# Successive Bayesian Reconstructor for Channel Estimation in Fluid Antenna Systems

Zijian Zhang, *Graduate Student Member, IEEE*, Jieao Zhu, *Graduate Student Member, IEEE*, Linglong Dai, *Fellow, IEEE*, and Robert W. Heath, Jr., *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 loss. 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 the maximum a posterior (MAP) estimator of FAS channels. 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, compressed sensing, Bayesian 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], fluid antenna systems (FASs) [3]–[5], also called movable antenna systems [6]–[8], 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 [7]. 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 [9]. Compared with the fixed antenna systems, FASs are promising to achieve higher transmission reliability [4] and larger-scale multiple-access [5] while using much fewer antennas [10].

### A. Prior Work

The original idea of reconfiguring antenna positions dated back to 2000s, when antenna selection (AS) was proposed to balance the hardware complexity and transmission performance of MIMO systems [11]. By connecting massive antennas with a few radio frequency (RF) chains through a switch network, AS-MIMO can carefully select a subset of antennas associated with the well-performed channels for transmission [12], [13]. Thanks to the advances in the microwave field, more position-flexible antennas were proposed [14]–[16], and their positions can be reconfigured in an almost continuous manner. For example, in [17], RF micro-electro-mechanical systems (MEMSs) were integrated into some radiation patches to reconfigure their positions. In [18], a fluid antenna made of liquid metals could move continuously in a non-conductive tube. In [19], pixel-like switches were utilized to design an electronically reconfigurable array, which allows massive radiation patches to "move" in a virtual manner. Compared with the AS-MIMO, these highly flexible structures can further explore the diversity and multiplexing gains of a limited space, which has attracted many attentions from the communication community in recent years [3]–[9].

Up to now, the existing works on FASs have covered a wide range, mainly focusing on the performance analysis [3]–[5], [20]–[22], port selection [23]–[25], beamforming design [6]–[8], [26]–[28], and channel estimation [29]–[31]. For example, the ergodic capacity and the lower-bound capacity of FASs were derived in [3]. Then, the multiplexing gains of multi-user FASs were analyzed in [5] and [32]. By combining machine learning methods and analytical approximations, the port selection for FASs was studied in [23]–[25] to realize fast antenna placements. To maximize the sum-rate in FASs, alternaing optimization (AO)-based algorithms that jointly optimize the antenna positions and beamformers were proposed in [8] and [26]. To reduce the power consumption of FASs, sequential convex approximation (SCA)-based algorithms for FAS beamforming were proposed in [7] and [27].

Despite the encouraging prospects of FASs, such as high capacity and low power consumption [4]–[6], these expected gains are hard to achieve in practice. The transmission performance of FASs heavily relies on the positions of fluid antennas [23]–[25]. To ensure favorable antenna placements,

The conference version of this paper will be presented at the IEEE WCNC'24, Dubai, United Arab Emirates, April 21–24, 2024 [1].

Zijian Zhang, Jieao Zhu, and Linglong Dai are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China, as well as the Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China (e-mail: {zhangzj20, 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, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA (e-mail: rwheathjr@ucsd.edu).

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2020YFB1807201), in part by the National Natural Science Foundation of China (Grant No. 62031019), and in part by the European Commission through the H2020-MSCA-ITN META WIRELESS Research Project under Grant 956256.

the channel state information (CSI) of available locations is essential, and most existing works have assumed the perfectly known CSI [6]–[8], [26]–[28]. 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 [4]. 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 [29]–[31]. 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 [29]–[31]. For example, in [29], a sequential linear minimum mean square error (SeLMMSE) method was proposed. In each estimation subframe, the fluid antennas move to some equally-spaced ports for channel measurements [29]. 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 [30], 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 [31], 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 [29], angular-domain sparsity [30], and known AoAs [31]. 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.

- To the best of our knowledge, Bayesian inference [33]–[35] is introduced into the FAS channel estimation for the first time. Fluid antennas can switch their locations among the ports to measure channels, resembling a sampling process. Since the ports are densely deployed, the FAS channels have strong correlation. These properties motivate us to build a kernel that characterizes inherent channel priors and then construct FAS channels by kernel-based sampling and regression. With an experiential kernel, Bayesian regression can reconstruct a continuous function by only sampling a few carefully-selected points, which provides a general framework to reconstruct FAS channels in a non-parametric manner.
- Based on the Bayesian regression, 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 a two-stage scheme. In the first stage, the sampling sequence, i.e., the index set of the sampled ports, is designed offline. By building an experiential kernel and following the criterion of maximum posterior variance, the entries of the sampling sequence can be determined in a successive manner. In the second stage, the ports associated with the sampling sequence are selected for online channel measurements. By combing the measured channels with the experiential kernel for process regression, the predictive mean of the regressed stochastic process can be viewed as the maximum a posterior (MAP) estimator of FAS channels.
- To verify the effectiveness, the minimum mean square error (MSE) achieved by the proposed S-BAR is analyzed, and numerical simulations are provided to compare the performances of different channel estimators. Our results show that, in both model-mismatched and model-matched cases, the proposed S-BAR outperforms the benchmark schemes with channel assumptions. Particularly, the estimation accuracy of S-BAR can increase by an order of magnitude. In addition, thanks to the hybrid offline and online implementation mechanism of S-BAR, the computational complexity of online employing S-BAR is only linear with the number of ports, which further verifies the high efficiency of the proposed S-BAR.

