2025 IEEE International Conference on Communications Workshops (ICC Workshops) | 979-8-3315-9624-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCWorkshops67674.2025.11162249

Channel Estimation for Dense Array Systems via Ice-Filling

Mingyao Cui*, Zijian Zhang†, Linglong Dai†, *Fellow*, IEEE, and Kaibin Huang*, *Fellow*, IEEE
*Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR
†Department of Electronic Engineering, Tsinghua University, Beijing, China
Emails: cuimy23@connect.hku.hk; zhangzj20@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn; huangkb@eee.hku.hk

***Abstract*—By deploying a large number of antennas with sub-half-wavelength spacing in a compact space, dense array systems (DASs) can fully unleash the multiplexing and diversity gains of limited apertures. To acquire these gains, accurate channel estimation is necessary but challenging due to the large antenna numbers. To overcome this obstacle, this paper reveals that a well-designed observation matrix, that exploits the high spatial correlation of DAS channels, is crucial for realizing near-optimal channel estimation. To materialize this vision, we propose to design the observation matrix by maximizing the mutual information between received pilots and channels. An ice-filling algorithm is proposed to solve the formulated problem. Its key idea is to properly assign the observation vectors with the eigenvectors of the channel covariance matrix in a greedy manner. Theoretical analysis reveals that the proposed algorithm is equivalently a quantized version of the water-filling principle, which guarantees its near-optimal channel estimation performance. Finally, numerical results confirm that our proposed designs outperform existing channel estimation methods significantly.**

***Index Terms*—Estimation theory, mutual-information maximization, dense array systems (DASs), observation matrix design.**

I. Introduction

In recent years, dense array systems (DASs) have attracted an increasing attention from the wireless communication community [1]. Different from a conventional antenna array with a half-wavelength antenna spacing $\lambda/2$, DASs deploy massive antennas with sub-wavelength (e.g., $\lambda/8$) spacing in a compact area. With a nearly continuous antenna arrangement, DASs promise to realize the ultimate control of electromagnetic waves, which can fully unleash the multiplexing and diversity gains of limited apertures [1]. Aligned with this vision, many dense-antenna transceiver architectures have emerged, such as continuous-aperture arrays, reconfigurable intelligent surfaces, and movable antennas [2], [3].

The full exploitation of the gain of DASs relies on the accurate estimation of channel state information (CSI), particularly when the number of radio frequency (RF) chains is much smaller than the number of antennas in DASs. Up to now, many estimators have been proposed to acquire the CSI of large antenna arrays [4]–[8]. For example, when the available pilot length is larger than the antenna number, the classical non-parametric algorithms can be used, such as the least square (LS) estimator. Leveraging the property of channel sparsity, compressed sensing (CS)-based channel estimators are studied to improve the estimation accuracy and reduce the pilot overhead, such as the orthogonal matching pursuit (OMP)-based estimator [4], the message passing (MP)-based estimator [5], and the gridless sparse signal reconstructor [6]. By tuning parameters using a large amount of channel data, the deep learning approaches are also utilized to realize data-driven and model-driven channel estimators [7]. Besides, beam alignment techniques [8], including beam sweeping and hierarchical beam training, have also been widely explored to acquire the implicit CSI with low pilot overhead.

Although existing channel estimators [4]–[8] can be directly adopted in DASs, they fail to fully exploit the high spatial correlation of DAS channels, which are far from optimal. Specifically, the observation matrices of existing schemes, used for receiving pilots, are either randomly generated or set as predefined codebooks, such as the Fourier matrix and identity matrix. However, the extremely-dense deployment of DAS antennas significantly increases the similarity of radio waves impinging on antenna ports [1]. This similarity makes the covariance matrices of DAS channels no longer diagonal but highly structured. It is believed that properly manipulating the observation matrices according to these structured covariance matrices could remarkably boost the channel estimation accuracy in DASs [9], which motivates our work.

To achieve this goal, we propose to design the DAS's observation matrix by maximizing the mutual information (MI) between the received pilots and channels. An ice-filling algorithm is proposed to address the formulated problem. It sequentially generates each column of the observation matrix by maximizing MI increment between two adjacent pilot transmissions. An important insight is that the ice-filling algorithm works like a quantized version of the well-known water-filling principle. Specifically, the generated observation vectors are selected from the eigenspace of channel covariance matrix. Moreover, *the number of times that one eigenvector of the channel covariance is assigned to the observation matrix* by ice-filling is shown to be *the quantization of the continuous power allocated by water-filling*. We prove that the quantization error is smaller than one, which ensures the near-optimality of the proposed ice-filling algorithm. Numerical simulations demonstrate that our proposed IF algorithm can outperform existing channel estimation methods significantly.

II. System Model

As illustrated in Fig. 1, we consider the narrowband channel estimation of an uplink single-input multiple-output (SIMO)

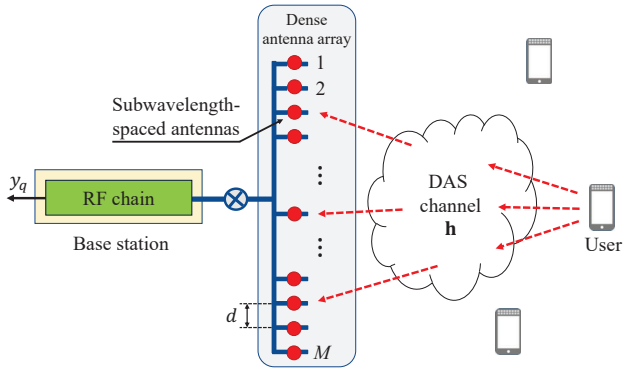


Fig. 1. An illustration of uplink channel estimation for a DAS.

