathrm{D,new}}]$, and the MSE matrix of the channel estimation error becomes:

$$\begin{aligned}&\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}_{\mathrm{new}}^{\mathrm{H}}\Delta\widetilde{\boldsymbol{H}}_{\mathrm{new}}\}\\=&\boldsymbol{R}_{\mathrm{H}}-\boldsymbol{R}_{\mathrm{H}}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}},\widetilde{\boldsymbol{F}}\widehat{\boldsymbol{X}}_{\mathrm{D,new}}](\boldsymbol{\Sigma}_{\mathrm{new}}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\boldsymbol{I})^{-1}\\&\times[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}},\widetilde{\boldsymbol{F}}\widehat{\boldsymbol{X}}_{\mathrm{D,new}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\\\triangleq&N_{\mathrm{R}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C,new}}.\end{aligned}\tag{47}$$

Then as for the data estimation, since different segments of data are precoded with different TPCs, the updated data equalizers with respect to different sequences are

$$\begin{aligned}\widetilde{\boldsymbol{G}}_{\mathrm{DE},1}=&(\boldsymbol{Y}_{\mathrm{new}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,new}}\boldsymbol{F})^{\mathrm{H}}[\sigma_n^2\boldsymbol{I}+\boldsymbol{I}\mathrm{Tr}(\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C,new}})\\&+\boldsymbol{Y}_{\mathrm{new}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,new}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,new}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{new}}^{\mathrm{H}}]^{-1}\end{aligned}\tag{48}$$

$$\begin{aligned}\widetilde{\boldsymbol{G}}_{\mathrm{DE},2}=&(\boldsymbol{Y}_{\mathrm{new}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,new}}\widetilde{\boldsymbol{F}})^{\mathrm{H}}[\sigma_n^2\boldsymbol{I}+\boldsymbol{I}\mathrm{Tr}(\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C,new}})\\&+\boldsymbol{Y}_{\mathrm{new}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,new}}\widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,new}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{new}}^{\mathrm{H}}]^{-1},\end{aligned}\tag{49}$$

and different data estimation errors may be derived, respectively. The schematic is depicted in Fig. 2. Note that, with more data segments harnessed as pilots, they are supposed to be recovered using different equalizers, correspondingly resulting in different estimation error models.

IV. Joint Quadratic Matrix Optimization

Inspired by the fact that the final goal is to recover the desired data, channel estimation is only an intermediate step, and hence the QMO-JO regime is proposed. A straightforward solution is to simply embed the channel estimation procedure into the data estimation procedure, where no explicit channel estimation is required. Then joint optimization is performed by optimizing the data estimation performance. Based on the previous discussions, from a pure mathematical viewpoint, the joint transceiver design has to solve the following equation array

$$\begin{aligned} \boldsymbol{Y}_{\mathrm{P}} &= \boldsymbol{H}\boldsymbol{X}_{\mathrm{P}} + \boldsymbol{N}_{\mathrm{P}}, \\ \boldsymbol{y}_{\mathrm{D}} &= \boldsymbol{H}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} + \boldsymbol{n}. \end{aligned} \tag{50}$$

Here, it is worth noting that $\boldsymbol{N}_{\mathrm{P}}$ and $\boldsymbol{n}$ represent a random matrix and random vector, respectively. In order to recover the desired signal accurately, the channel matrix $\boldsymbol{H}$ can be estimated based on the first equation. After performing linear channel estimation $\boldsymbol{G}_{\mathrm{CE,Q}}$, the estimated channel may be expressed as $\widehat{\boldsymbol{H}}_{\mathrm{Q}} = \boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}$, and thus the recovered signal can be written as:

$$\begin{aligned} \hat{\boldsymbol{s}} =& \boldsymbol{G}_{\mathrm{DE,Q}}(\widehat{\boldsymbol{H}}_{\mathrm{Q}} + \Delta\boldsymbol{H}_{\mathrm{Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} + \boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{n} \\ =& \boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} \\ &+ \boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{H} - \boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}} - \boldsymbol{N}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} \\ &+ \boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{n}, \end{aligned} \tag{51}$$

where $\boldsymbol{G}_{\mathrm{DE,Q}}$ is the linear equalizer matrix at the destination and $\boldsymbol{F}_{\mathrm{Q}}$ is the linear TPC matrix at source. The joint optimization of the estimator and transceiver, which aims for minimizing the data estimation MSE subject to a power constraint, is formulated as

$$\begin{aligned} \min_{\boldsymbol{\mathcal{A}}} \ & \mathbb{E}\{\|\boldsymbol{W}^{\frac{1}{2}}[\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} \\ & + \boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{H} - \boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}} - \boldsymbol{N}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} \\ & + \boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{n} - \boldsymbol{s}]\|^2\} - \log|\boldsymbol{W}| \\ \text{s.t.} \ & \mathrm{Tr}(\boldsymbol{F}_{\mathrm{Q}}{\boldsymbol{F}_{\mathrm{Q}}}^{\mathrm{H}}) \leq P_{\mathrm{D}}, \end{aligned} \tag{52}$$

where $\boldsymbol{W}$ is the weighting matrix, $\mathbb{E}\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\} = \boldsymbol{I}$, and we have the set $\boldsymbol{\mathcal{A}} = \{(\boldsymbol{G}_{\mathrm{DE,Q}}, \boldsymbol{G}_{\mathrm{CE,Q}}, \boldsymbol{F}_{\mathrm{Q}})\}$. The weighted MSE minimization in (52) is more general than the sum MSE. In our work, $\boldsymbol{W}$ is a constant matrix and if $\boldsymbol{W}$ is taken as optimization variable as well, the optimization (52) is equivalent to the maximization of the mutual information in the data transmission phase. As derived in Appendix E, the objective function in (52) equals

$$\begin{aligned} & \mathbb{E}\{\|\boldsymbol{W}^{\frac{1}{2}}[\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} \\ & + \boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{H} - \boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}} - \boldsymbol{N}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} \\ & + \boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{n} - \boldsymbol{s}]\|^2\} - \log|\boldsymbol{W}| \\ =& \mathrm{Tr}[\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}} + \boldsymbol{W} \\ & - \boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}} - (\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}] \\ & + \sigma_n^2 \ \mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) \\ & + \mathrm{Tr}[\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}(\boldsymbol{I} - \boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})^{\mathrm{H}}] \\ & \times \mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) \\ & + \sigma_n^2 \ \mathrm{Tr}(\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}) \ \mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) \\ & - \log|\boldsymbol{W}| \\ \triangleq & \boldsymbol{\Phi}_{\mathrm{D,Q}}. \end{aligned} \tag{53}$$

It is readily seen that the joint optimization problem (52) exhibits quadratic properties with respect to any single matrix variable. This kind of optimization problems are termed as quadratic matrix optimization. In the following, the optimization problem (52) is solved by alternatively optimizing $\boldsymbol{G}_{\mathrm{DE,Q}}$, $\boldsymbol{G}_{\mathrm{CE,Q}}$, and $\boldsymbol{F}_{\mathrm{Q}}$ with the rest variables fixed. At each iteration, the closed-form optimization solutions of $\boldsymbol{G}_{\mathrm{DE,Q}}$, $\boldsymbol{G}_{\mathrm{CE,Q}}$, and $\boldsymbol{F}_{\mathrm{Q}}$ can be derived and thus the convergence of the proposed algorithm may be guaranteed.

1) Optimization of $\boldsymbol{G}_{\mathrm{CE,Q}}$: The optimization with respect to $\boldsymbol{G}_{\mathrm{CE,Q}}$ is unconstrained and the objective function (OF) is also convex with respect to $\boldsymbol{G}_{\mathrm{CE,Q}}$, i.e., it is a standard quadratic function. Then, the optimal $\boldsymbol{G}_{\mathrm{CE,Q}}$ can be solved directly based on its corresponding complex matrix derivative operation.