### 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 channel estimation problem is formulated. In Section III, existing solutions for FAS channel estimation are reviewed, and their potential challenges are discussed. In Section IV, the S-BAR is proposed, and its performance is analyzed. 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:* $\mathbb{C}$ denotes the set of complex numbers; $[\cdot]^{-1}$, $[\cdot]^{\dagger}$, $[\cdot]^{*}$, $[\cdot]^{\mathrm{T}}$, and $[\cdot]^{\mathrm{H}}$ 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 $\mathbf{x}$; $\mathbf{X}(i,j)$, $\mathbf{X}(j,:)$ and $\mathbf{X}(:,j)$ denote the $(i,j)$-th entry, the $j$-th row, and the $j$-th column of matrix $\mathbf{X}$, respectively; $\mathrm{Tr}(\cdot)$ denotes the trace of its argument; $\mathsf{P}(\cdot|\cdot)$ is the conditional probability density function; $\mathsf{E}(\cdot)$ is the expectation of its argument; $\Re\{\cdot\}$ denotes the real part of the argument; $\ln(\cdot)$ denotes the natural logarithm of its argument; $\dim(\cdot)$ is the dimensional 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

[1]Simulation codes will be provided to reproduce the results in this article: http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html.

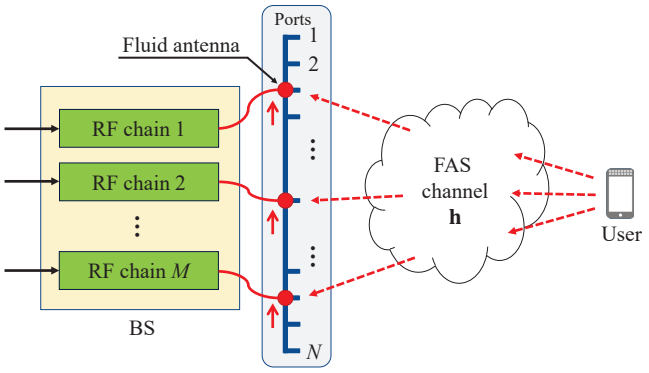



Fig. 1. An illustration of channel estimation for an FAS, where one $N$-port BS equipped with $M$ fluid antennas receives the pilots from a user in the uplink.

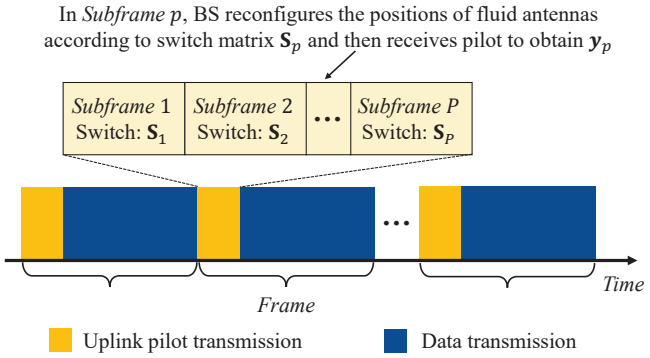



Fig. 2. The frame structure for FAS channel reconstruction.

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; and $\mathbf{0}_L$ is an all-zero vector or matrix with dimension $L$.

## II. 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 narrowband 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):** Binary indicator $\mathbf{S}_p \in \{0,1\}^{M\times N}$ is defined as the switch matrix of multiple fluid antennas in timeslot $p$. The $(m,n)$-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 row of $\mathbf{S}_p$ has one 1 entry, and all 1 entries in $\mathbf{S}_p$ are not in the same column, i.e., $\|\mathbf{S}_p(m,:)\| = 1$ for all $m \in \{1,\cdots,M\}$, $\|\mathbf{S}_p(:,n)\| \in \{0,1\}$ for all $n \in \{1,\cdots,N\}$, and $\mathbf{S}_p\mathbf{S}_p^{\mathrm{H}} = \mathbf{I}_M$.

