Pilot Sequence Design and Channel Estimation for Backscatter Communications with Multiple Antennas

Jixiang Chen, Yue Rong, *Senior Member, IEEE*, Quansheng Guan, *Senior Member, IEEE*, Dong Li, *Senior Member, IEEE*

***Abstract*—Backscatter communication (BackCom) technology that takes advantage of the radio frequency signals to facilitate the communications of passive devices has attracted much attention in recent years. To enhance its communication performance, the multiple-input and multiple-output (MIMO) technology has been introduced to BackCom. Channel estimation is crucial for the MIMO BackCom system. However, the optimization for the pilot training sequences has not been studied. In this paper, we propose a pilot sequence design algorithm for MIMO BackCom systems with spatially correlated antennas, which can estimate both the direct link and the backscatter link channel information. We derive the optimal structure of the source and the tag pilot sequences which achieves the minimum mean-squared error (MSE) of channel estimation. Then, we optimize the power allocation between the source pilot sequences. Simulation results show that our proposed algorithm can estimate channel efficiently and achieve better sum MSE performance than the benchmark without power allocation.**

***Index Terms*—Backscatter communications, pilot design, channel estimation, MIMO**

I. Introduction

Backscatter communication (BackCom) is a newly emerging paradigm, which utilizes the radio frequency signal as the carrier to reduce the power consumption. In BackCom systems, battery-free tags communicate with a reader by reflecting radio frequency (RF) signals. Due to its passive and low-cost tag, BackCom has been seen as an attractive technology for the Internet of Things (IoT) [1]–[5]. However, the strength of the backscattered signal in BackCom is weak due to a large path loss and tag's low reflection efficiency [6]–[9].

The multi-antenna technology is efficient in improving the weak signal in BackCom. The papers [10]–[16] studied the BackCom system with a multi-antenna tag or reader to facilitate the signal detection. The dual-antenna receiver is efficient in eliminating the interference from unknown sources [10], [11]. For example, the receiver uses the received signals at two antennas to detect the backscattered symbols by simply using the received signal of one antenna to divide that of another antenna. This method can detect binary modulated symbols without channel state information (CSI).

To further increase the communication rate or decrease communication errors, the multiple-input and multiple-output (MIMO) technology has been introduced to BackCom [17]–[26]. In [17], the authors designed the tag's signalling matrix to obtain the gain from a multi-antenna source. The work [18] presented a unified generalized space-time shift keying (GSTSK)-backscatter architecture for BackCom. The work [19], [20] studied spatial modulation-based backscatter communication. Generalized quadrature space-time modulation for backscatter communication was proposed in [21]. Furthermore, the work [23] maximized the sum capacity of primary and secondary transmissions by designing the beamforming. Most modulation/detection and beamforming schemes for MIMO BackCom require CSI. Thus, obtaining accurate CSI, i.e., channel estimation, is imperative for a MIMO BackCom system.

However, MIMO BackCom channel estimation received little attention. In [27], the authors design a channel estimation method that can estimate both the direct link channel and the backscatter link channel simultaneously with multiple sources and tags. However, the optimization for channel training sequences for MIMO BackCom has not been explored. In this paper, we extend [27] by considering spatially correlated antennas and study the optimal pilot sequence design and channel estimation algorithm for a general MIMO BackCom system. We consider a general system setup that three nodes (source, tag, reader) in a BackCom system are all equipped with multiple antennas. Furthermore, antennas on the source, the tag and the reader have correlations. Training sequences are sent by both the source and the tag, and the reader estimates both the source-tag-reader channel and the source-reader channel simultaneously. We aim at deriving the optimal pilot sequences structure and power resource allocation for both the source and the tag. We would like to note that it is not easy to obtain optimal training sequences jointly, because different to relay or reconfigurable intelligent surface (RIS)-aided communication systems, the source and the tag signals are independent of each other in BackCom systems. Moreover, the signal received at the reader is not a simple

This work was supported by the National Natural Science Foundation of China under Grants U23A20281, 62341129, and the Science and Technology Planning Project of Guangdong Province of China under Grant 2023A0505050097. (Corresponding author: Yue Rong.)

Jixiang Chen, and Quansheng Guan are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: eejxchen@mail.scut.edu.cn; eeqshguan@scut.edu.cn).

Yue Rong is with the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Bentley, WA 6102, Australia (e-mail: y.rong@curtin.edu.au).

Dong Li is with the School of Computer Science and Engineering, Macau University of Science and Technology, Macau 999078, China (e-mail: dli@must.edu.mo).

multiplication of the source and tag training sequences. To tackle this challenge, we rewrite the received signal expression which combines the source and the tag signals together. But this expression has a complicated channel matrix, which brings new challenges to our channel training algorithm design.

The main contributions of this paper are summarized as follows.

- We derive the structure of the optimal pilot sequences for both the tag and the source that minimizes the sum mean-squared error (MSE) of channel estimation. Specifically, we reveal that the eigenvector matrix of the correlation matrix of the MIMO channel is matched with the optimal source pilot matrix. Moreover, in the proposed algorithm, the power resource allocation optimization for the source pilot sequences is investigated.
- Interestingly, we show that the optimal pilot sequence at the tag is an orthogonal sequence, which does not depend on the channel statistics. This makes the implementation of the tag sequence feasible in practice. The optimization of the source pilot sequence can be run at either the source or the reader, which has a higher computation capability than the tag.
- Simulation results show that our channel training algorithm can achieve a much better MSE performance compared to the traditional equal power allocation scheme in which the training sequences at the source are orthogonal.

Notations: Scalars are denoted by lowercase letters, while vectors and matrices are represented by bold lowercase letters and uppercase letters, respectively. $\mathbf{I}_N$ denotes the identity matrix of size N. $\mathcal{CN}(\mu, \sigma^2)$ denotes the complex Gaussian distribution with mean μ and variance σ^2. $\otimes$ stands for the matrix Kronecker product [28]. $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote matrix conjugate, transpose and Hermitian transpose, respectively. $Bdiag[\cdot]$ denotes a block diagonal matrix. $tr(\cdot)$ stands for matrix trace, and $E[\cdot]$ denotes statistical expectation. $[\cdot]_m$ stands for the m-th column of a matrix and $\{\cdot\}_m$ stands for the m-th row of a matrix. $[\cdot]_{i,j}$ stands for the (i, j)-th element of a matrix. $(\cdot)^{-1}$ denotes matrix inversion. $vec(\cdot)$ denotes the vectorization operator which is realized by stacking all column vectors of a matrix on top of each other. $\lceil x \rceil$ stands for the ceiling function for x. $O(f(n))$ means computational complexity order as a function of n.

The structure of this paper is listed as follows. Section II describes the channel and signal model for MIMO BackCom. In Section III, we propose a channel estimation algorithm for MIMO BackCom. We provide simulation results in Section IV. Finally, conclusions are given in Section V.

II. MIMO System for BackCom

In this section, we introduce the system model for MIMO BackCom.

A. System Model

Consider a MIMO BackCom system consisting of a tag, a reader and a RF source as illustrated in Fig. 1. There are N_s, N_t and N_r antennas in the source, the tag and the reader respectively. By adjusting its impedance, the tag can harvest energy powering its circuit and backscatter symbols. For example, the RF signal is reflected as the load impedance is mismatched, while the RF signal is harvested when the load impedance is matched.

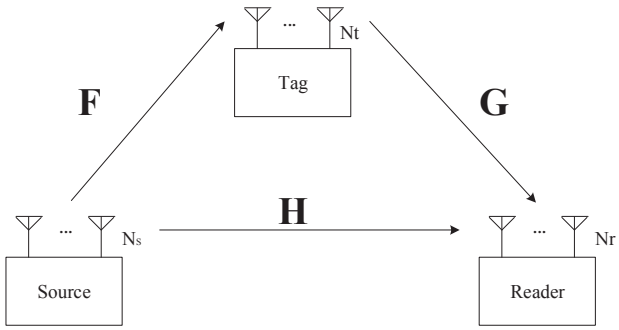

Fig. 1. System model of MIMO BackCom. The signal broadcasted by the source can be received by both the tag and the reader. The tag modulates the incident signal by varying its load impedance, and backscatters the modulated signal to the reader.

The source and the tag need to send designed training sequences to the reader to facilitate the channel estimation. Thus, during our channel training phase, the cooperation from the source is required[1]. Then, in the data transmission phase, the source transmits modulated symbols to its users normally. In this phase, the reader can detect backscattered symbols without cooperation of the source.

