

Successive Bayesian Reconstructor for Channel Estimation in Fluid Antenna Systems

Zijian Zhang^{ID}, *Graduate Student Member, IEEE*, Jieao Zhu^{ID}, *Graduate Student Member, IEEE*,
Linglong Dai^{ID}, *Fellow, IEEE*, and Robert W. Heath Jr.^{ID}, *Fellow, IEEE*

Abstract—Fluid antenna systems (FASs) can reconfigure their antenna locations freely within a spatially continuous space. To keep favorable antenna positions, the channel state information (CSI) acquisition for FASs is essential. While some techniques have been proposed, most existing FAS channel estimators require several channel assumptions, such as slow variation and angular-domain sparsity. When these assumptions are not reasonable, the model mismatch may lead to unpredictable performance losses. In this paper, we propose the successive Bayesian reconstructor (S-BAR) as a general solution to estimate FAS channels. Unlike model-based estimators, the proposed S-BAR is prior-aided, which builds the experiential kernel for CSI acquisition. Inspired by Bayesian regression, the key idea of S-BAR is to model the FAS channels as a stochastic process, whose uncertainty can be successively eliminated by kernel-based sampling and regression. In this way, the predictive mean of the regressed stochastic process can be viewed as a Bayesian channel estimator. Simulation results verify that, in both model-mismatched and model-matched cases, the proposed S-BAR can achieve higher estimation accuracy than the existing schemes.

Index Terms—Fluid antenna system (FAS), movable antenna (MA), dense array system (DAS), compressed sensing, channel estimation, Gaussian process regression.

I. INTRODUCTION

DIFFERENT from the conventional multiple-input multiple-output (MIMO) system where all antennas are fabricated and fixed on an array [2], [3], [4], fluid antenna systems (FASs) [5], [6], [7], also called movable antenna (MA) systems [8], [9], [10], introduce a movable structure where a few fluid antennas can switch their locations freely within a given space. In contrast to the MIMO whose antenna

spacing is usually subwavelength order of magnitude, the spacing of the available locations (referred to as “ports”) for fluid antennas can approach infinitesimal [9]. This almost continuously movable feature allows FASs to keep antennas always at favorable positions, which can fully explore the diversity and multiplexing gains of a given space [11]. Compared to the fixed antenna systems, FASs can maintain high transmission reliability [6] and large-scale multiple-access [7] while using much fewer antennas and radio frequency (RF) components [12]. As a result, FAS has become a promising solution to effectively reduce the hardware cost and complexity of transceivers [13].

A. Prior Work

The original idea of reconfiguring antenna positions dated back to 2000s, when reconfigurable antennas were proposed to balance the hardware complexity and transmission performance of MIMO systems [14], [15], [16]. By connecting massive antennas with a few RF chains through a switch network, reconfigurable antennas can carefully select a subset of antennas associated with the well-performed channels for transmission [17], [18]. Thanks to the advances in the microwave field, more position-flexible fluid antennas were proposed [19], [20]. Due to their compact hardware structures based on cheap materials, the cost and complexity of fluid antennas are low. For example, in [21], RF micro-electro-mechanical systems (MEMSs) were integrated into some radiation patches to reconfigure their positions. In [22], a fluid antenna made of liquid metals could move continuously in a non-conductive tube. In [23], pixel-like switches were utilized to design an electronically reconfigurable surface. By properly configuring multiple analog switches that link massive patches, the positions of radiation points can be reconfigured in several microseconds. As a result, these highly flexible structures are promising to explore the diversity and multiplexing gains of a limited space [5], [6], [7], [8], [9], [10], [11].

Up to now, the existing works on FASs have covered a wide range, mainly focusing on the performance analysis [5], [6], [7], [24], [25], [26], port selection [27], [28], [29], beamforming design [8], [9], [10], [30], [31], [32], and channel estimation [33], [34], [35]. For example, the ergodic capacity and the lower-bound capacity of FASs were derived in [5]. Then, the multiplexing gains of multi-user FASs were analyzed in [7] and [36]. By combining machine learning methods and analytical approximations, the port selection for FASs

Received 21 January 2024; revised 23 August 2024; accepted 5 December 2024. Date of publication 18 December 2024; date of current version 12 March 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFB3811503 and in part by the National Natural Science Foundation of China under Grant 62325106 and Grant 62031019. An earlier version of this paper was presented at the IEEE WCNC’24, Dubai, United Arab Emirates, April 21–24, 2024 [DOI: 10.1109/WCNC57260.2024.10571302]. The associate editor coordinating the review of this article and approving it for publication was V. Raghavan. (Corresponding author: Linglong Dai.)

Zijian Zhang, Jieao Zhu, and Linglong Dai are with the Department of Electronic Engineering and Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China (e-mail: zhangzj20@mails.tsinghua.edu.cn; zja21@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn).

Robert W. Heath Jr. is with the Department of Electrical and Computer Engineering, University of California at San Diego, San Diego, CA 92093 USA (e-mail: rwheathjr@ucsd.edu).

Digital Object Identifier 10.1109/TWC.2024.3515135

was studied in [27], [28], and [29] to realize fast antenna placements. To maximize the sum-rate in FASs, alternating optimization (AO)-based algorithms that jointly optimize the antenna positions and beamformers were proposed in [10] and [30]. To reduce the power consumption of FASs, sequential convex approximation (SCA)-based algorithms for FAS beamforming were proposed in [9] and [31].

Despite the encouraging prospects of FASs, such as high capacity and low power consumption [6], [7], [8], these expected gains are hard to achieve in practice. The transmission performance of FASs heavily relies on the positions of fluid antennas [27], [28], [29]. To ensure favorable antenna placements, the channel state information (CSI) of available locations is essential, and most existing works have assumed the perfectly known CSI [8], [9], [10], [30], [31], [32]. However, the channel estimation for FASs is challenging. Particularly, to fully obtain the performance gains provided by antenna mobility, the fluid antennas should move almost continuously in a given space [6]. Different from conventional MIMO which has large antenna spacing, the available locations (i.e., the ports) of fluid antennas are quite dense, leading to very high-dimensional port channels. Within a coherence time, it requires a large number of pilots for CSI acquisition in FASs. Besides, limited by the hardware structure of FASs, only a few ports can be connected to RF chains for pilot reception within the coherence time, which exacerbates the difficulty of channel estimation [33], [34], [35]. As a result, FASs can not acquire the precise CSI in real-time, which finally bottlenecks their performances in practice.

To tackle the above issue, some pilot-reduced FAS channel estimators have been proposed [33], [34], [35]. For example, in [33], a sequential linear minimum mean square error (SeLMMSE) method was proposed. In each subframe, the fluid antennas move to some equally-spaced ports for channel measurements [33]. By assuming the channels are slow-varying in a short distance, the channels of those unmeasured ports are assumed to be equal to those of their nearby measured ports. In [34], by assuming the angular-domain sparsity of FAS channels, a compressed sensing (CS)-based estimator was proposed. Employing orthogonal matching pursuit (OMP), this estimator can estimate the channel parameters including the angles of departure (AoDs), angles of arrival (AoAs), and gains of multiple paths. In [35], by assuming the perfectly known AoAs of all channel paths, the gains of FAS channels can be estimated by a least-square (LS)-based method. Although these methods can achieve channel estimation for FASs, they heavily rely on some channel assumptions, such as the spatially slow variation [33], angular-domain sparsity [34], and known AoAs [35]. When these assumptions are not reasonable, the model mismatch may lead to unpredictable performance losses.

B. Our Contributions

In this paper, we propose a successive Bayesian reconstructor (S-BAR) as a general solution to FAS channel estimation.¹ Our contributions are summarized as follows.

¹Simulation codes are provided to reproduce the results in this article: <http://oa.ee.tsinghua.edu.cn/dailiinglong/publications/publications.html>

- **Bayesian regression in FASs:** To the best of our knowledge, the idea of Bayesian regression [37] is introduced into FAS channel estimation for the first time. Firstly, fluid antennas can switch their locations among ports to measure channels, resembling a sampling process. Secondly, since the ports are densely deployed, the FAS channels have strong correlation. These two properties motivate us to build a kernel that characterizes the channel correlations and then construct channels by kernel-based sampling and regression. By carefully selecting kernels and ports for channel measurements, the idea of Bayesian regression is suitable to reconstruct FAS channels in a non-parametric manner.
- **General FAS channel estimator:** The S-BAR is proposed as a general solution to estimate FAS channels without assuming channel models. The key idea is to model the FAS channels as a stochastic process, whose uncertainty can be successively eliminated by kernel-based sampling and regression. The proposed S-BAR is realized in two stages. In the first stage, following the greedy-sampling idea of Bayesian regression, the switch matrix for FASs is designed by maximizing the mutual-information increment (MII) between two adjacent sampling points. In the second stage, according to the designed switch matrix, the channel observations of the select ports are combined with the kernels for process regression. Finally, the regressed stochastic process can be viewed as the Bayesian estimator of FAS channels.
- **Performance analysis and simulations:** The minimum mean square error (MSE) achieved by the proposed S-BAR is derived, which only depends on the eigenvalues of channel covariance and noise power. Then, numerical simulations are provided to compare the performances of different estimators. Our results show that, in both model-mismatched and model-matched cases, the proposed S-BAR outperforms the benchmark schemes with channel assumptions, and the estimation accuracy of S-BAR can increase by an order of magnitude. In addition, thanks to the hybrid offline and online implementation of S-BAR, the computational complexity of online employing S-BAR is only linear with the number of ports.

C. Organization and Notation

Organization: The rest of this paper is organized as follows. In Section II, the system model of an FAS is introduced, and the existing solutions of FAS channel estimation are reviewed. In Section III, the principle and the implementation of the proposed S-BAR are illustrated. In Section IV, the performance of S-BAR is analyzed, and the kernel selection for S-BAR is discussed. In Section V, simulations are carried out to compare the estimation performances of different schemes. Finally, in Section VI, conclusions are drawn and future works are discussed.

Notation: $[\cdot]^{-1}$, $[\cdot]^{\dagger}$, $[\cdot]^*$, $[\cdot]^T$, and $[\cdot]^H$ denote the inverse, pseudo-inverse, conjugate, transpose, and conjugate-transpose operations, respectively; $\|\cdot\|$ denotes the l_2 -norm of the argument; $\mathbf{x}(i)$ denotes the i -th entry of vector $\mathbf{x}$; $\mathbf{X}(i, j)$, $\mathbf{X}(j, :)$

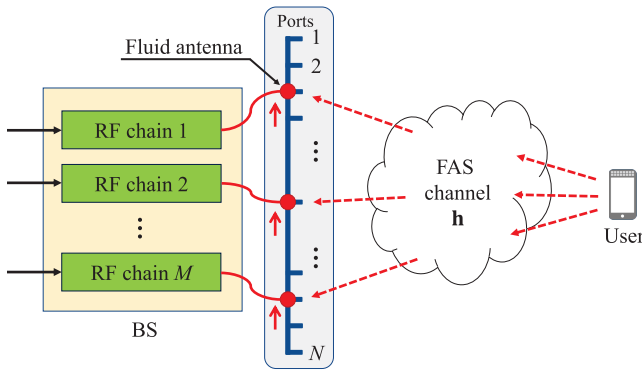


Fig. 1. An illustration of channel estimation for an FAS, where one N -port base station (BS) equipped with M fluid antennas receives pilots from a user in the uplink.

and $\mathbf{X}(:, j)$ denote the (i, j) -th entry, the j -th row, and the j -th column of matrix $\mathbf{X}$, respectively; $[K]$ denotes the integer set $\{1, \dots, K\}$; $\text{Tr}(\cdot)$ denotes the trace of its argument; $\mathcal{P}(\cdot|\cdot)$ is the conditional probability density function; $\mathbb{E}_{\mathbf{x}}(\cdot)$ is the expectation operator with respect to (w.r.t) the random vector $\mathbf{x}$; $\Re\{\cdot\}$ denotes the real part of the argument; $\ln(\cdot)$ denotes the natural logarithm of its argument; $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathcal{GP}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ respectively denote the complex Gaussian distribution and complex Gaussian process with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$; $\mathcal{U}(a, b)$ denotes the uniform distribution between a and b ; $\mathbf{I}_L$ is an $L \times L$ identity matrix; $\mathbf{e}_L$ is a zero-one vector with its L -th element being one and the other elements being zero; and $\mathbf{0}_L$ is a zero vector or matrix with dimension L .

II. SYSTEM MODEL AND EXISTING METHODS

In this section, the system model is introduced in Subsection II-A. In Subsection II-B, we review two typical channel estimators for FASs, including the CS-based channel estimator and the AO-based estimator. Finally, in Subsection II-C, the challenges and the opportunities of FAS channel estimation are discussed.

A. System Model

This paper considers the narrowband channel estimation of an uplink FAS, which consists of an N -port BS equipped with M fluid antennas ($M \ll N$) and a single-antenna user.² As shown in Fig. 1, the N ports are uniformly distributed along a linear dimension at the receiver. Each fluid antenna is connected to an RF chain for pilot reception, and the location of each antenna can be switched to one of the N available port locations. Let $\mathbf{h} \in \mathbb{C}^N$ denote the channels of N ports, and let P denote the number of transmit pilots within a coherence-time frame. As shown in Fig. 2, in each timeslot (subframe), fluid antennas can switch their positions to receive pilots. To characterize the locations of M fluid antennas in timeslot p , we introduce the definition of switch matrix as follows:

Definition 1 (Switch Matrix and Its Constraint): Binary indicator $\mathbf{S}_p \in \{0, 1\}^{N \times M}$ is defined as the switch matrix

²By using orthogonal pilots at different users, the proposed scheme in this paper can be extended to the multi-user case without difficulty.