system. The base station (BS) employs an M-antenna DAS, which comprises an analog combining structure supported by one RF chain, to receive the pilots from a single-antenna user [4]. Let $\mathbf{h} \in \mathbb{C}^{M\times 1}$ be the channel vector and Q the number of transmit pilots within a coherent time frame. The received signal $y_q \in \mathbb{C}$ at the BS in timeslot q is modeled as

$$y_q = \mathbf{w}_q^H \mathbf{h} s_q + \mathbf{w}_q^H \mathbf{z}_q, \tag{1}$$

where $\mathbf{w}_q \in \mathbb{C}^{M\times 1}$ is the observation vector at the BS, s_q the pilot transmitted by the user, and $\mathbf{z}_q \sim \mathcal{CN}\big(\mathbf{0}_M, \sigma^2\mathbf{I}_M\big)$ the additive white Gaussian noise. As the observation vector $\mathbf{w}_q$ affects both the desired signal $\mathbf{h}s_q$ and the noise $\mathbf{z}_q$, its power has no impact on the estimation accuracy. Therefore, without any loss of generality, we can normalize $\mathbf{w}_q$ to $\|\mathbf{w}_q\|_2^2 = 1$, which reshapes the noise distribution to $\mathbf{w}_q^H\mathbf{z}_q \sim \mathcal{CN}\big(0, \sigma^2\big)$. Considering the total Q-timeslot pilot transmission, the received signal vector $\mathbf{y} := [y_1, \cdots, y_Q]^{\mathrm{T}}$ is expressed as

$$\mathbf{y} = \mathbf{W}^H\mathbf{h} + \mathbf{z}, \tag{2}$$

where s_q is assumed to be 1 for all $q \in \{1, \cdots, Q\}$, $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_Q]$ stands for the observation matrix, and $\mathbf{z} := \big[\mathbf{w}_1^H\mathbf{z}_1, \cdots, \mathbf{w}_Q^H\mathbf{z}_Q\big]^{\mathrm{T}}$. As the power of $\mathbf{w}_q$ is normalized to 1, we have $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_Q, \sigma^2\mathbf{I}_Q)$.

This article adopts the classical Bayesian regression to recover $\mathbf{h}$ [9]. Suppose the channel follows a Gaussian process $\mathcal{CN}(\mathbf{0}_M, \mathbf{\Sigma_h})$, where $\mathbf{\Sigma_h} \in \mathbb{C}^{M\times M}$ characterizes the channel covariance. Then, the estimated channel, $\hat{\mathbf{h}}$, is determined by the posterior mean of $\mathbf{h}$ given observation $\mathbf{y}$:

$$\boldsymbol{\mu}_{\mathbf{h}|\mathbf{y}} = \mathbf{\Sigma_h}\mathbf{W}\big(\mathbf{W}^H\mathbf{\Sigma_h}\mathbf{W} + \sigma^2\mathbf{I}_Q\big)^{-1}\mathbf{y}, \tag{3}$$

which is also known as the minimum-mean-square-error (MMSE) estimator. The posterior covariance matrix is

$$\mathbf{\Sigma}_{\mathbf{h}|\mathbf{y}} = \mathbf{\Sigma_h} - \mathbf{\Sigma_h}\mathbf{W}\big(\mathbf{W}^H\mathbf{\Sigma_h}\mathbf{W} + \sigma^2\mathbf{I}_Q\big)^{-1}\mathbf{W}^H\mathbf{\Sigma_h}. \tag{4}$$

Notably, the posterier covariance, characterizing the channel estimation error, is a function of the prior covariance $\mathbf{\Sigma_h}$ and the observation matrix $\mathbf{W}$. Due to the extremely-high spatial correlation exhibited by DAS channels, the prior covariance $\mathbf{\Sigma_h}$ would remarkably deviate from the identity matrix. This deviation indicates that aligning the observation matrix $\mathbf{W}$ with the subspace of $\mathbf{\Sigma_h}$ could be greatly helpful in reducing the channel estimation error. Motivated by this fact, this work concentrates on the design of observation matrix $\mathbf{W}$

III. Mutual Information Maximization for Channel Estimation in DASs

In this section, we first formulate the observation matrix design problem from the view of MI maximization. Then, the water-filling principle is investigated to show the upper bound of the formulated problem.

A. Problem Formulation

We adopt the MI between the received signal $\mathbf{y}$ and the channel $\mathbf{h}$ to evaluate the quality of the designed observation matrix. There are two reasons for using the MI as a metric. First, MI tells the amount of information about the unknown channel $\mathbf{h}$ we can gain from the received signal $\mathbf{y}$ [10]. The second reason is that, according to the information-theoretic properties of Bayesian statistics [11], the maximization of MI is asymptotically the minimization of Cramer Rao Bound.

Given the distributions $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{\Sigma_h})$ and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_Q, \sigma^2\mathbf{I}_Q)$, the MI is formulated as

$$\max_{\mathbf{W}\in\mathcal{W}} I(\mathbf{y};\mathbf{h}) = \log\det\left(\mathbf{I}_Q + \frac{1}{\sigma^2}\mathbf{W}^H\mathbf{\Sigma_h}\mathbf{W}\right), \tag{5}$$

where $I(\cdot;\cdot)$ denotes the MI, $\det(\cdot)$ is the determinant of its argument, and the feasible set of $\mathbf{W}$ is represented as $\mathcal{W} = \{\mathbf{W} : \|\mathbf{w}_q\| = 1, \forall q\}$. Notice that we mainly focus on the case where both the amplitude and phase of the coefficients of $\mathbf{w}_q$ are adjustable. Extension to scenarios where only the phase of $\mathbf{w}_q$ is tunnable will be briefly discussed in Section IV-C.