B. Channel Model and Signal Model

We consider the channel from the source to the reader as a direct channel $\mathbf{H} \in \mathbb{C}^{N_r \times N_s}$, the channel from the source to the tag as a forward channel $\mathbf{F} \in \mathbb{C}^{N_t \times N_s}$, and the channel from the tag to the reader as a backward channel $\mathbf{G} \in \mathbb{C}^{N_r \times N_t}$, respectively. The multiplicative channel of $\mathbf{F}$ and $\mathbf{G}$ is referred to as the backscatter channel. In this paper, we assume that $\mathbf{H}, \mathbf{F}$ and $\mathbf{G}$ satisfy the Gaussian-kronecker model [29], which is also adopted in [30], [31][2], where $\mathbf{H}, \mathbf{F}$ and $\mathbf{G}$ are complex-valued Gaussian random matrices following

$$\begin{aligned} \mathbf{H} &\sim \mathcal{CN}\left(\mathbf{0}, \mathbf{T}_{HF} \otimes \mathbf{R}_H\right), \\ \mathbf{F} &\sim \mathcal{CN}\left(\mathbf{0}, \mathbf{T}_{HF} \otimes \mathbf{R}_F\right), \\ \mathbf{G} &\sim \mathcal{CN}\left(\mathbf{0}, \mathbf{T}_G \otimes \mathbf{R}_G\right). \end{aligned} \tag{1}$$

Matrices $\mathbf{T}_{HF}$ and $\mathbf{T}_G$ denote the $N_s \times N_s$ and $N_t \times N_t$ covariance matrix at the transmit side of $\mathbf{H}(\mathbf{F})$ and $\mathbf{G}$, respectively. Matrices $\mathbf{R}_H, \mathbf{R}_F$ and $\mathbf{R}_G$ denote the $N_r \times N_r$, $N_t \times N_t$ and $N_r \times N_r$ covariance matrix at the receive side of $\mathbf{H}, \mathbf{F}$ and $\mathbf{G}$, respectively. These matrices are fixed for a given setup and can, thus, be assumed to be known [30], [31]. In practice, given the setup of antennas, the channel covariance matrices can be calculated with high accuracy as shown in [29], if the

[1]The synchronization between the RF source signal and tag signal can be achieved similar to the approaches in [6], [40]–[42].

[2]Different to [30], [31], we consider a general MIMO setup for all nodes and the direct link channel is not blocked.

number of scatters is large. Then, we can equivalently rewrite the channel matrices as

$$\begin{aligned}\mathbf{H} &= \mathbf{B}_H\mathbf{H}_w\mathbf{A}^H,\\ \mathbf{F} &= \mathbf{B}_F\mathbf{F}_w\mathbf{A}^H,\\ \mathbf{G} &= \mathbf{B}_G\mathbf{G}_w\mathbf{A}_G^H,\end{aligned} \tag{2}$$

where $\mathbf{A}\mathbf{A}^H = \mathbf{T}_{HF}^T, \mathbf{B}_H\mathbf{B}_H^H = \mathbf{R}_H, \mathbf{B}_F\mathbf{B}_F^H = \mathbf{R}_F, \mathbf{A}_G\mathbf{A}_G^H = \mathbf{T}_G^T, \mathbf{B}_G\mathbf{B}_G^H = \mathbf{R}_G$. Matrices $\mathbf{H}_w, \mathbf{F}_w$ and $\mathbf{G}_w$ are $N_r \times N_s$, $N_t \times N_s$ and $N_r \times N_t$ Gaussian random matrices with independent and identically distributed $(i.i.d.)$ zero mean and unit variance entries. Assume that $\mathbf{H}_w, \mathbf{F}_w$ and $\mathbf{G}_w$ are statistically independent of each other. We assume that covariance matrices of channels to be estimated are known at the reader and these channels are block flat fading.

The reader receives both source signals and backscattered signals. Then, the received signals at the reader is [3]

$$\mathbf{y} = (\mathbf{H} + \mathbf{G}\alpha\mathbf{X}'\mathbf{F})\mathbf{s} + \mathbf{n}, \tag{3}$$

where α is a coefficient representing the scattering efficiency and antenna gain, $\mathbf{X}' \in \mathbb{C}^{N_t \times N_t}$ is the diagonal matrix representing the signals conveyed by the tag, $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is the source signal vector and $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ is the zero-mean additive white Gaussian noise (AWGN) with variance σ_n^2. The diagonal elements in $\mathbf{X}'$ satisfy $|x_i| \leq 1, i = 1, 2, \cdots, N_t$, where x_i is a complex number [17], [33], [34]. Let $\mathbf{X} = \alpha\mathbf{X}'$. Then, the received signals can be expressed as

$$\mathbf{y} = (\mathbf{H} + \mathbf{G}\mathbf{X}\mathbf{F})\mathbf{s} + \mathbf{n}. \tag{4}$$

We would like to note that the system model (4) is different to that of the RIS system. In BackCom systems, $\mathbf{X}$ carries information transmitted by the tag to the reader, while in RIS systems, $\mathbf{X}$ acts as a precoding matrix to aid the transmission of the source information $\mathbf{s}$.

III. Pilot Design Algorithm

In this section, we propose a channel estimation algorithm designing training sequences for both the tag and the source to estimate the direct channel and the backscatter channel simultaneously.

A. Optimal Training Sequences Structure

The expression in (4) is inconvenient to be used to jointly estimate both channels. To solve this problem, we change the received signal form to

$$\mathbf{y} = \mathbf{Q}\mathbf{t} + \mathbf{n}, \tag{5}$$

where $\mathbf{t} = \begin{bmatrix}\mathbf{s}\\ \mathbf{x}\otimes\mathbf{s}\end{bmatrix} \in \mathbb{C}^{(N_t+1)N_s \times 1}$, $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ contains the diagonal elements of $\mathbf{X}$, and

$$\mathbf{Q} = \begin{bmatrix}\mathbf{H}, & \mathbf{g}_1\mathbf{f}_1, & \ldots, & \mathbf{g}_{N_t}\mathbf{f}_{N_t}\end{bmatrix} \in \mathbb{C}^{N_r \times (N_t+1)N_s}, \tag{6}$$

where $\mathbf{f}_i \in \mathbb{C}^{1\times N_s}$, $\mathbf{g}_i \in \mathbb{C}^{N_r \times 1}$, $i = 1, 2, \cdots, N_t$ are the rows of $\mathbf{F}$ and the columns of $\mathbf{G}$, respectively. Thus, we have $\mathbf{F} = \begin{bmatrix}\mathbf{f}_1^T, & \mathbf{f}_2^T, & \ldots, & \mathbf{f}_{N_t}^T\end{bmatrix}^T$ and $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_{N_t}]$. Note that the transformation from (4) to (5) is equivalent.

The number of unknowns in $\mathbf{Q}$ is much larger than the number of entries in $\mathbf{t}$. Thus, more than one time slot is required to estimate the channels. First, the $N_r \times 1$ received signal vector at the reader at the l-th time slot is expressed as

$$\begin{aligned}\mathbf{y}_l &= (\mathbf{H} + \mathbf{G}\mathbf{X}_l\mathbf{F})\mathbf{s}_l + \mathbf{n}_l.\\ &= \mathbf{Q}\mathbf{t}_l + \mathbf{n}_l,\end{aligned} \tag{7}$$

where $\mathbf{s}_l$, $\mathbf{t}_l$, and $\mathbf{n}_l$ are the corresponding $\mathbf{s}$, $\mathbf{t}$, and $\mathbf{n}$ in (5) at the l-th time slot. Assume that the source transmits an $N_s \times L$ training signals matrix $\mathbf{S}$ where L is the length of the training time slots. Second, the received signal in total L time slots can be expressed as

$$\mathbf{Y} = \mathbf{Q}\mathbf{T} + \mathbf{N}, \tag{8}$$

where $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \cdots, \mathbf{n}_L]$ is the noise matrix at the reader and $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_L]$.

Denoting $\mathbf{x}_l = \left[x_{l,1}, x_{l,2}, \cdots, x_{l,N_t}\right]^T$, then $\mathbf{T}$ can be written as

$$\mathbf{T} = \begin{bmatrix}\mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_L\\ x_{1,1}\mathbf{s}_1 & x_{2,1}\mathbf{s}_2 & \cdots & x_{L,1}\mathbf{s}_L\\ \vdots & \vdots & \ddots & \vdots\\ x_{1,N_t}\mathbf{s}_1 & x_{2,N_t}\mathbf{s}_2 & \cdots & x_{L,N_t}\mathbf{s}_L\end{bmatrix}. \tag{9}$$

Consider the following eigenvalue decomposition (EVD)

$$\begin{aligned}\mathbf{T}_{HF}^T &= \mathbf{U}'\mathbf{\Lambda}'\mathbf{U}'^H,\\ \mathbf{T}_G^T &= \mathbf{U}'_G\mathbf{\Lambda}'_G\mathbf{U}'^H_G.\end{aligned} \tag{10}$$

Then we have $\mathbf{A}^H = \mathbf{\Pi}'\mathbf{\Lambda}'^{\frac{1}{2}}\mathbf{U}'^H$, $\mathbf{A}_G^H = \mathbf{\Pi}'_G\mathbf{\Lambda}'^{\frac{1}{2}}_G\mathbf{U}'^H_G$, where $\mathbf{\Pi}'$ and $\mathbf{\Pi}'_G$ are arbitrary $N_s \times N_s$ and $N_t \times N_t$ unitary matrix, respectively.