In order to optimize the channel estimator $\boldsymbol{G}_{\mathrm{CE,Q}}$, the OF of (52) with respect to $\boldsymbol{G}_{\mathrm{CE,Q}}$ is reformulated as

$$\begin{aligned} & \mathbb{E}\{\|\boldsymbol{W}^{\frac{1}{2}}[\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} \\ & + \boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{H} - \boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}} - \boldsymbol{N}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s} \\ & + \boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{n} - \boldsymbol{s}]\|^2\} - \log|\boldsymbol{W}| \\ =& \mathrm{Tr}[\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}} \\ & - \boldsymbol{F}_{\mathrm{Q}}\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}} - (\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}] \\ & + \sigma_n^2\mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}})\mathrm{Tr}(\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}) \\ & + \mathrm{Tr}[\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{\Psi}\boldsymbol{X}_{\mathrm{P}} \\ & - \boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{\Psi}\boldsymbol{X}_{\mathrm{P}} - \boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{\Psi}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}] \\ & \times \mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) + c_p, \end{aligned} \tag{54}$$

where c_p is the term independent of $\boldsymbol{G}_{\mathrm{CE,Q}}$. Then, the optimal solution of $\boldsymbol{G}_{\mathrm{CE,Q}}$ may be shown to be:

$$\begin{aligned} \boldsymbol{G}_{\mathrm{CE,Q}} =& [\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}} + \mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) \\ & \times (\sigma_n^2\boldsymbol{I} + \boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{\Psi}\boldsymbol{X}_{\mathrm{P}})]^{-1}[\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}} \\ & + \mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{\Psi}\boldsymbol{X}_{\mathrm{P}}]^{\mathrm{H}}(\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}})^{-1}. \end{aligned} \tag{55}$$

$$\boldsymbol{\Sigma}_{\mathrm{new}} = \begin{bmatrix} \boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}} & \boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\widetilde{\boldsymbol{F}}\widehat{\boldsymbol{X}}_{\mathrm{D,new}} \\ \widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}} + \phi_1\boldsymbol{I} & \widetilde{\boldsymbol{X}}_{\mathrm{D}}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\widetilde{\boldsymbol{F}}\widehat{\boldsymbol{X}}_{\mathrm{D,new}} \\ \widehat{\boldsymbol{X}}_{\mathrm{D,new}}^{\mathrm{H}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \widehat{\boldsymbol{X}}_{\mathrm{D,new}}^{\mathrm{H}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}\widetilde{\boldsymbol{X}}_{\mathrm{D}} & \widehat{\boldsymbol{X}}_{\mathrm{D,new}}^{\mathrm{H}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\widetilde{\boldsymbol{F}}\widehat{\boldsymbol{X}}_{\mathrm{D,new}} + \phi_2\boldsymbol{I} \end{bmatrix},$$

$$\phi_1 = \mathrm{Tr}(\widetilde{\boldsymbol{\Phi}}_{\mathrm{D}}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}), \quad \phi_2 = \mathrm{Tr}(\widetilde{\boldsymbol{\Phi}}_{\mathrm{D,new}}\widetilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\widetilde{\boldsymbol{F}}). \tag{46}$$

2) Optimization of $\boldsymbol{G}_{\mathrm{DE,Q}}$: The optimization of $\boldsymbol{G}_{\mathrm{DE,Q}}$ is similar to $\boldsymbol{G}_{\mathrm{CE,Q}}$. Specifically, the optimization is an unconstrained convex problem. Following a similar logic, the optimal solution of $\boldsymbol{G}_{\mathrm{DE,Q}}$ is obtained as

$$\begin{aligned}\boldsymbol{G}_{\mathrm{DE,Q}} =&(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}\left((\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}\right.\\ &+\{\mathrm{Tr}[\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})^{\mathrm{H}}]\\ &\left.+\sigma_n^2\mathrm{Tr}(\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}})+\sigma_n^2\}\boldsymbol{I}\right)^{-1}, \end{aligned} \tag{56}$$

which is in nature an LMMS equalizer. Defining the following auxiliary matrix variable $\boldsymbol{\Phi}_{\mathrm{C,Q}}$

$$\begin{aligned}\boldsymbol{\Phi}_{\mathrm{C,Q}} =&(\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}-\boldsymbol{I})^{\mathrm{H}}\boldsymbol{\Psi}(\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}-\boldsymbol{I})\\ &+\sigma_n^2\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}},\end{aligned} \tag{57}$$

the optimization problem (53) is equivalent to

$$\begin{aligned}\min_{\boldsymbol{F}_{\mathrm{Q}},\boldsymbol{G}_{\mathrm{CE,Q}}}\ & \mathrm{Tr}\left[\boldsymbol{W}\left(\frac{\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}}{\sigma_n^2+\mathrm{Tr}(\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}{}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{C,Q}})}+\boldsymbol{I}\right)^{-1}\right]\\ \mathrm{s.t.}\ \ & \boldsymbol{\Phi}_{\mathrm{C,Q}}=(\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}-\boldsymbol{I})^{\mathrm{H}}\boldsymbol{\Psi}(\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}-\boldsymbol{I})\\ &\quad+\sigma_n^2\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}},\\ & \mathrm{Tr}(\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}})\le P_{\mathrm{D}}.\end{aligned} \tag{58}$$

It is worth emphasizing that the optimization of $\boldsymbol{G}_{\mathrm{DE,Q}}$ has been omitted and only $\boldsymbol{G}_{\mathrm{CE,Q}}$ and $\boldsymbol{F}_{\mathrm{Q}}$ have to be optimized in an alternating manner, after which the optimal $\boldsymbol{G}_{\mathrm{DE,Q}}$ is obtained in (56). It is also plausible that the optimization of $\boldsymbol{F}_{\mathrm{Q}}$ in (58) can be solved efficiently based on the framework of matrix-monotonic optimization. Moreover, the convergence speed of the proposed alternating algorithm has been accelerated.

3) Optimization of $\boldsymbol{F}_{\mathrm{Q}}$: Similar to Section III-B, the optimal linear TPC $\boldsymbol{F}_{\mathrm{Q}}$ is

$$\begin{aligned}\boldsymbol{F}_{\mathrm{Q}} =&\sqrt{\frac{P_{\mathrm{D}}}{\mathrm{Tr}[(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{C,Q}})^{-1}\boldsymbol{F}_{\mathrm{eq,Q}}\boldsymbol{F}_{\mathrm{eq,Q}}^{\mathrm{H}}]}}\\ &\times(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{C,Q}})^{-\frac{1}{2}}\boldsymbol{F}_{\mathrm{eq,Q}},\end{aligned} \tag{59}$$

where the rectangular matrix $\boldsymbol{F}_{\mathrm{eq,Q}}$ equals

$$\boldsymbol{F}_{\mathrm{eq,Q}}=\boldsymbol{U}_{\mathrm{Q}}\left[\left(\frac{1}{\sqrt{\mu_{\mathrm{Q}}}}\boldsymbol{\Lambda}_{\mathrm{Q}}^{-\frac{1}{2}}-\boldsymbol{\Lambda}_{\mathrm{Q}}^{-1}\right)^{+}\right]^{\frac{1}{2}}, \tag{60}$$

and μ_{Q} is the Lagrange multiplier guaranteeing $\mathrm{Tr}(\boldsymbol{F}_{\mathrm{eq,Q}}\boldsymbol{F}_{\mathrm{eq,Q}}^{\mathrm{H}})=P_{\mathrm{D}}$ satisfied. The matrices $\boldsymbol{U}_{\mathrm{Q}}$ and $\boldsymbol{\Lambda}_{\mathrm{Q}}$ are defined based on the following EVD

$$\begin{aligned}&(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{C,Q}})^{-\frac{1}{2}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}(\sigma_{\mathrm{n}}^2\boldsymbol{I}+P_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{C,Q}})^{-\frac{1}{2}}\\ =&\boldsymbol{U}_{\mathrm{Q}}\boldsymbol{\Lambda}_{\mathrm{Q}}\boldsymbol{U}_{\mathrm{Q}}^{\mathrm{H}},\end{aligned} \tag{61}$$

where $\boldsymbol{\Lambda}_{\mathrm{Q}}$ is a diagonal matrix with nonzero diagonal elements in a non-increasing order and $\boldsymbol{U}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{U}_{\mathrm{Q}}=\boldsymbol{I}$.

4) Discussion: The main differences between FM-JO and QMO-JO are whether to find an explicit channel estimation solution and the update criterion of $\boldsymbol{G}_{\mathrm{CE}}$. Explicitly, in FM-JO, $\boldsymbol{G}_{\mathrm{CE}}$ is firstly optimized to minimize the MSE of channel estimation error before the data estimation and channel refinement phases. By contrast, in QMO-JO, $\boldsymbol{G}_{\mathrm{CE,Q}}$ is jointly optimized with $\boldsymbol{G}_{\mathrm{DE,Q}}$ and $\boldsymbol{F}_{\mathrm{Q}}$ to minimize the MSE of data estimation error. It seems that FM-JO before channel refinement is a special case of QMO-JO, where $\boldsymbol{G}_{\mathrm{CE,Q}}$ is fixed, and only $\boldsymbol{G}_{\mathrm{DE,Q}}$ and $\boldsymbol{F}_{\mathrm{Q}}$ are optimized jointly. Furthermore, the optimal $\boldsymbol{G}_{\mathrm{CE}}$ in FM-JO is almost the same as $\boldsymbol{G}_{\mathrm{CE,Q}}$ in QMO-JO at high SNR, which tends to be $\boldsymbol{X}_{\mathrm{P}}^{-1}$. It will also be demonstrated in Section V. At the same time, following the idea of channel refinement in FM-JO, this may also be conducted in QMO-JO for improving the system performance, which we term as QMO-FM-JO. Section V will explore whether there exists a further beneficial performance improvement. First, we jointly optimize the transceiver as in Section IV with the knowledge of transmitted pilot signals and received pilot signals for minimizing the error of data estimation $\boldsymbol{\Phi}_{\mathrm{D,Q}}$.

