Two-Dimensional Ice Filling Based Channel Estimation in Densifying MIMO Systems

Zijian Zhang* and Mingyao Cui†
*Department of Electronic Engineering, Tsinghua University, Beijing, China
†Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR
Emails: zhangzj20@mails.tsinghua.edu.cn; cuimy23@connect.hku.hk

***Abstract*— Densifying multiple-input multiple-output (MIMO) has attracted much attention in recent years. The strong correlations among densifying antennas provide sufficient prior knowledge about channel state information (CSI), which inspires the careful design of observation matrices (e.g., transmit precoders and receive combiners) to boost channel estimation performance. To achieve this, this work proposes to jointly design the combiners and precoders by maximizing the mutual information between the received pilots and densifying MIMO channels. Particularly, a two-dimensional ice-filling (2DIF) algorithm is proposed, which is motivated by the fact that the eigenspace of MIMO channel covariance can be decoupled into two sub-eigenspaces. By properly setting the precoder and the combiner as the eigenvectors from these two sub-eigenspaces, the 2DIF promises to generate near-optimal observation matrices for channel estimation. Simulation results demonstrate that, the proposed 2DIF method outperforms the state-of-the-art schemes in channel estimation accuracy.**

I. Introduction

Densifying multiple-input multiple-output (MIMO) has attracted considerable attention in recent years. The antenna spacing of densifying MIMO is much smaller than the half wavelength $\lambda/2$, e.g., $\lambda/6$ [1] or even $\lambda/23$ [2]. By densely arranging subwavelength-spaced antennas in a compact space, many dense-antenna transceiver architectures have emerged, such as holographic MIMO (H-MIMO), superdirective antenna arrays, reconfigurable intelligent surfaces (RISs), and fluid antenna systems. Utilizing the extensive channel observations facilitated by a multitude of antennas, densifying MIMO is anticipated to achieve significant array gains and multiplexing-diversity gains [3]. Furthermore, densifying MIMO can mitigate the effects of grating lobes and offer enhanced performance for large oblique angles of incidence [4]. Some studies have also highlighted their capabilities to realize super-directivity or super-bandwidth [5] in wireless transmissions.

To ensure the transmission performance of MIMO systems, an indispensable technology is the acquisition of channel state information (CSI). To date, numerous channel estimators have been proposed, such as the classical least square (LS) and minimum mean square error (MMSE) estimators. Some compressed sensing (CS)-based methods are used to further enhance the estimation accuracy, such as the approximate message passing (AMP)-based estimator. Although many channel estimators can be adopted in densifying MIMO systems, they often exhibit a non-negligible performance gap compared to the optimal estimator. This is because most existing estimators overlook the strong correlations among densifying MIMO antennas. Since the antenna spacing of densifying MIMO is very small, the channels associated with close-by antennas are spatially similar [6], which leads to the highly structured covariance matrices of channels. Such an informative covariance matrix can provide appreciable prior knowledge for the featured design of observation matrices (e.g., combiners and precoders), thus improving the accuracy of CSI acquisition.

To exploit the strong channel correlations for improved CSI acquisition, our prior work [7] proposes an ice filling (IF) based observation matrix design in dense array systems (DASs), which is inspired by the idea of Gaussian Process Regression (GPR). By maximizing the mutual information (MI) characterized by the covariance matrix, the IF algorithm can sequentially produce the observation vectors of receivers in a pilot-by-pilot manner [7]. However, IF method is only feasible to design the *vector-form receive combiner* in single-input multiple-output (SIMO) system with a single-antenna transmitter and a multi-antenna single-radio frequency (RF)-chain receiver [7]. For a general densifying MIMO system with multiple antennas and RF chains at both transceivers, *the receive and transmit channel covariance pair*, as well as the *matrix-form receive combiner and transmit precoder pair* are coupled together. As IF scheme fails to tackle these couplings, the full exploitation of densifying MIMO's channel covariance for designing observation matrices is still a challenge.

To fill in this gap, this work proposes a two-dimensional ice filling (2DIF) based observation matrix design, whose core idea is to jointly design precoders and combiners by maximizing the MI between channels and pilots. Specifically, to overcome the design challenges imposed by the coupled matrix-form combiners and precoders in observation matrices, the proposed 2DIF based design employs a greedy method to jointly produce the combiners and precoders in a block-by-block way. Firstly, we prove that the eigenspace of the channel covariance can be decoupled into two sub-eigenspaces, which are associated with the correlations of transmitter antennas and receiver antennas, respectively. Then, utilizing the eigenspace invariance, the near-optimal observation matrix can be obtained by properly setting the precoder and the combiner as the eigenvectors from these two sub-eigenspaces, which can be realized by a linear search algorithm. Finally, we provide an intuitive and insightful explanation for 2DIF to clarify its physical significance. Similar to the water-filling precoding that maximizes the MIMO capacity, the implementation of 2DIF can be viewed as a two-dimensional ice-filling process.

II. System Model and Problem Formulation

This paper considers the uplink channel estimation of an densifying MIMO system, consisting of an N_{R}-antenna base station (BS) equipped with N_{RF} RF chains and an N_{T}-antenna

ICC 2025 - IEEE International Conference on Communications | 979-8-3315-0521-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICC52391.2025.11161466

user. The antennas at transceivers are densely arranged with sub-wavelength antenna spacing d. We define $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ as the wireless channel and Q as the number of transmit pilots within a coherence-time frame. The received signal $\mathbf{y}_q \in \mathbb{C}^{N_{\mathrm{RF}}}$ at the BS in timeslot q is modeled as

$$\mathbf{y}_q = \mathbf{W}_q^H \mathbf{H} \mathbf{v}_q s_q + \mathbf{W}_q^H \mathbf{z}_q = \left(\mathbf{v}_q^T \otimes \mathbf{W}_q^H\right) \mathbf{h} s_q + \mathbf{W}_q^H \mathbf{z}_q,$$

where $\mathbf{h} \equiv \mathrm{vec}(\mathbf{H})$, s_q is the pilot symbol, and $\mathbf{z}_q \sim \mathcal{CN}\left(\mathbf{0}_M, \sigma^2 \mathbf{I}_M\right)$ is the additive white Gaussian noise (AWGN). Vector $\mathbf{v}_q \in \mathbb{C}^{N_{\mathrm{T}}}$ denotes the precoder at the user. For the transmitter, the user equipment typically employs a fully digital precoder with a moderate number of antennas, N_T. Thereby, the coefficient of $\mathbf{v}_q$ can be freely configured as long as the total power constraint is satisfied: $\|\mathbf{v}_q\|^2 = P$, where P is the maximum transmit power per pilot. For the receiver, $\mathbf{W}_q := \mathbf{A}_q \mathbf{D}_q \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{RF}}}$ is the hybrid combiner at the BS, with $\mathbf{A}_q \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{RF}}}$ and $\mathbf{D}_q \in \mathbb{C}^{N_{\mathrm{RF}} \times N_{\mathrm{RF}}}$ being the analog and digital combiners, respectively.

Without loss of generality, we assume that $s_q = 1, \forall q \in \{1, \cdots, Q\}$. Considering the total Q timeslots, we arrive at

$$\mathbf{y} = \mathbf{X}^H \mathbf{h} + \mathbf{z}, \tag{1}$$

where $\mathbf{y} := \left[\mathbf{y}_1^T, \cdots, \mathbf{y}_Q^T\right]^T$, $\mathbf{z} := \left[\mathbf{z}_1^H \mathbf{W}_1, \cdots, \mathbf{z}_Q^H \mathbf{W}_Q\right]^H$, $\mathbf{X} = [\mathbf{X}_1, \cdots, \mathbf{X}_Q]$, and $\mathbf{X}_q := \mathbf{v}_q^* \otimes \mathbf{W}_q$ is defined as the observation matrix for each pilot. This paper aims at accurately estimating $\mathbf{h}$ from $\mathbf{y}$ by jointly designing combiners $\{\mathbf{W}_q\}_{q=1}^Q$ and precoders $\{\mathbf{v}_q\}_{q=1}^Q$.