By utilizing $\mathbf{U}'\mathbf{U}'^H = \mathbf{I}_{N_s}$, we rewrite (8) as

$$\mathbf{Y} = \mathbf{Q}'\mathbf{T}' + \mathbf{N}, \tag{11}$$

where

$$\begin{aligned}\mathbf{Q}' &= \begin{bmatrix}\mathbf{H}\mathbf{U}', & (\mathbf{g}_1\mathbf{f}_1)\mathbf{U}', & \ldots, & (\mathbf{g}_{N_t}\mathbf{f}_{N_t})\mathbf{U}'\end{bmatrix},\\ \mathbf{T}' &= \begin{bmatrix}\mathbf{U}'^H\mathbf{s}_1 & \mathbf{U}'^H\mathbf{s}_2 & \cdots & \mathbf{U}'^H\mathbf{s}_L\\ \mathbf{U}'^H(x_{1,1}\mathbf{s}_1) & \mathbf{U}'^H(x_{2,1}\mathbf{s}_2) & \cdots & \mathbf{U}'^H(x_{L,1}\mathbf{s}_L)\\ \mathbf{U}'^H(x_{1,N_t}\mathbf{s}_1) & \mathbf{U}'^H(x_{2,N_t}\mathbf{s}_2) & \cdots & \mathbf{U}'^H(x_{L,N_t}\mathbf{s}_L)\end{bmatrix}.\end{aligned} \tag{12}$$

According to $\mathbf{A}\mathbf{D}\mathbf{B} = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{D})$ [28], we obtain

$$vec(\mathbf{Y}) = (\mathbf{T}'^T \otimes \mathbf{I}_{N_r})vec(\mathbf{Q}') + vec(\mathbf{N}). \tag{13}$$

Then, we rewrite (13) as

$$vec(\mathbf{Y}) = \mathbf{M}\mathbf{q} + \mathbf{n}', \tag{14}$$

[3]Here are some differences between the channel estimation of backscatter communications and RIS-aided communications. First, in RIS-aided communications, the direct link (source-receiver) is usually assumed to be blocked while this is not the case in backscatter communications. Second, the RIS has its unique array geometry and this knowledge can be exploited during the channel estimation phase while is not considered in backscatter communications. Third, due to the close distance between the RIS and source, the channel between them is a line-of-sight channel. Since the knowledge of array geometry is not exploited, our method can be seen as a general channel estimation method for RIS-based communications.

where $\mathbf{M} = \mathbf{T'}^T \otimes \mathbf{I}_{N_r}, \mathbf{q} = vec(\mathbf{Q'}), \mathbf{n'} = vec(\mathbf{N})$.

Due to its simplicity, we apply a linear minimum mean-squared error (MMSE) estimator [35] at the reader to estimate $\mathbf{q}$. Then, we obtain

$$\widehat{\mathbf{q}} = \mathbf{W}^H vec(\mathbf{Y}), \tag{15}$$

where $\mathbf{W}$ is the weight matrix of the MMSE estimator and $\widehat{\mathbf{q}}$ is the estimation of $\mathbf{q}$. Since a linear estimator is used in (15), the length of channel training time slots satisfies $L \geq (N_t+1)N_s$. Then, the MSE of estimating $\widehat{\mathbf{q}}$ can be written as

$$\begin{aligned} MSE &= E\left[tr\left(\left(\widehat{\mathbf{q}}-\mathbf{q}\right)\left(\widehat{\mathbf{q}}-\mathbf{q}\right)^H\right)\right] \\ &= tr\Big(\Big(\mathbf{W}^H\mathbf{M}-\mathbf{I}_{N_r\times(N_t+1)N_s}\Big)\mathbf{R}_q\Big(\mathbf{W}^H\mathbf{M}-\mathbf{I}_{N_r\times(N_t+1)N_s}\Big)^H \\ &\quad +\mathbf{W}^H\mathbf{R}_n\mathbf{W}\Big), \end{aligned} \tag{16}$$

where $\mathbf{R}_q = E[\mathbf{q}\mathbf{q}^H] = E[vec(\mathbf{Q'})vec(\mathbf{Q'})^H]$, and $\mathbf{R}_n = E[\mathbf{n}\mathbf{n}^H] = E[vec(\mathbf{N})vec(\mathbf{N})^H]$. Next, we are going to obtain $\mathbf{R}_q$, $\mathbf{R}_n$ and $\mathbf{W}$.

Theorem 1: The covariance matrix $\mathbf{R}_q$ is given by

$$\mathbf{R}_q = Bdiag\left[\mathbf{\Lambda'}\otimes\mathbf{R}_H, \mathbf{\Lambda}'_1\otimes\mathbf{R}_G, \cdots, \mathbf{\Lambda}'_{N_t}\otimes\mathbf{R}_G\right], \tag{17}$$

where the m-th element of $\mathbf{\Lambda}'_{n_t}$ is ${\lambda'}_m b_{n_t}$, $b_{n_t} = tr\left(\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\{\mathbf{B}_F\}_{n_t}^H\left[\mathbf{A}_G^H\right]_{n_t}^H\right), n_t = 1, 2, \cdots, N_t$.

Proof: Please see Appendix A.

Meanwhile, the covariance matrix of $\mathbf{n'}$ is

$$\mathbf{R}'_n = E[\mathbf{n'}\mathbf{n'}^H] = \sigma_n^2\mathbf{I}_{LN_r}, \tag{18}$$

and the matrix $\mathbf{W}$ with the minimized MSE in (16) is given by

$$\mathbf{W} = \left(\mathbf{M}\mathbf{R}_q\mathbf{M}^H + \mathbf{R}_n\right)^{-1}\mathbf{M}\mathbf{R}_q. \tag{19}$$

We substitute (19) back into (16), and utilize the matrix inversion equality of $(\mathbf{D}+\mathbf{BCA})^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{B}\left(\mathbf{A}\mathbf{D}^{-1}\mathbf{B}+\mathbf{C}^{-1}\right)\mathbf{A}\mathbf{D}^{-1}$ [36], and then the MSE of estimating of $\widehat{\mathbf{q}}$ is given by

$$MSE = tr\left(\left(\mathbf{R}_q^{-1} + \mathbf{M}^H\mathbf{R}_n^{-1}\mathbf{M}\right)^{-1}\right). \tag{20}$$

To optimize the MSE, the power constraint should be considered. From (20) and the constraint on the passive tag and the source, the optimal training sequences can be derived by solving the following optimization problem

$$\min_{x_{i,j},\mathbf{s}_i} tr\left(\left(\mathbf{R}_q^{-1} + \mathbf{M}^H\mathbf{R}_n^{-1}\mathbf{M}\right)^{-1}\right) \tag{21}$$

$$s.t.\ \ tr\left(E[\mathbf{s}_i\mathbf{s}_i^H]\right) \leq P_s, i = 1, 2, \cdots, L, \tag{22}$$

$$|x_{i,j}| \leq 1, i = 1, 2, \cdots, L, j = 1, 2, \cdots N_t, \tag{23}$$

where P_s is the average source transmission power available.

Let us introduce the training matrices for the source and the tag as

$$\begin{aligned} \mathbf{S} &= [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_L] \\ &= \begin{bmatrix} s_{1,1} & s_{2,1} & \cdots & s_{L,1} \\ s_{1,2} & s_{2,2} & \cdots & s_{L,2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1,N_s} & s_{2,N_s} & \cdots & s_{L,N_s} \end{bmatrix}, \end{aligned} \tag{24}$$

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_L \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{L,1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,N_t} & x_{2,N_t} & \cdots & x_{L,N_t} \end{bmatrix}. \end{aligned} \tag{25}$$

The optimal structure of $\mathbf{T}$ for both the tag and the source is given by the following theorem.

Theorem 2: The optimal $\mathbf{T}$ can be constructed by partition the L time slots in the channel training phase into N_t+1 groups with equal length of N_s. For clarity, we introduce $\mathbf{s}_{(i-1)N_s+j}$ and $\mathbf{x}_{(i-1)N_s+j}$ as the source and tag training sequences, respectively, where $i = 1, \cdots, N_t+1$ is the group index, and $j = 1, \cdots, N_s$ is the index of the training sequence in each group. For each group, the source sequences satisfy $\Sigma_{j=1}^{N_s}\mathbf{s}_{(i-1)N_s+j}\mathbf{s}_{(i-1)N_s+j}^H = 1/(N_t+1)\mathbf{U'}\mathbf{\Sigma}\mathbf{U'}^H$, $i = 1, 2, \cdots, N_t+1$, where $\mathbf{\Sigma}$ is a diagonal matrix. The tag training sequences satisfy $\mathbf{x}_{(i-1)N_s+j} = \bar{\mathbf{x}}_i$, $j = 1, \cdots, N_s$, and $\bar{\mathbf{X}}\bar{\mathbf{X}}^H = (N_t+1)\mathbf{I}_{N_t+1}$, where $\bar{\mathbf{X}} = \begin{bmatrix} 1, & \cdots, & 1 \\ \bar{\mathbf{x}}_1, & \cdots, & \bar{\mathbf{x}}_{N_t+1} \end{bmatrix}$. That is, the tag training sequences remain the same within one group and are orthogonal between groups.