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

$$\mathbf{y}_p = \mathbf{S}_p\mathbf{h}s_p + \mathbf{z}_p, \tag{1}$$

where $s_p$ 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 $s_p = 1$ for all $p \in \{1,\cdots,P\}$. Considering the total $P$ timeslots for pilot transmission, we arrive at

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{z}, \tag{2}$$

where $\mathbf{y} := \left[\mathbf{y}_1^{\mathrm{T}},\cdots,\mathbf{y}_P^{\mathrm{T}}\right]^{\mathrm{T}}$, $\mathbf{S} := \left[\mathbf{S}_1^{\mathrm{T}},\cdots,\mathbf{S}_P^{\mathrm{T}}\right]^{\mathrm{T}}$, and $\mathbf{z} := \left[\mathbf{z}_1^{\mathrm{T}},\cdots,\mathbf{z}_P^{\mathrm{T}}\right]^{\mathrm{T}}$. Our goal is to reconstruct the $N$-dimensional channel $\mathbf{h}$ according to the $PM$-dimensional noisy pilot $\mathbf{y}$ ($PM \ll N$).

## III. Existing Methods for FAS Channel Estimation

In this section, we review two pilot-reduced channel estimators for FASs, which can reconstruct high-dimensional channel $\mathbf{h}$ according to the low-dimensional pilot $\mathbf{y}$. Specifically, utilizing the sparse representation of FAS channels, the CS-based channel estimator is firstly reviewed in Subsection III-A. Then, following the maximum likelihood (ML) principle, the AO-based channel estimator is reviewed in Subsection III-B. Finally, the challenges and opportunities of FAS channel estimation are discussed.

### A. Compressed Sensing Based Channel Estimation

To reconstruct the high-dimensional FAS channel $\mathbf{h}$ according to 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 [36]. For our considered FAS channel estimation, by viewing switch matrix $\mathbf{S}$ as a virtual analog precoder, many CS-based methods can be adopted to reconstruct channel $\mathbf{h}$ with reduced pilots. Specifically, the realization is explained as follows.

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

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}}\left[1, e^{\mathrm{j}\frac{2\pi}{\lambda}d\cos(\theta)},\cdots,e^{\mathrm{j}\frac{2\pi}{\lambda}(N-1)d\cos(\theta)}\right]^{\mathrm{T}}, \tag{3}$$

where $\lambda$ 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

$$\mathbf{h} = \sqrt{\frac{N}{CR}}\sum_{c=1}^{C}\sum_{r=1}^{R} g_{c,r}\mathbf{a}(\theta_{c,r}), \tag{4}$$

[2]This paper focuses on the single-user uniform linear array (ULA) case. However, all schemes in this paper can be easily extended to the multi-user case by using orthogonal pilots. Besides, they can be easily extended to the uniform planar array (UPA) case by considering the UPA channel model.

**Algorithm 1** FAS-OMP reconstructor

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

1: Employ randomly generated $\mathbf{S}$ at the BS, and then obtain the received pilots: $\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{z}$
2: Initialization: $\mathbf{r} = \mathbf{y}$, $\Upsilon = \varnothing$, $\bar{\mathbf{h}} = \mathbf{0}_N$
3: **for** $l \in \{1, \cdots, L\}$ **do**
4: Update correlation matrix: $\mathbf{\Gamma} = \mathbf{\Psi}^{\mathrm{H}}\mathbf{r}$
5: Find new support: $\upsilon^\star = \underset{\upsilon}{\operatorname{argmax}} |\mathbf{\Gamma}(\upsilon)|$
6: Update support set: $\Upsilon = \Upsilon \overset{\upsilon}{\cup} \{\upsilon^\star\}$
7: Orthogonal projection: $\bar{\mathbf{h}}(\Upsilon) = \mathbf{\Psi}(:, \Upsilon)^\dagger \mathbf{y}$
8: Update residual: $\mathbf{r} = \mathbf{r} - \mathbf{\Psi}(:, \Upsilon)\bar{\mathbf{h}}(\Upsilon)$
9: **end for**
10: Channel estimation: $\hat{\mathbf{h}} = \mathbf{F}(:, \Upsilon)\bar{\mathbf{h}}(\Upsilon)$
11: **return** Reconstructed FAS channel $\hat{\mathbf{h}}$

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 ($L \ll N$). Then, the received signal $\mathbf{y}$ in (2) can be rewritten as

$$\mathbf{y} = \mathbf{S}\mathbf{F}\tilde{\mathbf{h}} + \mathbf{z} = \mathbf{\Psi}\tilde{\mathbf{h}} + \mathbf{z}, \tag{5}$$

where $\mathbf{\Psi} = \mathbf{S}\mathbf{F}$ 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 $\|\mathbf{S}(m,:)\| = 1$ for all $m \in \{1, \cdots, M\}$, $\|\mathbf{S}(:,n)\| \in \{0, 1\}$ for all $n \in \{1, \cdots, N\}$, and $\mathbf{S}\mathbf{S}^{\mathrm{H}} = \mathbf{I}_{PM}$.