We consider the Saleh-Valenzuela (SV) multi-cluster channel model. The uniform linear arrays (ULAs) are deployed at the transceivers for the ease of discussion. Let $\mathbf{a}(\theta) \in \mathbb{C}^{N_{\mathrm{R}}}$ and $\mathbf{b}(\varphi) \in \mathbb{C}^{N_{\mathrm{T}}}$ denote the steering vectors at the receiver and transmitter, respectively, given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N_{\mathrm{R}}}} [1, e^{\mathrm{j}\frac{2\pi}{\lambda} d \cos(\theta)}, \cdots, e^{\mathrm{j}\frac{2\pi}{\lambda}(N_{\mathrm{R}}-1) d \cos(\theta)}]^{\mathrm{T}}, \tag{2}$$

$$\mathbf{b}(\varphi) = \frac{1}{\sqrt{N_{\mathrm{T}}}} [1, e^{\mathrm{j}\frac{2\pi}{\lambda} d \cos(\varphi)}, \cdots, e^{\mathrm{j}\frac{2\pi}{\lambda}(N_{\mathrm{T}}-1) d \cos(\varphi)}]^{\mathrm{T}}, \tag{3}$$

where θ and ϕ respectively refer to the angle-of-arrival (AoA) and the angle-of-departure (AoD), and λ is the wavelength. Assuming that the number of clusters is C each consisting of R rays, the SV channel $\mathbf{H}$ is represented as

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{T}} N_{\mathrm{R}}}{CR}} \sum\nolimits_{c=1}^{C} \sum\nolimits_{r=1}^{R} g_{c,r} \mathbf{a}(\theta_{c,r}) \mathbf{b}^H(\varphi_{c,r}), \tag{4}$$

where $g_{c,r}$, $\theta_{c,r}$, $\varphi_{c,r}$ are the complex path gain, AoA, and AoD associated with the r-th ray in the c-th cluster, respectively. According to the standard 3GPP setting [8], the path gains $\{g_{c,r}\}_{c=1,r=1}^{C,R}$ are usually modeled as independently and identically distributions (i.i.d.) with zero mean and normalized power, i.e., $\mathsf{E}(g_{c,r}) = 0$ and $\mathsf{E}(|g_{c,r}|^2) = 1$; and the AoAs $\{\theta_{c,r}\}_{c=1,r=1}^{C,R}$ and the AoDs $\{\varphi_{c,r}\}_{c=1,r=1}^{C,R}$ associated with the same cluster are correlated (depending on the angle spread), while those associated with the different clusters are i.i.d [8].

As the antennas of densifying MIMO are packed with a small spacing, the channels across close-by antennas are strongly correlated. Define the covariance of channel $\mathbf{h}$ as $\mathsf{E}\left(\mathbf{h}\mathbf{h}^H\right) = \mathbf{\Sigma}_{\mathbf{h}} \in \mathbb{C}^{N_{\mathrm{R}} N_{\mathrm{T}} \times N_{\mathrm{R}} N_{\mathrm{T}}}$, which is also called the *kernel* of channel. The high-correlation property of wireless channel indicates that the kernel $\mathbf{\Sigma}_{\mathbf{h}}$ is structural and underdetermined, which can provide prior knowledge to achieve high-accuracy channel estimation. To materialize this vision, we follow the idea of GPR to design the optimal estimator and the optimal observation matrix. Specifically, the channel is assumed to be sampled from the Gaussian process $\mathcal{CN}(\mathbf{0}_{N_{\mathrm{R}} N_{\mathrm{T}}}, \mathbf{\Sigma}_{\mathbf{h}})$. Define $\mathbf{\Xi} = \sigma^2 \mathrm{blkdiag}\left(\mathbf{W}_1^H \mathbf{W}_1, \cdots, \mathbf{W}_Q^H \mathbf{W}_Q\right)$ as the covariance matrix of the noise $\mathbf{z}$. Given $\mathbf{y}$, the posterior mean and the posterior covariance of $\mathbf{h}$ are expressed as

$$\boldsymbol{\mu}_{\mathbf{h}|\mathbf{y}} = \mathbf{\Sigma}_{\mathbf{h}} \mathbf{X} \left(\mathbf{X}^H \mathbf{\Sigma}_{\mathbf{h}} \mathbf{X} + \mathbf{\Xi}\right)^{-1} \mathbf{y}, \tag{5}$$

$$\mathbf{\Sigma}_{\mathbf{h}|\mathbf{y}} = \mathbf{\Sigma}_{\mathbf{h}} - \mathbf{\Sigma}_{\mathbf{h}} \mathbf{X} \left(\mathbf{X}^H \mathbf{\Sigma}_{\mathbf{h}} \mathbf{X} + \mathbf{\Xi}\right)^{-1} \mathbf{X}^H \mathbf{\Sigma}_{\mathbf{h}}, \tag{6}$$

which give rise to the optimal channel estimator $\boldsymbol{\mu}_{\mathbf{h}|\mathbf{y}}$ and the associated estimation error $\mathbf{\Sigma}_{\mathbf{h}|\mathbf{y}}$, respectively. Notably, the posterior covariance $\mathbf{\Sigma}_{\mathbf{h}|\mathbf{y}}$ is largely dependent on the observation matrix $\mathbf{X}$. Thereby, well-designed combiners and precoders, $\{\mathbf{W}_q\}_{q=1}^Q$ and $\{\mathbf{v}_q\}_{q=1}^Q$, can substantially reduce the channel estimation error. Motivated by this fact, GPR attempts to produce the observation matrices to gain as much information of $\mathbf{h}$ as possible from the received signal $\mathbf{y}$. Following this idea, our objective is to maximize the MI between $\mathbf{y}$ and $\mathbf{h}$, which is formulated as:

$$\max_{\mathbf{W} \in \mathcal{W}, \|\mathbf{v}_q\|_2^2 = P} I(\mathbf{y}; \mathbf{h}) = \log_2 \det\left(\mathbf{I}_{N_{\mathrm{RF}} Q} + \mathbf{\Xi}^{-1} \mathbf{X}^H \mathbf{\Sigma}_{\mathbf{h}} \mathbf{X}\right), \tag{7}$$

where $\mathcal{W}$ stands for the feasible set of hybrid combiners, which depends on the implementation of analog combiners. Generally, there are two implementations of the analog combiners: 1) the amplitude-and-phase controllable combiner and 2) the phase-only controllable combiner. In this paper, the derivations and results in the former case can be easily extended to the later case by adopting some well-performed hybrid precoding algorithms, e.g., [9], to solve $\min_{\mathbf{A}_q, \mathbf{D}_q} \left\|\mathbf{A}_q \mathbf{D}_q - \mathbf{W}_q^{\mathrm{opt}}\right\|$. Thus, we focus on the former case, i.e., $\mathbf{W}_q \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{RF}}}$, for ease of discussion.