Proof: Let us introduce the EVD of $\mathbf{R}_H = \mathbf{U}_H\mathbf{\Lambda}_H\mathbf{U}_H^H$ and $\mathbf{R}_G = \mathbf{U}_G\mathbf{\Lambda}_G\mathbf{U}_G^H$. Then, we can rewrite (17) as

$$\mathbf{R}_q = \mathbf{U}_q Bdiag\left[\mathbf{\Lambda'}\otimes\mathbf{\Lambda}_H, \mathbf{\Lambda}'_1\otimes\mathbf{\Lambda}_G, \cdots, \mathbf{\Lambda}'_{N_t}\otimes\mathbf{\Lambda}_G\right]\mathbf{U}_q^H, \tag{26}$$

where $\mathbf{U}_q = Bdiag\left[\mathbf{I}_{N_s}\otimes\mathbf{U}_H, \mathbf{I}_{N_s}\otimes\mathbf{U}_G, \cdots, \mathbf{I}_{N_s}\otimes\mathbf{U}_G\right]$. Meanwhile, we have

$$\begin{aligned} \mathbf{M}^H\mathbf{M} &= \left(\mathbf{T'}^T\otimes\mathbf{I}_{N_r}\right)^H\left(\mathbf{T'}^T\otimes\mathbf{I}_{N_r}\right) = \left(\mathbf{T'}^*\mathbf{T'}^T\right)\otimes\mathbf{I}_{N_r} \\ &= \left(\mathbf{T'}\mathbf{T'}^H\right)^T\otimes\mathbf{I}_{N_r}. \end{aligned} \tag{27}$$

Substituting (18), (26) and (27) into (20), the MSE can be equivalently rewritten as (28) shown at the top of the next page, where the second and third rows are obtained because $\mathbf{U}_q$ is a block diagonal unitary matrix. Then, by denoting $\mathbf{D}_H = {\mathbf{\Lambda'}}^{-1}\otimes\mathbf{\Lambda}_H^{-1}, \mathbf{D}_{G,n_t} = {\mathbf{\Lambda}'_{n_t}}^{-1}\otimes\mathbf{\Lambda}_G^{-1}, n_t = 1, 2, \cdots, N_t$, (28) can be equivalently rewritten as (29) shown at the top of the next page. To minimize the MSE in (29), $\left(\mathbf{T'}\mathbf{T'}^H\right)^T\otimes\mathbf{I}_{N_r}$ is required to be a diagonal matrix [37], which means $\mathbf{T'}\mathbf{T'}^H$ needs to be diagonal. Substituting (12) into $\mathbf{T'}\mathbf{T'}^H$, we obtain (30) shown at the top of the next page.

To make $\mathbf{T'}\mathbf{T'}^H$ a diagonal matrix, the off-diagonal matrices in (30) should be all-zero matrix. Moreover, the diagonal sub-matrices should be diagonal. This indicates that the training sequences $\mathbf{S}$ and $\mathbf{X}$ should satisfy the following conditions

$$
\begin{aligned}
MSE &= tr\left(\left[\left(\mathbf{U}_q Bdiag\left[\mathbf{\Lambda}'\otimes\mathbf{\Lambda}_H, \mathbf{\Lambda}'_1\otimes\mathbf{\Lambda}_G, \cdots, \mathbf{\Lambda}'_{N_t}\otimes\mathbf{\Lambda}_G\right]\mathbf{U}_q^H\right)^{-1} + \left(\sigma_n^2\mathbf{I}_{N_r(N_t+1)N_s}\right)^{-1}\left(\left(\mathbf{T}'\mathbf{T}'^H\right)^T\otimes\mathbf{I}_{N_r}\right)\right]^{-1}\right)\\
&= tr\left(\mathbf{U}_q^H\left[\left(\mathbf{U}_q Bdiag\left[\mathbf{\Lambda}'\otimes\mathbf{\Lambda}_H, \mathbf{\Lambda}'_1\otimes\mathbf{\Lambda}_G, \cdots, \mathbf{\Lambda}'_{N_t}\otimes\mathbf{\Lambda}_G\right]\mathbf{U}_q^H\right)^{-1} + \left(\sigma_n^2\mathbf{I}_{N_r(N_t+1)N_s}\right)^{-1}\left(\left(\mathbf{T}'\mathbf{T}'^H\right)^T\otimes\mathbf{I}_{N_r}\right)\right]^{-1}\mathbf{U}_q\right)\\
&= tr\left(\left[\left(Bdiag\left[\mathbf{\Lambda}'\otimes\mathbf{\Lambda}_H, \mathbf{\Lambda}'_1\otimes\mathbf{\Lambda}_G, \cdots, \mathbf{\Lambda}'_{N_t}\otimes\mathbf{\Lambda}_G\right]\right)^{-1} + \mathbf{U}_q^H\left(\sigma_n^2\mathbf{I}_{N_r(N_t+1)N_s}\right)^{-1}\left(\left(\mathbf{T}'\mathbf{T}'^H\right)^T\otimes\mathbf{I}_{N_r}\right)\mathbf{U}_q\right]^{-1}\right)
\end{aligned}
\tag{28}
$$

$$
MSE = tr\left(\left[\left[\begin{array}{cccc}\mathbf{D}_H & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{D}_{G,1} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_{G,N_t}\end{array}\right] + \left(\sigma_n^2\mathbf{I}_{N_r(N_t+1)N_s}\right)^{-1}\left(\left(\mathbf{T}'\mathbf{T}'^H\right)^T\otimes\mathbf{I}_{N_r}\right)\right]^{-1}\right) \tag{29}
$$

$$
\mathbf{T}'\mathbf{T}'^H = \left[\begin{array}{cccc}\mathbf{U}'^H\mathbf{s}_1 & \mathbf{U}'^H\mathbf{s}_2 & \cdots & \mathbf{U}'^H\mathbf{s}_L\\ \mathbf{U}'^H(x_{1,1}\mathbf{s}_1) & \mathbf{U}'^H(x_{2,1}\mathbf{s}_2) & \cdots & \mathbf{U}'^H(x_{L,1}\mathbf{s}_L)\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{U}'^H(x_{1,N_t}\mathbf{s}_1) & \mathbf{U}'^H(x_{2,N_t}\mathbf{s}_2) & \cdots & \mathbf{U}'^H(x_{L,N_t}\mathbf{s}_L)\end{array}\right]\left[\begin{array}{cccc}\mathbf{U}'^H\mathbf{s}_1 & \mathbf{U}'^H\mathbf{s}_2 & \cdots & \mathbf{U}'^H\mathbf{s}_L\\ \mathbf{U}'^H(x_{1,1}\mathbf{s}_1) & \mathbf{U}'^H(x_{2,1}\mathbf{s}_2) & \cdots & \mathbf{U}'^H(x_{L,1}\mathbf{s}_L)\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{U}'^H(x_{1,N_t}\mathbf{s}_1) & \mathbf{U}'^H(x_{2,N_t}\mathbf{s}_2) & \cdots & \mathbf{U}'^H(x_{L,N_t}\mathbf{s}_L)\end{array}\right]^H =
$$
$$
\left[\begin{array}{cccc}\left[\mathbf{U}'^H\left(\mathbf{s}_1\mathbf{s}_1^H+\cdots+\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}'\right] & \cdots & \cdots & \left[\mathbf{U}'^H\left(x_{1,N_t}\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,N_t}\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}'\right]\\ \left[\mathbf{U}'^H\left(x_{1,1}\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,1}\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}'\right] & \ddots & \cdots & \left[\mathbf{U}'^H\left(x_{1,1}x_{1,N_t}^*\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,1}x_{L,N_t}^*\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}'\right]\\ \vdots & \cdots & \ddots & \vdots\\ \left[\mathbf{U}'^H\left(x_{1,N_t}\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,N_t}\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}'\right] & \cdots & \cdots & \left[\mathbf{U}'^H\left(x_{1,N_t}x_{1,N_t}^*\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,N_t}x_{L,N_t}^*\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}'\right]\end{array}\right] \tag{30}
$$

$$
\mathbf{U}'^H\left(\mathbf{s}_1\mathbf{s}_1^H+\mathbf{s}_2\mathbf{s}_2^H+\cdots+\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}' = \mathbf{\Sigma}, \tag{31}
$$

$$
\begin{aligned}
&\mathbf{U}'^H\left(x_{1,n_t}x_{1,n_t}^*\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,n_t}x_{L,n_t}^*\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}' = \mathbf{\Sigma}_{n_t},\\
&n_t = 1,2,\cdots,N_t,
\end{aligned}
\tag{32}
$$

$$
\begin{aligned}
&\mathbf{U}'^H\left(x_{1,n_t}\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,n_t}\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}' = \mathbf{0},\\
&n_t = 1,2,\cdots,N_t,
\end{aligned}
\tag{33}
$$

$$
\begin{aligned}
&\mathbf{U}'^H\left(x_{1,n_t}x_{1,n_t'}^*\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,n_t}x_{L,n_t'}^*\mathbf{s}_L\mathbf{s}_L^H\right)\mathbf{U}' = \mathbf{0},\\
&n_t = 1,2,\cdots,N_t, n_t'\neq n_t, n_t' = 1,2,\cdots,N_t.
\end{aligned}
\tag{34}
$$

Since the increase of $|x_{l,n_t}|, l = 1,2,\cdots,L, n_t = 1,2,\cdots N_t$, leads to the decrease in MSE, we can deduce that $|x_{l,n_t}| = 1, l = 1,2,\cdots,L, n_t = 1,2,\cdots N_t$. Thus, we have $x_{1,n_t}x_{1,n_t}^H\mathbf{s}_1\mathbf{s}_1^H+\cdots+x_{L,n_t}x_{L,n_t}^H\mathbf{s}_L\mathbf{s}_L^H = \mathbf{s}_1\mathbf{s}_1^H+\mathbf{s}_2\mathbf{s}_2^H+\cdots+\mathbf{s}_L\mathbf{s}_L^H = \mathbf{U}'\mathbf{\Sigma}\mathbf{U}'^H = \mathbf{U}'\mathbf{\Sigma}_{n_t}\mathbf{U}'^H$. That is $\mathbf{\Sigma} = \mathbf{\Sigma}_{n_t}$.