Then, following the logic in Section III, the recovered data signals are taken into account for refining the estimation of the channel. The updated channel estimator $\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}$ may be expressed as

$$\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}=(\boldsymbol{\Sigma}_{\mathrm{Q}}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\boldsymbol{I})^{-1}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}_{\mathrm{Q}}\widehat{\boldsymbol{X}}_{\mathrm{D,Q}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}, \tag{62}$$

where $\widehat{\boldsymbol{X}}_{\mathrm{D,Q}}=\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{D}}$ and

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathrm{Q}}=&\begin{bmatrix}\boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \boldsymbol{X}_{\mathrm{P}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}_{\mathrm{Q}}\widehat{\boldsymbol{X}}_{\mathrm{D,Q}}\\ \widehat{\boldsymbol{X}}_{\mathrm{D,Q}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{Q}}{}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{X}_{\mathrm{P}} & \widehat{\boldsymbol{X}}_{\mathrm{D,Q}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}_{\mathrm{Q}}\widehat{\boldsymbol{X}}_{\mathrm{D,Q}}+\sigma_3\boldsymbol{I}\end{bmatrix},\\ \sigma_3 =&\mathrm{Tr}(\boldsymbol{\Phi}_{\mathrm{D,Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\boldsymbol{F}_{\mathrm{Q}}).\end{aligned} \tag{63}$$

The corresponding estimation error of the refined estimated channel may be shown to be

$$\begin{aligned}\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}_{\mathrm{Q}}^{\mathrm{H}}\Delta\widetilde{\boldsymbol{H}}_{\mathrm{Q}}\}=&\boldsymbol{R}_{\mathrm{H}}-\boldsymbol{R}_{\mathrm{H}}[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}_{\mathrm{Q}}\widehat{\boldsymbol{X}}_{\mathrm{D,Q}}](\boldsymbol{\Sigma}_{\mathrm{Q}}+\sigma_{\mathrm{n}}^2N_{\mathrm{R}}\boldsymbol{I})^{-1}\\ &\times[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}_{\mathrm{Q}}\widehat{\boldsymbol{X}}_{\mathrm{D,Q}}]^{\mathrm{H}}\boldsymbol{R}_{\mathrm{H}}\\ \triangleq&\widetilde{\boldsymbol{\Phi}}_{\mathrm{C,Q}}.\end{aligned} \tag{64}$$

As a result, the updated equalizer associated with $\boldsymbol{Y}=[\boldsymbol{Y}_{\mathrm{P}},\boldsymbol{Y}_{\mathrm{D}}]$ is

$$\begin{aligned}\widetilde{\boldsymbol{G}}_{\mathrm{DE,Q}}=&(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}[\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}\\ &+\sigma_n^2\boldsymbol{I}+\boldsymbol{I}\mathrm{Tr}(\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}{}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C,Q}})]^{-1},\end{aligned} \tag{65}$$

and error of data estimation is

$$\mathrm{Tr}\left[\left(\frac{\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}}{\sigma_n^2+\mathrm{Tr}(\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}{}^{\mathrm{H}}\widetilde{\boldsymbol{\Phi}}_{\mathrm{C,Q}})}+\boldsymbol{I}\right)^{-1}\right]. \tag{66}$$

The QMO-FM-JO procedure is illustrated in Algorithm 1, where both QMO-JO and FM-JO may be viewed as the special parts or particular steps of it.

In addition, inspired by the connection between FM-JO and QMO-JO, the channel refinement and data re-estimation may be solved together, which only minimizes the MSE of data re-estimation error with the aid of the estimated data via jointly optimizing $\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}$ and $\widetilde{\boldsymbol{G}}_{\mathrm{DE,Q}}$. Upon defining $\boldsymbol{X}=[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{X}_{\mathrm{D}}]$, $\overline{\boldsymbol{X}}=[\boldsymbol{X}_{\mathrm{P}},\boldsymbol{F}_{\mathrm{Q}}\widehat{\boldsymbol{X}}_{\mathrm{D,Q}}]$, the OF is written as

$$\begin{aligned}&\mathbb{E}\{\|\boldsymbol{W}^{\frac{1}{2}}[\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\boldsymbol{Y}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s}+\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}(\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\\ &-\boldsymbol{N}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{s}+\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\boldsymbol{n}-\boldsymbol{s}]\|^2\}-\log|\boldsymbol{W}|\\ =&\mathrm{Tr}[\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}(\boldsymbol{Y}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})(\boldsymbol{Y}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}^{\mathrm{H}}+\boldsymbol{W}\end{aligned}$$

Algorithm 1 Joint optimization of the channel estimator and data transceiver along with channel estimation refinement

1. Initialize $\boldsymbol{G}_{\mathrm{CE,Q}}^{(0)}$ and $i = 1$.

repeat

2. Update $\boldsymbol{F}_{\mathrm{Q}}^{(i)}$ with fixed $\boldsymbol{G}_{\mathrm{CE,Q}}^{(i-1)}$ in (59).
3. Update $\boldsymbol{G}_{\mathrm{CE,Q}}^{(i)}$ with fixed $\boldsymbol{F}_{\mathrm{Q}}^{(i)}$ in (55).

until the convergence of objective function in (52)

4. Design $\boldsymbol{G}_{\mathrm{DE,Q}}$ in (56) and obtain recovered data $\widehat{\boldsymbol{X}}_{\mathrm{D,Q}}$.
5. Re-estimate channel $\widetilde{\boldsymbol{H}}$ with updated $\widetilde{\boldsymbol{G}}_{\mathrm{CE,Q}}$ in (62).
6. Re-estimate data $\widetilde{\boldsymbol{X}}_{\mathrm{D,Q}}$ with updated $\widetilde{\boldsymbol{G}}_{\mathrm{DE,Q}}$ in (65) in turn.

$$
\begin{aligned}
&- \boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\boldsymbol{Y}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}} - (\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\boldsymbol{Y}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}] \\
&+ \sigma_n^2 \ \mathrm{Tr}(\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}^{\mathrm{H}}) + \mathrm{Tr}(\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}^{\mathrm{H}}) \\
&\times \mathrm{Tr}[\boldsymbol{\Psi}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}} + \boldsymbol{\Psi}\overline{\boldsymbol{X}}\ \overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}^{\mathrm{H}}\overline{\boldsymbol{X}}^{\mathrm{H}} \\
&- \boldsymbol{\Psi}\overline{\boldsymbol{X}}\ \overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}} - \boldsymbol{\Psi}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}^{\mathrm{H}}\overline{\boldsymbol{X}}^{\mathrm{H}} \\
&+ \mathrm{Tr}(\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}^{\mathrm{H}})\boldsymbol{\Psi}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{\Phi}_{\mathrm{D,Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}] \\
&+ \sigma_n^2 \ \mathrm{Tr}(\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}^{\mathrm{H}}) \ \mathrm{Tr}(\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}^{\mathrm{H}}) \\
&- \log|\boldsymbol{W}|. \qquad (67)
\end{aligned}
$$

We then iteratively update $\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}$ and $\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}$ until the convergence of the OF. Fixing $\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}$, the optimal solution of $\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}$ becomes:

$$
\begin{aligned}
\overline{\boldsymbol{G}}_{\mathrm{DE,Q}} =& (\boldsymbol{Y}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}}((\boldsymbol{Y}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})(\boldsymbol{Y}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}})^{\mathrm{H}} \\
&+ \{\mathrm{Tr}[\boldsymbol{\Psi}(\boldsymbol{I} - \overline{\boldsymbol{X}}\ \overline{\boldsymbol{G}}_{\mathrm{CE,Q}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}(\boldsymbol{I} - \overline{\boldsymbol{X}}\ \overline{\boldsymbol{G}}_{\mathrm{CE,Q}})^{\mathrm{H}}] \\
&+ \mathrm{Tr}(\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}^{\mathrm{H}})\mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{\Phi}_{\mathrm{D,Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}) \\
&+ \sigma_n^2 \mathrm{Tr}(\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}^{\mathrm{H}}) + \sigma_n^2\}\boldsymbol{I})^{-1}, \qquad (68)
\end{aligned}
$$

and the solution of $\overline{\boldsymbol{G}}_{\mathrm{CE,Q}}$ with fixed $\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}$ becomes:

$$
\begin{aligned}
\overline{\boldsymbol{G}}_{\mathrm{CE,Q}} =& [\mathrm{Tr}(\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}^{\mathrm{H}})([\sigma_n^2 + \mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{\Phi}_{\mathrm{D,Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}})]\boldsymbol{I} \\
&+ \overline{\boldsymbol{X}}^{\mathrm{H}}\boldsymbol{\Psi}\overline{\boldsymbol{X}}) + \boldsymbol{Y}^{\mathrm{H}}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}^{\mathrm{H}}\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\boldsymbol{Y}]^{-1} \\
&\times [\mathrm{Tr}(\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}^{\mathrm{H}})\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}}\boldsymbol{\Psi}\overline{\boldsymbol{X}} \\
&+ \boldsymbol{F}_{\mathrm{Q}}\boldsymbol{W}\overline{\boldsymbol{G}}_{\mathrm{DE,Q}}\boldsymbol{Y}]^{\mathrm{H}}(\boldsymbol{F}_{\mathrm{Q}}\boldsymbol{F}_{\mathrm{Q}}^{\mathrm{H}})^{-1}. \qquad (69)
\end{aligned}
$$

5) Complexity Analysis: The computational complexity of the proposed algorithm mainly comes from the matrix multiplication and matrix decompositions. The specific computational complexity of each procedure is illustrated in Table III. Observe that the computational complexity of channel refinement and data re-estimation is highly dependent on the length of the estimated data, which is exploited as pilots.