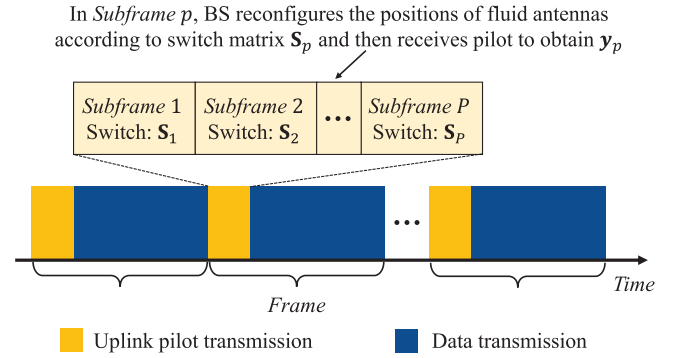


Fig. 2. The frame structure for FAS channel reconstruction.

of multiple fluid antennas in timeslot p . The (n, m) -th entry being 1 (or 0) means that the m -th antenna is (or not) located at the n -th port. Constrained by the hardware structure, M of N ports should be selected in each timeslot. To satisfy this constraint, here we constrain that each column of $\mathbf{S}_p$ has one 1 entry, and all 1 entries in $\mathbf{S}_p$ are not in the same row, i.e.,

$$\mathcal{S} := \left\{ \|\mathbf{S}_p(:, m)\| = 1, \|\mathbf{S}_p(n, :)\| \in \{0, 1\}, \forall m \in [M], \forall n \in [N], \forall p \in [P] \right\}. \quad (1)$$

Thus, we have $\mathbf{S}_p^H \mathbf{S}_p = \mathbf{I}_M$ for all $p \in [P]$.

Utilizing Definition 1, the signal vector $\mathbf{y}_p \in \mathbb{C}^M$ received at the BS in timeslot p can be modeled as

$$\mathbf{y}_p = \mathbf{S}_p^H \mathbf{h} x_p + \mathbf{z}_p, \quad (2)$$

where x_p is the pilot transmitted by the user and $\mathbf{z}_p \sim \mathcal{CN}(\mathbf{0}_M, \sigma^2 \mathbf{I}_M)$ is the additive white Gaussian noise (AWGN) at M selected ports. Without loss of generality, we assume that $x_p = 1$ for all $p \in [P]$. Considering the total P timeslots for pilot transmission, we arrive at

$$\mathbf{y} = \mathbf{S}^H \mathbf{h} + \mathbf{z}, \quad (3)$$

where $\mathbf{y} := [\mathbf{y}_1^H, \dots, \mathbf{y}_P^H]^H$, $\mathbf{S} := [\mathbf{S}_1, \dots, \mathbf{S}_P]$, and $\mathbf{z} := [\mathbf{z}_1^H, \dots, \mathbf{z}_P^H]^H$. Our goal is to reconstruct the N -dimensional channel $\mathbf{h}$ according to the PM -dimensional noisy pilot $\mathbf{y}$ ($PM \ll N$). To achieve this goal, existing works usually focus on CS-based and/or AO-based methods, which are reviewed in the next subsection.

B. Existing Methods for FAS Channel Estimation

1) Compressed Sensing Based Estimator: To reconstruct the high-dimensional FAS channel $\mathbf{h}$ from the low-dimensional pilot $\mathbf{y}$, a well-known method is the CS. Under the assumption of spatially-sparse channels, the CS-based channel estimator can achieve considerable performance in MIMO systems with hybrid structure [38]. For our considered FAS channel estimation, by viewing switch matrix $\mathbf{S}$ as a virtual analog precoder, CS-based methods can be similarly adopted.

Specifically, let $\mathbf{a}(\theta)$ denote the steering vector as a function of incident angle θ , which is defined as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{j\frac{2\pi}{\lambda} d \cos(\theta)}, \dots, e^{j\frac{2\pi}{\lambda} (N-1)d \cos(\theta)} \right]^T, \quad (4)$$

Algorithm 1 FAS-OMP Reconstructor

Input: Number of pilots P , spatial sparsity L .
Output: Reconstructed FAS channel $\hat{\mathbf{h}}$.

- 1: Employ randomly generated $\mathbf{S} \in \mathcal{S}$ at the BS
- 2: Initialization: $\mathbf{r} = \mathbf{y}$, $\mathcal{T} = \emptyset$, $\bar{\mathbf{h}} = \mathbf{0}_N$
- 3: **for** $l \in \{1, \dots, L\}$ **do**
- 4: Update correlation matrix: $\mathbf{\Gamma} = \mathbf{\Psi}^H \mathbf{r}$
- 5: Find new support: $v^* = \arg\max |\mathbf{\Gamma}(v)|$
- 6: Update support set: $\mathcal{T} = \mathcal{T} \cup \{v^*\}$
- 7: Orthogonal projection: $\bar{\mathbf{h}}(\mathcal{T}) = \mathbf{\Psi}(:, \mathcal{T})^\dagger \mathbf{y}$
- 8: Update residual: $\mathbf{r} = \mathbf{r} - \mathbf{\Psi}(:, \mathcal{T}) \bar{\mathbf{h}}(\mathcal{T})$
- 9: **end for**
- 10: **return** Reconstructed channel $\hat{\mathbf{h}} = \mathbf{F}(:, \mathcal{T}) \bar{\mathbf{h}}(\mathcal{T})$

where λ is the signal wavelength and d is the port spacing. Assuming that the channel is spatially sparse with C clusters each contributing R rays, the FAS channel $\mathbf{h}$ can be approximately modeled as

$$\mathbf{h} = \sqrt{\frac{N}{CR}} \sum_{c=1}^C \sum_{r=1}^R g_{c,r} \mathbf{a}(\theta_{c,r}), \quad (5)$$

where $g_{c,r}$ and $\theta_{c,r}$ are the complex path gain and the incident angle of the r -th ray in the c -th cluster, respectively. Let $\mathbf{F} \in \mathbb{C}^{N \times N}$ denote the discrete Fourier transform (DFT) matrix, thus $\mathbf{h}$ can be transformed into its angular-domain representation $\tilde{\mathbf{h}}$. Under the assumption of spatial sparsity, the angular-domain channel $\tilde{\mathbf{h}}$ only takes L significant values in a few of its entries ($L \ll N$). In this context, the received signal $\mathbf{y}$ in (3) can be rewritten as

$$\mathbf{y} = \mathbf{S}^H \mathbf{F} \tilde{\mathbf{h}} + \mathbf{z} = \mathbf{\Psi} \tilde{\mathbf{h}} + \mathbf{z}, \quad (6)$$

where $\mathbf{\Psi} = \mathbf{S}^H \mathbf{F}$ is the sensing matrix. Under the principle of uniform sampling, we can assume that the elements in $\mathbf{S}$ are randomly selected from $\{0, 1\}$ subject to $\mathbf{S} \in \mathcal{S}$.

One can find that, (6) is a standard observation equation for sparse signal reconstruction, which can be addressed by the some existing CS-based algorithms. For example, in [34], an FAS-OMP reconstructor is proposed to estimate the channel parameters in FASs. After iteratively selecting L columns of sensing matrix $\mathbf{\Psi}$ for sparse representation, the parameters of $\mathbf{h}$ including path gains and incident angles can be estimated. By substituting the estimated channel parameters into (5), the FAS-OMP reconstructor in [34] can be modified to acquire channel $\mathbf{h}$ explicitly, as summarized in **Algorithm 1**.

2) *Alternating Optimization Based Estimator*: The CS-based channel estimators usually operate based on a gridded codebook, e.g., the DFT matrix $\mathbf{F}$. When the real channel parameter values fall between the grid points, the non-ideal grid sampling may introduce estimation error [39]. To improve the estimation accuracy, the channel parameters can be further optimized to cope with the errors introduced by on-grid sampling [40]. Given the spatial sparsity L , this goal can be achieved by finding the path gains $\mathbf{g} := [g_1, \dots, g_L]^T$ and incident angles $\boldsymbol{\theta} := [\theta_1, \dots, \theta_L]^T$ that maximize the likelihood function $P(\mathbf{y}|\mathbf{g}, \boldsymbol{\theta})$, i.e., the maximum likelihood (ML) estimator. Following this idea, an AO-based method

Algorithm 2 FAS-ML Reconstructor

Input: Number of pilots P , spatial sparsity L .
Output: Reconstructed FAS channel $\hat{\mathbf{h}}$.

- 1: Employ randomly generated $\mathbf{S} \in \mathcal{S}$ at the BS
- 2: Initialize $\hat{\mathbf{g}}$ and $\hat{\boldsymbol{\theta}}$ via a CS-based algorithm, e.g., OMP
- 3: Initialize step length by $\zeta = 1$
- 4: **while** no convergence of $P(\mathbf{y}|\hat{\mathbf{g}}, \hat{\boldsymbol{\theta}})$ **do**
- 5: Update path gain: $\hat{\mathbf{g}} = \mathbf{B}^\dagger(\hat{\boldsymbol{\theta}}) \mathbf{y}$
- 6: Update angle: $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \zeta \Re \left\{ \frac{\partial \|\mathbf{y} - \mathbf{B}(\hat{\boldsymbol{\theta}}) \hat{\mathbf{g}}\|^2}{\partial \boldsymbol{\theta}} \right\}_{\mathbf{g}=\hat{\mathbf{g}}, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$
- 7: Update step length: $\zeta = \zeta/2$
- 8: **end while**
- 9: **return** Reconstructed channel $\hat{\mathbf{h}} = \sqrt{\frac{N}{L}} \sum_{l=1}^L \hat{g}_l \mathbf{a}(\hat{\theta}_l)$

called FAS-ML reconstructor is summarized in **Algorithm 2**, which is explained as follows.

Assume that channel $\mathbf{h}$ has the spatial sparsity as shown in (5). For the considered problem of channel estimation, the logarithm likelihood function can be written as

$$\ln(P(\mathbf{y}|\mathbf{g}, \boldsymbol{\theta})) = -N \ln(\sigma^2 \pi) - \frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{B}(\boldsymbol{\theta}) \mathbf{g}\|^2, \quad (7)$$

wherein $\mathbf{B}(\boldsymbol{\theta}) := \sqrt{\frac{N}{L}} \mathbf{S}^H [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_L)]$. Then, an ML-based estimation can be achieved by solving

$$\{\mathbf{g}^*, \boldsymbol{\theta}^*\} = \underset{\mathbf{g}, \boldsymbol{\theta}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{B}(\boldsymbol{\theta}) \mathbf{g}\|^2. \quad (8)$$

However, due to the coupled variables $\mathbf{g}$ and $\boldsymbol{\theta}$, the optimal $\mathbf{g}$ and $\boldsymbol{\theta}$ are hard to be obtained simultaneously. As a compromise, $\mathbf{g}$ and $\boldsymbol{\theta}$ are usually updated in an alternating way to approach the sub-optimal solution [41]. By fixing one and optimize the other variables, the likelihood function $P(\mathbf{y}|\mathbf{g}, \boldsymbol{\theta})$ can be maximized through an AO process. Finally, the FAS-ML reconstructor in **Algorithm 2** can be obtained. After sufficient iterations, all variable updates make $P(\mathbf{y}|\mathbf{g}, \boldsymbol{\theta})$ monotonically increase, thus **Algorithm 2** is always convergent.

C. Challenge and Opportunity

Challenge: Despite the feasibility of existing FAS channel estimators, two inherent drawbacks have bottlenecked their estimation accuracy. First, most existing estimators are parametric algorithms, which usually follow the assumption of spatially sparse channels [34]. In practical scenarios, the model mismatch may lead to unpredictable performance loss. Second, most existing estimators are based on the randomly generated zero-one distributed $\mathbf{S}$, i.e., the switch matrix. Different from the MIMO analog precoder whose elements are all unit-modulus [38], only a few elements in $\mathbf{S}$ are one while the others are all zero. It implies that, the received pilot $\mathbf{y}$ only contains the information of a few channels associated with the selected ports. Consequently, high pilot overhead (i.e., a large P) may be required to accurately reconstruct channel $\mathbf{h}$.

Opportunity: Different from MIMO that measures all channels simultaneously, fluid antennas switch their locations among the ports to measure channels in a successive way, which resembles a sampling process. Besides, unlike the conventional MIMO whose typical antenna spacing is usually

$\lambda/2$, the port spacing of FASs d is usually smaller, such as $\lambda/10$ [5], [6], [7]. Thus, the channels of several ports closer to each other are strongly correlated [42]. It suggests that, when the channel of a port is sampled (measured), the uncertainty of the channels associated with its close-by ports can be partially eliminated. These properties inspire that, by carefully designing a sampling sequence and conducting regression, the FAS channels can be reconstructed in a non-parametric way.

III. PROPOSED SUCCESSIVE BAYESIAN RECONSTRUCTOR

By building experiential kernels for an objective function, Bayesian regression can determine the sampling strategy and reconstruct the objective functions via kernel-based regression. Thanks to the sampler-like fluid antennas and strongly correlated channels in FASs, Bayesian regression perfectly matches the problem of FAS channel estimation.

Following this idea, in this section, we propose the S-BAR as a high-accuracy solution to FAS channel estimation [37]. Firstly, the Bayesian regression is introduced in Subsection III-A. Then, in Subsection III-B, the proposed S-BAR is illustrated. Finally, in Subsection III-C, the practical implementation of S-BAR is discussed.

A. Bayesian Linear Regression

Without making any prior assumptions, the attempt to recover the function $f(\mathbf{x})$ from a limited number of samples appears to be a challenging endeavor. Modeling $f(\mathbf{x})$ as a realization of a stochastic process offers an elegant means of specifying function properties in a non-parametric manner. Under this framework, Bayesian linear regression, also called Gaussian process regression (GPR) or Kriging method [43], has become a popular solution. Specifically, function $f(\mathbf{x})$ can be modeled as a sample of Gaussian process $\mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, where any finite subset follows a consistent multivariate Gaussian distribution. It is completely specified by its mean function $\mu(\mathbf{x})$, which can be assumed to be zero, and its kernel function $k(\mathbf{x}, \mathbf{x}')$, which encodes smoothness properties of recovered $f(\mathbf{x})$. For clarity, here we summarize the Bayesian linear regression in **Algorithm 3**, and the detailed explanations are provided as follows.