Conditions (33) and (34) can be satisfied as follows. First, we partition the training interval L into W groups of equal length N_g. We will show the optimality of equal partition and determine the optimal W and N_g. Let $\mathbf{s}_{(i-1)N_g+j}$ and $\mathbf{x}_{(i-1)N_g+j}$ denote the j-th source and tag training sequence within the i-th group, respectively, $i = 1,\ldots,W$, $j = 1,\ldots,N_g$. By setting $\mathbf{x}_{(i-1)N_g+j} = \bar{\mathbf{x}}_i$, $j = 1,\ldots,N_g$, and

$$
\sum_{j=1}^{N_g}\mathbf{s}_{(i-1)N_g+j}\mathbf{s}_{(i-1)N_g+j}^H = \frac{N_g}{L}\mathbf{U}'\mathbf{\Sigma}\mathbf{U}'^H, \quad i = 1,\ldots,W, \tag{35}
$$

the left-hand side of (33) and (34) can be written as

$$
\begin{aligned}
&\mathbf{U}'^H\left(\sum_{l=1}^{L}x_{l,n_t}\mathbf{s}_l\mathbf{s}_l^H\right)\mathbf{U}'\\
&= \mathbf{U}'^H\left(\sum_{i=1}^{W}\bar{x}_{i,n_t}\sum_{j=1}^{N_g}\mathbf{s}_{(i-1)N_g+j}\mathbf{s}_{(i-1)N_g+j}^H\right)\mathbf{U}'\\
&= \frac{N_g}{L}\mathbf{\Sigma}\sum_{i=1}^{W}\bar{x}_{i,n_t},
\end{aligned}
\tag{36}
$$

$$
\begin{aligned}
&\mathbf{U}'^H\left(\sum_{l=1}^{L}x_{l,n_t}x_{l,n_t'}^*\mathbf{s}_l\mathbf{s}_l^H\right)\mathbf{U}'\\
&= \mathbf{U}'^H\left(\sum_{i=1}^{W}\bar{x}_{i,n_t}\bar{x}_{i,n_t'}^*\sum_{j=1}^{N_g}\mathbf{s}_{(i-1)N_g+j}\mathbf{s}_{(i-1)N_g+j}^H\right)\mathbf{U}'\\
&= \frac{N_g}{L}\mathbf{\Sigma}\sum_{i=1}^{W}\bar{x}_{i,n_t}\bar{x}_{i,n_t'}^*.
\end{aligned}
\tag{37}
$$

From (36) and (37), we can see that (33) and (34) hold if

$$MSE = tr\left(\left[\left[\begin{array}{cccc} \mathbf{D}_H & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{G,1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_{G,N_t} \end{array}\right] + \left(\sigma_n^2 \mathbf{I}_{N_r(N_t+1)N_s}\right)^{-1} \left(\left(\mathbf{T}'\mathbf{T}'^H\right)^T \otimes \mathbf{I}_{N_r}\right)\right]^{-1}\right) =$$

$$tr\left(\left[\left[\begin{array}{cccc} \mathbf{\Lambda}'^{-1} \otimes \mathbf{\Lambda}_H^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_1'^{-1} \otimes \mathbf{\Lambda}_G{}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Lambda}_{N_t}'{}^{-1} \otimes \mathbf{\Lambda}_G{}^{-1} \end{array}\right] + \left(\sigma_n^2 \mathbf{I}_{N_r(N_t+1)N_s}\right)^{-1} \left[\begin{array}{cccc} \mathbf{\Sigma} \otimes \mathbf{I}_{N_r} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \otimes \mathbf{I}_{N_r} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma} \otimes \mathbf{I}_{N_r} \end{array}\right]^T\right]^{-1}\right) \tag{38}$$

$\bar{\mathbf{X}} = \left[\begin{array}{ccc} 1, & \ldots, & 1 \\ \bar{\mathbf{x}}_1, & \ldots, & \bar{\mathbf{x}}_W \end{array}\right]$ has orthogonal rows. Together with $|x_{l,n_t}| = 1$, this leads to $\bar{\mathbf{X}}\bar{\mathbf{X}}^H = (N_t+1)\mathbf{I}_{N_t+1}$, which requires $W \geq N_t + 1$. To minimize the training overhead, we choose $W = N_t + 1$.

To make $\mathbf{\Sigma}$ a general diagonal matrix with full rank of N_s, we can see from (35) that there should be $N_g \geq N_s$. To minimize the training overhead, we choose $N_g = N_s$ for each training group. □

With the designed training sequences, the MSE is equivalently changed to (38) shown on the next page.

B. Optimal Power Allocation

Based on *Theorem 2*, the optimal training sequence design problem (21)-(23) can be equivalently converted to the problem of optimizing $\tilde{\mathbf{\Sigma}}$ as below

$$\min_{\tilde{\mathbf{\Sigma}}} (38) \tag{39}$$

$$s.t. \quad tr(\tilde{\mathbf{\Sigma}}) \leq P_s, \tag{40}$$

$$\tilde{\mathbf{\Sigma}} \geq 0, \tag{41}$$

where for the matrix $\tilde{\mathbf{\Sigma}} = \mathbf{U}'^H \left(\frac{1}{N_s}\Sigma_{i=1}^{N_s} \mathbf{s}_i \mathbf{s}_i^H\right) \mathbf{U}'$, $\tilde{\mathbf{\Sigma}} \geq 0$ means it is a positive semi-definite (PSD) matrix. Note that based on *Theorem 2*, $\tilde{\mathbf{\Sigma}}$ is a diagonal matrix.

To facilitate solving the problem (39)-(41), we convert the matrix variable to scalar variables. Denote σ_{n_s} as the n_s-th diagonal element in $\tilde{\mathbf{\Sigma}}$. Then we get

$$\min_{\sigma_{ns}} \sum_{n_s=1}^{N_s} \sum_{n_r=1}^{N_r} \left(\frac{1}{\lambda'_{n_s}\lambda_{H,n_r}} + \frac{L\sigma_{n_s}}{\sigma_n^2}\right)^{-1} + \sum_{n_t=1}^{N_t} \sum_{n_s=1}^{N_s} \sum_{n_r=1}^{N_r} \left(\frac{1}{\lambda'_{n_s,n_t}\lambda_{G,n_r}} + \frac{L\sigma_{n_s}}{\sigma_n^2}\right)^{-1}, \tag{42}$$

$$s.t. \quad \sum_{n_s=1}^{N_s} \sigma_{n_s} \leq P_s, \tag{43}$$

$$\sigma_{n_s} \geq 0, n_s = 1, 2, \cdots, N_s. \tag{44}$$

where λ'_{n_s} is the n_s-th diagonal element of $\mathbf{\Lambda}'$, λ'_{n_s,n_t} is the n_s-th diagonal element of $\mathbf{\Lambda}'_{n_t}$, λ_{H,n_r} is the n_r-th diagonal element of $\mathbf{\Lambda}_H$ and λ_{G,n_r} is the n_r-th diagonal element of $\mathbf{\Lambda}_G$.

The optimal σ_{n_s} can be efficiently obtained through the Karush-Kuhn-Tucker (KKT) optimality conditions of the problem (42)-(44) [38]. The Lagrangian function is derived as (45) shown at the top of this page, where μ and η_{n_s} are the Lagrange multipliers and satisfy $\mu \geq 0$ and $\eta_{n_s} \geq 0, n_s = 1, 2, \cdots, N_s$. Then, the gradient conditions are given by (46) shown at the top of this page. Meanwhile, the complementary slackness conditions are given by

$$\mu\left(P_s - \sum_{n_s=1}^{N_s} \sigma_{n_s}\right) = 0, \tag{47}$$

$$\eta_{n_s}\sigma_{n_s} = 0, n_s = 1, 2, \cdots, N_s. \tag{48}$$

Since the object function in (42) is a monotonically decreasing function of σ_{n_s}, the bi-section search can be applied to obtain μ and σ_{n_s}.

C. Training Sequences

Based on *Theorem 2*, the optimal training sequences at different antennas of the tag (i.e., $\bar{\mathbf{X}}$) should be orthogonal to each other. Orthogonal training sequences including Hadamard matrix, modified Zadoff–Chu (ZC) sequences and discrete Fourier transform (DFT) matrix provided in [27] can be applied in our pilot training algorithm. In the following, we introduce Hadamard matrix and DFT matrix for constructing the training sequences for the tag. The details about ZC sequences are neglected here since it has similar advantages and disadvantages for tags [27].