The observation matrix design problem in (5) resembles the point-to-point multiple-input-multiple-output (MIMO) precoding problem. Their major difference lies in the constraints imposed on $\mathbf{W}$. For MIMO precoding, each column of $\mathbf{W}$ refers to a precoding vector associated with one dedicated RF chain. The optimization of precoding vectors, leveraging multiple RF chains, typically enforces the *total-power-constraint* on $\mathbf{W}$, i.e., $\|\mathbf{W}\|_F^2 = Q$. In terms of the observation matrix design, each column of $\mathbf{W}$ signifies an observation vector used for receiving one pilot signal. As discussed in Section II, since each observation vector $\mathbf{w}_q$ amplifies both the desired signal and the noise, it is subject to the *pilot-wise power constraint*, i.e., $\|\mathbf{w}_q\|_2^2 = 1$. The distinction in the observation matrix constraint differentiates the discussed problem in (5) from the classical MIMO precoding.

B. Water-Filling Inspired Ideal Observation Matrix Design

Prior to addressing problem (5), we would like to consider an ideal (but practically unachievable) situation to obtain an upper bound of (5). Specifically, we temporarily make the assumption that the noise vector $\mathbf{z}$ in (2) is independent of the observation matrix $\mathbf{W}$ and its distribution is always $\mathcal{CN}(\mathbf{0}_Q, \sigma^2\mathbf{I}_Q)$ regardless of the power of $\mathbf{w}_q$. Under this assumption, the pilot-wise power constraint, $\|\mathbf{w}_q\|_2^2 = 1$, is relaxed to the total power constraint, $\|\mathbf{W}\|_F^2 = \sum_{q=1}^{Q}\|\mathbf{w}_q\|_2^2 = Q$. In this case, problem (5) is identical to the point-to-point MIMO precoding problem, which can be optimally solved by the water-filling principle.

Let the eigenvalue decomposition (EVD) of $\mathbf{\Sigma_h}$ be $\mathbf{U}_K\mathbf{\Lambda}_K\mathbf{U}_K^H$, where K is the rank of $\mathbf{\Sigma_h}$, $\mathbf{U}_K = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_K]$, and $\mathbf{\Lambda}_K = \text{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_K\}$ with $\lambda_1 \geq \cdots \geq \lambda_K > 0$. The ideal observation matrix $\mathbf{W}^{\text{Ideal}}$ is given as $\mathbf{W}^{\text{Ideal}} = \mathbf{U}_K\mathbf{P}$. Matrix $\mathbf{P} \in \mathbb{R}^{K\times Q}$ represents the power allocation matrix expressed as $\mathbf{P} = [\mathbf{P}_K, \mathbf{0}_{K\times(Q-K)}]$ with $\mathbf{P}_K = \text{diag}\{\sqrt{p_1}, \cdots, \sqrt{p_K}\}$. The power p_k allocated to each eigenvector is calculated by the water-filling principle:

$$p_k = (\beta - \sigma^2/\lambda_k)^+, \forall k \in \{1, 2, \cdots, K\}, \tag{6}$$

where $(x)^+ = \max\{x, 0\}$, and the water-level β is properly selected to meet the power constraint: $\sum_{k=1}^K p_k = Q$. As a result, $\mathbf{W}^{\text{Ideal}}$ offers an upper bound of (5).

IV. Proposed Ice-Filling Algorithm

This section elaborates on the observation matrix design in practical scenarios. First, we propose an ice-filling algorithm to solve problem (5), which is summarized in **Algorithm 1**. Then, its relationship with the water-filling principle is revealed and its near-optimality is proved.

A. Ice-Filling for Observation Matrix Design

In practical uplink channel estimation where the noise vector $\mathbf{z}$ is amplified by the observation vector $\mathbf{w}_q$, the pilot-wise power constraint, $\|\mathbf{w}_q\| = 1$, should be considered. In this case, it is intractable to find the optimal solution to (5). To overcome this challenge, we use a greedy method to obtain the observation vectors pilot-by-pilot. Specifically, define $\mathbf{W}_t = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_t]$ as the observation matrix for timeslots $1 \sim t$, where $t < Q$, and denote $\mathbf{y}_t = \mathbf{W}_t^H\mathbf{h} + \mathbf{z}_t$ as the corresponding received signal. Given the current observation matrix $\mathbf{W}_t$, our algorithm aims to determine the next observation vector, $\mathbf{w}_{t+1}$, by maximizing the MI increment from timeslot t to $t+1$, i.e., $\max_{\mathbf{w}_{t+1}} I(\mathbf{y}_{t+1}; \mathbf{h}) - I(\mathbf{y}_t; \mathbf{h})$. As proved in [12], the MI increment is expressed as

$$I(\mathbf{y}_{t+1}; \mathbf{h}) - I(\mathbf{y}_t; \mathbf{h}) = \log_2\left(1 + \frac{1}{\sigma^2}\mathbf{w}_{t+1}^H\mathbf{\Sigma}_t\mathbf{w}_{t+1}\right), \tag{7}$$

where $\mathbf{\Sigma}_t$ represents the posterior covariance matrix of $\mathbf{h}$ given the current observation $\mathbf{y}_t$, i.e.,

$$\mathbf{\Sigma}_t = \mathbf{\Sigma_h} - \mathbf{\Sigma_h}\mathbf{W}_t\left(\mathbf{W}_t^H\mathbf{\Sigma_h}\mathbf{W}_t + \sigma^2\mathbf{I}_t\right)^{-1}\mathbf{W}_t^H\mathbf{\Sigma_h}. \tag{8}$$