III. 2DIF Based Observation Matrix Design

A. Precoder/Combiner Design Using Greedy Method

Observing problem (7), one can find that the MI $I(\mathbf{y}; \mathbf{h})$ is non-concave. Due to the coupled term $\mathbf{v}_q^T \otimes \mathbf{W}_q^H$ in $\mathbf{X}$ and the colored-noise covariance matrix $\mathbf{\Xi}$ introduced by combiners $\{\mathbf{W}_q\}_{q=1}^Q$, the global optimal solution to (7) is hard to obtain. To address this issue, we adopt a greedy method to generate $\{\mathbf{W}_q\}_{q=1}^Q$ and $\{\mathbf{v}_q\}_{q=1}^Q$ in a pilot-by-pilot manner. Specifically, we define $\bar{\mathbf{X}}_t = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_t]$ as the overall observation matrix for timeslots $1 \sim t$, where $t \leq Q$. Let $\bar{\mathbf{y}}_t = \bar{\mathbf{X}}_t^H \mathbf{h} + \bar{\mathbf{z}}_t$ denote the corresponding received signal, wherein $\bar{\mathbf{y}}_t = [\mathbf{y}_1^T, \mathbf{y}_2^T, \cdots, \mathbf{y}_t^T]^T$ and $\bar{\mathbf{z}}_t := \left[\mathbf{z}_1^H \mathbf{W}_1, \cdots, \mathbf{z}_t^H \mathbf{W}_t\right]^H$. Given $\{\mathbf{W}_q\}_{q=1}^t$ and $\{\mathbf{v}_q\}_{q=1}^t$ in the first t timeslots, our greedy strategy aims to find the combiner $\mathbf{W}_{t+1}$ and the precoder $\mathbf{v}_{t+1}$ in the next timeslot, which maximize the MI increment from timeslot t to $t+1$, i.e.,

$$\max_{\|\mathbf{v}_{t+1}\|^2 = P, \mathbf{W}_{t+1}} \Delta I_{t+1} := I(\bar{\mathbf{y}}_{t+1}; \mathbf{h}) - I(\bar{\mathbf{y}}_t; \mathbf{h}). \tag{8}$$

Algorithm 1 2DIF Based Combiner and Precoder Design

Input: Number of pilots Q, kernel $\mathbf{\Sigma_h}$.
Output: Designed precoders $\{\mathbf{v}_q^{\rm opt}\}_{q=1}^Q$ and combiners $\{\mathbf{W}_q^{\rm opt}\}_{q=1}^Q$.
1: Rewrite kernel as $\mathbf{\Sigma_h} = \mathbf{\Sigma}_{\rm T} \otimes \mathbf{\Sigma}_{\rm R}$
2: Find the eigenvectors $[\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{N_{\rm T}}]$ and the corresponding eigenvalues $[\alpha_1, \alpha_2, \cdots, \alpha_{N_{\rm T}}]$ of $\mathbf{\Sigma}_{\rm T}$
3: Find the eigenvectors $[\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_{N_{\rm R}}]$ and the corresponding eigenvalues $[\beta_1, \beta_2, \cdots, \beta_{N_{\rm R}}]$ of $\mathbf{\Sigma}_{\rm R}$
4: Initialize: $[\lambda_{1,1}^0, \lambda_{1,2}^0, \cdots, \lambda_{N_{\rm T},N_{\rm R}}^0] = [\alpha_1\beta_1, \alpha_1\beta_2, \cdots, \alpha_{N_{\rm T}}\beta_{N_{\rm R}}]$
5: **for** $t = 0, \cdots, Q-1$ **do**
6: Find optimal $n_{\rm T}^{\rm opt}$ and $\{n_{{\rm R},k}^{\rm opt}\}_{k=1}^{N_{\rm RF}}$ via **Algorithm 2**
7: Eigenvector-assignment: $\mathbf{v}_{t+1}^{\rm opt} = \sqrt{P}\mathbf{a}_{n_{\rm T}^{\rm opt}}^*$ and $\mathbf{W}_{t+1}^{\rm opt} = [\mathbf{b}_{n_{{\rm R},1}^{\rm opt}}, \cdots, \mathbf{b}_{n_{{\rm R},N_{\rm RF}}^{\rm opt}}]$
8: Eigenvalue-update for all $n_{\rm T} \in \{1, \cdots, N_{\rm T}\}$ and $n_{\rm R} \in \{1, \cdots, N_{\rm R}\}$ via
9: $\lambda_{n_{\rm T},n_{\rm R}}^{t+1} = \begin{cases} \frac{\lambda_{n_{\rm T},n_{\rm R}}^t \sigma^2}{P\lambda_{n_{\rm T},n_{\rm R}}^t + \sigma^2}, & n_{\rm T} = n_{\rm T}^{\rm opt} \ \&\ n_{\rm R} \in \{n_{{\rm R},k}^{\rm opt}\}_{k=1}^{N_{\rm RF}}, \\ \lambda_{n_{\rm T},n_{\rm R}}^t, & \text{else.} \end{cases}$
10: **end for**
11: **return** Precoders $\{\mathbf{v}_q^{\rm opt}\}_{q=1}^Q$ and combiners $\{\mathbf{W}_q^{\rm opt}\}_{q=1}^Q$.

Particularly, $\Delta I_1 := I(\mathbf{y}_1; \mathbf{h})$. For clarity, we summarize the proposed design strategy in **Algorithm 1**, and the sequential designs of $\{\mathbf{W}_q\}_{q=1}^Q$ and $\{\mathbf{v}_q\}_{q=1}^Q$ are illustrated as follows.

Given the optimal precoder $\{\mathbf{v}_q\}_{q=1}^t$ and combiner $\{\mathbf{W}_q\}_{q=1}^t$, we then consider designing $\mathbf{v}_{t+1}$ and $\mathbf{W}_{t+1}$. As proved in Appendix A in this paper's Supplementary Materials [10], ΔI_{t+1}, can be equivalently rewritten as

$$\Delta I_{t+1} = \log_2 \det(\mathbf{I}_{N_{\rm RF}} + \sigma^{-2}\mathbf{X}_{t+1}^H \mathbf{\Sigma}_t \mathbf{X}_{t+1}), \tag{9}$$

wherein

$$\mathbf{\Sigma}_t = \mathbf{\Sigma_h} - \mathbf{\Sigma_h}\bar{\mathbf{X}}_t \big(\bar{\mathbf{X}}_t^H \mathbf{\Sigma_h} \bar{\mathbf{X}}_t + \sigma^2 \mathbf{I}_{N_{\rm RF}t}\big)^{-1} \bar{\mathbf{X}}_t^H \mathbf{\Sigma_h} \tag{10}$$

is the posterior kernel of channel $\mathbf{h}$ given the observation $\bar{\mathbf{y}}_t$. In particular, we have $\mathbf{\Sigma}_0 := \mathbf{\Sigma_h}$. Since the reformulated problem (9) is still intricate, we seek to simplify it by deriving the following two lemmas.

Lemma 1: In MIMO systems, the covariance of the vectored channel $\mathbf{\Sigma_h}$ can be rewritten as the form of a Kronecker product of two kernels, i.e.,

$$\mathbf{\Sigma_h} = \mathbf{\Sigma}_{\rm T} \otimes \mathbf{\Sigma}_{\rm R}, \tag{11}$$

where $\mathbf{\Sigma}_{\rm T} \in \mathbb{C}^{N_{\rm T} \times N_{\rm T}}$ and $\mathbf{\Sigma}_{\rm R} \in \mathbb{C}^{N_{\rm R} \times N_{\rm R}}$ characterize the channel correlation among transmit antennas and that among receive antennas, respectively.

Proof 1: See Appendix A in Supplementary Materials [10].

Lemma 2: Introducing the orthogonality constraints $\mathbf{W}_q^H \mathbf{W}_q = \mathbf{I}_{N_{\rm RF}}$ for all $q \in \{1, \cdots, Q\}$ does not influence the optimal value of MI $I(\mathbf{y}; \mathbf{h})$ in (7).