TABLE III
COMPUTATIONAL COMPLEXITY OF THE PROPOSED ALGORITHM

Algorithm/Procedure		Computational Complexity
Conventional method	Channel Estimation	$\mathcal{O}(N^3)$
	Data Estimation	$\mathcal{O}(N^3)$
FM-JO	Channel Refinement	$\mathcal{O}((T_{\mathrm{D}} + T_{\mathrm{P}})^3)$
	Data Re-estimation	$\mathcal{O}(N^2(T_{\mathrm{D}} + T_{\mathrm{P}}))$
QMO-JO		$\mathcal{O}(N^3)$

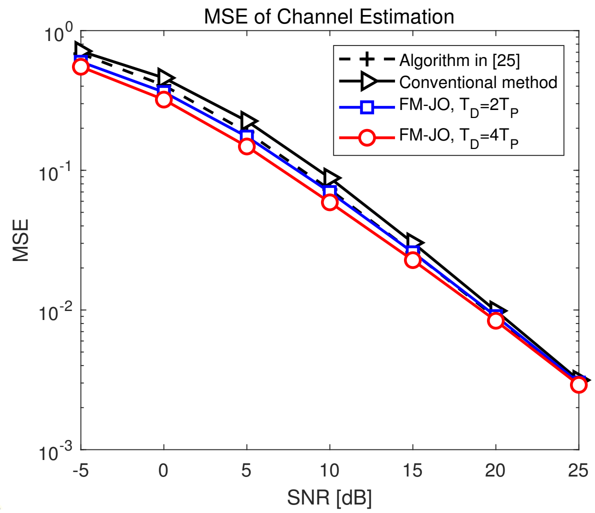

Fig. 3. The MSE of channel estimation associated with $N_{\mathrm{T}} = N_{\mathrm{R}} = r = 64$. The results of the conventional method are evaluated from the MSE metrics in Eq. (6) and those of FM-JO with different T_{D} are from the MSE metrics in Eq. (26).

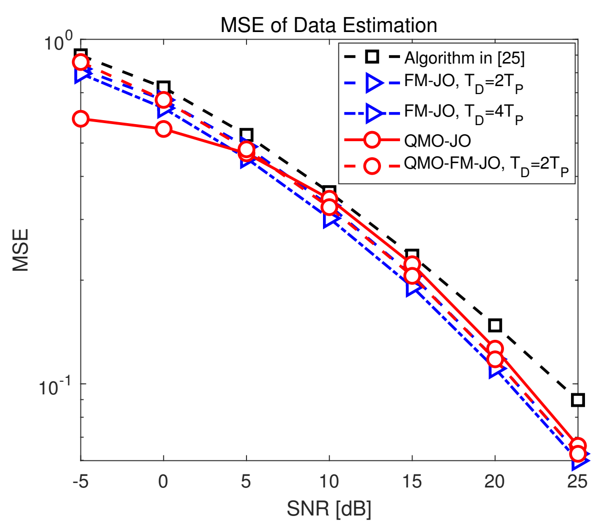

Fig. 4. The MSE of data estimation associated with $N_{\mathrm{T}} = N_{\mathrm{R}} = r = 64$. The results of FM-JO, QMO-JO and QMO-FM-JO are evaluated from the MSE metrics in Eq. (33), Eq. (53) and Eq. (66), respectively.

V. Numerical Results and Discussions

In this section, some simulation results are given to access the performance of the proposed joint optimization of the channel estimator and data transceiver. The most widely used exponential correlation model, i.e., $[\boldsymbol{\Psi}]_{i,j} = \theta^{|i-j|}$ [17] is adopted for the transmit correlation matrix, and θ is set as 0.4 without loss of generality. The transmitted pilot signals are orthogonal, data signals are QPSK modulated, and the simulation results are averaged over 10^4 independent random realizations. The length of the transmitted symbols to compute the BER is set to 10^4. Algorithm in [25] is chosen as the baseline algorithm. A single round of channel estimation refinement is adapted in the simulation.

First, we investigate the scenario of $N_{\mathrm{T}} = N_{\mathrm{R}} = 64$, and discuss the influence of different levels of noise on the

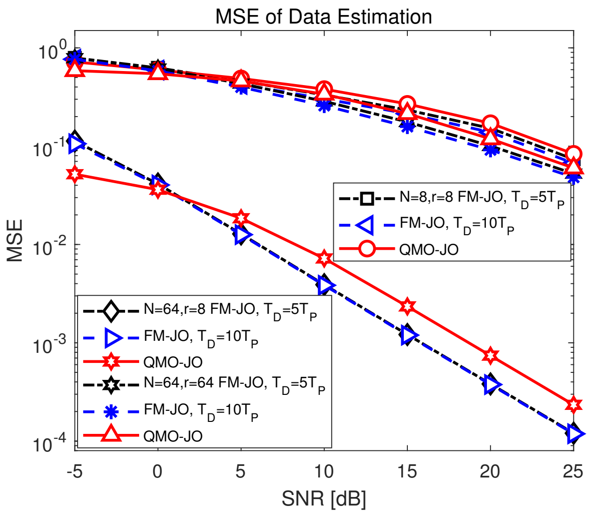

Fig. 5. The MSE of data estimation under different numbers of antennas and data streams.

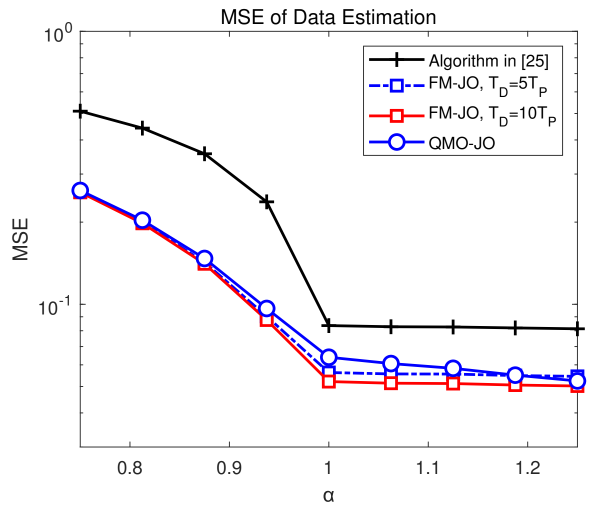

Fig. 6. The MSE of data estimation versus α associated with $N_{\mathrm{T}} = N_{\mathrm{R}} = 64$ at SNR=25dB.

performance of the proposed scheme. Note that, $\mathbb{E}\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\} = \boldsymbol{I}$ and the SNR mentioned here is $10\lg(\frac{1}{\sigma_n^2})$ dB. In Fig. 3, the MSE of FM-JO is always better than that of the conventional method, which only optimizes $\boldsymbol{G}_{\mathrm{CE}}$ once. This shows that the FM-JO beneficially exploits the transmitted data for improving the accuracy of channel estimation. When the length of the data used for channel refinement is $T_{\mathrm{D}} = 4T_{\mathrm{P}}$, the performance gap between the conventional method and FM-JO becomes obvious in Fig. 3 and Fig. 4, compared to the situation of $T_{\mathrm{D}} = 2T_{\mathrm{P}}$. This is because the longer transmitted data provides more information for the channel estimator and enhances the performance of data recovery, albeit at higher computational complexity. Although the algorithm in [25] also used data-aided channel estimation, its performance improvement was lower than that of FM-JO relying on our precoder design.

Additionally, as illustrated in Fig. 4, the MSE of QMO-JO is lower than that of FM-JO at low SNRs. Since the data estimation is inaccurate due to the poor channel estimation at low SNRs, the channel refinement is inefficient, while the QMO-JO succeeds in jointly optimizing the parameters for minimizing the MSE of data estimation, thereby attaining improved system performance. At a high SNR, the MSE of FM-JO is superior to that of QMO-JO, since it processes more accurate estimated data as pilots. Furthermore, QMO-FM-JO combines the advantages of FM-JO and QMO-JO. At a low SNR, the MSE of QMO-FM-JO is almost the same as that of FM-JO, since the inefficient channel estimation neutralizes the improvements caused by the QMO-JO. Upon increasing the SNRs, the QMO-FM-JO starts to perform better than the FM-JO, and they perform similarly at a high SNR, due to the specific characteristics of QMO-JO.