Notably, the initial covariance, $\mathbf{\Sigma}_0$, is exactly the prior covariance, $\mathbf{\Sigma_h}$, since there is no observation when $t = 0$. According to (7), the optimal $\mathbf{w}_{t+1}$ can be attained as

$$\mathbf{w}_{t+1} = \underset{\|\mathbf{w}\|=1}{\text{argmax}}\, \mathbf{w}^H\mathbf{\Sigma}_t\mathbf{w}. \tag{9}$$

Problem (9) is a standard Rayleigh quotient problem. Its optimal solution is clearly the principal eigenvector of $\mathbf{\Sigma}_t$, i.e., $\mathbf{\Sigma}_t\mathbf{w}_{t+1} = \lambda(\mathbf{\Sigma}_t)\mathbf{w}_{t+1}$, where $\lambda(.)$ denotes the largest eigenvalue of its argument. As a result, the columns of the observation matrix $\mathbf{W}$ can be sequentially generated by updating the posterior covariance $\mathbf{\Sigma}_t$ and the observation vector $\mathbf{w}_t$ using (8) and (9) alternatively. Nevertheless, equations (8) and (9) require to repeatedly calculate the inverse of $\mathbf{W}_t^H\mathbf{\Sigma_h}\mathbf{W}_t + \sigma^2\mathbf{I}_t$ and the EVD of $\mathbf{\Sigma}_t$. To avoid these computationally inefficient operations, we attempt to seek the simplest expressions of $\mathbf{w}_t$ and $\mathbf{\Sigma}_t$. To this end, we first use the inverse of block matrix to obtain an equivalent expression of $\mathbf{\Sigma}_t$ in (8):[1]

$$\mathbf{\Sigma}_t = \mathbf{\Sigma}_{t-1} - \frac{\mathbf{\Sigma}_{t-1}\mathbf{w}_t\mathbf{w}_t^H\mathbf{\Sigma}_{t-1}}{\mathbf{w}_t^H\mathbf{\Sigma}_{t-1}\mathbf{w}_t + \sigma^2}. \tag{10}$$

Equation (10) allows us to directly update $\mathbf{\Sigma}_t$ from $\mathbf{\Sigma}_{t-1}$ and $\mathbf{w}_t$ without the need for matrix inversion. Then, if we further consider the fact that the optimal $\mathbf{w}_t$ is the principal eigenvector of $\mathbf{\Sigma}_{t-1}$, we can arrive at the most concise expression of $\mathbf{\Sigma}_t$ as follows, which lays the foundation of ice-filling.

Theorem 1: If $\mathbf{w}_t$ is the principal eigenvector of $\mathbf{\Sigma}_{t-1}$, then $\mathbf{\Sigma}_t$ in (10) can be rewritten as

$$\mathbf{\Sigma}_t = \mathbf{\Sigma}_{t-1} - \frac{\lambda^2(\mathbf{\Sigma}_{t-1})}{\lambda(\mathbf{\Sigma}_{t-1}) + \sigma^2}\mathbf{w}_t\mathbf{w}_t^H. \tag{11}$$

Proof: Applying the property of the principal eigenvector $\mathbf{\Sigma}_{t-1}\mathbf{w}_t = \lambda(\mathbf{\Sigma}_{t-1})\mathbf{w}_t$, we get $\mathbf{w}_t^H\mathbf{\Sigma}_{t-1}\mathbf{w}_t = \lambda(\mathbf{\Sigma}_{t-1})$ and $\mathbf{\Sigma}_{t-1}\mathbf{w}_t\mathbf{w}_t^H\mathbf{\Sigma}_{t-1} = \lambda^2(\mathbf{\Sigma}_{t-1})\mathbf{w}_t\mathbf{w}_t^H$, which together with (10) give rise to (11). ■

Two crucial conclusions can be drawn from **Theorem 1**.

Remark 1: *Given that $\mathbf{w}_t$ is the principal eigenvector of $\mathbf{\Sigma}_{t-1}$, **Theorem 1** indicates that the posterior covariances, $\mathbf{\Sigma}_t$ and $\mathbf{\Sigma}_{t-1}$, share the identical eigenspace. By further considering the generality of t, the eigenvectors, $\{\mathbf{u}_k\}_{k=1}^K$, of the prior covariance, $\mathbf{\Sigma}_0 = \mathbf{\Sigma_h}$, are inherited by all posterior covariances, $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \cdots, \mathbf{\Sigma}_{Q-1}$. Thereafter, all observation vectors are picked from the eigenspace of $\mathbf{\Sigma_h}$ and we only need to compute the EVD of $\mathbf{\Sigma_h}$ for one time.*

Remark 2: *In terms of the eigenvalues when updating $\mathbf{\Sigma}_{t-1}$ to $\mathbf{\Sigma}_t$, only the principal eigenvalue, $\lambda(\mathbf{\Sigma}_{t-1})$, of $\mathbf{\Sigma}_{t-1}$ is squeezed to an eigenvalue of $\mathbf{\Sigma}_t$ given by*

$$\lambda(\mathbf{\Sigma}_{t-1}) - \frac{\lambda^2(\mathbf{\Sigma}_{t-1})}{\lambda(\mathbf{\Sigma}_{t-1}) + \sigma^2} = \frac{\lambda(\mathbf{\Sigma}_{t-1})\sigma^2}{\lambda(\mathbf{\Sigma}_{t-1}) + \sigma^2}, \tag{12}$$

while the other eigenvalues are preserved.