Proof 2: See Appendix B in Supplementary Materials [10].

Utilizing **Lemma 1** and **Lemma 2**, the optimal precoder $\mathbf{v}_{t+1}$ and combiner $\mathbf{W}_{t+1}$ at the $(t+1)$-th timeslot can be obtained by solving

$$\begin{aligned} \max_{\mathbf{v}_{t+1}, \mathbf{W}_{t+1}} \quad & \log_2 \det(\mathbf{I}_{N_{\rm RF}} + \sigma^{-2}\mathbf{X}_{t+1}^H \mathbf{\Sigma}_t \mathbf{X}_{t+1}) \\ \text{s.t.} \quad & \|\mathbf{v}_{t+1}\|^2 = P, \\ & \mathbf{W}_{t+1}^H \mathbf{W}_{t+1} = \mathbf{I}_{N_{\rm RF}}, \end{aligned} \tag{12}$$

where $\mathbf{X}_{t+1} := \mathbf{v}_{t+1}^* \otimes \mathbf{W}_{t+1}$ refers to the Kronecker constraint imposed on the observation matrix $\mathbf{X}_{t+1}$.

It is easy to prove that, when $t = 0$, the kernel $\mathbf{\Sigma}_0 := \mathbf{\Sigma_h}$ can be decoupled into $\mathbf{\Sigma}_{\rm T} \otimes \mathbf{\Sigma}_{\rm R}$ such that $\mathbf{v}_1$ and $\mathbf{W}_1$ in $\mathbf{X}_1$ can be obtained by selecting the appropriate eigenvectors of $\mathbf{\Sigma}_{\rm T}$ and $\mathbf{\Sigma}_{\rm R}$ to maximize the MI increment in (12). We attempt to use the similar idea to solve for $\mathbf{v}_{t+1}$ and $\mathbf{W}_{t+1}$. To this end, we first define the EVD: $\mathbf{\Sigma}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H$ where $\mathbf{U}_t \in \mathbb{C}^{N_{\rm T}N_{\rm R} \times N_{\rm T}N_{\rm R}}$. Notice that the constraints $\|\mathbf{v}_{t+1}\|^2 = P$ and $\mathbf{W}_{t+1}^H \mathbf{W}_{t+1} = \mathbf{I}_{N_{\rm RF}}$ make the overall matrix observation $\mathbf{X}_{t+1}$ orthogonal, i.e., $\mathbf{X}_{t+1}^H \mathbf{X}_{t+1} = P\mathbf{I}_{\rm RF}$. If we temporarily omit the Kronecker constraint $\mathbf{X}_{t+1} = \mathbf{v}_{t+1}^* \otimes \mathbf{W}_{t+1}$ and try to solve (12) by considering the orthogonal constraint $\mathbf{X}_{t+1}^H \mathbf{X}_{t+1} = P\mathbf{I}_{\rm RF}$ only, it becomes evident that the global optimal solution to (12) is $\mathbf{X}_{t+1} = \sqrt{P}\mathbf{U}_t(:, [1, \cdots, N_{\rm RF}])$. Motivated by this discovery, a natural question arises: *when the Kronecker constraint holds, is it possible to set $\mathbf{X}_{t+1}$ as the principal eigenvectors of $\mathbf{\Sigma}_t$ by properly designing $\mathbf{v}_{t+1}$ and $\mathbf{W}_{t+1}$?* Addressing this question is crucial, thus we would like to investigate it by analyzing the impacts and feasibility of setting $\mathbf{X}_{t+1}$ as the principal eigenvectors of $\mathbf{\Sigma}_t$.

i) Impacts: We first need to exploit its influence on the evolution rule of the posterior kernel $\mathbf{\Sigma}_{t+1}$. The following lemma characterizes the relationship between $\mathbf{\Sigma}_{t+1}$ and $\mathbf{\Sigma}_t$.

Lemma 3: Let $\lambda_n(\cdot)$ denote the n-th largest eigenvalue of the matrix in its argument, e.g., $\lambda_n(\mathbf{\Sigma}_t) = \mathbf{\Lambda}_t(n, n)$ for $\mathbf{\Sigma}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H$. If $\mathbf{X}_{t+1} = \sqrt{P}\mathbf{U}_t(:, [1, \cdots, N_{\rm RF}])$, then the EVD of $\mathbf{\Sigma}_{t+1}$ can be derived from $\mathbf{\Sigma}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H$ by

$$\begin{aligned} \mathbf{\Sigma}_{t+1} = & \mathbf{U}_t {\rm diag}\bigg(\frac{\lambda_1(\mathbf{\Sigma}_t)\sigma^2}{P\lambda_1(\mathbf{\Sigma}_t) + \sigma^2}, \frac{\lambda_2(\mathbf{\Sigma}_t)\sigma^2}{P\lambda_2(\mathbf{\Sigma}_t) + \sigma^2}, \cdots, \\ & \frac{\lambda_{N_{\rm RF}}(\mathbf{\Sigma}_t)\sigma^2}{P\lambda_{N_{\rm RF}}(\mathbf{\Sigma}_t) + \sigma^2}, \lambda_{N_{\rm RF}+1}(\mathbf{\Sigma}_t), \lambda_{N_{\rm RF}+2}(\mathbf{\Sigma}_t), \\ & \cdots, \lambda_{N_{\rm R}N_{\rm T}}(\mathbf{\Sigma}_t) \bigg) \mathbf{U}_t^H. \end{aligned} \tag{13}$$

Proof 3: See Appendix D in Supplementary Materials [10].

Lemma 3 reveals that, when $\mathbf{X}_{t+1} = \sqrt{P}\mathbf{U}_t(:, [1, \cdots, N_{\rm RF}])$, the posterior covariance matrix $\mathbf{\Sigma}_{t+1}$ shares the same eigenvector space, $\mathbf{U}_t$, as $\mathbf{\Sigma}_t$. The only difference on their EVDs is that the $N_{\rm RF}$-largest eigenvalues of $\mathbf{\Sigma}_t$, i.e., $\{\lambda_n(\mathbf{\Sigma}_t)\}_{n=1}^{N_{\rm RF}}$, are replaced by $\{\frac{\lambda_n(\mathbf{\Sigma}_t)\sigma^2}{P\lambda_n(\mathbf{\Sigma}_t)+\sigma^2}\}_{n=1}^{N_{\rm RF}}$ in $\mathbf{\Sigma}_{t+1}$. Considering the generality of t, we can conclude that, the eigenvectors of channel covariance $\mathbf{\Sigma}_0 := \mathbf{\Sigma_h}$ are preserved by all subsequent posterior kernels $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$, $\cdots$, and $\mathbf{\Sigma}_{Q-1}$. In other words, the only difference among $\mathbf{U}_0$, $\mathbf{U}_1$, $\cdots$, $\mathbf{U}_{Q-1}$ is their column arrangement orders.

ii) Feasibility: Given the eigenvector-preserving property in **Lemma 3**, the feasibility of setting $\mathbf{X}_{t+1}$ as the principal eigenvectors of $\mathbf{\Sigma}_t$ lies in the feasibility of setting $\mathbf{X}_{t+1}$ as the eigenvectors of $\mathbf{\Sigma}_0 = \mathbf{\Sigma_h}$. To assess this feasibility, we examine the structure of the eigenspace of $\mathbf{\Sigma_h}$ below.