Denote L_H as the number of rows or columns of a Hadamard matrix. Taking $L_H = 4$ as an example, a 4×4 Hadamard matrix $\mathbf{C}_H$ is given by

$$\mathbf{C}_H = \left[\begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{array}\right]. \tag{49}$$

The all-1 row in Hadamard matrix is excluded since it is considered as the inherent training sequences for the source as can be seen from (25) and (30). Since the row size of Hadamard matrix is the integer power of two and one-row cannot be used, the number of training time slots of our system is usually designed to be larger than the shortest length $(N_t + 1)N_s$ to match the size of the Hadamard matrix. The number L_H can be determined by $L_H = 2^{\lceil \log_2(N_t+1) \rceil}$.

$$\mathcal{L} = \sum_{n_s=1}^{N_s}\sum_{n_r=1}^{N_r}\left(\frac{1}{\lambda'_{n_s}\lambda_{H,n_r}} + \frac{L\sigma_{n_s}}{\sigma_n^2}\right)^{-1} + \sum_{n_t=1}^{N_t}\sum_{n_s=1}^{N_s}\sum_{n_r=1}^{N_r}\left(\frac{1}{\lambda'_{n_s,n_t}\lambda_{G,n_r}} + \frac{L\sigma_{n_s}}{\sigma_n^2}\right)^{-1} + \mu\left(\sum_{n_s=1}^{N_s}\sigma_{n_s} - P_s\right) - \sum_{n_s=1}^{N_s}\eta_{n_s}\sigma_{n_s} \tag{45}$$

$$-\sum_{n_r=1}^{N_r}\frac{\frac{L}{\sigma_n^2}}{\left(\frac{1}{\lambda'_{n_s}\lambda_{H,n_r}} + \frac{L\sigma_{n_s}}{\sigma_n^2}\right)^2} - \sum_{n_t=1}^{N_t}\sum_{n_r=1}^{N_r}\frac{\frac{L}{\sigma_n^2}}{\left(\frac{1}{\lambda'_{n_s,n_t}\lambda_{G,n_r}} + \frac{L\sigma_{n_s}}{\sigma_n^2}\right)^2} + \mu - \eta_{n_s} = 0, n_s = 1, 2, \cdots, N_s \tag{46}$$

Meanwhile, we consider that a DFT matrix is used at the tag. For example,

$$\mathbf{C}_D = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_\tau & \cdots & W_\tau^{\tau-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & W_\tau^K & \cdots & W_\tau^{(\tau-1)K} \end{bmatrix}, \tag{50}$$

where $W_\tau = e^{-j2\pi/\tau}$. The number of rows equals $K = N_t + 1$ and the number of columns is $\tau = L$.

In summary, the DFT matrix based sequences can achieve the shortest estimation duration, but it increases the complexity for tags. Hadamard matrix based sequences employ only two values simplifying the hardware but may increase the shortest estimation duration.

Remarks: Practical implementation: For the tag, only two impedance is required if Hadamard matrix is adopted, which is easy to implement. The DFT matrix is more suitable for semi-passive or active tags [27]. There is no optimization for the tag's training symbols. It means that the tag does not need to receive any optimization information from other nodes. This keeps the tag simple. As for the source symbol optimization, it is affordable for the source or the reader since it is usually a plug-in device.

Time slot overhead: The time slot overhead is minimum. Similar to [27], our channel estimation method is one-shot. The one-shot method means that we estimate the channels for all tag's antenna simultaneously because we treat the source as a hidden tag and the training symbols are orthogonal. It has the minimum overhead.

Computational complexity: The solution of the problem (21)-(23) is calculated by the source or the reader. Thus, the tag has no additional burden. To obtain the optimal training sequence, $\mathbf{U}'$ and $\mathbf{\Sigma}$ are necessary as shown in *Theorem 2*. The matrix $\mathbf{U}'$ is obtained by the EVD of $\mathbf{T}_{HF}$. The computational complexity is $O(N_s^3)$. As for $\mathbf{\Sigma}$, it is calculated by solving the problem (42)-(44) and the corresponding complexity is $O(N_s)$.

Algorithm optimality: The proposed solution is globally optimal as *Theorem 2* shows the optimal structure of $\mathbf{S}$ and $\mathbf{X}$ minimizing the MSE. In the case of calculating $\mathbf{\Sigma}$ by solving the problem (42)-(44), the KKT optimality conditions assure the globally optimality of the proposed solution.

Characteristics of estimator: The proposed channel estimation method is based on the MMSE estimator. First, this estimator is a long-term solution, which requires the statistical information of the channel and the noise. If the MIMO channel or noise is highly dynamic or complex, e.g., in mmWave RIS systems, an artificial intelligence (AI) enabled method is more efficient for estimation [43]. Second, the estimator is linear. If the signal form of the multi-antenna system is non-linear, e. g., squared amplitude, neural network can be applied to improve channel estimation [44]. Finally, the computational complexity of the MMSE estimator increases significantly with a large number of antennas. For backscatter communications, the number of antennas of the system node, especially the tag, is usually small. When the receiver and the source are equipped with a large number of antennas, and extend the scenario to RIS channel estimations, an AI-enabled method can help to reduce computational complexity [45].

IV. Simulation Results and Discussions

In this section, simulation results are provided to evaluate the performance of the proposed pilot sequence design and channel estimation algorithm.

We set matrices $\mathbf{H}_w, \mathbf{F}_w$ and $\mathbf{G}_w$ as $N_r \times N_s$, $N_t \times N_s$ and $N_r \times N_t$ Gaussian random matrices with $i.i.d.$ $\mathcal{CN}(0,1)$ entries. The coefficient α is set as 0.5. The signal-to-noise ratio (SNR) is defined as P_s/σ_n^2 and is set as 2 dB unless stated otherwise. The commonly used exponential Toeplitz structure [29] is adopted as the channel covariance matrices and there is $[\mathbf{T}_{HF}]_{m,n} = \rho_s^{|m-n|}$, $[\mathbf{R}_H]_{m,n} = \rho_r^{|m-n|}$, $[\mathbf{R}_F]_{m,n} = \rho_t^{|m-n|}$, $[\mathbf{T}_G]_{m,n} = \rho_t^{|m-n|}$,$[\mathbf{R}_G]_{m,n} = \rho_r^{|m-n|}$. We set $\rho_s = 0.9$, $\rho_t = 0.8$ and $\rho_r = 0.7$, unless stated otherwise. The tag uses a DFT matrix as its training sequences. The average source power P_s is set to 1. The length L is determined by $L = (N_t + 1)N_s$. The number of antennas N_s, N_t and N_r are set as 2, 3 and 6, respectively, unless stated otherwise. The final MSE is normalized by multiplying $1/(N_s N_r)$. We set the equal power scheme as the benchmark in which $\sigma_{n_s}, n_s = 1, 2, \cdots, N_s$ are equal to each other and their sum is P_s. The other channel estimation methods, such as compressed sensing or deep learning-based approaches [39] have their limitations. For example, comprseed sensing is not applicable since the channels in our model are not sparse. Also, deep learning-based approaches may not guarantee the global optimality. We also simulate [27] with the setting that a DFT matrix based sequences are assigned to the source and the tag. The method in [27] is abbreviated as 'ortho' in the legend.

We first evaluate the MSE of the channel training algorithm versus the SNR in Fig. 2. With the increase of the SNR, the MSE decreases. The optimal power allocation scheme has a lower MSE than the equal power allocation scheme especially in the low SNR level, validating the advantage of

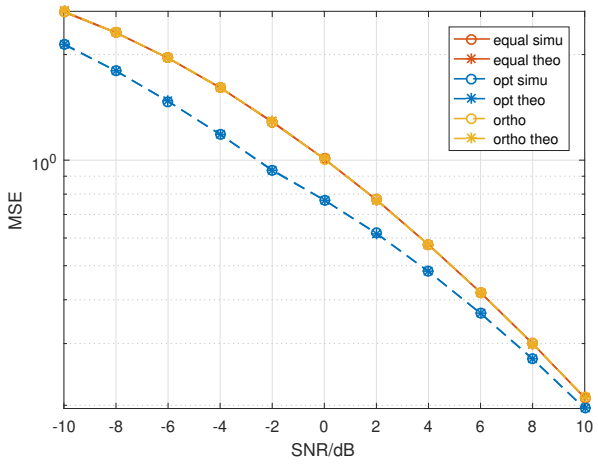

Fig. 2. MSE versus SNR

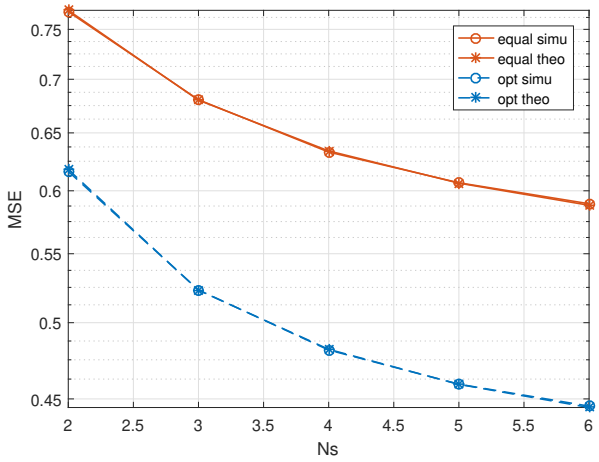

Fig. 4. MSE versus N_s

our proposed scheme. They have almost the same MSE in the high SNR region. This is because the allocated power of the optimal scheme is almost equal in this situation. The proposed method in [27] has the same MSE performance compared to the equal power scheme because $\mathbf{\Sigma}$ for [27] is also a scaled identity matrix. This scaled identity matrix is derived from the orthogonal training sequences of the source. The simulation curves are consistent with the theoretical results, showing our analysis is correct.

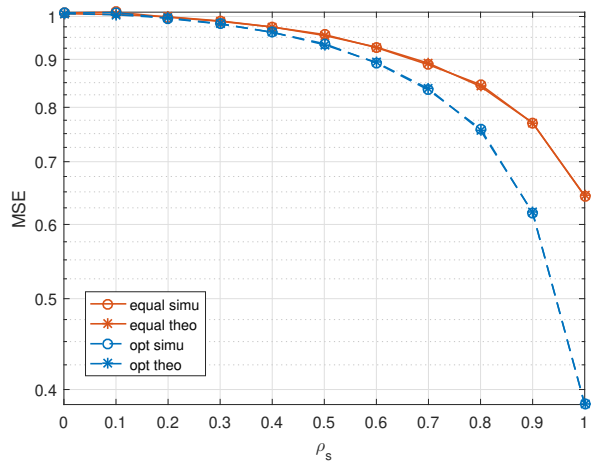

Fig. 3. MSE versus ρ_s

The MSE of the channel training algorithm versus the correlation coefficient ρ_s is illustrated in Fig. 3. With the increase of ρ_s, the MSE decreases. This phenomenon indicates that dependent antennas have a better estimation performance. With the increase of ρ_s, the MSE gap between two algorithms widens.