In timeslot t , consider a prior $\mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ over $f(\mathbf{x})$. Let $\gamma^t := [\gamma^1, \dots, \gamma^t]^T$ denote t noisy measurements for points in $\mathcal{A}^t := \{\mathbf{x}^1, \dots, \mathbf{x}^t\}$, where $\gamma^i = f(\mathbf{x}^i) + n_i$ with $n_i \sim \mathcal{CN}(0, \delta^2)$. The joint probability distribution of $f(\mathbf{x})$ and γ^t satisfies

$$\begin{bmatrix} f(\mathbf{x}) \\ \gamma^t \end{bmatrix} \sim \mathcal{CN}\left(\begin{bmatrix} \mu(\mathbf{x}) \\ \mu^t \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}, \mathbf{x}) & (\mathbf{k}^t(\mathbf{x}))^H \\ \mathbf{k}^t(\mathbf{x}) & \mathbf{K}^t + \delta^2 \mathbf{I}_t \end{bmatrix}\right), \quad (9)$$

where $\mathbf{k}^t(\mathbf{x}) := [k(\mathbf{x}^1, \mathbf{x}), \dots, k(\mathbf{x}^t, \mathbf{x})]^T$; $\mu^t := [\mu(\mathbf{x}^1), \dots, \mu(\mathbf{x}^t)]^T$; and the (i, j) -th entry of $\mathbf{K}^t \in \mathbb{C}^{t \times t}$ is $k(\mathbf{x}^i, \mathbf{x}^j)$, for all $i, j \in \{1, \dots, t\}$. It is easy to prove that, given γ^t , the posterior over $f(\mathbf{x})$ is also a Gaussian process, with its mean and covariance being:

$$\mu^t(\mathbf{x}) = \mu(\mathbf{x}) + (\mathbf{k}^t(\mathbf{x}))^H (\mathbf{K}^t + \delta^2 \mathbf{I}_t)^{-1} (\gamma^t - \mu^t), \quad (10)$$

$$k^t(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - (\mathbf{k}^t(\mathbf{x}))^H (\mathbf{K}^t + \delta^2 \mathbf{I}_t)^{-1} \mathbf{k}^t(\mathbf{x}'). \quad (11)$$

Algorithm 3 Bayesian Linear Regression

Input: Definition domain $\mathcal{S}$, prior $\mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, tolerance threshold ε .

Output: Reconstructed function $\hat{f}(\mathbf{x})$.

- 1: Initialization: $t = 0$, $\mathcal{A}^0 = \emptyset$, $k^0(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}')$
 - 2: **while** $k^t(\mathbf{x}, \mathbf{x}) > \varepsilon$ for an $\mathbf{x} \in \mathcal{S}$ **do**
 - 3: Timeslot update: $t = t + 1$
 - 4: Sample selection: $\mathbf{x}^t = \arg \max_{\mathbf{x} \in \mathcal{S}/\mathcal{A}^{t-1}} k^{t-1}(\mathbf{x}, \mathbf{x})$
 - 5: Update measured points: $\mathcal{A}^t = \mathcal{A}^{t-1} \cup \{\mathbf{x}^t\}$
 - 6: Measurement: $\gamma^t = f(\mathbf{x}^t) + n_t$
 - 7: Posterior update: Calculate mean $\mu^t(\mathbf{x})$ and covariance $k^t(\mathbf{x}, \mathbf{x}')$ by (10) and (11) for all $\mathbf{x}, \mathbf{x}' \in \mathcal{S}$, respectively
 - 8: **end while**
 - 9: Reconstruction: $\hat{f}(\mathbf{x}) = \mu^t(\mathbf{x})$
 - 10: **return** Reconstructed function $\hat{f}(\mathbf{x})$
-

Then, the next candidate point to be sampled, i.e., $\mathbf{x}^{t+1}$, can be determined based on the updated posterior. From the information perspective, the points to be sampled should be as uncorrelated as possible. In the case of one-by-one sampling, sampling the point with the maximum posterior variance can obtain the most information. Therefore, by assuming that $\mathbf{x} \in \mathcal{S}$, $\mathbf{x}^{t+1}$ can be chosen according to

$$\mathbf{x}^{t+1} = \arg \max_{\mathbf{x} \in \mathcal{S}/\mathcal{A}^t} k^t(\mathbf{x}, \mathbf{x}), \quad (12)$$

where $/$ is the set difference. By updating $\mathcal{A}^t$ and γ^t accordingly, the variance function $k^t(\mathbf{x}, \mathbf{x})$ decreases asymptotically, which means that the uncertainty of $f(\mathbf{x})$ is gradually eliminated. After reaching the tolerance threshold ε , the posterior mean $\mu^t(\mathbf{x})$ in (10) can be viewed as the Bayesian estimator of $f(\mathbf{x})$, which completes the algorithm.

B. Working Principle of the Proposed S-BAR

In each pilot timeslot, M fluid antennas can move positions and measure channels, thus the channel estimation of FASs is similar to a successive sampling process. Since the port spacing is very short, the FAS channels are strongly correlated. These features inspire us to recover $\mathbf{h}$ through optimizing the reconstructor $\hat{\mathbf{h}}$ and the switch matrix $\mathbf{S}$ jointly, which can be formulated as

$$\min_{\mathbf{h}, \mathbf{S} \in \mathcal{S}} \mathbb{E}_{\mathbf{h}, \mathbf{z}} \left(\|\mathbf{h} - \hat{\mathbf{h}}\|^2 \right), \quad (13)$$

It is obvious that (13) is a Bayesian estimation problem. To efficiently solve this problem, one can model $\mathbf{h}$ as a sample of Gaussian process $\mathcal{GP}(\mathbf{0}_N, \Sigma)$. In particular, semidefinite Hermitian matrix $\Sigma \in \mathbb{C}^{N \times N}$ is called the kernel, which characterizes (but does not need to be) the prior covariance of $\mathbf{h}$. Given the distribution of channels and noises, the joint probability distribution of $\mathbf{h}$ and $\mathbf{y}$ satisfies

$$\begin{bmatrix} \mathbf{h} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{CN}\left(\begin{bmatrix} \mathbf{0}_N \\ \mathbf{0}_{PM} \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma \mathbf{S} \\ \mathbf{S}^H \Sigma & \mathbf{S}^H \Sigma \mathbf{S} + \sigma^2 \mathbf{I}_{PM} \end{bmatrix}\right), \quad (14)$$

Therefore, for given $\mathbf{y}$, the posterior mean and the posterior covariance of $\mathbf{h}$ can be calculated by:

$$\boldsymbol{\mu}_{\mathbf{h}|\mathbf{y}} = \boldsymbol{\Sigma} \mathbf{S} (\mathbf{S}^H \boldsymbol{\Sigma} \mathbf{S} + \sigma^2 \mathbf{I}_{PM})^{-1} \mathbf{y}, \quad (15)$$

$$\boldsymbol{\Sigma}_{\mathbf{h}|\mathbf{y}} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma} \mathbf{S} (\mathbf{S}^H \boldsymbol{\Sigma} \mathbf{S} + \sigma^2 \mathbf{I}_{PM})^{-1} \mathbf{S}^H \boldsymbol{\Sigma}. \quad (16)$$

Given a switch matrix $\mathbf{S}$, it is straightforward to prove that the optimal reconstructor $\hat{\mathbf{h}}$ is exactly the posterior mean, i.e., $\hat{\mathbf{h}} = \boldsymbol{\mu}_{\mathbf{h}|\mathbf{y}}$, and the MSE in (13) can be rewritten as the trace of posterior covariance, i.e., $\mathbb{E}_{\mathbf{h}, \mathbf{z}} (\|\mathbf{h} - \hat{\mathbf{h}}\|^2) = \text{Tr}(\boldsymbol{\Sigma}_{\mathbf{h}|\mathbf{y}})$. Thus, problem (13) can be solved by finding a proper switch matrix $\mathbf{S}$ that minimizes $\text{Tr}(\boldsymbol{\Sigma}_{\mathbf{h}|\mathbf{y}})$, i.e.,

$$\mathbf{S} = \underset{\mathbf{S} \in \mathcal{S}}{\text{argmin}} \text{Tr}(\boldsymbol{\Sigma}_{\mathbf{h}|\mathbf{y}}). \quad (17)$$

Unfortunately, due to the constraint $\mathbf{S} \in \mathcal{S}$ and the non-convex objective $\text{Tr}(\boldsymbol{\Sigma}_{\mathbf{h}|\mathbf{y}})$, problem (17) is still hard to solve.

To overcome the above challenge, we adopt the greedy-sampling idea from Bayesian linear regression to design $\mathbf{S}$. The core technique is to greedily optimize $\mathbf{S}$ in a column-by-column manner by maximizing the mutual-information increment (MII) between channels and received pilots. Specifically, let $\mathbf{S}_t := [\mathbf{s}_1, \dots, \mathbf{s}_t] \in \{0, 1\}^{N \times t}$ denote the first- t columns of $\mathbf{S}$ where $1 \leq t \leq PM$. Accordingly, the first- t received signals can be written as $\mathbf{y}_t = \mathbf{S}_t^H \mathbf{h} + \mathbf{z}_t$, wherein $\mathbf{z}_t$ is the first- t AWGN. Given $\mathbf{S}_t$, we aim to design the $(t+1)$ -th column $\mathbf{s}_{t+1}$ by maximizing the MII between received pilots and channels from t to $t+1$, which can be formulated as

$$\max_{\mathbf{s}_{t+1} \in \mathcal{S}} I(\mathbf{y}_{t+1}; \mathbf{h}) - I(\mathbf{y}_t; \mathbf{h}), \quad (18)$$

where the mutual information $I(\mathbf{y}_t; \mathbf{h})$ is given by

$$I(\mathbf{y}_t; \mathbf{h}) = \log_2 \det \left(\mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{S}_t^H \boldsymbol{\Sigma} \mathbf{S}_t \right). \quad (19)$$

Furthermore, we introduce the following lemma to simplify the problem formulation:

Lemma 1 (Equivalent Problem of MII Maximization):

Given $\mathbf{S}_t$, the MII maximization problem in (18) can be equivalently rewritten as

$$\mathbf{s}_{t+1} = \underset{\mathbf{s} \in \mathcal{S}}{\text{argmax}} \mathbf{s}^H \boldsymbol{\Sigma}_t \mathbf{s}, \quad (20)$$

where $\boldsymbol{\Sigma}_t$ is written as

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma} - \boldsymbol{\Sigma} \mathbf{S}_t (\mathbf{S}_t^H \boldsymbol{\Sigma} \mathbf{S}_t + \sigma^2 \mathbf{I}_t)^{-1} \mathbf{S}_t^H \boldsymbol{\Sigma}, \quad (21)$$

which is exactly the posterior covariance of $\mathbf{h}$ for given $\mathbf{y}_t$. In particular, we have $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{PM} = \boldsymbol{\Sigma}_{\mathbf{h}|\mathbf{y}}$.

Proof: Constructive proof is given in Appendix A. ■

Recalling the hardware constraint of FASs $\mathcal{S}$, the zero-one switch vector $\mathbf{s}_{t+1}$ should only have one 1 entry, such that the corresponding fluid antenna is positioned at one port. Thus, one may feel that problem (20) can be optimally solved by finding the sampling index $n_{t+1} = \arg \max_{n \in [N]} \boldsymbol{\Sigma}_t(n, n)$ and then setting $\mathbf{s}_{t+1} = \mathbf{e}_{n_{t+1}}$. However, in fact, selecting index from the full port set $[N]$ is not correct, because multiple antennas cannot be located at the same port in a single timeslot. That is to say, all 1 entries in switch matrix $\mathbf{S}_p$ cannot be in the same row, as illustrated in **Definition 1**.

Algorithm 4 Proposed Successive Bayesian Reconstructor

Input: Number of pilots P , kernel $\boldsymbol{\Sigma}$.

Output: Reconstructed FAS channel $\hat{\mathbf{h}}$.

```

1: # Stage 1 (Offline Design):
2: Initialization:  $t = 0$ ,  $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}$ 
3: for  $p \in \{1, \dots, P\}$  do
4:   Initialize sampling sequence at pilot  $p$ :  $\Omega_p = \emptyset$ 
5:   for  $m \in \{1, \dots, M\}$  do
6:     Posterior covariance update: Calculate  $\boldsymbol{\Sigma}_t$  by (24)
7:     Candidate selection:  $n_{t+1} = \arg \max_{n \in [N]/\Omega_p} \boldsymbol{\Sigma}_t(n, n)$ 
8:     Switch vector design:  $\mathbf{s}_{t+1} = \mathbf{e}_{n_{t+1}}$ 
9:     The  $p$ -th sequence update:  $\Omega_p = \Omega_p \cup \{n_{t+1}\}$ 
10:    Counter update:  $t = t + 1$ 
11:   end for
12: end for
13: Merge switch vectors:  $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_{PM}]$ 
14: Weight calculation:  $\mathbf{W} = (\mathbf{S}^H \boldsymbol{\Sigma} \mathbf{S} + \sigma^2 \mathbf{I}_{PM})^{-1} \mathbf{S}^H \boldsymbol{\Sigma}$ 
15: # Stage 2 (Online Regression):
16: Employ the designed switch matrix  $\mathbf{S}$  at the BS, and then
   obtain the received pilot:  $\mathbf{y} = \mathbf{S}^H \mathbf{h} + \mathbf{z}$ 
17: Channel reconstruction:  $\hat{\mathbf{h}} = \mathbf{W}^H \mathbf{y}$ 
18: return Reconstructed FAS channel  $\hat{\mathbf{h}}$ 
```

To address the port conflict issue while optimizing $\mathbf{s}_{t+1}$, Ω_p is introduced to represent the sampling sequence, which is composed of the indexes of the selected ports in the p -th timeslot. Ω_p is initialized as an empty set. Once a port is selected, it will be added to Ω_p , and then the next sampling index can be selected from the difference set $[N]/\Omega_p$. Thus, the optimal solution to problem (20) can be obtained by

$$n_{t+1} = \underset{n \in [N]/\Omega_p}{\text{argmax}} \boldsymbol{\Sigma}_t(n, n), \quad (22a)$$

$$\mathbf{s}_{t+1} = \mathbf{e}_{n_{t+1}}. \quad (22b)$$

In this way, constraint $\mathcal{S}$ can be naturally satisfied during the design process. This ensures that the designed $\mathbf{S}$ is practically implementable in FAS systems. After determining n_{t+1} , we can update Ω_p by $\Omega_p \cup \{n_{t+1}\}$ and repeat the above process. When the number of indexes in Ω_p is equal to the number of fluid antennas M , we can update Ω_p in (22) by Ω_{p+1} to focus on the port selection in the $(p+1)$ -th timeslot. After obtaining all PM columns of switch matrix $\mathbf{S} := [\mathbf{s}_1, \dots, \mathbf{s}_{PM}]$, one can obtain the channel reconstructor by

$$\hat{\mathbf{h}} = \boldsymbol{\Sigma} \mathbf{S} (\mathbf{S}^H \boldsymbol{\Sigma} \mathbf{S} + \sigma^2 \mathbf{I}_{PM})^{-1} \mathbf{y}, \quad (23)$$

which completes the algorithm.