As a result, updating the covariance matrix, $\mathbf{\Sigma}_t$, is equivalent to squeezing the eigenvalues within the identical eigenspace. Our proposed ice-filling algorithm is thus intrinsically an

Algorithm 1 Ice-Filling-Based Observation Matrix Design

Input: Number of pilots Q, covariance $\mathbf{\Sigma_h}$.
Output: Designed observation matrix $\mathbf{W}$.
1: Find the eigenvectors $[\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_K]$ and the corresponding eigenvalues $[\lambda_1, \lambda_2, \cdots, \lambda_K]$ of $\mathbf{\Sigma}_0 = \mathbf{\Sigma_h}$
2: Initialize: $[\lambda_1^0, \lambda_2^0, \cdots, \lambda_K^0] = [\lambda_1, \lambda_2, \cdots, \lambda_K]$
3: **for** $t = 0, \cdots, Q-1$ **do**
4: $k_t = \arg\max_{k\in\{1,2,\cdots,K\}}\{\lambda_k^t\}$
5: Eigenvector-assignment: $\mathbf{w}_{t+1} = \mathbf{u}_{k_t}$
6: Eigenvalue-update: $\lambda_{k_t}^{t+1} = \frac{\lambda_{k_t}^t\sigma^2}{\lambda_{k_t}^t + \sigma^2}$
7: Eigenvalue-preserve: $\lambda_k^{t+1} = \lambda_k^t$ for $k \neq k_t$
8: **end for**
9: Construct observation matrix: $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_Q]$
10: **return** Designed observation matrix $\mathbf{W}$

[1]Detailed proof of (10) can be found in Appendix A of [13].

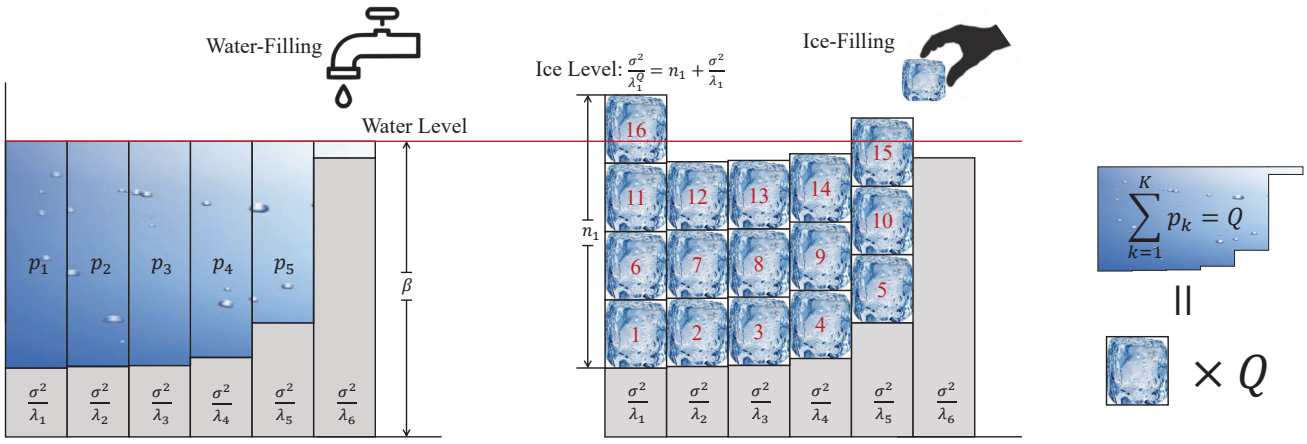


Fig. 2. Comparison between water-filling and ice-filling. The rank of the prior covariance, the number of antennas, and the total pilot length are set as $K = 6$, $M = 128$, and $Q = 16$, respectively. The number inside each ice block represents the order in which the ice block is filled.

eigenvector-assignment-process, as presented in **Algorithm 1**. We define $\{\lambda_1^t, \lambda_2^t, \cdots, \lambda_K^t\}$ as the eigenvalues of the posterior covariance $\mathbf{\Sigma}_t$, which are initialized as the eigenvalues of the prior covariance $\mathbf{\Sigma_h}$ in Step 2. In each timeslot, we find the largest eigenvalue from $\{\lambda_1^t, \lambda_2^t, \cdots, \lambda_K^t\}$ and index it by k_t in Step 4. Then, the corresponding eigenvector $\mathbf{u}_{k_t}$ of $\mathbf{\Sigma}_h$, which is equivalent to the principal eigenvector of $\mathbf{\Sigma}_t$, is assigned to $\mathbf{w}_{t+1}$ in Step 5. Finally, according to **Theorem 1**, the selected eigenvalue $\lambda_{k_t}^t$ is squeezed to $\frac{\lambda_{k_t}^t \sigma^2}{\lambda_{k_t}^t + \sigma^2}$, while the other eigenvalues of $\mathbf{\Sigma}_t$ are preserved, in order to attain the eigenvalues of the next covariance $\mathbf{\Sigma}_{t+1}$. These steps are repeatedly executed until all observation vectors are generated.

B. Ice-Filling Versus Water-Filling

We now put forward a more insightful interpretation to the ice-filling algorithm. Generally speaking, it can be viewed as a quantization of the water-filling principle in (6). Specifically, the observation matrices generated by both the water-filling and ice-filling schemes fall within the eigenspace of $\mathbf{\Sigma_h}$. Their distinction is that the ice-filling algorithm transforms the *continuous power-allocation process* of water-filling into an *discrete eigenvector-assignment process*. To see it more clearly, we introduce the concept of "pilot reuse frequency" as follows.

Definition 1: The pilot reuse frequency $n_k^t \in \mathbb{Z}^+, \forall k \in \{1, 2, \cdots, K\}, \forall t \in \{1, 2, \cdots, Q\}$, is defined as the number of times that the k-th eigenvector, $\mathbf{u}_k$, of $\mathbf{\Sigma_h}$ is selected as the observation vector by **Algorithm 1** during timeslots $1 \sim t$, which satisfies $\sum_{k=1}^{K} n_k^t = t$. For ease of expression, we denote $n_k := n_k^Q$ as the final pilot reuse frequency.