The MSE of the channel training algorithm versus the number of the source antennas is presented in Fig. 4. With the increase in the number of source antennas, the MSE decreases. This is because more pilot signals are backscattered with the increase of N_s.

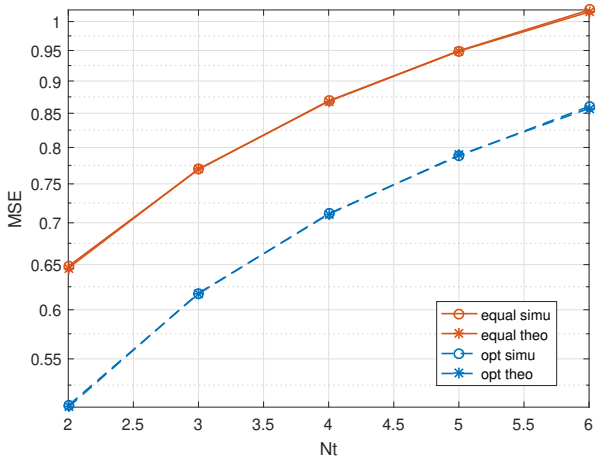

Fig. 5. MSE versus N_t

The impact of the number of tag antennas on the MSE of the channel training algorithm is illustrated in Fig. 5. With the increase in the number of tag antennas, the MSE increases. With the increase of N_t, the interference between tags increases and then the estimation performance gets worse.

The MSE of the channel training algorithm versus the number of reader antennas is presented in Fig. 6. With the increase of N_r, the MSE decreases. With more signals received at the reader, the estimation performance is improved. It informs us that we can increase the number of reader antennas to enhance the estimation accuracy.

V. Conclusion

In this paper, we have proposed and investigated the performance of the pilot sequence design and channel estimation algorithm for MIMO BackCom systems. The proposed algorithm can efficiently estimate the direct link and the backscatter link CSI for MIMO BackCom systems and outperforms the channel training algorithm without power allocation. The optimal training sequences can be efficiently implemented in practice.

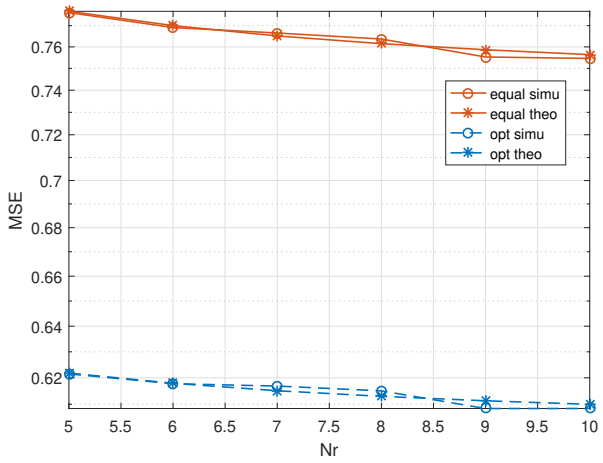

Fig. 6. MSE versus N_r

Appendix A
Proof of Theorem 1

The following lemma will be used in the sequel.

Lemma 1 [32]: For $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Theta} \otimes \mathbf{\Phi})$, we have $E[\mathbf{H}\mathbf{A}\mathbf{H}^H] = tr(\mathbf{A}\mathbf{\Theta}^T)\mathbf{\Phi}$ and $E[\mathbf{H}^H\mathbf{A}\mathbf{H}] = tr(\mathbf{\Phi}\mathbf{A})\mathbf{\Theta}^T$.

Let $\widetilde{\mathbf{H}} = \mathbf{H}\mathbf{U}', \widetilde{\mathbf{g}_1\mathbf{f}_1} = (\mathbf{g}_1\mathbf{f}_1)\mathbf{U}', \cdots, \widetilde{\mathbf{g}_{N_t}\mathbf{f}_{N_t}} = (\mathbf{g}_{N_t}\mathbf{f}_{N_t})\mathbf{U}'$.

First, the m-th column of $\widetilde{\mathbf{H}}$ is given by

$$\begin{aligned}[\widetilde{\mathbf{H}}]_m &= [\mathbf{B}_H\mathbf{H}_w\mathbf{A}^H\mathbf{U}']_m = [\mathbf{B}_H\mathbf{H}_w\mathbf{\Pi}'\mathbf{\Lambda}'^{\frac{1}{2}}\mathbf{U}'^H\mathbf{U}']_m \\ &= \lambda'^{\frac{1}{2}}_m\mathbf{B}_H\mathbf{H}_w[\mathbf{\Pi}']_m,\end{aligned} \tag{51}$$

and λ'_m is the m-th diagonal element in $\mathbf{\Lambda}'$. Based on Lemma 1, we get

$$\begin{aligned}&E\left[[\widetilde{\mathbf{H}}]_m[\widetilde{\mathbf{H}}]_m^H\right] \\ &= E\left[\lambda'^{\frac{1}{2}}_m\mathbf{B}_H\mathbf{H}_w[\mathbf{\Pi}']_m\left(\lambda'^{\frac{1}{2}}_m\mathbf{B}_H\mathbf{H}_w[\mathbf{\Pi}']_m\right)^H\right] \\ &= \lambda'_m E\left[\mathbf{B}_H\mathbf{H}_w[\mathbf{\Pi}']_m[\mathbf{\Pi}']_m^H\mathbf{H}_w^H\mathbf{B}_H^H\right] \\ &= \lambda'_m tr([\mathbf{\Pi}']_m[\mathbf{\Pi}']_m^H)\mathbf{B}_H\mathbf{B}_H^H \\ &= \lambda'_m\mathbf{R}_H.\end{aligned} \tag{52}$$

Next, the m-th column of $\widetilde{\mathbf{g}_{n_t}\mathbf{f}_{n_t}}$ is calculated by

$$\begin{aligned}&\mathbf{g}_{n_t}\mathbf{f}_{n_t}\mathbf{U}' \\ &= [\mathbf{G}]_{n_t}\{\mathbf{F}\}_{n_t}\mathbf{U}' \\ &= \left[\mathbf{B}_G\mathbf{G}_w\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\mathbf{F}_w\mathbf{A}^H\}_{n_t}\mathbf{U}' \\ &= \left[\mathbf{B}_G\mathbf{G}_w\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\mathbf{F}_w\mathbf{\Pi}'\mathbf{\Lambda}'^{\frac{1}{2}}\mathbf{U}'^H\}_{n_t}\mathbf{U}' \\ &= \mathbf{B}_G\mathbf{G}_w\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\mathbf{F}_w\mathbf{\Pi}'\mathbf{\Lambda}'^{\frac{1}{2}},\end{aligned} \tag{53}$$

and

$$\begin{aligned}\left[\mathbf{g}_{n_t}\mathbf{f}_{n_t}\mathbf{U}'\right]_m &= \left[\mathbf{B}_G\mathbf{G}_w\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\mathbf{F}_w\mathbf{\Pi}'\mathbf{\Lambda}'^{\frac{1}{2}}\right]_m \\ &= \lambda'^{\frac{1}{2}}_m\mathbf{B}_G\mathbf{G}_w\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\mathbf{F}_w[\mathbf{\Pi}']_m.\end{aligned} \tag{54}$$

Thus, based on Lemma 1, $E\left[\left[\widetilde{\mathbf{g}_{n_t}\mathbf{f}_{n_t}}\right]_m\left[\widetilde{\mathbf{g}_{n_t}\mathbf{f}_{n_t}}\right]_m^H\right]$ can be derived as (55) shown at the top of next page, where $b_{n_t} = tr\left(\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}{}^H\left[\mathbf{A}_G^H\right]_{n_t}{}^H\right)$.

The off-diagonal entries in this matrix are zeros because Matrices $\mathbf{H}_w, \mathbf{F}_w$ and $\mathbf{G}_w$ are Gaussian random matrices with $(i.i.d.)$ zero mean entries. With (52) and (55), $\mathbf{R}_q$ is derived in (17). □

References

[1] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, "Ambient backscatter communications: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2889–2922, 4th Quart., 2018.

[2] C. Xu, L. Yang and P. Zhang, "Practical backscatter communication systems for battery-free Internet of Things: A tutorial and survey of recent research," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 16-27, Sep. 2018.

[3] W. Wu, X. Wang, A. Hawbani, L. Yuan, and W. Gong, "A survey on ambient backscatter communications: Principles, systems, applications, and challenges," *Comput. Netw.*, vol. 216, Oct. 2022.

[4] T. Jiang, Y. Zhang, W. Ma, M. Peng, Y. Peng, M. Feng and G. Liu, "Backscatter communication meets practical battery-free Internet of Things: A survey and outlook," *IEEE Commun. Surv. Tuts.*, vol. 25, no. 3, pp. 2021–2051, 3rd-quarter, 2023.

[5] B. Gu, D. Li, H. Ding, G. Wang, C. Tellambura, "Breaking the interference and fading gridlock in backscatter communications: State-of-the-art, design challenges, and future directions," *Early Access*, 3rd-quarter, 2024.