Then we discuss the effect of different numbers of antennas and data streams. Fig. 5 shows that when the number of data streams is equal to the number of antennas, the $N_{\mathrm{T}} = N_{\mathrm{R}} = 64$ scheme performs slightly better than $N_{\mathrm{T}} = N_{\mathrm{R}} = 8$. Since the sum power of the precoder is constant, having more antennas means that less power is allocated for each antenna, hence resulting in a modest improvement of the performance. In addition, when the number of data streams is set as 8, the scenario of $N_{\mathrm{T}} = N_{\mathrm{R}} = 64$ is superior to $N_{\mathrm{T}} = N_{\mathrm{R}} = 8$, thus improving the spectral efficiency of the communication system. However, the improvement of FM-JO with $T_{\mathrm{D}} = 10T_{\mathrm{P}}$ is almost the same as for $T_{\mathrm{D}} = 5T_{\mathrm{P}}$. This indicates that when the number of antennas is high, more data is required for channel refinement to achieve higher accuracy.

Let us denote the ratio of pilot sequence length to the number of transmit antennas by α. The above simulations are based on the assumption that the training sequence lengths are equal to the number of transmit antennas, i.e. $\alpha = 1$. Fig. 6 illustrates the situation when the SNR is 25dB. The proposed algorithms show excellent performance compared to the baseline algorithm, even when there is a shortage of pilots. The estimated data in FM-JO may be reused as pilots for efficiently refining the channel estimation and the QMO-JO scheme jointly optimizes the channel and data estimation. Furthermore, the performances of the proposed algorithms are similar, when insufficient pilots are used. Observe that, the FM-JO associated with $T_{\mathrm{D}} = 10T_{\mathrm{P}}$ performs better in the data-aided mode. Additionally, it also demonstrates that increasing the length of pilots beyond the number of antennas slightly improves the MSE of data estimation at a high SNR, but not sufficiently for justifying the increased pilot overhead.

Let us now consider the ultimate BER criterion to illustrate the performance of data detection. The baseline algorithm is that the TPC $\boldsymbol{F}$ is simply a DFT matrix, and $\boldsymbol{G}_{\mathrm{CE}}$ and $\boldsymbol{G}_{\mathrm{DE}}$ are optimized as in (8) and (17), respectively. Fig. 7 shows the performance of QMO-JO vs. SNR, where QMO-JO is seen to outperform the baseline and gradually approaches the performance based on perfect CSI. Furthermore, Fig. 8 indicates that QMO-JO compensates well for the shortage of pilots.

Fig. 9 shows the convergence of the proposed algorithms. Observed that they both converge within 3 iterations. As for

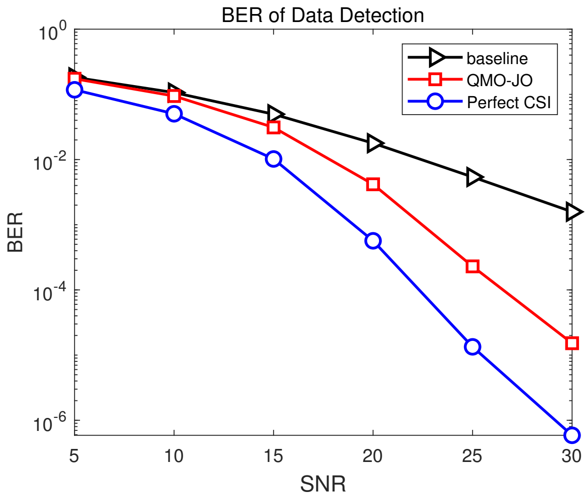

Fig. 7. The BER of data detection in QMO-JO versus SNR associated with $N_\mathrm{T} = N_\mathrm{R} = 16$.

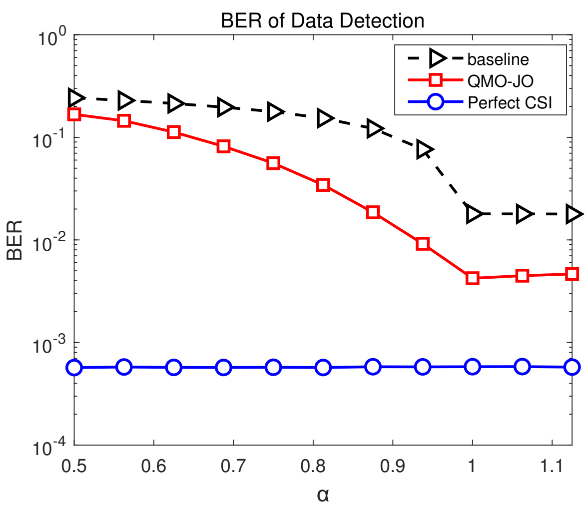

Fig. 8. The BER of data detection in QMO-JO versus α with $N_\mathrm{T} = N_\mathrm{R} = 16$ when the SNR is 20dB.

FM-JO, we have shown theoretically that the refined channel aided by the data exhibits a lower MSE than the conventional method solely relying on pilots. It also shows that an attractive performance vs. computational complexity trade-off may be attained after a single iteration. As for QMO-JO, since we were able to derive the optimal solution in every iteration with all other variables fixed, it is seen to converge to a fixed point and thus a reduced MSE is achieved.

It may be concluded that, at high SNR, FM-JO is preferred for channel estimation and data recovery in order to achieve improved performance, but this is achieved at an increased computation overhead. By contrast, at low SNR or in face of insufficient training resources, QMO-JO has the edge. This is because $\boldsymbol{G}_\mathrm{CE}$ that minimizes the channel estimation error might not be optimal in terms of minimizing the MSE of data estimation at low SNRs, particularly leading to avalanche-like error proliferation during channel refinement.

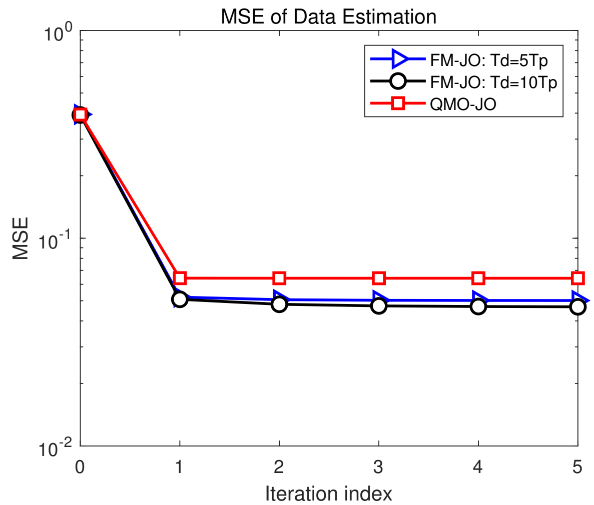

Fig. 9. The convergence of data estimation based on FM-JO with the length of data $T_\mathrm{D} = 5T_\mathrm{P}$ and $T_\mathrm{D} = 10T_\mathrm{P}$, and QMO-JO, when the number of antennas is 64×64 and we have SNR=25dB.

VI. Summery and Conclusions

In this paper, the joint optimization of the channel estimator and data transceiver was considered. The joint optimization relied on solving a pair of linear matrix equalities in the presence of noise terms in each. The channel estimation and data transmission phases were jointly designed. The channel estimation accuracy affects the data transmission and estimated data can also be reused as pilots to improve the channel estimation accuracy. In the FM-JO solution, the estimation error models the impact of channel estimation on data estimation and of data estimation on channel estimation, which is investigated. In the QMO-JO algorithm, the channel estimation phase is incorporated into the data estimation and then a joint optimization problem is formulated. Closed-form solutions are derived and connections of these algorithms are also investigated. Finally, the simulation results have illustrated the performance benefits of the proposed algorithms.

Appendix A
Derivations for $\mathbb{E}\{\mathrm{Tr}(\Delta\boldsymbol{X}_\mathrm{D}\Delta\boldsymbol{X}_\mathrm{D}^\mathrm{H})\}$