Definition 1 allows us to derive the analytical expressions of the eigenvalues $\{\lambda_k^t\}_{k=1}^K$ of $\mathbf{\Sigma}_t$. Specifically, the eigenvalue update rule in step 6 of **Algorithm 1** can be rewritten as:

$$\lambda_{k_t}^{t+1} = \frac{\lambda_{k_t}^t \sigma^2}{\lambda_{k_t}^t + \sigma^2} \Leftrightarrow \frac{\sigma^2}{\lambda_{k_t}^{t+1}} = 1 + \frac{\sigma^2}{\lambda_{k_t}^t}. \tag{13}$$

Equation (13) indicates that whenever the k-th eigenvector is selected as an observation vector, the value of σ^2/λ_k^t increases by 1. For timeslot t, the k-th eigenvector has been selected by n_k^t times according to **Definition 1**. Thus, we can naturally obtain the relationship between n_k^t and $\frac{\sigma^2}{\lambda_k^t}$:

$$\frac{\sigma^2}{\lambda_k^t} = n_k^t + \frac{\sigma^2}{\lambda_k} \Leftrightarrow n_k^t = \frac{\sigma^2}{\lambda_k^t} - \frac{\sigma^2}{\lambda_k}, \ k \in \{1, \cdots, K\}. \tag{14}$$

The comparison between equations (14) and (6) allows us to interpret the ice-filling algorithm as a quantization of the water-filling algorithm, which is elaborated below.

Interpretation of ice-filling: As shown in Fig. 2, the water-filling principle allocates Q units of water (power) to a vessel with K channels, each having a unique base level σ^2/λ_k, $k \in \{1, 2, \cdots, K\}$. By controlling the *uniform* water level β, the optimal power p_k in (6) is determined by the gap between the base level and the water level. In contrast, the ice-filling algorithm transforms the total Q units of water into Q ice blocks, each containing one unit of power and being used for one pilot transmission. Our algorithm starts from an empty vessel and fills one ice block onto the channel having the deepest base surface $\frac{\sigma^2}{\lambda_{k_0}}$. This operation is equivalent to finding the largest eigenvalue λ_{k_0} from $\{\lambda_1, \cdots, \lambda_K\}$. Then, the eigenvector $\mathbf{u}_{k_0}$ is assigned to the first observation vector $\mathbf{w}_1$, and ice level of this channel increases from $\frac{\sigma^2}{\lambda_{k_0}}$ to $\frac{\sigma^2}{\lambda_{k_0}^1} = \frac{\sigma^2}{\lambda_{k_0}} + 1$. In the subsequent time slots, the remaining $Q-1$ ice blocks are filled onto the channels locating at the deepest base surfaces or ice levels, indexed by $k_t = \arg\min_k \{\frac{\sigma^2}{\lambda_k^t}\}$, one by one. The corresponding eigenvectors $\mathbf{u}_{k_t}$ are used for pilot transmission, and the ice levels increase according to (14). Consequently, the final pilot reuse frequencies $\{n_k\}_{k=1}^K$ are determined by the number of ice blocks on top of each channel, and the ice levels are given by $\{\frac{\sigma^2}{\lambda_k^Q}\}_{k=1}^K = \{n_k + \frac{\sigma^2}{\lambda_k}\}_{k=1}^K$.

The above interpretation reveals that the continuous power p_k is quantized into the pilot reuse frequency n_k by ice-filling. To see this quantization nature more clearly, we prove an upper bound of quantization error in the following theorem.

Theorem 2: Considering the k-th pilot reuse frequency, n_k, and the k-th optimal power, p_k, we have

$$|n_k - p_k| < 1. \tag{15}$$

Proof: (See Appendix A). ■

Remark 3: ***Theorem 2*** *rigorously proves that the deviation of n_k from p_k is less than 1, rendering the pilot reuse frequency a good approximation of the optimal power allocation. Particularly, when the total pilot length Q is large, the relative error between n_k and p_k tends to zero because $\frac{|n_k-p_k|}{p_k} < \frac{1}{p_k} \overset{Q\to+\infty}{=} 0$. This guarantees the near-optimality of the proposed ice-filling algorithm.*

C. Extension to Analog Combiner with Phase Shifters

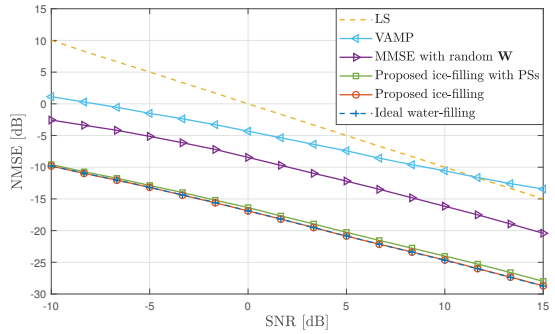
We now briefly discuss the extension of ice-filling algorithm to the analog combining architecture realized by phase-shifters (PSs). In this context, only the phase of the coefficients of $\mathbf{w}_q$ is tunable, i.e., $|w_{q,m}| = 1/\sqrt{M}$, $m \in \{1, 2, \cdots, M\}$. The most straightforward method is to project the designed observation matrix by **Algorithm 1**, $\mathbf{W}$, onto the feasible set of PS combiner: $\mathbf{W}^{\mathrm{PS}} = \frac{1}{\sqrt{M}} e^{j\angle \mathbf{W}}$.