[6] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," in *Proc. ACM SIGCOMM*, Jun. 2013, pp. 39-50.

[7] B. Kellogg, A. Parks, S. Gollakota, J. R. Smith and D. Wetherall, "Wi-Fi backscatter: Internet connectivity for RF-powered devices," in *Proc. ACM SIGCOMM*, Aug. 2014, pp. 607-618.

[8] D. Li, "Two birds with one stone: Exploiting decode-and-forward relaying for opportunistic ambient backscattering," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1405-1416, Mar. 2020.

[9] D. Li,"Hybrid active and passive antenna selection for backscatter-assisted MISO systems," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7258-7269, Nov. 2020.

[10] A. N. Parks, A. Liu, S. Gollakota and J. R. Smith, "Turbocharging ambient backscatter communication," in *Proc. ACM SIGCOMM*, pp. 619-630, Aug. 2014.

[11] T. Kim and W. Lee, "Anyscatter: Eliminating technology dependency in ambient backscatter systems," in *IEEE INFOCOM 2020*, pp. 288-296, July 2020.

[12] G. Yang, Q. Zhang, and Y.-C. Liang, "Cooperative ambient backscatter communications for green Internet-of-Things," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1116-1130, Apr. 2018.

[13] Q. Tao, C. Zhong, X. Chen, H. Lin, and Z. Zhang, "Maximum eigenvalue detector for multiple antenna ambient backscatter communication systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12411–12415, 2019.

[14] Q. Tao, C. Zhong, X. Chen, H. Lin, and Z. Zhang, "Optimal detection for ambient backscatter communication systems with multi-antenna reader under complex Gaussian illuminator," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11 371–11 383, 2020.

[15] C. Chen, G. Wang, H. Guan, Y. Liang, and C. Tellambura, "Transceiver design and signal detection in backscatter communication systems with multiple-antenna tags," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3273–3288, 2020.

[16] C. Chen, G. Wang, P. D. Diamantoulakis, R. He, G. K. Karagiannidis, and C. Tellambura, "Signal detection and optimal antenna selection for ambient backscatter communications with multi-antenna tags," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 466–479, 2020.

[17] X. Wang, H. Yiğitler and R. Jäntti, "Gaining from multiple ambient sources: Signaling matrix for multi-antenna backscatter devices", *IEEE Wireless Commun. Lett.*, vol. 12, no. 3, pp. 491-495, 2023.

[18] Z. Niu, L. Xiao, W. Ma, M. Peng and T. Jiang, "Generalized space-time architecture for ambient backscatter communication," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 1912-1925, Apr. 2023.

[19] Z. Niu, W. Ma, W. Wang, and T. Jiang, "Spatial modulation-based ambient backscatter: Bring energy self-sustainability to massive Internet of everything in 6G," *China Commun.*, vol. 17, no. 12, pp. 52-65, Dec. 2020.

[20] J. Chen, H. Yu, Q. Guan, G. Yang, and Y.-C. Liang, "Spatial modulation based multiple access for ambient backscatter networks," *IEEE Commun. Lett.*, vol. 26, no. 1, Jan. 2022.

$$
\begin{aligned}
&E\left[\left[\widetilde{\mathbf{g}_{n_t}\mathbf{f}_{n_t}}\right]_m\left[\widetilde{\mathbf{g}_{n_t}\mathbf{f}_{n_t}}\right]_m^H\right] = E\left[\left[(\mathbf{g}_{n_t}\mathbf{f}_{n_t})\mathbf{U}'\right]_m\left[(\mathbf{g}_{n_t}\mathbf{f}_{n_t})\mathbf{U}'\right]_m^H\right] \\
&= E\left[\lambda'^{\frac{1}{2}}_m\mathbf{B}_G\mathbf{G}_w\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\mathbf{F}_w[\mathbf{\Pi}']_m\left[\lambda'^{\frac{1}{2}}_m\mathbf{B}_G\mathbf{G}_w\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\mathbf{F}_w[\mathbf{\Pi}']_m\right]^H\right] \\
&= \lambda'_m E\left[\mathbf{B}_G\mathbf{G}_w\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\mathbf{F}_w[\mathbf{\Pi}']_m\left[\mathbf{B}_G\mathbf{G}_w\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\mathbf{F}_w[\mathbf{\Pi}']_m\right]^H\right] \\
&= \lambda'_m tr\left(\left[\mathbf{A}_G^H\right]_{n_t}\{\mathbf{B}_F\}_{n_t}\{\mathbf{B}_F\}_{n_t}{}^H\left[\mathbf{A}_G^H\right]_{n_t}^H\right)\mathbf{B}_G\mathbf{B}_G^H = \lambda'_m b_{n_t}\mathbf{R}_G
\end{aligned}
\tag{55}
$$

[21] J. Yang, L. Wang, X. Yang, "Generalized quadrature space-time modulation for ambient backscatter communication," *IEEE Commun. Lett.*, *Early access*, May 2024.

[22] J. Liu, J. Yu, R. Zhang, S. Wang, K. Yang, J. An, "Covert MIMO ambient backscatter communication," *IEEE Trans. Commun.*, vol. 72, no. 3, pp. 1746-1758, Nov. 2023.

[23] T. Wu, M. Jiang, Q. Zhang, Q. Li, and J. Qin, "Beamforming design in multiple-input-multiple-output symbiotic radio backscatter systems," *IEEE Commun. Lett.*, vol. 25, pp. 1949–1953, Jun 2021.

[24] W. Liu, S. Shen, D. H. K. Tsang, and R. D. Murch, "Enhancing ambient backscatter communication utilizing coherent and non-coherent space-time codes," *IEEE Trans. Wireless Commun.*, vol. 20, pp. 6884–6897, Oct. 2021.

[25] X. Liu, Z. Chi, W. Wang, Y. Yao, and T. Zhu, "VMscatter: A versatile MIMO backscatter," in *Proc. USENIX NSDI*, Santa Clara, CA, USA, pp. 895-909, Feb. 2020.

[26] A. H. Raghavendra, A. K. Kowshik, S. Gurugopinath, S. Muhaidat and C. Tellambura, "Generalized space shift keying for ambient backscatter communication," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5018-5029, Aug. 2022.

[27] F. Rezaei, D. Galappaththige, C. Tellambura, and A. Maaref, "Time-spread pilot-based channel estimation for backscatter networks," *IEEE Trans. Commun.*, vol. 72, pp. 434–449, Jan. 2024.

[28] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 772–781, Sep. 1978.

[29] D. S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 503–513, Mar. 2000.

[30] W.-X. Long, M. Moretti, A. Abrardo, L. Sanguinetti, R. Chen, "MMSE design of RIS-aided communications with spatially-correlated channels and electromagnetic interference," *IEEE Trans. Wireless Commun.*, Early access, Aug. 2024.

[31] O. T. Demir and E. Björnson, "Is channel estimation necessary to select phase-shifts for RIS-assisted massive MIMO?", *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9537-9552, Nov. 2022.

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

[33] J. D. Griffin and G. D. Durgin, "Gains for RF tags using multiple antennas," *IEEE Trans. Antennas Propag.*, vol. 56, no. 2, pp. 563–570, Feb. 2008.

[34] E. Denicke, H. Hartmann, N. Peitzmeier, and B. Geck, "Backscatter beamforming: A transponder for novel MIMO RFID transmission schemes," *IEEE J. Radio Freq. Identif.*, vol. 2, no. 2, pp. 80–85, Jun. 2018.

[35] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cilffs, NJ: Prentice-Hall, 1993.

[36] K. B. Petersen and M. S. Petersen, *The Matrix Cookbook* [Online]. Available: https://www.math.uwaterloo.ca/hwolkowi/matrixcookbook.pdf, ver. Nov. 15, 2012.

[37] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. New York, NY, USA: Springer, 2009.

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[39] C. Singhal and S. De, *Resource Allocation in Next-Generation Broadband Wireless Access Networks*. Hershey, PA, USA : IGI Global, 2017.

[40] J. Qian, A. N. Parks, J. R. Smith, F. Gao, and S. Jin, "IoT communications with M-PSK modulated ambient backscatter: Algorithm, analysis, and implementation," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 844–855, 2019.

[41] P. Zhang, D. Bharadia, K. Joshi, and S. Katti, "Hitchhike: Practical backscatter using commodity WiFi," in *Proceedings of the 14th ACM-Conf. Embedded Network Sensor Systems*, Standford,CA, Nov. 2016.

[42] M. Dunna, M. Meng, P.-H. Wang, C. Zhang, P. Mercier, and D. Bharadia, "SyncScatter: Enabling WiFi like synchronization and range for WiFi backscatter communication," in *Proc. 18th USENIX Sympos. Networked Systems Design and Implementation (NSDI 21)*, Apr. 2021.

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

[44] H. Sun, L. Zhu, W. Mei, and R. Zhang, "Power measurement based channel estimation for IRS-enhanced wireless coverage," IEEE Trans. Wireless Commun., vol. 23, no. 12, pp. 19183–19198, Dec. 2024.

[45] A. Abdallah, A. Celik, M. M. Mansour, and A. M. Eltawil, "RIS-aided mmWave MIMO channel estimation using deep learning and compressive sensing," IEEE Trans. Wireless Commun., vol. 22, no. 5, pp. 3503–3521, May 2023.