In this appendix, the quadratic expectation term in (16) is derived. Assuming $\mathbb{E}\{\boldsymbol{X}_\mathrm{D}\boldsymbol{X}_\mathrm{D}^\mathrm{H}\} = T_\mathrm{D}\boldsymbol{I}$ and performing statistical expectation operations over $\Delta\boldsymbol{X}_\mathrm{D}$, the term $\mathbb{E}\{\mathrm{Tr}(\Delta\boldsymbol{X}_\mathrm{D}\Delta\boldsymbol{X}_\mathrm{D}^\mathrm{H})\}$ equals to [44]

$$
\begin{aligned}
&\mathbb{E}\{\mathrm{Tr}(\Delta\boldsymbol{X}_\mathrm{D}\Delta\boldsymbol{X}_\mathrm{D}^\mathrm{H})\}\\
=&\mathbb{E}\left\{\mathrm{Tr}[(\widehat{\boldsymbol{X}}_\mathrm{D}-\boldsymbol{X}_\mathrm{D})(\widehat{\boldsymbol{X}}_\mathrm{D}-\boldsymbol{X}_\mathrm{D})^\mathrm{H}]\right\}\\
=&\mathbb{E}\Big\{\mathrm{Tr}([\boldsymbol{G}_\mathrm{DE}(\boldsymbol{Y}_\mathrm{P}\boldsymbol{G}_\mathrm{CE}+\Delta\boldsymbol{H})\boldsymbol{F}\boldsymbol{X}_\mathrm{D}+\boldsymbol{G}_\mathrm{DE}\boldsymbol{N}_\mathrm{D}-\boldsymbol{X}_\mathrm{D}]\\
&\times[\boldsymbol{G}_\mathrm{DE}(\boldsymbol{Y}_\mathrm{P}\boldsymbol{G}_\mathrm{CE}+\Delta\boldsymbol{H})\boldsymbol{F}\boldsymbol{X}_\mathrm{D}+\boldsymbol{G}_\mathrm{DE}\boldsymbol{N}_\mathrm{D}-\boldsymbol{X}_\mathrm{D}]^\mathrm{H})\Big\}\\
=&T_\mathrm{D}\mathbb{E}\Big\{\mathrm{Tr}([\boldsymbol{G}_\mathrm{DE}(\boldsymbol{Y}_\mathrm{P}\boldsymbol{G}_\mathrm{CE}+\Delta\boldsymbol{H})\boldsymbol{F}-\boldsymbol{I})]\\
&\times[\boldsymbol{G}_\mathrm{DE}(\boldsymbol{Y}_\mathrm{P}\boldsymbol{G}_\mathrm{CE}+\Delta\boldsymbol{H})\boldsymbol{F}-\boldsymbol{I}]^\mathrm{H}\Big\}\\
&+\sigma_n^2T_\mathrm{D}\mathrm{Tr}(\boldsymbol{G}_\mathrm{DE}\boldsymbol{G}_\mathrm{DE}^\mathrm{H})
\end{aligned}
$$

$$=T_{\rm D}{\rm Tr}\{\boldsymbol{I}-\boldsymbol{G}_{\rm DE}\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE}\boldsymbol{F}-(\boldsymbol{G}_{\rm DE}\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE}\boldsymbol{F})^{\rm H} + (\boldsymbol{G}_{\rm DE}[\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE}\boldsymbol{F}\boldsymbol{F}^{\rm H}\boldsymbol{G}_{\rm CE}^{\rm H}\boldsymbol{Y}_{\rm P}^{\rm H}+\sigma_n^2\boldsymbol{I} + \boldsymbol{I}{\rm Tr}(\boldsymbol{F}\boldsymbol{F}^{\rm H}\boldsymbol{\Phi}_{\rm C})]\boldsymbol{G}_{\rm DE}^{\rm H})\}. \tag{70}$$

Appendix B
Derivations for $\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\rm H}\Delta\widetilde{\boldsymbol{H}}\}$

Taking the statistical expectation over $\Delta\widetilde{\boldsymbol{H}}$, the quadratic term $\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\rm H}\Delta\widetilde{\boldsymbol{H}}\}$ equals to

$$\begin{aligned}
&\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\rm H}\Delta\widetilde{\boldsymbol{H}}\}\\
=&\mathbb{E}\{(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}-\boldsymbol{H})^{\rm H}(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}-\boldsymbol{H})\}\\
=&\mathbb{E}\{(\boldsymbol{H}([\boldsymbol{X}_{\rm P},\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\rm D}+\Delta\boldsymbol{X}_{\rm D})]\widetilde{\boldsymbol{G}}_{\rm CE}-\boldsymbol{I})+[\boldsymbol{N}_{\rm P},\boldsymbol{N}_{\rm D}]\widetilde{\boldsymbol{G}}_{\rm CE})^{\rm H}\\
&\times(\boldsymbol{H}([\boldsymbol{X}_{\rm P},\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\rm D}+\Delta\boldsymbol{X}_{\rm D})]\widetilde{\boldsymbol{G}}_{\rm CE}-\boldsymbol{I})+[\boldsymbol{N}_{\rm P},\boldsymbol{N}_{\rm D}]\widetilde{\boldsymbol{G}}_{\rm CE})\}\\
=&\mathbb{E}\{(\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}[\boldsymbol{X}_{\rm P},\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\rm D}+\Delta\boldsymbol{X}_{\rm D})]^{\rm H}-\boldsymbol{I})\boldsymbol{H}^{\rm H}\boldsymbol{H}\\
&\times([\boldsymbol{X}_{\rm P},\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\rm D}+\Delta\boldsymbol{X}_{\rm D})]\widetilde{\boldsymbol{G}}_{\rm CE}-\boldsymbol{I})\}+\sigma_{\rm n}^2N_{\rm R}\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}\widetilde{\boldsymbol{G}}_{\rm CE}\\
=&\mathbb{E}\{\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}[\boldsymbol{X}_{\rm P},\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\rm D}+\Delta\boldsymbol{X}_{\rm D})]^{\rm H}\boldsymbol{R}_{\rm H}\\
&\times[\boldsymbol{X}_{\rm P},\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\rm D}+\Delta\boldsymbol{X}_{\rm D})]\widetilde{\boldsymbol{G}}_{\rm CE}\}-\boldsymbol{R}_{\rm H}[\boldsymbol{X}_{\rm P},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}]\widetilde{\boldsymbol{G}}_{\rm CE}\\
&-\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}[\boldsymbol{X}_{\rm P},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}]^{\rm H}\boldsymbol{R}_{\rm H}+\boldsymbol{R}_{\rm H}+\sigma_{\rm n}^2N_{\rm R}\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}\widetilde{\boldsymbol{G}}_{\rm CE}.
\end{aligned}\tag{71}$$

The quadratic term in the first term of the fourth equality in (71) can be further reformulated into the following formula

$$\begin{aligned}
&\mathbb{E}\{[\boldsymbol{X}_{\rm P},\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\rm D}+\Delta\boldsymbol{X}_{\rm D})]^{\rm H}\boldsymbol{R}_{\rm H}[\boldsymbol{X}_{\rm P},\boldsymbol{F}(\widehat{\boldsymbol{X}}_{\rm D}+\Delta\boldsymbol{X}_{\rm D})]\}\\
=&\begin{bmatrix}\boldsymbol{X}_{\rm P}^{\rm H}\boldsymbol{R}_{\rm H}\boldsymbol{X}_{\rm P} & \boldsymbol{X}_{\rm P}^{\rm H}\boldsymbol{R}_{\rm H}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}\\ \widehat{\boldsymbol{X}}_{\rm D}^{\rm H}\boldsymbol{F}^{\rm H}\boldsymbol{R}_{\rm H}\boldsymbol{X}_{\rm P} & \widehat{\boldsymbol{X}}_{\rm D}^{\rm H}\boldsymbol{F}^{\rm H}\boldsymbol{R}_{\rm H}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}+\boldsymbol{I}{\rm Tr}(\boldsymbol{\Phi}_{\rm D}\boldsymbol{F}^{\rm H}\boldsymbol{R}_{\rm H}\boldsymbol{F})\end{bmatrix}\\
\triangleq&\boldsymbol{\Sigma}.
\end{aligned}\tag{72}$$

As a result, we finally have

$$\begin{aligned}
&\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\rm H}\Delta\widetilde{\boldsymbol{H}}\}\\
=&\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}\boldsymbol{\Sigma}\widetilde{\boldsymbol{G}}_{\rm CE}-\boldsymbol{R}_{\rm H}[\boldsymbol{X}_{\rm P},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}]\widetilde{\boldsymbol{G}}_{\rm CE}-\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}[\boldsymbol{X}_{\rm P},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}]^{\rm H}\boldsymbol{R}_{\rm H}\\
&+\boldsymbol{R}_{\rm H}+\sigma_{\rm n}^2N_{\rm R}\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}\widetilde{\boldsymbol{G}}_{\rm CE}.
\end{aligned}\tag{73}$$

Appendix C
Derivations for $\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\rm H}\Delta\widetilde{\boldsymbol{H}}\}$ with optimal $\widetilde{\boldsymbol{G}}_{\rm CE}$

The MSE of channel refinement relying on the optimal $\widetilde{\boldsymbol{G}}_{\rm CE}$ may be rewritten as

$$\begin{aligned}
&\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\rm H}\Delta\widetilde{\boldsymbol{H}}\}\\
=&\boldsymbol{R}_{\rm H}-\boldsymbol{R}_{\rm H}[\boldsymbol{X}_{\rm P},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}](\boldsymbol{\Sigma}+\sigma_{\rm n}^2N_{\rm R}\boldsymbol{I})^{-1}[\boldsymbol{X}_{\rm P},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}]^{\rm H}\boldsymbol{R}_{\rm H},
\end{aligned}\tag{74}$$

and we define

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{U}\\ \boldsymbol{V} & \boldsymbol{D}\end{bmatrix}=\boldsymbol{\Sigma}+\sigma_{\rm n}^2N_{\rm R}\boldsymbol{I}.\tag{75}$$