For better understanding, the proposed S-BAR is summarized in **Algorithm 4**, which is realized in a hybrid offline and online manner. More details about its two-stage implementation will be discussed in the next subsection.

C. Hybrid Offline and Online Implementation of S-BAR

In this subsection, the practical implementation of the proposed S-BAR is discussed. Firstly, from the derivations in the above subsection, we obtain the following three observations.

- Equation (23) indicates that, given $\mathbf{S}$, the channel reconstructor $\hat{\mathbf{h}}$ is the linear weighted sum of the received pilots $\mathbf{y}$, i.e., $\hat{\mathbf{h}} = \mathbf{W}^H \mathbf{y}$, where the weight $\mathbf{W} := (\mathbf{S}^H \mathbf{\Sigma} \mathbf{S} + \sigma^2 \mathbf{I}_{PM})^{-1} \mathbf{S}^H \mathbf{\Sigma}$ only relies on the kernel $\mathbf{\Sigma}$.
- Equation (21) shows that, posterior covariance $\mathbf{\Sigma}_t$ only relies on kernel $\mathbf{\Sigma}$, while it is unrelated to the received pilot $\mathbf{y}$ in instantaneous channel estimation.
- Equation (22) suggests that the FAS port selection only relies on the posterior covariance $\mathbf{\Sigma}_t$.

These observations reveal that, the switch matrix $\mathbf{S}$ and the weight $\mathbf{W}$ for reconstructing $\mathbf{h}$ can be designed offline and then deployed online for regression. Therefore, the implementation of the proposed S-BAR can be realized in the two stages as shown in **Algorithm 4**, so that the complexity of employing the S-BAR online can be significantly reduced. The details are explained as follows.

1) *Stage 1 (Offline Design)*: Since switch matrix is determined by the posterior covariance $\mathbf{\Sigma}_t$, while $\mathbf{\Sigma}_t$ only relies on the kernel $\mathbf{\Sigma}$. $\mathbf{S} \in \{0, 1\}^{N \times PM}$ and the weight $\mathbf{W} \in \mathbb{C}^{PM}$ for reconstructing $\mathbf{h} \in \mathbb{C}^N$ can be designed offline at the first stage. By updating $\mathbf{\Sigma}_t$ in (21) and n_{t+1} in (22) until $t = PM$, the switch matrix $\mathbf{S}$ can be designed in a column-by-column way, which determines all locations of M antennas in P pilot timeslots. To avoid the high computational complexity of matrix inversion in (21), the following lemma is derived to efficiently calculate $\mathbf{\Sigma}_t$ in a recursive way:

Lemma 2 (Recursion Formula of $\mathbf{\Sigma}_t$): Given $\mathbf{\Sigma}_{t-1}$ and n_t , the posterior covariance $\mathbf{\Sigma}_t$ can be calculated by

$$\mathbf{\Sigma}_t = \mathbf{\Sigma}_{t-1} - \frac{\mathbf{\Sigma}_{t-1}(:, n_t) \mathbf{\Sigma}_{t-1}(n_t, :)}{\mathbf{\Sigma}_{t-1}(n_t, n_t) + \sigma^2}. \quad (24)$$

Proof: Constructive proof is given in Appendix B. ■ Then, for given $\mathbf{S}$, the optimal weight for reconstructing $\mathbf{h}$ can be obtained by

$$\mathbf{W} = (\mathbf{S}^H \mathbf{\Sigma} \mathbf{S} + \sigma^2 \mathbf{I}_{PM})^{-1} \mathbf{S}^H \mathbf{\Sigma}. \quad (25)$$

2) *Stage 2 (Online Regression)*: Since *Stage 1* is realized offline, the switch matrix $\mathbf{S}$ and the weight $\mathbf{W}$ can be designed and saved at the BS in advance. This mechanism of offline design and online deployment can significantly reduce the computational complexity of employing S-BAR. In *Stage 2*, the scheme is then employed online for channel measurements. The M fluid antennas of the BS will move and receive pilots according to the designed $\mathbf{S}$, arriving at the noisy pilot $\mathbf{y}$. According to the Bayesian estimator in (15), channel $\mathbf{h}$ can be reconstructed by the weighted sum $\hat{\mathbf{h}} = \mathbf{W}^H \mathbf{y}$, which finally completes the proposed S-BAR in **Algorithm 4**.

IV. PERFORMANCE ANALYSIS AND KERNEL SELECTION

A. Performance Analysis of S-BAR

To evaluate the performance limits of our proposed S-BAR, in this subsection, the estimation accuracy and computational complexity of S-BAR are analyzed, respectively.

1) *Estimation Accuracy of S-BAR*: To quantitatively depict the estimation accuracy of S-BAR, we first define the MSE of reconstructing $\mathbf{h}$ as $E = \mathbb{E}_{\mathbf{h}, \mathbf{z}}(\|\hat{\mathbf{h}} - \mathbf{h}\|^2)$. By adopting some matrix techniques, we obtain the following lemma.

TABLE I
COMPUTATIONAL COMPLEXITY OF DIFFERENT SCHEMES

Scheme	Computational complexity
FAS-OMP	$\mathcal{O}(LPMN^2)$
FAS-ML	$\mathcal{O}(I_o PML(PM + N))$
S-BAR (Stage 1)	$\mathcal{O}(PM(P^2 M^2 + NPM + N^2))$
S-BAR (Stage 2)	$\mathcal{O}(PMN)$

Lemma 3 (MSE of S-BAR for Given Kernel $\mathbf{\Sigma}$): Assume $\mathbf{E}_{\mathbf{h}}(\mathbf{h}\mathbf{h}^H) = \mathbf{\Sigma}_{\text{cov}}$. Given a kernel $\mathbf{\Sigma}$ and a switch matrix $\mathbf{S}$ as the inputs of S-BAR, the MSE of reconstructing $\mathbf{h}$ can be derived as:

$$E = \text{Tr}(\mathbf{\Pi}^H (\mathbf{S}^H \mathbf{\Sigma}_{\text{cov}} \mathbf{S} + \sigma^2 \mathbf{I}_{PM}) \mathbf{\Pi}) - 2\Re(\text{Tr}(\mathbf{\Pi}^H \mathbf{S}^H \mathbf{\Sigma}_{\text{cov}})) + \text{Tr}(\mathbf{\Sigma}_{\text{cov}}), \quad (26)$$

where $\mathbf{\Pi}$ is a matrix function w.r.t $\mathbf{\Sigma}$, given by

$$\mathbf{\Pi} = (\mathbf{S}^H \mathbf{\Sigma} \mathbf{S} + \sigma^2 \mathbf{I}_{PM})^{-1} \mathbf{S}^H \mathbf{\Sigma}. \quad (27)$$

Proof: Constructive proof is given in Appendix C. ■

Lemma 3 characterizes the estimation accuracy of S-BAR for a given kernel $\mathbf{\Sigma}$. Particularly, it allows us to evaluate the estimation error of S-BAR when the selected kernel $\mathbf{\Sigma}$ and the real covariance $\mathbf{\Sigma}_{\text{cov}}$ are mismatched. Besides, we find that E is independent of the real mean of channels, i.e., $\mathbf{E}_{\mathbf{h}}(\mathbf{h})$, which means modeling $\mathbf{h}$ as a zero-mean process does not loss the generality. Then, as the fundamental limit, one may be concerned with the achievable minimum MSE of S-BAR, thus we introduce the following corollary.

Lemma 4 (Achievable Minimum MSE of S-BAR): Assume $\mathbf{E}(\mathbf{h}\mathbf{h}^H) = \mathbf{\Sigma}_{\text{cov}}$, where $\mathbf{\Sigma}_{\text{cov}}$ is rank- K . The achievable minimum MSE of reconstructing $\mathbf{h}$ via S-BAR, i.e. $E_{\min} = \min_{\mathbf{\Sigma}, \mathbf{S}} \mathbf{E}_{\mathbf{h}, \mathbf{z}}(\|\hat{\mathbf{h}} - \mathbf{h}\|^2)$, only depends on the eigenvalues of $\mathbf{\Sigma}_{\text{cov}}$ and the noise power σ^2 , which can be written as:

$$E_{\min} = \sum_{k=1}^K \frac{\lambda_k \sigma^2}{\lambda_k + \sigma^2}, \quad (28)$$

where $\{\lambda_1, \dots, \lambda_K\}$ are the K positive eigenvalues of $\mathbf{\Sigma}_{\text{cov}}$. The equality holds if and only if the kernel is exactly the real covariance of $\mathbf{h}$, i.e., $\mathbf{\Sigma} = \mathbf{\Sigma}_{\text{cov}}$, and the FAS channels are completely observed, i.e., $PM = N$.

Proof: Constructive proof is given in Appendix D. ■

Lemma 4 characterizes the lower bound of MSE while employing S-BAR for FAS channel estimation. Note that, when the dimension of $\mathbf{S}$ is $N \times N$, the reconstructor degenerates into a well-informed estimator. Furthermore, if the kernel $\mathbf{\Sigma}$ is chosen to equal the real channel covariance $\mathbf{\Sigma}_{\text{cov}}$, the MSE of the Bayesian reconstructor will coincide with the well-known posterior covariance formula in linear estimation theory [44]. From the perspective of statistical signal processing, **Lemma 4** reveals the equivalence between a perfect kernel-based Bayesian reconstructor and a LMMSE estimator.

2) *Computational Complexity of S-BAR*: The entire procedures of the existing FAS-OMP reconstructor in **Algorithm 1** [34] and FAS-ML reconstructor in **Algorithm 2** should be employed online. Different from these schemes, the proposed S-BAR incorporates a hybrid offline and online implementation process. Specifically, the signal processing of S-BAR scheme is composed of two stages, the offline design at *Stage 1* and the online regression at *Stage 2*. At *Stage 1*, the computational complexity is dominated by the calculation of posterior covariance Σ_t and weight $\mathbf{W}$. According to (24) and (25), the complexity of *Stage 1* is $\mathcal{O}(P^2 M^2 (P^2 M^2 + NPM + N^2))$. At *Stage 2*, the computational complexity is from the weighted sum of received pilot $\mathbf{y}$, i.e., $\hat{\mathbf{h}} = \mathbf{W}^H \mathbf{y}$. Thus, the computational complexity of *Stage 2* is $\mathcal{O}(PMN)$. For comparison, here we summarize their computational complexities in Table I, wherein I_o denotes the number of iterations required by FAS-ML reconstructor.

Note that, although the complexity of *Stage 1* is high, *Stage 1* can be implemented offline in advance. Then, the calculated switch matrix $\mathbf{S}$ and weight $\mathbf{W}$ can be saved at the BS for the subsequent online regression in *Stage 2*. Therefore, in practical applications, the effective complexity of employing S-BAR is only linear to the number of ports N . Besides, we point out that the designs of $\mathbf{S}$ and $\mathbf{W}$ do not depend on the specific user, which suggests that the proposed S-BAR scheme can be extended to multi-user case without difficulty. In addition, $\hat{\mathbf{h}} = \mathbf{W}^H \mathbf{y}$ in *Stage 2* can be calculated in a parallel way, which means that the time complexity of S-BAR is dimension-independent. These encouraging features further enhance the expansibility and the practicality of our proposed S-BAR.

B. Kernel Selection of S-BAR

Selecting a proper kernel Σ for S-BAR is essential in building an effective regression model. Σ determines the shape and flexibility of the proposed S-BAR, which in turn affects its ability to capture patterns and make accurate reconstruction. Considering the localized correlation property of FAS channels, an ideal kernel should assign higher similarity to nearby ports and decrease influence rapidly with distance. Since the channel covariance does not change so frequently as channels, a well-designed kernel Σ can work for a long time.

1) *Covariance Kernel*: Since the mathematical significance of kernel is the prior covariance, an ideal approach is to use the real covariance of $\mathbf{h}$ as the kernel for reconstruction, i.e., $\Sigma_{\text{cov}} = \mathbf{E}(\mathbf{h}\mathbf{h}^H)$. Before employing S-BAR, we can train an approximated Σ_{cov} based on the CSI knowledge in some existing channel datasets, which can be obtained by

$$\Sigma_{\text{cov}} \approx \frac{1}{R} \sum_{r=1}^R \mathbf{h}_r \mathbf{h}_r^H, \quad (29)$$

where $\mathbf{h}_r$ is the r -th channel for training kernels and R is the number of training timeslots.

2) *Experiential Kernel*: In some scenarios where the explicit CSI is difficult to acquire, training an experiential kernel based on received pilots is more practical. Let $\mathbf{x}_n$ denote the position of the n -th port. After balancing the complexity and practicality, two experiential kernels are recommended:

- **Exponential Kernel**: The exponential kernel Σ_{exp} , also known as the Laplacian kernel, is the most popular choice in Bayesian regression, which is given by

$$\Sigma_{\text{exp}}(n, n') = \exp(-\eta \|\mathbf{x}_n - \mathbf{x}_{n'}\|^2) \quad (30)$$

for all $n, n' \in \{1, \dots, N\}$, where $\eta > 0$ is an adjustable hyperparameter. The exponential kernel is not sensitive to outliers, thus it is suitable to reconstruct channels without obvious regularity.