Note that, (5) is a standard observation equation for sparse signal reconstruction. Thus, some existing CS-based algorithms, such as OMP, iterative hard thresholding (IHT), and vector approximate message passing (VAMP), can be utilized for (5) to reconstruct $\mathbf{h}$ from $\mathbf{y}$. For example, in [30], 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. In our considered scenario, by substituting the estimated channel parameters into (4), the FAS-OMP reconstructor in [30] can be modified to acquire channel $\mathbf{h}$ explicitly. For clarity, here we summarize this estimator in **Algorithm 1**.

### *B. Alternating Optimization Based Channel Estimation*

The CS-based channel estimators usually operate based on a gridded codebook, e.g., the DFT matrix $\mathbf{F}$, which leads to the limited resolution in estimating channels. When the real channel parameter values fall between the grid points, the non-ideal grid sampling may introduce estimation error [37]. To improve the estimation accuracy, the channel parameters can be further optimized to cope with the estimation error introduced by on-grid sampling [38]. Given the spatial sparsity $L$, this goal can be achieved by finding the path gains $\mathbf{g} := [g_1, \cdots, g_L]^{\mathrm{T}}$ and incident angles $\boldsymbol{\theta} := [\theta_1, \cdots, \theta_L]^{\mathrm{T}}$ that maximize the likelihood function $\mathsf{P}(\mathbf{y}|\mathbf{g}, \boldsymbol{\theta})$, i.e., the ML estimator. Following this idea, an AO-based method called FAS-ML reconstructor is summarized in **Algorithm 2**, and its details are explained as follows.

**Algorithm 2** FAS-ML reconstructor

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

1: Employ randomly generated $\mathbf{S}$ at the BS, and then obtain the received pilots: $\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{z}$
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 $\mathsf{P}(\mathbf{y}|\hat{\mathbf{g}}, \hat{\boldsymbol{\theta}})$ **do**
5: Update path gain: $\hat{\mathbf{g}} = \mathbf{B}^\dagger(\hat{\boldsymbol{\theta}})\mathbf{y}$
6: Update angle: $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \zeta\Re\left\{\frac{\partial\|\mathbf{y}-\mathbf{B}(\boldsymbol{\theta})\mathbf{g}\|^2}{\partial\boldsymbol{\theta}}|_{\mathbf{g}=\hat{\mathbf{g}},\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right\}$
7: Update step length: $\zeta = \zeta/2$
8: **end while**
9: Channel estimation: $\hat{\mathbf{h}} = \sqrt{\frac{N}{L}}\sum_{l=1}^{L}\hat{g}_l\mathbf{a}(\hat{\theta}_l)$
10: **return** Reconstructed FAS channel $\hat{\mathbf{h}}$

Assume that channel $\mathbf{h}$ has the spatial sparsity as shown in (4). For the considered problem of channel estimation, the logarithm likelihood function can be written as

$$\ln\left(\mathsf{P}\left(\mathbf{y}|\mathbf{g}, \boldsymbol{\theta}\right)\right) = -N\ln\left(\sigma^2\pi\right) - \frac{1}{\sigma^2}\left\|\mathbf{y} - \mathbf{B}\left(\boldsymbol{\theta}\right)\mathbf{g}\right\|^2, \tag{6}$$

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

$$\{\mathbf{g}^\star, \boldsymbol{\theta}^\star\} = \underset{\mathbf{g},\boldsymbol{\theta}}{\operatorname{argmax}}\left\|\mathbf{y} - \mathbf{B}\left(\boldsymbol{\theta}\right)\mathbf{g}\right\|^2. \tag{7}$$

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 [39]. By fixing one and optimize the other variables, the likelihood function $\mathsf{P}\left(\mathbf{y}|\mathbf{g}, \boldsymbol{\theta}\right)$ can be maximized through an AO process. Then, the FAS-ML reconstructor in **Algorithm 2** can be obtained.