Corollary 1: The EVD of the prior covariance $\mathbf{\Sigma_h}$ can be written as

$$\begin{aligned} \mathbf{\Sigma_h} &= \mathbf{U}_0 \mathbf{\Lambda}_0 \mathbf{U}_0^H \\ &= (\mathbf{U}_{\rm T} \otimes \mathbf{U}_{\rm R})(\mathbf{\Lambda}_{\rm T} \otimes \mathbf{\Lambda}_{\rm R})\left(\mathbf{U}_{\rm T}^H \otimes \mathbf{U}_{\rm R}^H\right) \end{aligned}$$

$$= \sum_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}} \sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}} \alpha_{n_{\mathrm{T}}} \beta_{n_{\mathrm{R}}} \left(\mathbf{a}_{n_{\mathrm{T}}} \otimes \mathbf{b}_{n_{\mathrm{R}}}\right) \left(\mathbf{a}_{n_{\mathrm{T}}}^{H} \otimes \mathbf{b}_{n_{\mathrm{R}}}^{H}\right),$$

where $\mathbf{a}_{n_{\mathrm{T}}} := \mathbf{U}_{\mathrm{T}}(:, n_{\mathrm{T}})$ denotes the n_{T}-th eigenvector of $\mathbf{\Sigma}_{\mathrm{T}}$ and $\mathbf{b}_{n_{\mathrm{R}}} := \mathbf{U}_{\mathrm{R}}(:, n_{\mathrm{R}})$ denotes the n_{R}-th eigenvector of $\mathbf{\Sigma}_{\mathrm{R}}$. In particular, $\{\alpha_{n_{\mathrm{T}}} \beta_{n_{\mathrm{R}}}\}_{n_{\mathrm{T}}=1, n_{\mathrm{R}}=1}^{N_{\mathrm{T}}, N_{\mathrm{R}}}$ and $\{\mathbf{a}_{n_{\mathrm{T}}} \otimes \mathbf{b}_{n_{\mathrm{R}}}\}_{n_{\mathrm{T}}=1, n_{\mathrm{R}}=1}^{N_{\mathrm{T}}, N_{\mathrm{R}}}$ are the eigenvalues and the corresponding eigenvectors of $\mathbf{\Sigma}_{\mathbf{h}}$, respectively.

Proof 4: See Appendix E in Supplementary Materials [10]. Recalling that $\mathbf{X}_{t+1} := \mathbf{v}_{t+1}^{*} \otimes \mathbf{W}_{t+1}$, we find that the analytical form of $\mathbf{X}_{t+1}$ perfectly matches that of the eigenvectors of $\mathbf{\Sigma}_{\mathbf{h}}$, i.e., $\{\mathbf{a}_{n_{\mathrm{T}}} \otimes \mathbf{b}_{n_{\mathrm{R}}}\}_{n_{\mathrm{T}}=1, n_{\mathrm{R}}=1}^{N_{\mathrm{T}}, N_{\mathrm{R}}}$. This encouraging fact inspires us that, the desired optimal $\mathbf{X}_{t+1}$ to solve (12) can be achieved by setting $\mathbf{v}_{t+1}$ and $\mathbf{W}_{t+1}$ to the appropriate eigenvectors from $\left\{\sqrt{P}\mathbf{a}_{n_{\mathrm{T}}}^{*}\right\}_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}}$ and $\{\mathbf{b}_{n_{\mathrm{R}}}\}_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}$, respectively. Analogously, all our desired $\{\mathbf{X}_q\}_{q=1}^{Q}$ can be obtained by this successive process. As a result, the problem is transformed into: *How to select appropriate eigenvectors such that the objective function in problem (12) is maximized?*

B. Eigenvector Selection

In this subsection, the problem of eigenvector selection is investigated. According to **Lemma 3** and **Corollary 1**, all posterior kernels $\{\mathbf{\Sigma}_t\}_{t=0}^{Q-1}$ can be rewritten as the form of

$$\mathbf{\Sigma}_t = \sum_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}} \sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}} \lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t} \left(\mathbf{a}_{n_{\mathrm{T}}} \otimes \mathbf{b}_{n_{\mathrm{R}}}\right) \left(\mathbf{a}_{n_{\mathrm{T}}}^{H} \otimes \mathbf{b}_{n_{\mathrm{R}}}^{H}\right), \quad (14)$$

where $\lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t}$ is the $(n_{\mathrm{T}}, n_{\mathrm{R}})$-th eigenvalue of $\mathbf{\Sigma}_t$ associated with the eigenvector $\mathbf{a}_{n_{\mathrm{T}}} \otimes \mathbf{b}_{n_{\mathrm{R}}}$. In particular, $\lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{0} = \alpha_{n_{\mathrm{T}}} \beta_{n_{\mathrm{R}}}$, and its update from t to $t+1$ is expressed by

$$\lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t+1} = \begin{cases} \dfrac{\lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t} \sigma^2}{P \lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t} + \sigma^2}, & \mathbf{a}_{n_{\mathrm{T}}}^{*} \otimes \mathbf{b}_{n_{\mathrm{R}}} \text{ is selected,} \\ \lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t}, & \text{else.} \end{cases} \quad (15)$$

Based on the above derivation, we prove the following lemma to further simplify the original problem (12):

Lemma 4: When $\mathbf{v}_{t+1} \in \left\{\sqrt{P}\mathbf{a}_{n_{\mathrm{T}}}^{*}\right\}_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}}$ and the columns of $\mathbf{W}_{t+1}$ are belonging to $\{\mathbf{b}_{n_{\mathrm{R}}}\}_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}$, problem (12) can be transformed into an eigenvector selection problem, written as

$$\begin{aligned} \max_{n_{\mathrm{T}}, \{n_{\mathrm{R},k}\}_{k=1}^{N_{\mathrm{RF}}}} \quad & \sum_{k=1}^{N_{\mathrm{RF}}} \log_2(1 + P \lambda_{n_{\mathrm{T}}, n_{\mathrm{R},k}}^{t} \sigma^{-2}) \\ \text{s.t.} \quad & n_{\mathrm{T}} \in \{1, \cdots, N_{\mathrm{T}}\}, \\ & n_{\mathrm{R},k} \in \{1, \cdots, N_{\mathrm{R}}\}, \forall k, \\ & n_{\mathrm{R},k} \neq n_{\mathrm{R},k'}, \forall k \neq k'. \end{aligned} \quad (16)$$

Proof 5: See Appendix F in Supplementary Materials [10].

Problem (16) aims to find one eigenvector index n_T on the transmitter side and N_{RF} different eigenvector indices $\{n_{\mathrm{R},k}\}_{k=1}^{N_{\mathrm{RF}}}$ on the receiver side, such that the selected eigenvalues $\{\lambda_{n_{\mathrm{T}}, n_{\mathrm{R},k}}^{t}\}_{k=1}^{N_{\mathrm{RF}}}$ can maximize the objective (16). A linear search algorithm is proposed to solve (16) optimally, as summarized in **Algorithm 2**. The key idea is to fix an n_{T} and then find N_{RF}-largest values from $\{\lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t}\}_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}$ to calculate the objective $\sum_{k=1}^{N_{\mathrm{RF}}} \log_2(1 + P \lambda_{n_{\mathrm{T}}, n_{\mathrm{R},k}}^{t} / \sigma^2)$. After traversing all $n_{\mathrm{T}} \in \{1, \cdots, N_{\mathrm{T}}\}$, the optimal $n_{\mathrm{T}}^{\mathrm{opt}}$ and $\{n_{\mathrm{R},k}^{\mathrm{opt}}\}_{k=1}^{N_{\mathrm{RF}}}$ can be obtained from the indices of the maximum objective.