- **Bessel Kernel**: The Bessel kernel Σ_{bes} is well-suited to model complex-valued data with oscillatory or periodic patterns. The kernel is given by

$$\Sigma_{\text{bes}}(n, n') = J_0(\eta \|\mathbf{x}_n - \mathbf{x}_{n'}\|) \quad (31)$$

for all $n, n' \in \{1, \dots, N\}$, J_0 is the zero-order Bessel function of the first kind. Σ_{bes} has the flexibility to adapt to data that exhibits repeating fluctuations, thus it is suitable to reconstruct the channels with periodic patterns.

The hyperparameter η is the key factor that influences the performance of experiential kernels. We propose to determine the value of η via an ML-based method, so that the selected experiential kernels can well approximate the real channel covariance. Specifically, we assume that R channel realizations are utilized to train an experiential kernel Σ before employing S-BAR. By viewing Σ as a function of hyperparameter η , the ML estimator of η can be written as

$$\{\Sigma^*, \eta^*\} = \underset{\Sigma \in \{\Sigma_{\text{exp}}, \Sigma_{\text{bes}}\}, \eta > 0}{\text{argmax}} \sum_{r=1}^R \ln(P(\mathbf{y}_r | \eta)), \quad (32)$$

wherein the likelihood function is given by

$$P(\mathbf{y}_r | \eta) = \frac{\exp\left(-\mathbf{y}_r^H (\mathbf{S}_r^H \Sigma \mathbf{S}_r + \sigma^2 \mathbf{I}_{PM})^{-1} \mathbf{y}_r\right)}{\pi^{PM} \det(\mathbf{S}_r^H \Sigma \mathbf{S}_r + \sigma^2 \mathbf{I}_{PM})} \quad (33)$$

and Σ can be selected as Σ_{exp} or Σ_{bes} ; $\mathbf{y}_r := \mathbf{S}_r \mathbf{h}_r + \mathbf{z}_r \in \mathbb{C}^{PM}$ denotes the received pilot associated with the r -th training channel $\mathbf{h}_r$; and $\mathbf{S}_r$ is randomly selected from $\mathbf{S}$. Since η is a positive scalar and Σ only has two candidates, a one-dimensional search can be adopted to obtain the optimal solution to problem (32).

V. SIMULATION RESULTS

In this section, simulation results are carried out to verify the effectiveness of the proposed S-BAR. Firstly, the simulation setup and benchmark schemes are specified in Subsection V-A. Then, the estimation behavior of S-BAR is analyzed in Subsection V-B. Next, The impact of noise on the estimation accuracy is shown in Subsection V-C, and the influence of pilot overhead on the estimation accuracy is studied in Subsection V-D. Finally, the impact of electromagnetic (EM) coupling on estimation accuracy is investigated in Subsection V-E.

A. Simulation Setup and Benchmarks

Since we have assumed the normalized transmit power, the receiver signal-to-noise ratio (SNR) is defined as $\text{SNR} = \frac{\mathbf{E}(\|\mathbf{h}\|^2)}{\sigma^2}$, of which the default value is set to 20 dB. Let $\hat{\mathbf{h}}$

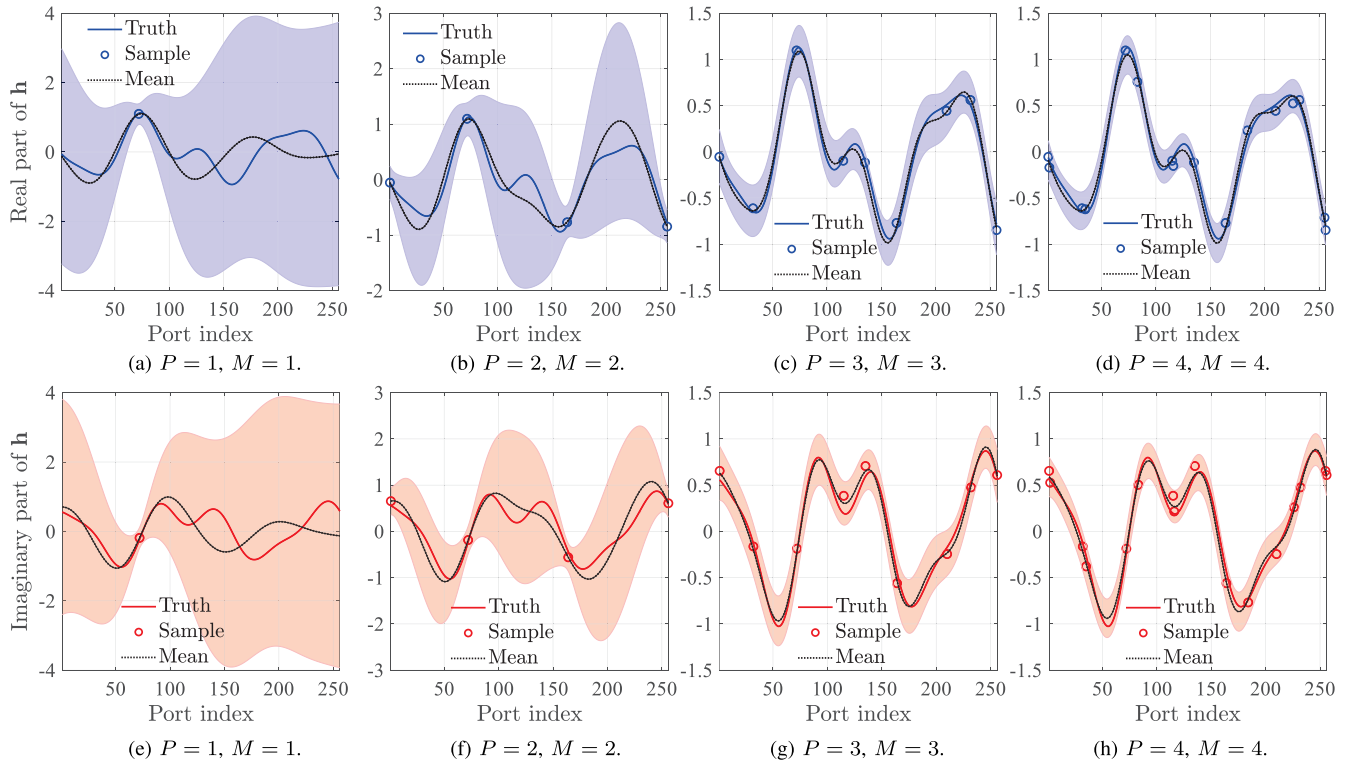


Fig. 3. An illustration of employing S-BAR scheme to estimate FAS channel $\mathbf{h}$. (a)-(d) illustrate the real part of $\mathbf{h}$ versus the index of ports. (e)-(h) illustrate the imaginary part of $\mathbf{h}$ versus the index of ports.

TABLE II
SIMULATION PARAMETERS OF FAS CHANNELS

Parameter	QuaDRiGa [45]	SSC in (5)
Carrier frequency f_c	3.5 GHz	3.5 GHz
Number of clusters	23	9
Number of rays	20	100
Path gains	[46, Table 7.7.1-2]	$\mathcal{CN}(0, 1)$
Incident angles	[46, Table 7.7.1-2]	$\mathcal{U}(-\pi, +\pi)$
Max. Angle spread	5°	5°
Path delays	[46, Table 7.7.1-2]	\
Max. Doppler shift	10 Hz	\

denote the estimated value of channel $\mathbf{h}$. The performance is evaluated by the normalized mean square error (NMSE), which is defined as $\text{NMSE} = \mathbb{E}_{\mathbf{h}, \mathbf{z}} \left(\frac{\|\mathbf{h} - \hat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2} \right)$.

1) *Simulation Setup*: Otherwise particularly specified, the number of FAS ports is set to $N = 256$ and that of fluid antennas is set to $M = 4$. The length of the fluid antenna array is set to $W = 10\lambda$, thus the port spacing should be $d = \frac{W}{N-1}$. The number of pilots is set to $P = 10$. To account for both the model-mismatched case and model-matched case for the existing parametric estimators, the simulations are provided based on both the QuaDRiGa channel model in [45] and the spatially-sparse clustered (SSC) channel model in (5). The channel parameters in 3GPP TR 38.901 [46] is used to generate the QuaDRiGa channels, and the main values are set as shown in Table II. For the kernel settings, we set the number of training timeslots as $R = 100$ to train kernels.

The hyperparameter η is obtained via (32) to generate the exponential kernel Σ_{exp} and the Bessel kernel Σ_{bes} .

2) *Simulation Schemes*: We consider the following four FAS channel estimators for simulations:

- **FAS-OMP**: Set the spatial sparsity L as the twice of the number of clusters C . Then, **Algorithm 1**, i.e., the modified FAS-OMP reconstructor in [34], is employed to reconstruct $\mathbf{h}$.
- **FAS-ML**: Set the spatial sparsity as $L = 2C$. Then, the FAS-ML reconstructor in **Algorithm 2** is employed to estimate channel $\mathbf{h}$. In particular, the path gains and incident angles of $\mathbf{h}$ are initialized via OMP.
- **SeLMMSE**: The SeLMMSE scheme proposed in [33] is adopted to estimate channel $\mathbf{h}$. It can be achieved by measuring channels of PM equally-spaced ports and then using zero-order interpolation to reconstruct $\mathbf{h}$.
- **Proposed S-BAR (Σ)**: Given a kernel Σ , **Algorithm 4**, i.e., the proposed S-BAR, is employed to estimate $\mathbf{h}$.

B. Estimation Behavior of the Proposed S-BAR

To better understand the working principle of the proposed S-BAR for channel estimation, we plot Fig. 3 to intuitively show its behavior for different system parameters. The QuaDRiGa channel model is considered to generate channel $\mathbf{h}$. The ideal covariance kernel Σ_{cov} is used as the input of S-BAR. Fig. 3 (a)-(d) show the real part of $\mathbf{h}$ as a function of the port index, and Fig. 3 (e)-(g) show the imaginary part of $\mathbf{h}$ as a function of the port index. Particularly, the curve “Truth” denotes the real channel $\mathbf{h}$, and the circle marks denote the sampled (measured) channels, which are selected by the

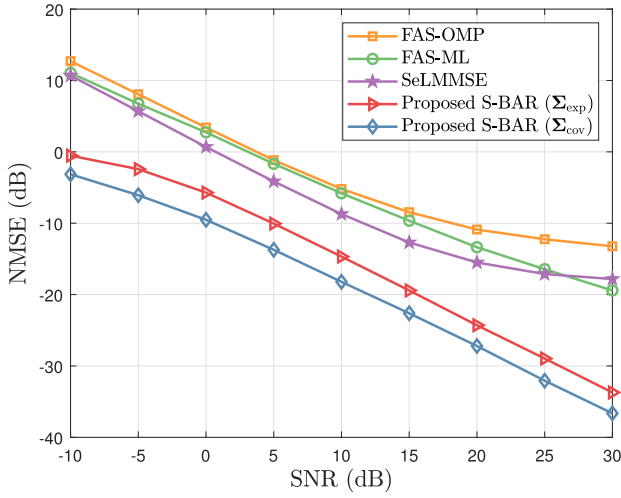


Fig. 4. The NMSE as a function of the receiver SNR under the assumption of QuaDRiGa channel model.

switch matrix $\mathbf{S}$ designed via S-BAR. The black dotted line “Mean” denotes the posterior mean of Bayesian regression $\mu_{\mathbf{h}|\mathbf{y}}$, i.e., the estimated channel $\hat{\mathbf{h}}$. The highlighted shadows in the figures represent the confidence interval of $\mathbf{h}$, defined as $[\mu_{\mathbf{h}|\mathbf{y}}(n) - 3\Sigma_{\mathbf{h}|\mathbf{y}}(n, n), \mu_{\mathbf{h}|\mathbf{y}}(n) + 3\Sigma_{\mathbf{h}|\mathbf{y}}(n, n)]$ for the n -th port channel.

From this figure, we have the following observations. Firstly, as P and M increase, the confidence interval, i.e., the vertical height of shadows, is gradually reduced. It indicates that, more pilots or antennas allow more sampling points for channel reconstruction, which can better eliminate the uncertainty of FAS channels. When the posterior variance $\Sigma_{\mathbf{h}|\mathbf{y}}(n, n)$ becomes sufficiently small, the posterior mean μ_{Ω} can well approximate $\mathbf{h}$. Secondly, Fig. 3 (c) and Fig. 3 (g) show that, nine samples are enough for the posterior mean μ_{Ω} to well approximate a 256-dimensional $\mathbf{h}$. This result suggests that, benefiting from the prior knowledge of strongly correlated channels, very few pilots and fluid antennas are sufficient to accurately reconstruct channels. Thirdly, one can note that the samples in Fig. 3 (d) and Fig. 3 (h) are close. The reason is that, the inflection regions of a function are more informative than its flat regions, which indicates that the inflection parts require more dense samples to realize well regression. That’s why the shadow heights in inflection regions are larger than those in flat regions, as shown in Fig. 3 (c) and Fig. 3 (g). Although some ports in inflection regions have been selected as samples, they are still hard to sufficiently eliminate the channel uncertainty of these regions. Thereby, in Fig. 3 (d) and Fig. 3 (h), some samples in inflection regions are selected to further eliminate the channel uncertainty of these regions.