### *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 follow the assumption of spatially sparse channels [30]. In the scenarios with rich scatters, the model mismatch of channels 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 [36], only a few elements in $\mathbf{S}$ are one while the others are all zero. It means that, the received pilot $\mathbf{y}$ is not sufficiently informative, which only contains the information of a few channels associated with the selected ports. From the mathematical perspective, the channel estimation in FASs is equivalent to the channel reconstruction, whose performance is determined by the effect

of interpolation/fitting operations. Consequently, high pilot overhead (i.e., a large $P$) may be required to accurately reconstruct the high-dimensional 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 $\lambda/10$ [3]–[5]. Thus, the channels of several ports closer to each other are strongly correlated. It suggests that, when the channel of a port is sampled (measured), the uncertainty of the channels associated with its nearby 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, which is exactly the key idea of Bayesian regression. To achieve this goal, we propose the S-BAR as a promising solution, which is illustrated in the next section.

## IV. 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 [33]–[35]. Firstly, the Bayesian regression is introduced in Subsection IV-A. Then, in Subsection IV-B, the proposed S-BAR is illustrated. Subsequently, in Subsection IV-C, the performance of S-BAR is analyzed. Finally, in Subsection IV-D, the kernel selection 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) [35] or Kriging method [40], has become a popular solution. Specifically, function $f(\mathbf{x})$ can be modeled as a sample of Gaussian process $\mathcal{GP}\left(\mu\left(\mathbf{x}\right), k\left(\mathbf{x},\mathbf{x}'\right)\right)$, where any finite subset follows a consistent multivariate Gaussian distribution. It is completely specified by its mean function $\mu\left(\mathbf{x}\right)$, which can be assumed to be zero, and its kernel function $k\left(\mathbf{x},\mathbf{x}'\right)$, 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 $\mathcal{GP}\left(\mu\left(\mathbf{x}\right), k\left(\mathbf{x},\mathbf{x}'\right)\right)$ over $f(\mathbf{x})$. Let $\boldsymbol{\gamma}^t := [\gamma^1, \cdots, \gamma^t]^{\mathrm{T}}$ denote $t$ noisy measurements for points in $\mathcal{A}^t := \{\mathbf{x}^1, \cdots, \mathbf{x}^t\}$, where $\gamma^i = f(\mathbf{x}^i) + n_i$ with $n_i \sim \mathcal{CN}\left(0, \delta^2\right)$. The joint probability distribution of $f(\mathbf{x})$ and $\boldsymbol{\gamma}^t$ satisfies

$$\left[\begin{array}{c} f\left(\mathbf{x}\right) \\ \boldsymbol{\gamma}^t \end{array}\right] \sim \mathcal{CN}\left(\left[\begin{array}{c} \mu\left(\mathbf{x}\right) \\ \boldsymbol{\mu}^t \end{array}\right], \left[\begin{array}{cc} k\left(\mathbf{x},\mathbf{x}\right) & \left(\mathbf{k}^t(\mathbf{x})\right)^{\mathrm{H}} \\ \mathbf{k}^t(\mathbf{x}) & \mathbf{K}^t + \delta^2\mathbf{I}_t \end{array}\right]\right), \tag{8}$$

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

$$\mu^t\left(\mathbf{x}\right) = \mu\left(\mathbf{x}\right) + \left(\mathbf{k}^t(\mathbf{x})\right)^{\mathrm{H}}\left(\mathbf{K}^t + \delta^2\mathbf{I}_t\right)^{-1}\left(\boldsymbol{\gamma}^t - \boldsymbol{\mu}^t\right), \tag{9}$$

$$k^t\left(\mathbf{x},\mathbf{x}'\right) = k\left(\mathbf{x},\mathbf{x}'\right) - \left(\mathbf{k}^t(\mathbf{x})\right)^{\mathrm{H}}\left(\mathbf{K}^t + \delta^2\mathbf{I}_t\right)^{-1}\mathbf{k}^t(\mathbf{x}'). \tag{10}$$

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

$$\mathbf{x}^{t+1} = \underset{\mathbf{x}\in\mathcal{S}/\mathcal{A}^t}{\arg\max}\ k^t\left(\mathbf{x},\mathbf{x}\right), \tag{11}$$

where $/$ is the set difference. By updating $\mathcal{A}^t$ and $\boldsymbol{\gamma}^t$ accordingly, the variance function $k^t\left(\mathbf{x},\mathbf{x}\right)$ decreases asymptotically, which means that the uncertainty of $f(\mathbf{x})$ is gradually eliminated. After reaching the tolerance threshold $\varepsilon$, the posterior mean $\mu^t\left(\mathbf{x}\right)$ in (9) can be viewed as the MAP estimator of $f(\mathbf{x})$, which completes the algorithm.