Algorithm 2 Linear Search for Eigenvalue Selection

Input: Eigenvalues $\{\lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t}\}_{n_{\mathrm{T}}=1, n_{\mathrm{R}}=1}^{N_{\mathrm{T}}, N_{\mathrm{R}}}$ in timeslot t.
Output: Optimal eigenvalue indices $n_{\mathrm{T}}^{\mathrm{opt}}$ and $\{n_{\mathrm{R},k}^{\mathrm{opt}}\}_{k=1}^{N_{\mathrm{RF}}}$ that maximize $\sum_{k=1}^{N_{\mathrm{RF}}} \log_2(1 + \lambda_{n_{\mathrm{T}}, n_{\mathrm{R},k}}^{t} / \sigma^2)$.
Initialize indexes: $n_{\mathrm{T}}^{\mathrm{opt}} = 1$ and $[n_{\mathrm{R},1}^{\mathrm{opt}}, \cdots, n_{\mathrm{R},N_{\mathrm{RF}}}^{\mathrm{opt}}] = [1, \cdots, N_{\mathrm{RF}}]$.
Initialize the maximum objective: $\zeta_{\max} = \sum_{k=1}^{N_{\mathrm{RF}}} \log_2(1 + P \lambda_{n_{\mathrm{T}}^{\mathrm{opt}}, n_{\mathrm{R},k}^{\mathrm{opt}}}^{t} / \sigma^2)$
for $n_{\mathrm{T}} = 1, \cdots, N_{\mathrm{T}}$ **do**
  Find the N_{RF}-largest values from $\{\lambda_{n_{\mathrm{T}}, n_{\mathrm{R}}}^{t}\}_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}$, and then denote their second indexes as $\{n_{\mathrm{R},k}\}_{k=1}^{N_{\mathrm{RF}}}$.
  if $\sum_{k=1}^{N_{\mathrm{RF}}} \log_2(1 + P \lambda_{n_{\mathrm{T}}, n_{\mathrm{R},k}}^{t} / \sigma^2) > \zeta_{\max}$ **then**
    Update the optimal indexes by $n_{\mathrm{T}}^{\mathrm{opt}} = n_{\mathrm{T}}$ and $[n_{\mathrm{R},1}^{\mathrm{opt}}, \cdots, n_{\mathrm{R},N_{\mathrm{RF}}}^{\mathrm{opt}}] = [n_{\mathrm{R},1}, \cdots, n_{\mathrm{R},N_{\mathrm{RF}}}]$
    Update the maximum objective: $\zeta_{\max} = \sum_{k=1}^{N_{\mathrm{RF}}} \log_2(1 + P \lambda_{n_{\mathrm{T}}^{\mathrm{opt}}, n_{\mathrm{R},k}^{\mathrm{opt}}}^{t} / \sigma^2)$
  end if
end for
return Optimal $n_{\mathrm{T}}^{\mathrm{opt}}$ and $\{n_{\mathrm{R},k}^{\mathrm{opt}}\}_{k=1}^{N_{\mathrm{RF}}}$.

Finally, the optimal precoder and the optimal combiner are thereby expressed as

$$\mathbf{v}_{t+1}^{\mathrm{opt}} = \sqrt{P}\mathbf{a}_{n_{\mathrm{T}}^{\mathrm{opt}}}^{*} \text{ and } \mathbf{W}_{t+1}^{\mathrm{opt}} = \left[\mathbf{b}_{n_{\mathrm{R},1}^{\mathrm{opt}}}, \cdots, \mathbf{b}_{n_{\mathrm{R},N_{\mathrm{RF}}}^{\mathrm{opt}}}\right], \quad (17)$$

respectively, which generate a feasible observation matrix $\mathbf{X}_{t+1}^{\mathrm{opt}} = \mathbf{v}_{t+1}^{\mathrm{opt}\,*} \otimes \mathbf{W}_{t+1}^{\mathrm{opt}}$ at timeslot $t+1$.

To summarize, the eigenvalue updating rule in (15), as well as the eigenvector selection method stated in **Algorithm 2** and (17), allow us to calculate all observation matrices.

C. Insightful Interpretation to 2DIF

In this subsection, we provide insightful explanations to the proposed 2DIF algorithm to clarify its physical significance.

1) Ideal water-filling: To better understand the proposed observation matrix design, we first interpret problem (7) from the view of *water-filling*. Specifically, the orthogonal property of $\mathbf{W}_q$ proved in **Lemma 2** allows us to replace the noise covariance matrix $\mathbf{\Xi}$ in (7) by $\sigma^2 \mathbf{I}_{N_{\mathrm{RF}}Q}$ without affecting the optimal $I(\mathbf{y}; \mathbf{h})$. Then, by further relaxing the constraints $\mathbf{W}_q^T \mathbf{W}_q = \mathbf{I}_{N_{\mathrm{RF}}}$ and $\|\mathbf{v}_q\|^2 = P$, we focus only on the total power constraint imposed on the overall observation matrix $\mathbf{X}$, i.e., $\mathrm{Tr}\left(\mathbf{X}\mathbf{X}^H\right) = P N_{\mathrm{RF}} Q$. In this case, the optimal value of problem (7) is shown to have an upper bound:

$$\begin{aligned} \max_{\mathbf{X}} \quad & \log_2 \det(\mathbf{I}_{N_{\mathrm{RF}}Q} + \sigma^{-2} \mathbf{X}^H \mathbf{\Sigma}_{\mathbf{h}} \mathbf{X}) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{X}\mathbf{X}^H) = P N_{\mathrm{RF}} Q. \end{aligned} \quad (18)$$

Notably, this upper bound is equivalent to the channel capacity of a point-to-point MIMO system equipped with $N_{\mathrm{T}} N_{\mathrm{R}}$ transmit antennas and $N_{\mathrm{RF}} Q$ receive antennas. Thereafter, the overall observation matrix can be optimally solved as $\mathbf{X}^{\mathrm{ideal}} = \mathbf{U}_0(:, [1, \cdots, N_{\mathrm{RF}} Q])\, \mathbf{P}$,where $\mathbf{P} = \mathrm{diag}\left(\sqrt{p_1}, \cdots, \sqrt{p_{N_{\mathrm{RF}}Q}}\right)$ is the power allocation matrix. The power allocated to the n-th eigenvector is determined by the water-filling principle, i.e., $p_n = (\beta - \sigma^{-2} \lambda_n(\mathbf{\Sigma}_{\mathbf{h}}))^{+}$, where the water-level β is adjusted to satisfy the total power constraint $\mathrm{Tr}\left(\mathbf{X}^{\mathrm{ideal}} (\mathbf{X}^{\mathrm{ideal}})^H\right) = \sum_{n=1}^{N_{\mathrm{RF}}Q} p_n = P N_{\mathrm{RF}} Q$.