C. NMSE Versus the Receiver SNR

We plot the NMSE as a function of the SNR in Fig. 4 and Fig. 5, which follow the assumptions of QuaDRiGa model and SSC model in (5), respectively. Note that, the FAS-OMP and FAS-ML reconstructors have assumed the SSC model in (5). Due to the lack of obvious regularity for QuaDRiGa channels, the exponential kernel Σ_{exp} is selected as the input of S-BAR in the QuaDRiGa case. Due to the

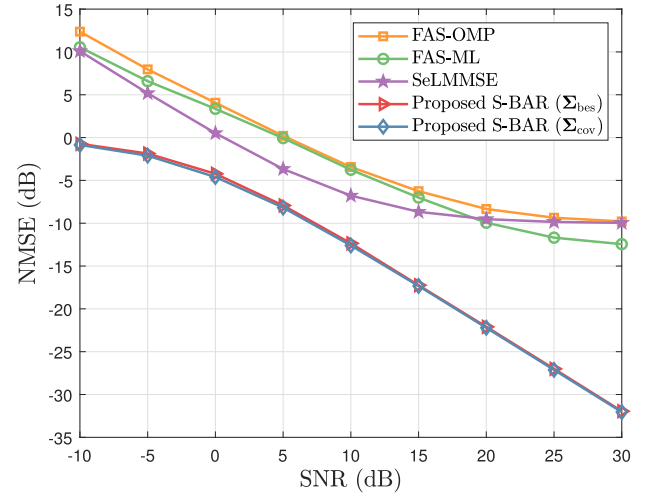


Fig. 5. The NMSE as a function of the receiver SNR under the assumption of SSC channel model.

periodic patterns of SSC channels in the spatial domain, the Bessel kernel Σ_{bes} is selected as the input of S-BAR in the SSC case [6], [7]. To show the performance of S-BAR in the ideal case, the pre-trained covariance kernel Σ_{cov} is considered in both cases. From these two figures, we have the following observations.

Firstly, as the receiver SNR rises, the NMSEs of all estimators decrease rapidly. Particularly, one can find that the proposed S-BAR achieves the highest estimation accuracy for both SSC channels and QuaDRiGa channels. The reason is that, the existing methods do not utilize the prior knowledge of FAS channels for estimation. For FAS-OMP and FAS-ML reconstructors, the measured ports are randomly selected, which means that the information provided by the measured channels may not capture all patterns of $\mathbf{h}$. For SeLMMSE, although channel correlation is partially exploited, the unmeasured channels are obtained by zero-order interpolation. From the statistical perspective, the potential estimation errors of the unmeasured channels are not considered by SeLMMSE. In contrast, the proposed S-BAR has incorporated the effect of prior information into its estimator, which can reduce the potential estimation errors of all channels simultaneously. Besides, S-BAR does not assume a specific channel model, which implies that it is not influenced by the model mismatch as the FAS-OMP and FAS-ML reconstructors.

Secondly, we observe that the S-BAR enabled by the exponential kernel Σ_{exp} and Bessel kernel Σ_{bes} can achieve similar performance as the covariance kernel Σ_{cov} . In particular, the curves of the two S-BAR schemes are almost coincident in the SSC channel case. Recall that Σ_{exp} and Σ_{bes} are generated by experiential parameters, while Σ_{cov} is trained from real channel data. This observation indicates that, even if the real channel covariance Σ_{cov} is totally unknown, the experiential kernel Σ_{exp} and Σ_{bes} can still enable S-BAR to achieve considerable performance. In other words, to estimate channels accurately, the proposed S-BAR only needs the “virtual” prior knowledge provided by carefully selected experiential kernels, while the “real” channel covariance is actually not necessary.

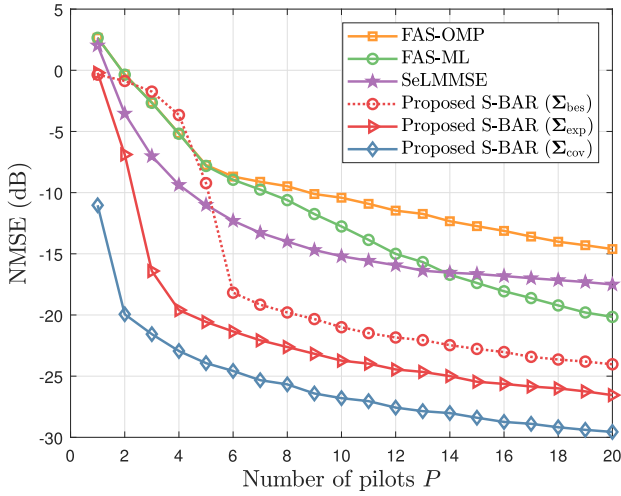


Fig. 6. The NMSE as a function of the number of pilots P under the assumption of QuaDRiGa channel model.

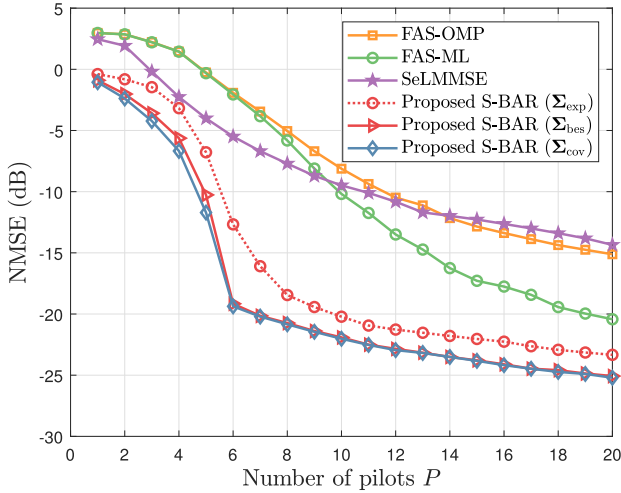


Fig. 7. The NMSE as a function of the number of pilots P under the assumption of SSC channel model.

D. NMSE Versus the Number of Pilots

We plot the NMSE as a function of the number of pilots P in Fig. 6 and Fig. 7, which follow the assumptions of QuaDRiGa model and SSC model, respectively. In addition, to show the influence of non-preferred kernels on S-BAR in the two channel cases, we have also added the baseline that Σ_{bes} is selected as the input of S-BAR in Fig. 6 and the baseline that Σ_{exp} is selected as the input of S-BAR in Fig. 7. These new baselines are expressed as dotted red lines in the figures.

We observe from these two figures that, to achieve the same estimation accuracy, the proposed S-BAR scheme consumes a much lower pilot overhead than the benchmark schemes. For example, to achieve an NMSE of -15 dB in the SSC channel case, the numbers of pilots required by FAS-OMP, FAS-ML, SeLMMSE, and the proposed S-BAR with the preferred kernel input are $P = 20, 13, 20$, and 5 , respectively. We can conclude that, compared to the state-of-art schemes, the proposed S-BAR scheme can reduce the pilot overhead of about 50%. Particularly, in Fig. 6, one can find that the curve “Proposed S-BAR (Σ_{cov})” begins from NMSE = -10 dB when $P = 1$.

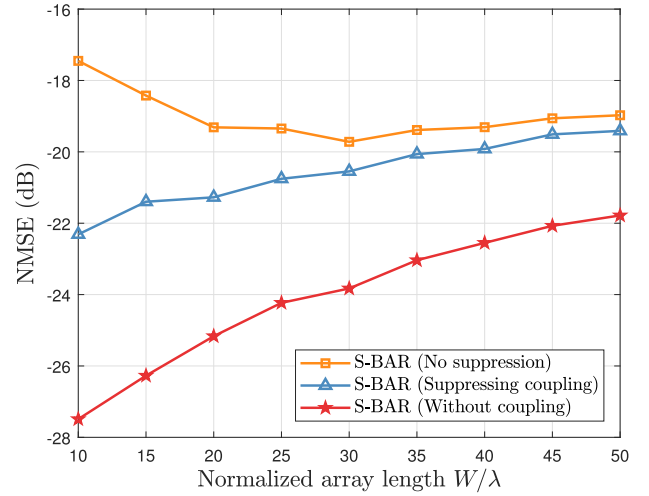


Fig. 8. The NMSE as a function of the normalized array length W/λ .

It implies that, aided by kernel Σ_{cov} , the acquired information of four observations ($P \times M = 4$) is already sufficient for S-BAR to accurately recover the channels of most ports.

Besides, we also find that, even if the kernel input is not preferred, the proposed S-BAR can still achieve satisfactory performance. For example, when Σ_{bes} is selected as the input of S-BAR in the QuaDRiGa channel case, the NMSE for S-BAR can achieve -21 dB when $P = 10$, which is 6 dB lower than that for the best-performed SeLMMSE. In the SSC channel case, when Σ_{exp} is selected as the input of S-BAR, its NMSE can achieve -20 dB, which is about 10 dB lower than the NMSE for the best-performed FAS-ML. These results have further verified the effectiveness of our proposed S-BAR.

Finally, in the SSC channel case, the NMSE for S-BAR decreases quickly when $P < 6$, while it decreases slowly when $P > 6$. The reason is that, when $P < 6$, the proposed S-BAR has not captured all spatial features of SSC channels, thus the estimation error is mainly caused by the incomplete regression. When $P > 6$, with sufficient channel sampling, the channel information captured by S-BAR can well describe the periodic feature of SSC channels. In this case, the estimation error is mainly caused by the measurement noise. On the other hand, in the QuaDRiGa channel case, this interesting phenomenon is not observed. The reason is that, unlike the SSC channels, the spatial periodicity is not obvious for the QuaDRiGa channels, which makes it difficult for S-BAR to fully capture all patterns of channels via a few samples. It explains why Σ_{bes} is not the preferred kernel for the QuaDRiGa channel case.

E. Impact of EM Coupling on Estimation Accuracy

For some FASs whose multiple antennas can move physically, the effect of EM coupling among fluid antennas can be avoided by forcing that the antenna spacing is always larger than $\lambda/2$ [31], [32]. In contrast, a few FASs are realized by pixel-like switch networks [23]. For such FAS structures, the EM coupling among ports cannot be ignored, thus only the coupled channels can be estimated. To take this effect into account, we introduce the EM coupling model in [47]. To be specific, coupling matrix $\mathbf{C}_{em} \in \mathbb{C}^{N \times N}$ is used to

modify the FAS channel $\mathbf{h}$, which can be obtained by field tests or full-wave simulations. As a typical example, here we consider dipole antennas with $\lambda/2$ length, $\lambda/100$ width, and $d := \frac{W}{N-1}$ spacing. With these parameters, the desired $\mathbf{C}_{\text{em}}$ can be directly generated by the Matlab Antenna Toolbox. In this context, the equivalent channel can be modeled as $\bar{\mathbf{h}} = \mathbf{C}_{\text{em}}^{1/2} \mathbf{h}$, and the equivalent kernel can be derived as $\mathbf{C}_{\text{em}}^{1/2} \Sigma (\mathbf{C}_{\text{em}}^{1/2})^H$. To show the effectiveness of S-BAR against the EM coupling effect, three schemes are compared:

- **No suppression:** Under the influence of EM coupling, the proposed S-BAR ignores coupling and estimates the equivalent channel $\bar{\mathbf{h}}$ based on the original kernel Σ_{cov} .
- **Suppressing coupling:** Under the influence of EM coupling, the proposed S-BAR is employed to estimate $\bar{\mathbf{h}}$ base on the modified kernel $\mathbf{C}_{\text{em}}^{1/2} \Sigma_{\text{cov}} (\mathbf{C}_{\text{em}}^{1/2})^H$.
- **Without coupling:** Assume that the inter-port EM coupling can be fully eliminated by hardware isolation, i.e., $\mathbf{C}_{\text{em}} = \mathbf{I}_N$. The proposed S-BAR is employed to estimate ideal $\mathbf{h}$ based on the ideal kernel Σ_{cov} .

Since the EM coupling depends on the port spacing d , we fix M and then plot the NMSE versus the normalized array length W/λ in Fig. 8, where the QuaDRiGa channels are considered. We observe that, despite the EM coupling among ports, the proposed S-BAR can still achieve considerable estimation performances, and the NMSE is still lower than -15 dB. It is interesting to find that, as the port spacing increases, the NMSE for the scheme “No suppression” decreases first and then rises. The reason is that, the estimation performance of S-BAR is influenced by the two factors, i.e., spatial correlation and EM coupling. When $W < 30\lambda$, the main factor that limits the estimation performance of “No suppression” scheme is EM coupling. As the port spacing increases, the effect of EM coupling is reduced, leading to its higher estimation accuracy. When $W > 30\lambda$, the main factor becomes the spatial correlation. As the port spacing increases, the spatial correlation among ports becomes weaker. Since the accuracy improvement of S-BAR is realized by utilizing the strong correlation among ports, its performance gains become less accordingly. In contrast, the effect of EM coupling has been eliminated by “Suppressing coupling” through considering the equivalent kernel $\mathbf{C}_{\text{em}}^{1/2} \Sigma_{\text{cov}} (\mathbf{C}_{\text{em}}^{1/2})^H$, and $\mathbf{C}_{\text{em}}$ is not taken into account by “Without coupling” case. Thus, the accuracy of these two estimators mainly depends on the spatial correlation. As the port spacing increases, the weaker spatial correlation becomes the bottleneck limiting their performance, which leads to their continuously rising NMSEs.

VI. CONCLUSION

In this paper, we have proposed S-BAR as a general solution to estimate channels in FASs. Specifically, this paper is the first attempt to introduce Bayesian inference into the FAS channel estimation. Different from the existing FAS channel estimators relying on channel assumptions, the general S-BAR utilizes the experiential kernel to estimate channels in a non-parametric way. Inspired by the Bayesian regression, the proposed S-BAR can select a few informative channels for measurement and combine them with the experiential kernel to reconstruct

high-dimensional FAS channels. Simulation results reveal that, in both model-mismatched and model-matched cases, the proposed S-BAR can achieve higher estimation accuracy than the existing schemes relying on channel assumptions.

For the follow-up works, the extension to wideband channel estimation for FASs will be interesting. Since the channels of multiple carriers share the same ports, the port selection should balance the uncertainty of the channels on different subcarriers. Besides, thanks to the generality of Bayesian regression, the proposed S-BAR can be extended to realize channel estimation in the general phased-array case [48]. In addition, since reconfigurable intelligent surfaces (RISs) can provide additional control degrees of freedom (DoFs) for FAS channels [49], the cooperation between RISs and FASs may be able to improve the accuracy of CSI acquisitions [50].