**Algorithm 3** Bayesian Linear Regression

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

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

```
Initialization: t = 0, A^0 = ∅, k^0(x, x') = k(x, x')
while k^t(x, x) > ε for an x ∈ S do
    Timeslot update: t = t + 1
    Sample selection: x^t = argmax_{x∈S/A^{t-1}} k^{t-1}(x, x)
    Update measured points: A^t = A^{t-1} ∪ {x^t}
    Measurement: γ^t = f(x^t) + n_t
    Posterior update: Calculate mean μ^t(x) and covariance
    k^t(x, x') by (9) and (10) for all x, x' ∈ S, respectively
end while
Reconstruction: f̂(x) = μ^t(x)
return Reconstructed function f̂(x)
```

### B. Proposed Successive Bayesian Reconstructor (S-BAR)

In each pilot timeslot, $M$ fluid antennas 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 Bayesian regression. To reconstruct FAS channels via kernel-based sampling and regression, we model $\mathbf{h}$ as a sample of Gaussian pro-

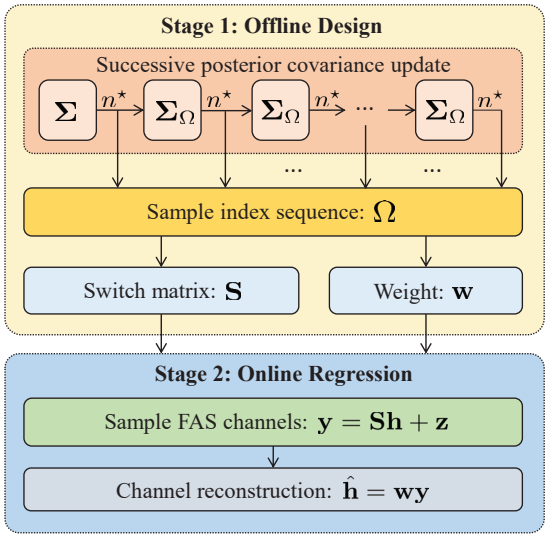



Fig. 3. A flowchart of the proposed S-BAR.

cess $\mathcal{GP}(\mathbf{0}_N, \mathbf{\Sigma})$. In particular, semidefinite Hermitian matrix $\mathbf{\Sigma} \in \mathbb{C}^{N\times N}$ is called the kernel[3], which characterizes (but does not need to be) the prior covariance of $\mathbf{h}$. For better understanding, here we plot the flowchart of the proposed S-BAR in Fig. 3, and the pseudo-code of S-BAR is summarized in **Algorithm 4**. For clarity, the algorithmic principle of S-BAR is firstly introduced as follows.

*1) Algorithmic Principle:* At some moment, let $\Omega$ denote the index sequence of the measured channels and $\mathbf{y}_\Omega \in \mathbb{C}^{\dim(\Omega)}$ denote the corresponding received pilot. The noisy pilot $\mathbf{y}_\Omega$ is from $\mathbf{y}_\Omega = \mathbf{h}(\Omega) + \mathbf{z}_\Omega$ with $\mathbf{z}_\Omega \sim \mathcal{CN}(\mathbf{0}_{\dim(\Omega)}, \sigma^2\mathbf{I}_{\dim(\Omega)})$ being the AWGN. In this case, the joint probability distribution of $\mathbf{h}$ and $\mathbf{y}_\Omega$ satisfies

$$\left[\mathbf{h}^{\mathrm{T}}\mathbf{y}_\Omega^{\mathrm{T}}\right]^{\mathrm{T}} \sim \mathcal{CN}\left(\begin{bmatrix}\mathbf{0}_N \\ \mathbf{0}_{\dim(\Omega)}\end{bmatrix}, \begin{bmatrix}\mathbf{\Sigma} & \mathbf{\Sigma}(:,\Omega) \\ \mathbf{\Sigma}(\Omega,:) & \mathbf{\Sigma}(\Omega,\Omega) + \sigma^2\mathbf{I}_{\dim(\Omega)}\end{bmatrix}\right). \tag{12}$$

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

$$\boldsymbol{\mu}_\Omega = \mathbf{\Sigma}(:,\Omega)\left(\mathbf{\Sigma}(\Omega,\Omega) + \sigma^2\mathbf{I}_{\dim(\Omega)}\right)^{-1}\mathbf{y}_\Omega, \tag{13}$$

$$\mathbf{\Sigma}_\Omega = \mathbf{\Sigma} - \left(\mathbf{\Sigma}(\Omega,:)\right)^{\mathrm{H}}\left(\mathbf{\Sigma}(\Omega,\Omega) + \sigma^2\mathbf{I}_{\dim(\Omega)}\right)^{-1}\mathbf{\Sigma}(\Omega,:). \tag{14}$$

Similar to the candidate selection in (11), for given $\Omega$, the next candidate channel to be measured can be determined by finding the index of the largest posterior variance, i.e.,

$$n^\star = \underset{n\in\{1,\cdots,N\}/\Omega}{\arg\max}\ \mathbf{\Sigma}_\Omega(n,n). \tag{15}$$

Subsequently, we can update $\Omega$ by $\Omega \cup \{n^\star\}$ and repeat the above process until the posterior mean $\boldsymbol{\mu}_\Omega$ can well approximate $\mathbf{h}$.