V. Simulation Results

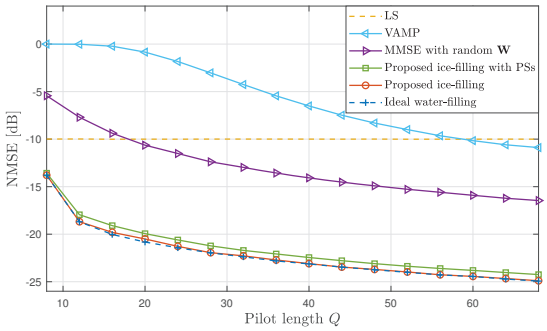
In this section, simulation results are provided to verify the effectiveness of the proposed ice-filling design. The performance is evaluated by the normalized mean square error (NMSE), defined as $\mathrm{NMSE} = \mathsf{E}(\|\mathbf{h} - \hat{\mathbf{h}}\|^2/\|\mathbf{h}\|^2)$.

1) Simulation Setup: The standard 3GPP TR 38.901 channel model is used for simulations. The number of clusters is set to 23 and that of rays per cluster is set to 20. Each cluster has an angular spread of 5° and a delay spread of 30 ns. The path gains are generated from the distribution $\mathcal{CN}(0, 1)$. The uniform planar array (UPA) is considered. The number of antennas is set to $M = 128$, and the numbers of horizontal antennas and vertical antennas are set to $M_x = 16$ and $M_y = 8$, respectively. The antenna spacing is set to $\frac{\lambda}{8}$ by default. The signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = \mathsf{E}\left(\|\mathbf{h}\|^2\right)/\sigma^2$, of which the default value is 10 dB.

Six benchmark schemes are considered for comparison. **1) LS:** The LS method is applicable when the pilot length is larger than the number of antennas. In this scheme, we set the pilot length as $Q = M$ and the observation matrix as a Fourier matrix. **2) MMSE with random $\mathbf{W}$**: The classical Bayesian regression in (3) is used to recover the channel, where an randomly generated observation matrix $\mathbf{W}$ of Q-pilot length is used. **3) VAMP**: The vector approximate message passing (VAMP) [5], a state-of-the-art CS algorithm, is leveraged to estimate the sparse channel. **4) Proposed ice-filling**: The observation matrix, $\mathbf{W}$, designed by ice-filling is used to enhance the performance of classical Bayesian regression in (3). **5) Proposed ice-filling with PSs**: The ice-filling designed matrix is first projected as $\mathbf{W}^{\mathrm{PS}} = \frac{1}{\sqrt{M}} e^{j\angle \mathbf{W}}$, and then used for Bayesian regression in PS-based analog combiners. **6) Ideal water-filling**: We regard the observation matrix $\mathbf{W}^{\mathrm{Ideal}}$ generated by the water-filling principle as an ideal case. To implement it, the noise vector, $\mathbf{z}$, in (2) is ideally assumed to be independent of the observation matrix $\mathbf{W}^{\mathrm{Ideal}}$, and its distribution is fixed as $\mathcal{CN}(\mathbf{0}_Q, \sigma^2\mathbf{I}_Q)$.



(a) NMSE versus SNR



(b) NMSE versus pilot length Q

Fig. 3. The impact of SNR and pilot length on NMSE performance.

Note that the pilot length for the LS method is $Q = M = 128$, while that for the other schemes is $Q = 64$ by default.

2) Simulation Results: The impact of the pilot length and SNR on the NMSE performance is investigated in Fig. 3. In Fig. 3(a), the SNR ranges from -10 dB $\sim$ 15 dB. In Fig. 3(b), the pilot length, Q, increases from 8 to 68 (except for the LS scheme with a fixed $Q = 128$). Thanks to the careful design of observation matrix, the proposed ice-filling schemes remarkably outperform existing benchmarks under all considered SNRs and pilot lengths. In particular, a 10 dB NMSE gap between ice-filling and MMSE with random $\mathbf{W}$, as well as a 15 dB gap between ice-filling and VAMP, are visible. Moreover, the NMSE curves of ice-filling method tightly align with the ideal water-filling's curves, demonstrating the near-optimal performance of ice-filling. Another finding is that the NMSE gap between the ice-filling and ice-filling with PSs is no higher than 0.5 dB. This validates that our ice-filling algorithm can be smoothly generalized to PS-based analog combiners.

In Fig. 4, the curves of the achieved NMSE performance versus the ratio of antenna spacing and wavelength, i.e., $\frac{d}{\lambda}$, are plotted. We can observe that, as the normalized antenna spacing $\frac{d}{\lambda}$ decreases from $\frac{1}{2}$ to $\frac{1}{16}$, the NMSEs for all channel estimators excluding LS method declines rapidly. For example, the NMSE achieved by the proposed ice-filling algorithm is decreased by about 10 dB when the antenna spacing ranges from $\lambda/2$ to $\lambda/8$. This is because a smaller antenna spacing leads to stronger channel correlations, which results in a more informative covariance $\mathbf{\Sigma}_{\mathbf{h}}$ for more precise channel reconstruction.

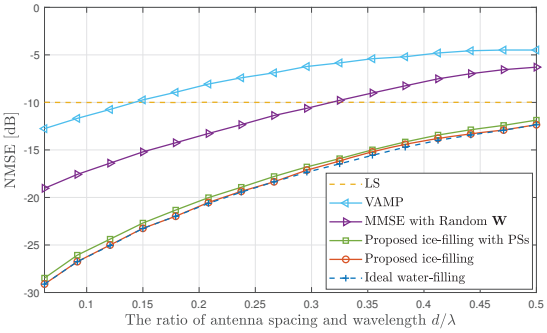


Fig. 4. NMSE versus the ratio of antenna spacing and wavelength.

VI. Conclusions

This paper incorporated the observation matrix design into DAS channel estimation. An ice-filling algorithm is proposed to generate the observation vectors. The algorithm was proved to be a quantization of water-filling principle with a quantization error smaller than 1, demonstrating its near-optimality.