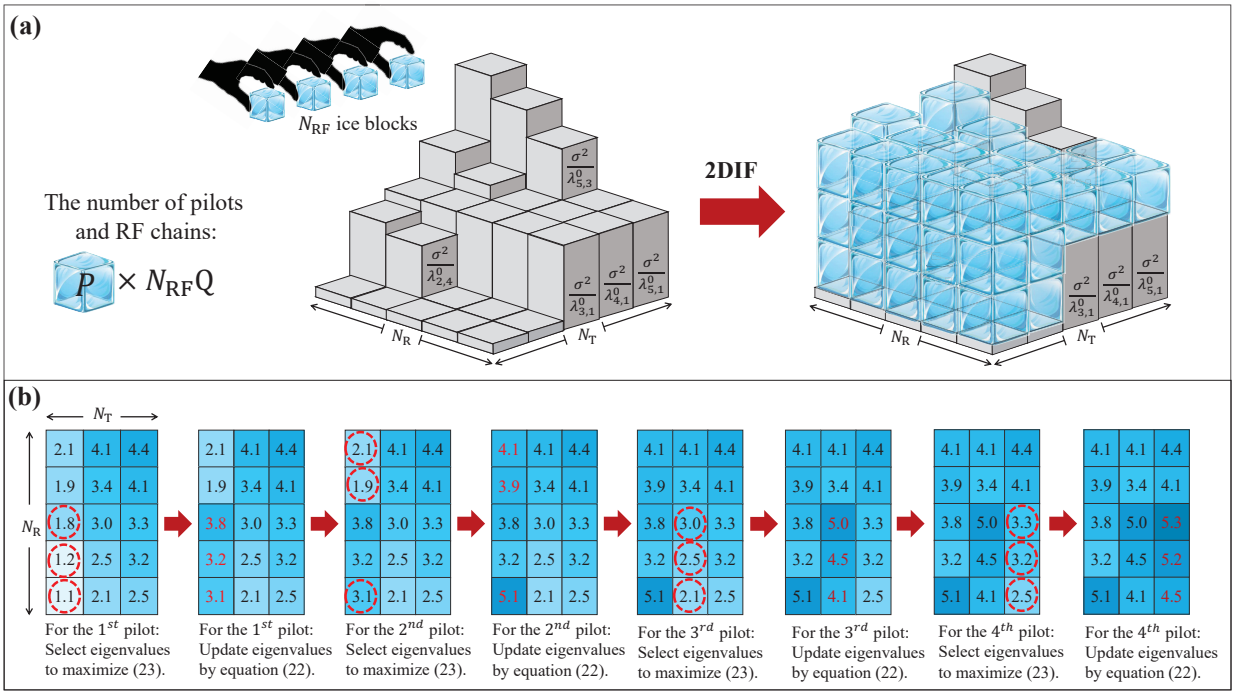

Fig. 1. (a) shows how the pilot allocation of the proposed 2DIF algorithm works to maximize the MI. (b) provides an example to show the implementation process of the proposed 2DIF, where $N_{\mathrm{R}} = 5$, $N_{\mathrm{T}} = 3$, $N_{\mathrm{RF}} = 3$, $P = 2$, and $Q = 4$. The number within the $(n_{\mathrm{T}}, n_{\mathrm{R}})$-th square is the ice-level $\frac{\sigma^2}{\lambda^t_{n_{\mathrm{T}},n_{\mathrm{R}}}}$.

Although the ideal observation matrix $\mathbf{X}^{\mathrm{ideal}}$, that maximizes the upper bound (18), might not be implementable in practice (as $\mathbf{X}^{\mathrm{ideal}}$ may violate the constraints $\mathbf{W}_q^T\mathbf{W}_q = \mathbf{I}_{N_{\mathrm{RF}}}$ and $\|\mathbf{v}_q\|^2 = P$), it can give us two pivotal intuitions. First, the observation matrix should align with the eigenspace $\{\mathbf{a}_{n_{\mathrm{T}}} \otimes \mathbf{b}_{n_{\mathrm{R}}}\}_{n_{\mathrm{T}}=1,n_{\mathrm{R}}=1}^{N_{\mathrm{T}},N_{\mathrm{R}}}$ of the full MIMO channel covariance, $\mathbf{\Sigma_h} = \mathbf{\Sigma}_{\mathrm{T}} \otimes \mathbf{\Sigma}_{\mathrm{R}}$. Second, we need to fill in more power (water) to the eigenvectors having larger eigenvalues $\{\lambda_n(\mathbf{\Sigma_h})\}_{n=1}^{N_{\mathrm{RF}}Q}$ (or equivalently lower base levels $\{\frac{\sigma^2}{\lambda_n(\mathbf{\Sigma_h})}\}_{n=1}^{N_{\mathrm{RF}}Q}$).

2) 2DIF versus water-filling: The proposed 2DIF algorithm materializes the above two intuitions under the practical constraints $\mathbf{W}_q^T\mathbf{W}_q = \mathbf{I}_{N_{\mathrm{RF}}}$ and $\|\mathbf{v}_q\|^2 = P$ via the eigenvector selection process in (16). The first intuition is automatically achieved by assigning $\mathbf{v}_{t+1}$ with a eigenvector from $\{\sqrt{P}\mathbf{a}^*_{n_{\mathrm{T}}}\}_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}}$ and assigning the columns of $\mathbf{W}_{t+1}$ with different eigenvectors from $\{\mathbf{b}_{n_{\mathrm{R}}}\}_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}$, as proved in **Lemma 4**. The second intuition is approximately accomplished via selecting eigenvectors that have lower base levels, $\{\frac{\sigma^2}{\lambda_n(\mathbf{\Sigma_h})}\}$, by more times. This is attributed to the fact that the maximization of (16) tends to select an eigenvalue combination $\{\lambda^t_{n_{\mathrm{T}},n_{\mathrm{R},k}}\}_{k=1}^{N_{\mathrm{RF}}}$ that has the lowest $\{\frac{\sigma^2}{\lambda^t_{n_{\mathrm{T}},n_{\mathrm{R},k}}}\}_{k=1}^{N_{\mathrm{RF}}}$ on average. To see this approximation more clearly, we rewrite the updating rule of the selected eigenvalue in (15) as

$$\underbrace{\frac{\sigma^2}{\lambda^{t+1}_{n_{\mathrm{T}},n_{\mathrm{R}}}}}_{\text{Updated ice-level}} = \underbrace{\frac{\sigma^2}{\lambda^{t}_{n_{\mathrm{T}},n_{\mathrm{R}}}}}_{\text{Current ice-level}} + \underbrace{P}_{\text{Height of an ice block}}. \tag{19}$$

(19) reveals that every time the eigenvector $\mathbf{a}_{n_{\mathrm{T}}} \otimes \mathbf{b}_{n_{\mathrm{R}}}$ is selected, the value of $\frac{\sigma^2}{\lambda^t_{n_{\mathrm{T}},n_{\mathrm{R}}}}$ increases by P. Similar to the water-filling process, the process described by (19) can be vividly interpreted as allocating an ice block having P-unit power to the $(n_{\mathrm{T}}, n_{\mathrm{R}})$-th orthogonal channel, where $\frac{\sigma^2}{\lambda^t_{n_{\mathrm{T}},n_{\mathrm{R}}}}$ is viewed as the ice level in the t-th timeslot. Due to the consideration of N_{RF} RF chains and $N_{\mathrm{T}} \times N_{\mathrm{R}}$ MIMO systems, in each timeslot, the 2DIF first selects N_{RF} orthogonal channels, which have the deepest ice-levels $\{\frac{\sigma^2}{\lambda^t_{n_{\mathrm{T}},n_{\mathrm{R},k}}}\}_{k=1}^{N_{\mathrm{RF}}}$ on average, from the total $N_{\mathrm{T}} \times N_{\mathrm{R}}$ channels. Then, the 2DIF will fill N_{RF} ice blocks (i.e., N_{RF} pilots) of height P onto them. As illustrated in Fig. 1 (a), after Q timeslots, the final ice-levels of all channels can have a similar height with the water-level β. In this case, the second intuition is approximately achieved.