Based on the inversion of the block matrix and the Woodbury identity, we arrive at:

$$\begin{aligned}
&[\boldsymbol{X}_{\rm P},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}](\boldsymbol{\Sigma}+\sigma_{\rm n}^2N_{\rm R}\boldsymbol{I})^{-1}[\boldsymbol{X}_{\rm P},\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}]^{\rm H}\\
=&\boldsymbol{X}_{\rm P}\boldsymbol{A}^{-1}\boldsymbol{X}_{\rm P}^{\rm H}+(\boldsymbol{X}_{\rm P}\boldsymbol{A}^{-1}\boldsymbol{U}-\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D})(\boldsymbol{D}-\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\\
&\times(\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{X}_{\rm P}^{\rm H}-\widehat{\boldsymbol{X}}_{\rm D}^{\rm H}\boldsymbol{F}^{\rm H})\\
=&\boldsymbol{X}_{\rm P}(\boldsymbol{X}_{\rm P}^{\rm H}\boldsymbol{R}_{\rm H}\boldsymbol{X}_{\rm P}+\sigma_{\rm n}^2N_{\rm R}\boldsymbol{I})^{-1}\boldsymbol{X}_{\rm P}^{\rm H}+\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}_{\rm C}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D}\\
&\times(\sigma_1\boldsymbol{I}+\sigma_{\rm n}^2N_{\rm R}\boldsymbol{I}+N_{\rm R}\widehat{\boldsymbol{X}}_{\rm D}^{\rm H}\boldsymbol{F}^{\rm H}\boldsymbol{\Phi}_{\rm C}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D})^{-1}\\
&\times(\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}_{\rm C}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D})^{\rm H}.
\end{aligned}\tag{76}$$

Then, the MSE of channel refinement is simplified to

$$\begin{aligned}
&\mathbb{E}\{\Delta\widetilde{\boldsymbol{H}}^{\rm H}\Delta\widetilde{\boldsymbol{H}}\}\\
=&N_{\rm R}\boldsymbol{\Phi}_{\rm C}-N_{\rm R}(\boldsymbol{\Phi}_{\rm C}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D})\\
&\times(\frac{\sigma_1}{N_{\rm R}}\boldsymbol{I}+\sigma_{\rm n}^2\boldsymbol{I}+\widehat{\boldsymbol{X}}_{\rm D}^{\rm H}\boldsymbol{F}^{\rm H}\boldsymbol{\Phi}_{\rm C}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D})^{-1}(\boldsymbol{\Phi}_{\rm C}\boldsymbol{F}\widehat{\boldsymbol{X}}_{\rm D})^{\rm H}.
\end{aligned}\tag{77}$$

Appendix D
Derivations for $\mathbb{E}\{{\rm Tr}(\Delta\widetilde{\boldsymbol{X}}_{\rm D}\Delta\widetilde{\boldsymbol{X}}_{\rm D}^{\rm H})\}$

Upon performing the statistical expectation over $\Delta\widetilde{\boldsymbol{X}}_{\rm D}$, the resultant quadratic term equals to

$$\begin{aligned}
&\mathbb{E}\{{\rm Tr}(\Delta\widetilde{\boldsymbol{X}}_{\rm D}\Delta\widetilde{\boldsymbol{X}}_{\rm D}^{\rm H})\}\\
=&\mathbb{E}\left\{{\rm Tr}\left[\left(\widetilde{\boldsymbol{X}}_{\rm D}-\boldsymbol{X}_{\rm D}\right)\left(\widetilde{\boldsymbol{X}}_{\rm D}-\boldsymbol{X}_{\rm D}\right)^{\rm H}\right]\right\}\\
=&\mathbb{E}\Big\{{\rm Tr}\Big(\Big[\widetilde{\boldsymbol{G}}_{\rm DE}(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}+\Delta\widetilde{\boldsymbol{H}})\boldsymbol{F}\boldsymbol{X}_{\rm D}+\widetilde{\boldsymbol{G}}_{\rm DE}\boldsymbol{N}_{\rm D}-\boldsymbol{X}_{\rm D}\Big]\\
&\times\Big[\widetilde{\boldsymbol{G}}_{\rm DE}(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}+\Delta\widetilde{\boldsymbol{H}})\boldsymbol{F}\boldsymbol{X}_{\rm D}+\widetilde{\boldsymbol{G}}_{\rm DE}\boldsymbol{N}_{\rm D}-\boldsymbol{X}_{\rm D}\Big]^{\rm H}\Big)\Big\}\\
=&T_{\rm D}\mathbb{E}\Big\{{\rm Tr}([\widetilde{\boldsymbol{G}}_{\rm DE}(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}+\Delta\widetilde{\boldsymbol{H}})\boldsymbol{F}-\boldsymbol{I}]\\
&\times[\widetilde{\boldsymbol{G}}_{\rm DE}(\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}+\Delta\widetilde{\boldsymbol{H}})\boldsymbol{F}-\boldsymbol{I}]^{\rm H})\Big\}+\sigma_n^2T_{\rm D}{\rm Tr}(\widetilde{\boldsymbol{G}}_{\rm DE}\widetilde{\boldsymbol{G}}_{\rm DE}^{\rm H})\\
=&T_{\rm D}{\rm Tr}[\boldsymbol{I}-\widetilde{\boldsymbol{G}}_{\rm DE}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}\boldsymbol{F}-(\widetilde{\boldsymbol{G}}_{\rm DE}\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}\boldsymbol{F})^{\rm H}\\
&+(\widetilde{\boldsymbol{G}}_{\rm DE}[\boldsymbol{Y}\widetilde{\boldsymbol{G}}_{\rm CE}\boldsymbol{F}\boldsymbol{F}^{\rm H}\widetilde{\boldsymbol{G}}_{\rm CE}^{\rm H}\boldsymbol{Y}^{\rm H}+\sigma_n^2\boldsymbol{I}\\
&+\boldsymbol{I}{\rm Tr}(\boldsymbol{F}\boldsymbol{F}^{\rm H}\widetilde{\boldsymbol{\Phi}}_{\rm C})]\widetilde{\boldsymbol{G}}_{\rm DE}^{\rm H})].
\end{aligned}\tag{78}$$

Appendix E
The Objective Function of Joint Optimization (52)

In this appendix, the objective function of the joint optimization problem (52) is derived. The detailed mathematical derivation procedure is formulated as

$$\begin{aligned}
&\mathbb{E}\{\|\boldsymbol{W}^{\frac{1}{2}}[\boldsymbol{G}_{\rm DE,Q}\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\boldsymbol{F}\boldsymbol{s}+\boldsymbol{G}_{\rm DE,Q}(\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\\
&-\boldsymbol{N}_{\rm P}\boldsymbol{G}_{\rm CE,Q})\boldsymbol{F}\boldsymbol{s}+\boldsymbol{G}_{\rm DE,Q}\boldsymbol{n}-\boldsymbol{s}]\|^2\}-\log|\boldsymbol{W}|\\
=&\mathbb{E}\{{\rm Tr}(\boldsymbol{W}^{\frac{1}{2}}[\boldsymbol{G}_{\rm DE,Q}\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\boldsymbol{F}\boldsymbol{s}+\boldsymbol{G}_{\rm DE,Q}(\boldsymbol{H}\\
&-\boldsymbol{H}\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE,Q}-\boldsymbol{N}_{\rm P}\boldsymbol{G}_{\rm CE,Q})\boldsymbol{F}\boldsymbol{s}+\boldsymbol{G}_{\rm DE,Q}\boldsymbol{n}-\boldsymbol{s}])\\
&\times(\boldsymbol{W}^{\frac{1}{2}}[\boldsymbol{G}_{\rm DE,Q}\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\boldsymbol{F}\boldsymbol{s}+\boldsymbol{G}_{\rm DE,Q}(\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\\
&-\boldsymbol{N}_{\rm P}\boldsymbol{G}_{\rm CE,Q})\boldsymbol{F}\boldsymbol{s}+\boldsymbol{G}_{\rm DE,Q}\boldsymbol{n}-\boldsymbol{s}])^{\rm H}\}-\log|\boldsymbol{W}|\\
=&\mathbb{E}\{{\rm Tr}([\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\rm DE,Q}(\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE,Q}+\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\\
&-\boldsymbol{N}_{\rm P}\boldsymbol{G}_{\rm CE,Q})\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}]\\
&\times[\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\rm DE,Q}(\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE,Q}+\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\\
&-\boldsymbol{N}_{\rm P}\boldsymbol{G}_{\rm CE,Q})\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}]^{\rm H})\}\\
&+\sigma_n^2{\rm Tr}(\boldsymbol{W}\boldsymbol{G}_{\rm DE,Q}\boldsymbol{G}_{\rm DE,Q}^{\rm H})-\log|\boldsymbol{W}|\\
=&\mathbb{E}\{{\rm Tr}([\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\rm DE,Q}(\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE,Q}+\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE,Q})\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}]\\
&\times[\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\rm DE,Q}(\boldsymbol{Y}_{\rm P}\boldsymbol{G}_{\rm CE,Q}+\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\rm P}\boldsymbol{G}_{\rm CE,Q})\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}]^{\rm H})\}\\
&+\mathbb{E}\{{\rm Tr}[(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\rm DE,Q}\boldsymbol{N}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\boldsymbol{F})\\
&\times(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\rm DE,Q}\boldsymbol{N}_{\rm P}\boldsymbol{G}_{\rm CE,Q}\boldsymbol{F})^{\rm H}]\}\\
&+\sigma_n^2{\rm Tr}(\boldsymbol{W}\boldsymbol{G}_{\rm DE,Q}\boldsymbol{G}_{\rm DE,Q}^{\rm H})-\log|\boldsymbol{W}|
\end{aligned}$$

$$=\mathbb{E}\{\mathrm{Tr}([\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}+\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}] \times[\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}+\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}]^{\mathrm{H}})\} + \sigma_n^2\mathrm{Tr}(\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}})\mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) + \sigma_n^2\mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) - \log|\boldsymbol{W}|. \quad (79)$$