APPENDIX A PROOF OF Lemma 1

Using some matrix partition operations, the mutual information $I(\mathbf{y}_{t+1}; \mathbf{h})$ can be rewritten as

$$\begin{aligned} I(\mathbf{y}_{t+1}; \mathbf{h}) &= \log_2 \det \left(\mathbf{I}_{t+1} + \frac{1}{\sigma^2} \mathbf{S}_{t+1}^H \Sigma \mathbf{S}_{t+1} \right) \\ &= \log_2 \det \begin{pmatrix} \mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{S}_t^H \Sigma \mathbf{S}_t & \frac{1}{\sigma^2} \mathbf{S}_t^H \Sigma \mathbf{s}_{t+1} \\ \frac{1}{\sigma^2} \mathbf{s}_{t+1}^H \Sigma \mathbf{S}_t & 1 + \frac{1}{\sigma^2} \mathbf{s}_{t+1}^H \Sigma \mathbf{s}_{t+1} \end{pmatrix} \\ &\stackrel{(a)}{=} \log_2 \det \begin{pmatrix} \mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{S}_t^H \Sigma \mathbf{S}_t & \frac{1}{\sigma^2} \mathbf{S}_t^H \Sigma \mathbf{s}_{t+1} \\ \mathbf{0}_{1 \times t} & 1 + \frac{1}{\sigma^2} \mathbf{s}_{t+1}^H \Sigma \mathbf{s}_{t+1} \end{pmatrix} \\ &\stackrel{(b)}{=} I(\mathbf{y}_t; \mathbf{h}) + \log_2 \left(1 + \frac{1}{\sigma^2} \mathbf{s}_{t+1}^H \Sigma \mathbf{s}_{t+1} \right), \end{aligned} \quad (34)$$

where (a) holds according to the matrix triangularization:

$$\begin{aligned} &\begin{bmatrix} \mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{S}_t^H \Sigma \mathbf{S}_t & \frac{1}{\sigma^2} \mathbf{S}_t^H \Sigma \mathbf{s}_{t+1} \\ \mathbf{0}_{1 \times t} & 1 + \frac{1}{\sigma^2} \mathbf{s}_{t+1}^H \Sigma \mathbf{s}_{t+1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_t & \mathbf{0}_t \\ -\mathbf{s}_{t+1}^H \Sigma \mathbf{S}_t (\mathbf{S}_t^H \Sigma \mathbf{S}_t + \sigma^2 \mathbf{I}_t)^{-1} & 1 \end{bmatrix} \\ &\quad \times \begin{bmatrix} \mathbf{I}_t + \frac{1}{\sigma^2} \mathbf{S}_t^H \Sigma \mathbf{S}_t & \frac{1}{\sigma^2} \mathbf{S}_t^H \Sigma \mathbf{s}_{t+1} \\ \frac{1}{\sigma^2} \mathbf{s}_{t+1}^H \Sigma \mathbf{S}_t & 1 + \frac{1}{\sigma^2} \mathbf{s}_{t+1}^H \Sigma \mathbf{s}_{t+1} \end{bmatrix}, \end{aligned} \quad (35)$$

and (b) holds according to the definition in (19). Clearly, maximizing $I(\mathbf{y}_{t+1}; \mathbf{h}) - I(\mathbf{y}_t; \mathbf{h}) = \log_2 \left(1 + \frac{1}{\sigma^2} \mathbf{s}_{t+1}^H \Sigma \mathbf{s}_{t+1} \right)$ is equivalent to finding a proper $\mathbf{s}$ that maximizes $\mathbf{s}^H \Sigma \mathbf{s}$, which completes the proof.

APPENDIX B PROOF OF Lemma 2

To obtain the recursion formula of Σ_t , the key step is to use block-matrix inversion formula in (36), shown at the top of the next page, to process the term $(\mathbf{S}_t^H \Sigma \mathbf{S}_t + \sigma^2 \mathbf{I}_t)^{-1}$ in (21). Letting $\Xi_t = \mathbf{S}_t^H \Sigma \mathbf{S}_t + \sigma^2 \mathbf{I}_t$, we have

$$\Xi_t = \begin{bmatrix} \Xi_{t-1} & \mathbf{S}_{t-1}^H \Sigma \mathbf{s}_t \\ \mathbf{s}_t^H \Sigma \mathbf{S}_{t-1} & \sigma^2 + \mathbf{s}_t^H \Sigma \mathbf{s}_t \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}. \quad (37)$$

Substituting the corresponding $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ in (37) into (36), the inverse of Ξ_t can be represented by the

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \quad (36)$$

combination of Ξ_{t-1}^{-1} , $\mathbf{S}_{t-1}$, and $\mathbf{s}_t$, written as

$$\Xi_t^{-1} = \begin{bmatrix} \Xi_{t-1}^{-1} + \frac{\Xi_{t-1}^{-1}\mathbf{S}_{t-1}^H\mathbf{\Sigma}\mathbf{s}_t\mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\Xi_{t-1}^{-1}}{\sigma^2 + \mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\mathbf{s}_t} & \frac{-\Xi_{t-1}^{-1}\mathbf{S}_{t-1}^H\mathbf{\Sigma}\mathbf{s}_t}{\sigma^2 + \mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\mathbf{s}_t} \\ \frac{-\mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\Xi_{t-1}^{-1}}{\sigma^2 + \mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\mathbf{s}_t} & \frac{1}{\sigma^2 + \mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\mathbf{s}_t} \end{bmatrix}. \quad (38)$$

Then, by taking Ξ_t^{-1} and $\mathbf{S}_t = [\mathbf{S}_{t-1} \ \mathbf{s}_t]$ back into the posterior covariance Σ_t in (21), we have

$$\begin{aligned} \Sigma_t &= \Sigma - \Sigma\mathbf{S}_t\Xi_t^{-1}\mathbf{S}_t^H\Sigma \\ &= \Sigma - \Sigma\mathbf{S}_{t-1}\Xi_{t-1}^{-1}\mathbf{S}_{t-1}^H\Sigma + \frac{1}{\sigma^2 + \mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\mathbf{s}_t} \\ &\quad \times \left(-(\Sigma - \Sigma\mathbf{S}_{t-1})\mathbf{s}_t\mathbf{s}_t^H(\Sigma - \Sigma\mathbf{S}_{t-1}) \right. \\ &\quad \left. + (\Sigma - \Sigma\mathbf{S}_{t-1})\mathbf{s}_t\mathbf{s}_t^H\Sigma + \Sigma\mathbf{s}_t\mathbf{s}_t^H(\Sigma - \Sigma\mathbf{S}_{t-1}) + \Sigma\mathbf{s}_t\mathbf{s}_t^H\Sigma \right) \\ &= \Sigma_{t-1} - \frac{\Sigma_{t-1}\mathbf{s}_t\mathbf{s}_t^H\Sigma_{t-1}}{\sigma^2 + \mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\mathbf{s}_t} \\ &\stackrel{(a)}{=} \Sigma_{t-1} - \frac{\Sigma_{t-1}(:, n_t)\Sigma_{t-1}(n_t, :)}{\Sigma_{t-1}(n_t, n_t) + \sigma^2}, \end{aligned} \quad (39)$$

where (a) holds since $\mathbf{s}_t^H\mathbf{\Sigma}\mathbf{S}_{t-1}\mathbf{s}_t = \Sigma_{t-1}(n_t, n_t)$ and $\Sigma\mathbf{s}_t = \Sigma_{t-1}(:, n_t)$. This completes the proof.

APPENDIX C PROOF OF Lemma 3

When $t = PM$, according to the definition of $\hat{\mathbf{h}}$ in (23), the square error of employing S-BAR can be rewritten as

$$\begin{aligned} \|\hat{\mathbf{h}} - \mathbf{h}\|^2 &\stackrel{(a)}{=} \left\| \left((\mathbf{S}^H\Sigma)^H (\mathbf{S}^H\Sigma\mathbf{S} + \sigma^2\mathbf{I}_{PM})^{-1} \mathbf{S}^H - \mathbf{I}_N \right) \mathbf{h} \right. \\ &\quad \left. + (\mathbf{S}^H\Sigma)^H (\mathbf{S}^H\Sigma\mathbf{S} + \sigma^2\mathbf{I}_{PM})^{-1} \mathbf{z} \right\|^2 \\ &\stackrel{(b)}{=} \left\| (\Pi^H\mathbf{S}^H - \mathbf{I}_N) \mathbf{h} + \Pi^H\mathbf{z} \right\|^2 \end{aligned} \quad (40)$$

where (a) holds according to (3); (b) holds according to (27). Next, recalling the properties $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_{PM}, \sigma^2\mathbf{I}_{PM})$ and $\mathbf{E}(\mathbf{h}\mathbf{h}^H) = \Sigma_{\text{cov}}$, the MSE can be derived as

$$\begin{aligned} E &= \mathbf{E} \left(\|\mu_\Omega - \mathbf{h}\|^2 \right) \\ &\stackrel{(c)}{=} \text{Tr} \left((\Pi^H\mathbf{S}^H - \mathbf{I}_N) \Sigma_{\text{cov}} (\Pi^H\mathbf{S}^H - \mathbf{I}_N)^H \right) \\ &\quad + \sigma^2 \text{Tr}(\Pi^H\Pi) = \text{Tr}(\Pi^H (\mathbf{S}^H\Sigma_{\text{cov}}\mathbf{S} + \sigma^2\mathbf{I}_{PM}) \Pi) \\ &\quad - 2\Re(\text{Tr}(\Pi^H\mathbf{S}^H\Sigma_{\text{cov}})) + \text{Tr}(\Sigma_{\text{cov}}), \end{aligned} \quad (41)$$

where (c) holds since $\|\mathbf{x}\|^2 = \text{Tr}(\mathbf{x}\mathbf{x}^H)$ for any vector $\mathbf{x}$. This completes the proof.

APPENDIX D PROOF OF Lemma 4

Observing (26), one can find that the MSE E is a quadratic form w.r.t matrix Π . Thus, the minimum MSE can be achieved by finding an Σ such that $\frac{\partial E}{\partial \Pi} = \mathbf{0}_{PM \times N}$ holds. According

to (26), the partial derivative of E with respect to Π can be derived as

$$\frac{\partial E}{\partial \Pi} = (\mathbf{S}^H\Sigma_{\text{cov}}\mathbf{S} + \sigma^2\mathbf{I}_{PM})^* \Pi^* - (\Sigma_{\text{cov}}\mathbf{S})^T. \quad (42)$$

By letting $\frac{\partial E}{\partial \Pi} = \mathbf{0}_{PM \times N}$ and substituting (27) into (42), the original proof can be reduced to prove that

$$\begin{aligned} &(\mathbf{S}^H\Sigma\mathbf{S} + \sigma^2\mathbf{I}_{PM})^{-1} \mathbf{S}^H\Sigma \\ &= (\mathbf{S}^H\Sigma_{\text{cov}}\mathbf{S} + \sigma^2\mathbf{I}_{PM})^{-1} \mathbf{S}^H\Sigma_{\text{cov}}. \end{aligned} \quad (43)$$

Since the proposed S-BAR does not depend on the value of noise power, (43) should hold for arbitrary σ^2 . By considering the special case when $\sigma^2 \rightarrow \infty$, (43) is equivalent to

$$\mathbf{S}^H(\Sigma - \Sigma_{\text{cov}}) = \mathbf{0}_{PM \times N} \quad (44)$$

Furthermore, since our derivation also does not impose the specific form of $\mathbf{S}$, $\Sigma = \Sigma_{\text{cov}}$ is necessary to make the equation always hold. By substituting $\Sigma = \Sigma_{\text{cov}}$ into (26), the MSE E can be rewritten as

$$\begin{aligned} E &= \text{Tr}(\Sigma_{\text{cov}}) - \text{Tr} \left(\Sigma_{\text{cov}}\mathbf{S}(\mathbf{S}^H\Sigma_{\text{cov}}\mathbf{S} + \sigma^2\mathbf{I}_{PM})^{-1} \mathbf{S}^H\Sigma_{\text{cov}} \right) \\ &= \text{Tr}(\Sigma_{\text{h|y}}) \big|_{\Sigma=\Sigma_{\text{cov}}} \end{aligned} \quad (45)$$

which is exactly the trace of the posterior covariance $\Sigma_{\text{h|y}}$ in (16) when $\Sigma = \Sigma_{\text{cov}}$. In Bayesian linear regression, the posterior variance of sample path, i.e., each diagonal entry of $\Sigma_{\text{h|y}}$, monotonically decreases as the number of samples increases [43]. Therefore, the MSE E achieves its minimum when all port channels are fully observed (although it may be impractical), i.e., $PM = N$. By substituting $PM = N$ into (45), we have

$$\begin{aligned} E_{\min} &= \text{Tr}(\Sigma_{\text{cov}}) - \text{Tr} \left(\Sigma_{\text{cov}}(\Sigma_{\text{cov}} + \sigma^2\mathbf{I}_N)^{-1} \Sigma_{\text{cov}} \right) \\ &\stackrel{(a)}{=} \text{Tr} \left((\Sigma_{\text{cov}}^{-1} + \sigma^{-2}\mathbf{I})^{-1} \right) \\ &\stackrel{(b)}{=} \sum_{k=1}^K \frac{1}{\frac{1}{\lambda_k} + \frac{1}{\sigma^2}} = \sum_{k=1}^K \frac{\lambda_k \sigma^2}{\lambda_k + \sigma^2}, \end{aligned} \quad (46)$$

where (a) holds by using the well-known Sherman-Morrison-Woodbury formula $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{CDA}^{-1}\mathbf{B})^{-1}\mathbf{CDA}^{-1}$ and (b) holds according to the eigenvalue decomposition $\Sigma_{\text{cov}} = \mathbf{U}\text{diag}(\lambda_1, \dots, \lambda_K)\mathbf{U}^H$ wherein $\mathbf{U} \in \mathbb{C}^{N \times K}$ is an orthogonal matrix. This completes the proof.

REFERENCES

- [1] Z. Zhang, J. Zhu, L. Dai, and R. W. Heath Jr., "Successive Bayesian reconstructor for FAS channel estimation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Dubai, United Arab Emirates, Apr. 2024, pp. 1–5.
- [2] Z. Gao, S. Liu, Y. Su, Z. Li, and D. Zheng, "Hybrid knowledge-data driven channel semantic acquisition and beamforming for cell-free massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 5, pp. 964–979, Sep. 2023.