IV. Simulation Results

We consider the channel estimation in a single-user densifying MIMO system. The 3GPP TR 38.901 channel model is utilized for simulations [8]. In specific, the carrier frequency is set to 3.5 GHz. The number of clusters is 23, and each cluster contributes 20 rays. The AoAs and the AoDs are randomly selected from $\mathcal{U}(-90°, +90°)$. For each ray in a cluster, the angle spread and the delay spread are randomly selected from $\mathcal{U}(-5°, +5°)$ and $\mathcal{U}(-30\text{ ns}, +30\text{ ns})$, respectively. The path gains are randomly selected from $\mathcal{CN}(0, 1)$. Otherwise specifically specified, the system parameters are set as: $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 64$, $N_{\mathrm{RF}} = 4$, and $Q = 48$, respectively. The antenna spacing is set to be $d = \frac{\lambda}{8}$. The signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = \frac{P}{\sigma^2}\mathsf{E}(\|\mathbf{h}\|^2)$, whose default value is set to 10 dB. The evaluation criterion is the normalized mean square error (NMSE), which is expressed as $\mathrm{NMSE} = \mathsf{E}(\frac{\|\mathbf{h}-\hat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2})$.

The following schemes are simulated for comparison: **1) LS:** The LS channel estimator is feasible only when $QN_{\mathrm{RF}} \geq N_{\mathrm{T}}N_{\mathrm{R}}$. We assume the pilot length is $Q = \lceil N_{\mathrm{T}}N_{\mathrm{R}}/N_{\mathrm{RF}} \rceil = 64$, and all combiners/precoders are generated from discrete Fourier transform (DFT) matrices. **2) MMSE**: Under the same

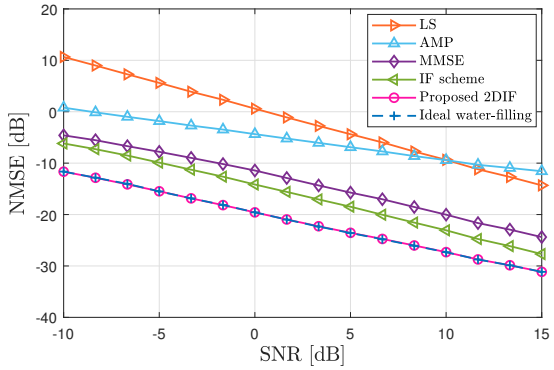

Fig. 2. The NMSE as a function of SNR for different schemes.

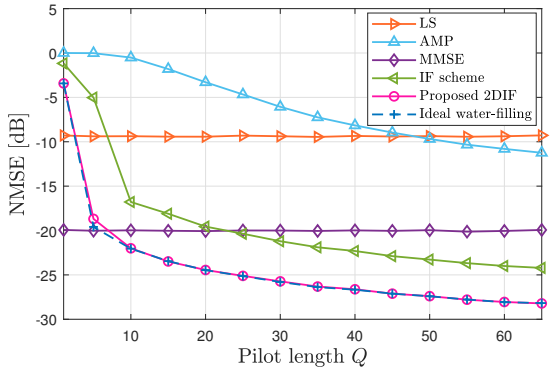

Fig. 3. The NMSE as a function of pilot length Q for different schemes.

setting of LS estimator, the MMSE estimator is implemented to recover channel via (5). **3) AMP**: The approximate message passing (AMP) method proposed in [11] is implemented to estimate channel. **4) IF scheme**: By viewing the considered MIMO system as N_{T} independent SIMO systems, the IF-based channel estimator [7] can be utilized to recover $\mathbf{H}$ in a column-by-column way. Note that, since IF scheme is only applicable to the single-RF-chain case, the pilot length used should be modified as $\lceil QN_{\mathrm{RF}} \rceil = 192$. **5) Proposed 2DIF**: The proposed 2DIF method in **Algorithm 1** is employed to design the precoders and combiners. Then, the estimator in (5) is employed to estimate channel. **6) Ideal water-filling**: The ideal (but may not be practically achievable) observation matrices $\{\mathbf{X}_q\}_{q=1}^{Q}$ are directly obtained by solving (18) via water-filling method. Then, (5) is employed to recover channel.

Firstly, we plot the NMSE as a function of SNR in Fig. 2. One can observe that, thanks to the carefully designed observation matrices/vectors, the proposed 2DIF scheme remarkably outperforms the benchmark schemes. The reason is that the proposed method fully exploits the correlations among the transceiver antennas for channel estimation. In particular, the NMSE for the proposed 2DIF scheme is about 5 dB lower than that for the IF scheme. It is because the IF schemes realizes the MIMO channel estimation by viewing it as N_{T} independent SIMO channel estimations, which ignores the spatial correlation of transmitter antennas. Besides, we note that the proposed 2DIF scheme achieves very similar performance to the ideal water-filling scheme. Then, the NMSE versus the number of pilots Q is provided in Fig. 3. One can find that the superiority of the proposed scheme still holds. Although the dimension of the estimated parameters is high as $N_{\mathrm{T}}N_{\mathrm{R}} = 256$, using a small number of pilots $Q = 20$, the NMSE for the proposed scheme can be lower than -20 dB. In contrast, even if the pilot length is longer than $Q = 60$, the conventional AMP estimator is still unable to achieve such high accuracy. It indicates that utilizing the correlation of compact antennas is of great significance for high-accuracy channel estimation.

V. Conclusions

By fully exploiting the channel correlations, this work has proposed a 2DIF based observation matrix design for channel estimation in densifying MIMO systems. The core idea is to jointly design precoders and combiners by maximizing the MI between channels and pilots. Particularly, we have proved that the eigenspace of the channel covariance can be decoupled into two sub-eigenspaces, and the near-optimal observation matrix can be obtained from these two sub-eigenspaces. We have also revealed that, similar to the water-filling precoding that maximizes the MIMO capacity, the implementation of 2DIF can be viewed as a two-dimensional ice-filling process. Simulation results have validated the superiority of our proposed methods compared to the conventional channel estimators.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 624B2123.

References

[1] D. González-Ovejero, G. Minatti, G. Chattopadhyay, and S. Maci, "Multibeam by metasurface antennas," *IEEE Trans. Antennas Propag.*, vol. 65, no. 6, pp. 2923–2930, Jun. 2017.

[2] M. Liu *et al.*, "Deeply subwavelength metasurface resonators for terahertz wavefront manipulation," *Adv. Opt. Mater.*, vol. 7, no. 21, p. 1900736, 2019.

[3] Z. Zhang and L. Dai, "Pattern-division multiplexing for multi-user continuous-aperture MIMO," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2350–2366, Aug. 2023.

[4] M. Di Renzo, D. Dardari, and N. Decarli, "LoS MIMO-arrays vs. LoS MIMO-surfaces," in *Proc. 17th European Conference on Antennas and Propagation (EuCAP'23)*, 2023, pp. 1–5.

[5] M. Akrout, V. Shyianov, F. Bellili, A. Mezghani, and R. W. Heath, "Super-wideband massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2414–2430, Aug. 2023.

[6] Y. Liu, M. Zhang, T. Wang, A. Zhang, and M. Debbah, "Densifying MIMO: Channel modeling, physical constraints, and performance evaluation for holographic communications," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 6, pp. 1504–1518, Jun. 2024.

[7] M. Cui, Z. Zhang, L. Dai, and K. Huang, "Ice-filling: Near-optimal channel estimation for dense array systems," *arXiv preprint arXiv:2404.06806*, Apr. 2024.

[8] 3GPP TR, "Study on channel model for frequencies from 0.5 to 100 GHz," *3GPP TR 38.901 version 14.0.0 Release*, Dec. 2019.

[9] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Mar. 2016.

[10] Z. Zhang and M. Cui, "Supplementary materials for 'Two-dimensional ice filling based channel estimation in densifying MIMO systems'," Oct. 2024. [Online]. Available: https://zhangzij15.github.io/pubs/Supp/ICC_supp.pdf

[11] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.