*2) Observations:* From the above equations, we obtain the following three observations.

[3]The selection of kernel $\mathbf{\Sigma}$ will be introduced in Subsection IV-D.

**Algorithm 4** Proposed Successive Bayesian Reconstructor

**Input:** Number of pilots $P$, kernel $\mathbf{\Sigma}$.
**Output:** Reconstructed FAS channel $\hat{\mathbf{h}}$.

*# Stage 1 (Offline Design):*
Initialization: $\Omega = \varnothing$, $\mathbf{S}_p = \mathbf{0}_{M\times N}$ for all $p \in \{1,\cdots,P\}$
**for** $p \in \{1,\cdots,P\}$ **do**
  **for** $m \in \{1,\cdots,M\}$ **do**
    Posterior covariance update: Calculate $\mathbf{\Sigma}_\Omega$ by (14)
    Candidate selection: $n^\star = \underset{n\in\{1,\cdots,N\}/\Omega}{\arg\max}\ \mathbf{\Sigma}_\Omega(n,n)$
    Switch matrix update: $\mathbf{S}_p(m,n^\star) = 1$
    Sequence update: $\Omega = \Omega \cup \{n^\star\}$
  **end for**
**end for**
Merge switch matrices: $\mathbf{S} := \left[\mathbf{S}_1^{\mathrm{T}},\cdots,\mathbf{S}_P^{\mathrm{T}}\right]^{\mathrm{T}}$
Weight calculation: $\mathbf{w} = (\mathbf{\Sigma}(\Omega,\Omega) + \sigma^2\mathbf{I}_{PM})^{-1}\mathbf{\Sigma}(\Omega,:)$
*# Stage 2 (Online Regression):*
Employ the designed switch matrix $\mathbf{S}$ at the BS, and then obtain the received pilot: $\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{z}$
Channel reconstruction: $\hat{\mathbf{h}} = \mathbf{w}^{\mathrm{H}}\mathbf{y}$
**return** Reconstructed FAS channel $\hat{\mathbf{h}}$

- Equation (13) indicates that, the posterior mean $\boldsymbol{\mu}_\Omega$ is the linear weighted sum of the received pilot $\mathbf{y}_\Omega$, i.e., $\boldsymbol{\mu}_\Omega = \mathbf{w}^{\mathrm{H}}\mathbf{y}_\Omega$, where the weight $\mathbf{w} := \left(\mathbf{\Sigma}(\Omega,\Omega) + \sigma^2\mathbf{I}_{\dim(\Omega)}\right)^{-1}\mathbf{\Sigma}(\Omega,:)$ only relies on the kernel $\mathbf{\Sigma}$.
- Equation (14) shows that, posterior covariance $\mathbf{\Sigma}_\Omega$ only relies on kernel $\mathbf{\Sigma}$, while it is unrelated to the received pilot $\mathbf{y}_\Omega$ in channel estimation.
- Equation (15) suggests that the FAS port selection only relies on the posterior covariance $\mathbf{\Sigma}_\Omega$.

These observations reveal that, the switch matrix $\mathbf{S}$ and the weight for reconstructing $\mathbf{h}$ can be designed offline and then deployed online for regression. In this way, the complexity of employing the S-BAR online can be significantly reduced. Therefore, the implementation of the proposed S-BAR can be realized in the two stages as shown in Fig. 3, and their details are explained as follows.