Acknowledgment

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

Appendix A: Proof of Theorem 2

Theorem 2 can be proved by the contradiction method. Suppose there exists an index k such that $n_k \geq p_k + 1$. Due to the constraint that $\sum_{k=1}^{K} n_k = \sum_{k=1}^{K} p_k = Q$, there must exists k' such that $k' \neq k$ and $n_{k'} < p_{k'}$, otherwise $\sum_{k=1}^{K} n_k$ will be larger than $\sum_{k=1}^{K} p_k$. Note that $p_{k'}$ is non-zero since $p_{k'} > n_{k'} \geq 0$. In this context, the water level can be expressed as $\beta = p_{k'} + \frac{\sigma^2}{\lambda_{k'}}$ and the k'-th ice level $\frac{\sigma^2}{\lambda_{k'}^Q}$ should be smaller than β because

$$\frac{\sigma^2}{\lambda_{k'}^Q} = n_{k'} + \frac{\sigma^2}{\lambda_{k'}} < p_{k'} + \frac{\sigma^2}{\lambda_{k'}} = \beta. \tag{16}$$

Then, consider the k-th ice level $\frac{\sigma^2}{\lambda_k^Q}$. Since $n_k \geq p_k + 1$, we have the following inequality:

$$\frac{\sigma^2}{\lambda_k^Q} = n_k + \frac{\sigma^2}{\lambda_k} \geq p_k + \frac{\sigma^2}{\lambda_k} + 1 = \left(\beta - \frac{\sigma^2}{\lambda_k}\right)^+ + \frac{\sigma^2}{\lambda_k} + 1 \overset{(a)}{\geq} \beta + 1, \tag{17}$$

where the inequality (a) holds because $\left(\beta - \frac{\sigma^2}{\lambda_k}\right)^+ \geq \beta - \frac{\sigma^2}{\lambda_k}$. Combining equations (16) and (17), we arrive at

$$\sigma^2/\lambda_k^Q > \sigma^2/\lambda_{k'}^Q + 1. \tag{18}$$

We introduce the following lemma to show the irrationality of inequality (18).

Lemma 1: For the k-th eigenvalue λ_k^t obtained at the t-th timeslot, if $n_k^t > 0$, we have the inequality

$$\sigma^2/\lambda_{k'}^t + 1 \geq \sigma^2/\lambda_k^t, \tag{19}$$

hold for all $k' \in \{1, 2, \cdots, K\}$.

Proof: Since $n_k^t > 0$, the k-th eigenvector is selected by the ice-filling algorithm at least once. Let $t' < t$ denote the latest timeslot before t when the k-th eigenvector is selected for pilot transmission such that $n_k^{t'} = n_k^t - 1$ and $n_k^{t'+1} = n_k^t$. Suppose there exists k' such that $\frac{\sigma^2}{\lambda_{k'}^t} < \frac{\sigma^2}{\lambda_k^t} - 1$, then we have

$$\frac{\sigma^2}{\lambda_{k'}^{t'}} \overset{(b)}{\leq} \frac{\sigma^2}{\lambda_{k'}^t} \overset{(c)}{<} \frac{\sigma^2}{\lambda_k} + n_k^t - 1 = \frac{\sigma^2}{\lambda_k} + n_k^{t'} = \frac{\sigma^2}{\lambda_k^{t'}}, \tag{20}$$

where (b) holds because the ice-level $\sigma^2/\lambda_{k'}^t$ is non-decreasing w.r.t the timeslot t, and (c) holds because $\frac{\sigma^2}{\lambda_k} + n_k^t = \frac{\sigma^2}{\lambda_k^t}$. The inequality (20) contradicts the eigenvector selection principle of the ice-filling algorithm that $\lambda_k^{t'}$ is the largest over the eigenvalue set $\{\lambda_1^{t'}, \cdots, \lambda_K^{t'}\}$. Therefore, $\sigma^2/\lambda_{k'}^t$ should be no less than $\sigma^2/\lambda_k^t - 1$ for all k', which completes the proof. ■

Comparing (18) and (19), it is clear that (18) contradicts **Lemma 1** given that $n_k \geq p_k + 1 > 0$. Therefore, n_k should be smaller than $p_k + 1$. We can use the similar contradiction method to prove that $n_k > p_k - 1$, which is omitted due to space constraint. As a result, we get $|n_k - p_k| < 1$.

References

[1] 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.*, Apr. 2023.

[2] W. Ma, L. Zhu, and R. Zhang, "Mimo capacity characterization for movable antenna systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3392–3407, Apr. 2024.

[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] Z. Gao, L. Dai, S. Han, C.-L. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Mar. 2018.

[5] 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.

[6] N. González-Prelcic, H. Xie, J. Palacios, and T. Shimizu, "Wideband channel tracking and hybrid precoding for mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2161–2174, Apr. 2021.

[7] H. Lei, J. Zhang, H. Xiao, X. Zhang, B. Ai, and D. W. K. Ng, "Channel estimation for XL-MIMO systems with polar-domain multi-scale residual dense network," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 1479–1484, Jan. 2024.

[8] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[9] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for gaussian process optimization in the bandit setting," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3250–3265, May 2012.

[10] C. Ouyang, Y. Liu, and H. Yang, "MIMO-ISAC: Performance analysis and rate region characterization," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 669–673, Apr. 2023.

[11] B. Clarke and A. Barron, "Information-theoretic asymptotics of bayes methods," *IEEE Trans. Inf. Theory*, vol. 36, no. 3, pp. 453–471, 1990.

[12] Z. Pi, "Optimal transmitter beamforming with per-antenna power constraints," in *Proc. 2012 IEEE Int. Conf. Commun. (ICC)*, 2012, pp. 3779–3784.

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