The first term in the final equality of (79) is further reformulated into

$$\mathbb{E}\{\mathrm{Tr}([\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}+\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}] \times[\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}+\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}]^{\mathrm{H}})\}$$
$$=\mathrm{Tr}[(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}}) \times(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}-\boldsymbol{W}^{\frac{1}{2}})^{\mathrm{H}}] + \mathbb{E}\{\mathrm{Tr}([\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{H}(\boldsymbol{I}-\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}] \times[\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{H}(\boldsymbol{I}-\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}]^{\mathrm{H}})\}$$
$$=\mathrm{Tr}[\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F})(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F})^{\mathrm{H}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}+\boldsymbol{W} - \boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}-(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F})^{\mathrm{H}}] + \mathrm{Tr}[\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})^{\mathrm{H}}] \times\mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) + \sigma_n^2\ \mathrm{Tr}(\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}})\mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}). \quad (80)$$

Therefore, the objective function of the joint optimization (52) can be reformulated into

$$\mathbb{E}\{\|\boldsymbol{W}^{\frac{1}{2}}[\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}\boldsymbol{s}+\boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{H}-\boldsymbol{H}\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}} - \boldsymbol{N}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}\boldsymbol{s}+\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{n}-\boldsymbol{s}]\|^2\}-\log|\boldsymbol{W}|$$
$$=\mathrm{Tr}[\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F})(\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F})^{\mathrm{H}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}+\boldsymbol{W} - \boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}-(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{Y}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F})^{\mathrm{H}}] + \sigma_n^2\ \mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}})+\mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) \times\mathrm{Tr}[\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{X}_{\mathrm{P}}\boldsymbol{G}_{\mathrm{CE,Q}})^{\mathrm{H}}] + \sigma_n^2\ \mathrm{Tr}(\boldsymbol{G}_{\mathrm{CE,Q}}\boldsymbol{F}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{G}_{\mathrm{CE,Q}}^{\mathrm{H}})\mathrm{Tr}(\boldsymbol{W}\boldsymbol{G}_{\mathrm{DE,Q}}\boldsymbol{G}_{\mathrm{DE,Q}}^{\mathrm{H}}) -\log|\boldsymbol{W}|. \quad (81)$$

References

[1] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, 2000.

[2] J. An, L. Wang, C. Xu, L. Gan, and L. Hanzo, "Optimal pilot power based channel estimation improves the throughput of intelligent reflective surface assisted systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16 202–16 206, 2020.

[3] Z. Wang, J. Zhang, B. Ai, C. Yuen, and M. Debbah, "Uplink performance of cell-free massive MIMO with multi-antenna users over jointly-correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7391–7406, 2022.

[4] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.

[5] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. on Inform. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[6] M. Biguesh and A. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 884–893, 2006.

[7] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, 2017.

[8] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[10] 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 Processing*, vol. 66, no. 13, pp. 3393–3406, 2018.

[11] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems," *IEEE Signal Processing Lett.*, vol. 15, pp. 202–205, 2008.

[12] W. Xu, H. A. Alshamary, T. Al-Naffouri, and A. Zaib, "Optimal joint channel estimation and data detection for massive SIMO wireless systems: A polynomial complexity solution," *IEEE Trans. Inform. Theory*, vol. 66, no. 3, pp. 1822–1844, 2020.

[13] S. Ma, W. Shen, J. An, and L. Hanzo, "Wideband channel estimation for IRS-aided systems in the face of beam squint," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6240–6253, 2021.

[14] 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.

[15] Z. Zhou, N. Ge, Z. Wang, and L. Hanzo, "Joint transmit precoding and reconfigurable intelligent surface phase adjustment: A decomposition-aided channel estimation approach," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1228–1243, 2021.

[16] Y. Teng, L. Jia, A. Liu, and V. K. N. Lau, "A hybrid pilot beamforming and channel tracking scheme for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6078–6092, 2021.

[17] C. Xing, S. Wang, S. Chen, S. Ma, H. V. Poor, and L. Hanzo, "Matrix-monotonic optimization – Part I: Single-variable optimization," *IEEE Trans. Signal Processing*, vol. 69, pp. 738–754, 2021.

[18] ——, "Matrix-monotonic optimization – Part II: Multi-variable optimization," *IEEE Trans. Signal Processing*, vol. 69, pp. 179–194, 2021.

[19] W. Yao, S. Chen, and L. Hanzo, "A transceiver design based on uniform channel decomposition and MBER vector perturbation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3153–3159, 2010.

[20] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, 2005.

[21] H. Shen, B. Li, M. Tao, and X. Wang, "MSE-based transceiver designs for the MIMO interference channel," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3480–3489, 2010.

[22] Q. Zhang, C. He, and L. Jiang, "Per-stream MSE based linear transceiver design for MIMO interference channels with CSI error," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1676–1689, 2015.

[23] J. L. Bing, L. Gopal, Y. Rong, C. W. R. Chiong, and Z. Zang, "Robust transceiver design for multi-hop AF MIMO relay multicasting from multiple sources," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1565–1576, 2021.

[24] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Processing*, vol. 62, no. 12, pp. 3111–3124, 2014.

[25] H. He, X. Yu, J. Zhang, S. Song, and K. B. Letaief, "Cell-free massive MIMO detection: A distributed expectation propagation approach," *arXiv preprint arXiv:2111.14022*, 2021.

[26] H. Vikalo, B. Hassibi, and P. Stoica, "Efficient joint maximum-likelihood channel estimation and signal detection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1838–1845, 2006.

[27] P. Zhang, S. Chen, and L. Hanzo, "Reduced-complexity near-capacity joint channel estimation and three-stage turbo detection for coherent space-time shift keying," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1902–1913, 2013.

[28] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, 2021.

[29] L. Wei, C. Huang, Q. Guo, Z. Yang, Z. Zhang, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Joint channel estimation and signal recovery for RIS-empowered multiuser communications," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4640–4655, 2022.

[30] J. An, C. Xu, L. Gan, and L. Hanzo, "Low-complexity channel estimation and passive beamforming for RIS-assisted MIMO systems relying on discrete phase shifts," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1245–1260, 2022.

[31] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.

[32] L. Bai, J. Liu, Q. Yu, J. Choi, and W. Zhang, “A collision resolution protocol for random access in massive MIMO,” *IEEE J. Select. Areas Commun.*, vol. 39, no. 3, pp. 686–699, 2021.

[33] D. Fan, F. Gao, Y. Liu, Y. Deng, G. Wang, Z. Zhong, and A. Nallanathan, “Angle domain channel estimation in hybrid millimeter wave massive MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, 2018.

[34] S. Gong, C. Xing, Z. Fei, and S. Ma, “Millimeter-wave secrecy beamforming designs for two-way amplify-and-forward MIMO relaying networks,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2059–2071, 2017.

[35] P. Singh, S. Srivastava, A. K. Jagannatham, and L. Hanzo, “Second-order statistics-based semi-blind techniques for channel estimation in millimeter-wave MIMO analog and hybrid beamforming,” *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6886–6901, 2020.

[36] J. Zheng, J. Zhang, E. Björnson, Z. Li, and B. Ai, “Cell-free massive MIMO-OFDM for high-speed train communications,” *IEEE J. Select. Areas Commun.*, vol. 40, no. 10, pp. 2823–2839, 2022.

[37] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, “Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, 2018.

[38] A. Pastore, M. Joham, and J. R. Fonollosa, “A framework for joint design of pilot sequence and linear precoder,” *IEEE Trans. on Inform. Theory*, vol. 62, no. 9, pp. 5059–5079, 2016.

[39] C. Xing, T. Yu, J. Song, Z. Zheng, L. Zhao, and L. Hanzo, “A unified joint optimization of training sequences and transceivers based on matrix-monotonic optimization,” *IEEE Trans. Veh. Technol.*, 2023.

[40] H. Sampath, P. Stoica, and A. Paulraj, “Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion,” *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, 2001.

[41] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, “Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback,” *IEEE Trans. Inform. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[42] L. Tong, B. Sadler, and M. Dong, “Pilot-assisted wireless transmissions: General model, design criteria, and signal processing,” *IEEE Signal Processing Mag.*, vol. 21, no. 6, pp. 12–25, 2004.

[43] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., 1993.

[44] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman & Hall/CRC, 2000.