- [3] K. Wang et al., "Knowledge and data dual-driven channel estimation and feedback for ultra-massive MIMO systems under hybrid field beam squint effect," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11240–11259, Sep. 2024.
- [4] Y. Wang, Z. Gao, D. Zheng, S. Chen, D. Gunduz, and H. V. Poor, "Transformer-empowered 6G intelligent networks: From massive MIMO processing to semantic communication," *IEEE Wireless Commun.*, vol. 30, no. 6, pp. 127–135, Dec. 2023.
- [5] K. K. Wong, A. Shojaefard, K.-F. Tong, and Y. Zhang, "Performance limits of fluid antenna systems," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2469–2472, Nov. 2020.
- [6] K.-K. Wong, A. Shojaefard, K.-F. Tong, and Y. Zhang, "Fluid antenna systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1950–1962, Mar. 2021.
- [7] K.-K. Wong and K.-F. Tong, "Fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4801–4815, Jul. 2022.
- [8] L. Zhu et al., "Movable-antenna array enhanced beamforming: Achieving full array gain with null steering," *IEEE Commun. Lett.*, vol. 27, no. 12, pp. 3340–3344, Dec. 2023.
- [9] L. Zhu, W. Ma, B. Ning, and R. Zhang, "Movable-antenna enhanced multiuser communication via antenna position optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7214–7229, Jul. 2024.
- [10] 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.
- [11] L. Zhu, W. Ma, and R. Zhang, "Movable antennas for wireless communication: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 62, no. 6, pp. 114–120, Jun. 2024.
- [12] K.-K. Wong, W. K. New, X. Hao, K.-F. Tong, and C.-B. Chae, "Fluid antenna system—Part I: Preliminaries," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 1919–1923, Aug. 2023.
- [13] K.-K. Wong, K.-F. Tong, Y. Shen, Y. Chen, and Y. Zhang, "Bruce lee-inspired fluid antenna system: Six research topics and the potentials for 6G," *Frontiers Commun. Netw.*, vol. 3, Mar. 2022, Art. no. 853416.
- [14] A. M. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, Jun. 2007.
- [15] V. Raghavan and A. M. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 345–364, Jan. 2011.
- [16] N. Gultati and K. R. Dandekar, "Learning state selection for reconfigurable antennas: A multi-armed bandit approach," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1027–1038, Mar. 2014.
- [17] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.
- [18] A. Alkhateeb, Y.-H. Nam, J. Zhang, and R. W. Heath, "Massive MIMO combining with switches," *IEEE Wireless Commun. Lett.*, vol. 5, no. 3, pp. 232–235, Jun. 2016.
- [19] K. N. Paracha, A. D. Butt, A. S. Alghamdi, S. A. Babale, and P. J. Soh, "Liquid metal antennas: Materials, fabrication and applications," *Sensors*, vol. 20, no. 1, p. 177, Dec. 2019.
- [20] C. Borda-Fortuny, L. Cai, K. F. Tong, and K.-K. Wong, "Low-cost 3D-printed coupling-fed frequency agile fluidic monopole antenna system," *IEEE Access*, vol. 7, pp. 95058–95064, 2019.
- [21] B. Cetiner, H. Jafarkhani, J.-Y. Qian, H. J. Yoo, A. Grau, and F. De Flaviis, "Multifunctional reconfigurable MEMS integrated antennas for adaptive MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 12, pp. 62–70, Dec. 2004.
- [22] G. J. Hayes, J.-H. So, A. Qusba, M. D. Dickey, and G. Lazzi, "Flexible liquid metal alloy (EGaIn) microstrip patch antenna," *IEEE Trans. Antennas Propag.*, vol. 60, no. 5, pp. 2151–2156, May 2012.
- [23] D. Rodrigo, B. A. Cetiner, and L. Jofre, "Frequency, radiation pattern and polarization reconfigurable antenna using a parasitic pixel layer," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3422–3427, Jun. 2014.
- [24] K. K. Wong, K. F. Tong, Y. Chen, and Y. Zhang, "Closed-form expressions for spatial correlation parameters for performance analysis of fluid antenna systems," *Electron. Lett.*, vol. 58, no. 11, pp. 454–457, Apr. 2022.
- [25] P. Mukherjee, C. Psomas, and I. Krikidis, "On the level crossing rate of fluid antenna systems," in *Proc. IEEE 23rd Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Oulu, Finland, Jul. 2022, pp. 1–5.
- [26] L. Tlebaldiyeva, G. Nauryzbayev, S. Arzykulov, A. Eltawil, and T. Tsiftsis, "Enhancing QoS through fluid antenna systems over correlated Nakagami- m fading channels," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Austin, TX, USA, Apr. 2022, pp. 78–83.
- [27] Z. Chai, K. Wong, K. Tong, Y. Chen, and Y. Zhang, "Port selection for fluid antenna systems," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1180–1184, May 2022.
- [28] K.-K. Wong, K.-F. Tong, Y. Chen, and Y. Zhang, "Fast fluid antenna multiple access enabling massive connectivity," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 711–715, Feb. 2023.
- [29] N. Waqar, K. Wong, K. Tong, A. Sharples, and Y. Zhang, "Deep learning enabled slow fluid antenna multiple access," *IEEE Commun. Lett.*, vol. 27, no. 3, pp. 861–865, Mar. 2023.
- [30] Z. Cheng, N. Li, J. Zhu, and C. Ouyang, "Sum-rate maximization for movable antenna enabled multiuser communications," 2023, *arXiv:2309.11135*.
- [31] Y. Wu, D. Xu, D. W. K. Ng, W. Gerstacker, and R. Schober, "Movable antenna-enhanced multiuser communication: Optimal discrete antenna positioning and beamforming," 2023, *arXiv:2308.02304*.
- [32] Z. Xiao, X. Pi, L. Zhu, X.-G. Xia, and R. Zhang, "Multiuser communications with movable-antenna base station: Joint antenna positioning, receive combining, and power control," *IEEE Trans. Wireless Commun.*, early access, Nov. 1, 2024, doi: [10.1109/TWC.2024.3486320](https://doi.org/10.1109/TWC.2024.3486320).
- [33] C. Skouroumounis and I. Krikidis, "Fluid antenna with linear MMSE channel estimation for large-scale cellular networks," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1112–1125, Feb. 2023.
- [34] W. Ma, L. Zhu, and R. Zhang, "Compressed sensing based channel estimation for movable antenna communications," *IEEE Commun. Lett.*, vol. 27, no. 10, pp. 2747–2751, Oct. 2023.
- [35] R. Wang, Y. Chen, Y. Hou, K. K. Wong, and X. Tao, "Estimation of channel parameters for port selection in millimeter-wave fluid antenna systems," in *Proc. IEEE/CIC Int. Conf. Commun.*, Chengdu, China, Aug. 2023, pp. 1–6.
- [36] K.-K. Wong, K.-F. Tong, Y. Chen, Y. Zhang, and C.-B. Chae, "Opportunistic fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7819–7833, Nov. 2023.
- [37] 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.
- [38] X. Gao, L. Dai, S. Han, and R. W. Heath Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
- [39] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.
- [40] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2663–2677, Apr. 2022.
- [41] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, (2014). *Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers*. [Online]. Available: https://stanford.edu/~boyd/papers/pdf/admm_distr_stats.pdf
- [42] J. An, C. Yuen, C. Huang, M. Debbah, H. Vincent Poor, and L. Hanzo, "A tutorial on holographic MIMO communications—Part I: Channel modeling and channel estimation," *IEEE Commun. Lett.*, vol. 27, no. 7, pp. 1664–1668, Jul. 2023.
- [43] M. A. Oliver and R. Webster, "Kriging: A method of interpolation for geographical information systems," *Int. J. Geograph. Inf. Syst.*, vol. 4, no. 3, pp. 313–332, 1990.
- [44] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [45] S. Jaekel, L. Raschkowski, K. Borner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.
- [46] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document TR 38.901, 3GPP, Dec. 2019.
- [47] T. Svantesson and A. Ranheim, "Mutual coupling effects on the capacity of multielement antenna systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, May 2001, pp. 2485–2488.
- [48] M. Cui, Z. Zhang, L. Dai, and K. Huang, "Ice-filling: Near-optimal channel estimation for dense array systems," 2024, *arXiv:2404.06806*.

- [49] Z. Zhang et al., "Active RIS vs. passive RIS: Which will prevail in 6G?" *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1707–1725, Mar. 2023.
- [50] A. Shojaeifard et al., "MIMO evolution beyond 5G through reconfigurable intelligent surfaces and fluid antenna systems," *Proc. IEEE*, vol. 110, no. 9, pp. 1244–1265, Sep. 2022.



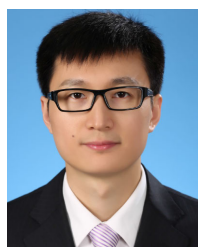
Zijian Zhang (Graduate Student Member, IEEE) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2020, where he is currently pursuing the Ph.D. degree.

He has authored several articles of *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, and *IEEE TRANSACTIONS ON COMMUNICATIONS*. His research interests include massive MIMO, holographic MIMO (H-MIMO), recon-

figurable intelligent surfaces (RISs), and fluid antenna systems (FASs). He received the National Scholarship in 2019 and 2024. In 2024, he received the Special Scholarship of Tsinghua University, which annually awards ten students out of more than 40,000 graduate students in Tsinghua University. He was listed in Stanford University's World's Top 2% Scientists in 2023.



Jieao Zhu (Graduate Student Member, IEEE) received the B.E. degree in electronic engineering and the B.S. degree in applied mathematics from Tsinghua University, Beijing, China, in 2021, where he is currently pursuing the Ph.D. degree with the Department of Electronic Engineering. His research interests include electromagnetic information theory (EIT), coding theory, and quantum computing. He has received the National Scholarship in 2018 and 2020 and the Excellent Graduates of Beijing in 2021.



Linglong Dai (Fellow, IEEE) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2003, the M.S. degree from China Academy of Telecommunications Technology, Beijing, China, in 2006, and the Ph.D. degree from Tsinghua University, Beijing, in 2011.

From 2011 to 2013, he was a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University, where he was an Assistant Professor from 2013 to 2016, an Associate Professor from 2016 to 2022, and he has been a

Professor since 2022. He has coauthored the book *MmWave Massive MIMO: A Paradigm for 5G* (Academic Press, 2016). He has authored or coauthored more than 100 IEEE journal articles and more than 60 IEEE conference papers. He also holds more than 20 granted patents. His current research interests include massive MIMO, reconfigurable intelligent surface (RIS), millimeter-wave and Terahertz communications, near-field communications, machine learning for wireless communications, and electromagnetic information theory. He has received five IEEE Best Paper Awards from the IEEE ICC 2013, the IEEE ICC 2014, the IEEE ICC 2017, the IEEE VTC 2017-Fall, the IEEE

ICC 2018, and the IEEE GLOBECOM 2023. He has also received the Tsinghua University Outstanding Ph.D. Graduate Award in 2011, Beijing Excellent Doctoral Dissertation Award in 2012, China National Excellent Doctoral Dissertation Nomination Award in 2013, the URSI Young Scientist Award in 2014, IEEE TRANSACTIONS ON BROADCASTING Best Paper Award in 2015, the *Electronics Letters* Best Paper Award in 2016, the National Natural Science Foundation of China for Outstanding Young Scholars in 2017, the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2017, the IEEE ComSoc Asia-Pacific Outstanding Paper Award in 2018, China Communications Best Paper Award in 2019, IEEE ACCESS Best Multimedia Award in 2020, the IEEE Communications Society Leonard G. Abraham Prize in 2020, the IEEE ComSoc Stephen O. Rice Prize in 2022, the IEEE ICC Best Demo Award in 2022, and the National Science Foundation for Distinguished Young Scholars in 2023. He was listed as a Highly Cited Researcher by Clarivate Analytics from 2020 to 2024.



Robert W. Heath Jr. (Fellow, IEEE) was with The University of Texas, Austin, most recently as Cockrell Family Regents Chair in engineering and the Director of UT SAVES from 2002 to 2020. From 2020 to 2023, he was the Lampe Distinguished Professor at North Carolina State University and the co-founder of 6GNC. He is the Charles Lee Powell Chair in wireless communications at the Department of Electrical and Computer Engineering, University of California at San Diego, San Diego, CA, USA. He is also the President and the CEO of MIMO

Wireless Inc. He has authored *Introduction to Wireless Digital Communication* (Prentice Hall, 2017) and *Digital Wireless Communication: Physical Layer Exploration Lab Using the NI USRP* (National Technology and Science Press, 2012), and co-authored *Millimeter Wave Wireless Communications* (Prentice Hall, 2014), and *Foundations of MIMO Communication* (Cambridge University Press, 2018).

Dr. Heath has been a co-author of a number award winning conference and journal papers, including recently the 2017 Marconi Prize Paper Award, the 2019 IEEE Communications Society Stephen O. Rice Prize, the 2020 IEEE Signal Processing Society Donald G. Fink Overview Paper Award, the 2021 IEEE Vehicular Technology Society Neal Shepherd Memorial Best Propagation Paper Award, and the 2022 IEEE Vehicular Technology Society Best Vehicular Electronics Paper Award. Other notable awards, include the 2017 EURASIP Technical Achievement award, the 2019 IEEE Kiyo Tomiyasu Award, the 2021 IEEE Vehicular Technology Society James Evans Avant Garde Award, and the 2025 IEEE/RSE James Clerk Maxwell Medal. In 2017, he was selected as a fellow of the National Academy of Inventors. In 2024, he was selected as a fellow of American Association for the Advancement of Science. He was a Member-at-Large on the IEEE Communications Society Board-of-Governors (2020–2022) and the IEEE Signal Processing Society Board-of-Governors (2016–2018). He was an Editor-in-Chief of *IEEE Signal Processing Magazine* from 2018 to 2020. He is also a Licensed Amateur Radio Operator, a Private Pilot, a Registered Professional Engineer in Texas.