*3) Stage 1 (Offline Design):* Since index sequence $\Omega$ is determined the posterior covariance $\mathbf{\Sigma}_\Omega$, while $\mathbf{\Sigma}_\Omega$ only relies on the kernel $\mathbf{\Sigma}$. The switch matrix $\mathbf{S} \in \{0,1\}^{PM\times N}$ 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 $\mathbf{\Sigma}_\Omega$ in (14) and $n^\star$ in (15) alternatingly until $\dim(\Omega) = PM$, index sequence $\Omega$ can collect all required indexes for the locations of $M$ antennas in $P$ pilot timeslots. Then, recall that indicator $\mathbf{S}_p(m,n) \in \{0,1\}$ determines whether the $m$-th fluid antenna is located at the position of the $n$-th port in timeslot $p$, thus we have $\mathbf{h}(\Omega) = \mathbf{S}\mathbf{h}$. To achieve the conversion from $\Omega$ to $\mathbf{S}$, we can initialize $\mathbf{S}$ as an all-zero matrix. Then, in each row of $\mathbf{S}$, fill in an 1 at the position associated with the selected port. Note that, this operation naturally satisfies $\|\mathbf{S}(m,:)\| = 1$ for all $m \in \{1,\cdots,M\}$, $\|\mathbf{S}(:,n)\| \in \{0,1\}$ for all $n \in \{1,\cdots,N\}$, and $\mathbf{S}\mathbf{S}^{\mathrm{H}} = \mathbf{I}_{PM}$. These properties ensure that the designed $\mathbf{S}$

is practically implementable in FAS systems. After obtaining the index sequence $\Omega$, the weight for reconstructing $\mathbf{h}$ can be obtained by

$$\mathbf{w} = \left(\mathbf{\Sigma}(\Omega,\Omega) + \sigma^2\mathbf{I}_{PM}\right)^{-1}\mathbf{\Sigma}(\Omega,:). \tag{16}$$

*4) 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 will 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 MAP estimator in (13), channel $\mathbf{h}$ can be reconstructed by the weighted sum $\hat{\mathbf{h}} = \mathbf{w}^{\mathrm{H}}\mathbf{y}$, which finally completes the proposed S-BAR in **Algorithm 4**.

### C. 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 $E = \mathsf{E}\left(\|\boldsymbol{\mu}_\Omega - \mathbf{h}\|^2\right)$. By adopting some matrix techniques, we obtain the following lemma.

**Lemma 1 (MSE of S-BAR for given kernel $\mathbf{\Sigma}$):** Assume $\mathsf{E}\left(\mathbf{h}\mathbf{h}^{\mathrm{H}}\right) = \mathbf{\Sigma}_{\mathrm{cov}}$. Given an index sequence of the measured channels $\Omega$ (corresponding to a switch matrix $\mathbf{S}$), the MSE of reconstructing $\mathbf{h}$ via S-BAR can be expressed as:

$$\begin{aligned} E =& \mathrm{Tr}\left(\mathbf{\Pi}^{\mathrm{H}}\left(\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}}\mathbf{S}^{\mathrm{H}} + \sigma^2\mathbf{I}_{PM}\right)\mathbf{\Pi}\right) - \\ & 2\Re\left(\mathrm{Tr}\left(\mathbf{\Pi}^{\mathrm{H}}\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}}\right)\right) + \mathrm{Tr}\left(\mathbf{\Sigma}_{\mathrm{cov}}\right), \end{aligned} \tag{17}$$

where $\mathbf{\Pi}$ is a matrix function with respect to $\mathbf{\Sigma}$, given by

$$\mathbf{\Pi} = \left(\mathbf{S}\mathbf{\Sigma}\mathbf{S}^{\mathrm{H}} + \sigma^2\mathbf{I}_{PM}\right)^{-1}\mathbf{S}\mathbf{\Sigma}. \tag{18}$$

*Proof:* Constructive proof is given in Appendix A. ∎

**Lemma 1** characterizes the estimation accuracy of S-BAR for the given kernel $\mathbf{\Sigma}$ and switch matrix $\mathbf{S}$. Particularly, $E$ in (17) can quantitatively describe the estimation error caused by the mismatch of the selected kernel $\mathbf{\Sigma}$ and the real covariance $\mathbf{\Sigma}_{\mathrm{cov}}$. We find that $E$ is independent of the real mean of $\mathbf{h}$, i.e., $\mathbb{E}(\mathbf{h})$, which explains why modeling $\mathbf{h}$ as a zero-mean process does not loss the generality [34]. Then, as the fundamental limit, one may be concerned with the achievable minimum MSE of S-BAR, thus we introduce the following lemma.

**Lemma 2 (Achievable Minimum MSE of S-BAR):** Assume $\mathsf{E}\left(\mathbf{h}\mathbf{h}^{\mathrm{H}}\right) = \mathbf{\Sigma}_{\mathrm{cov}}$. The achievable minimum MSE of reconstructing $\mathbf{h}$ via S-BAR, i.e. $E_{\min} = \min_{\mathbf{\Sigma},\Omega}\ \mathsf{E}(\|\boldsymbol{\mu}_\Omega - \mathbf{h}\|^2)$, can be written as:

$$E_{\min} = \mathrm{Tr}\left(\mathbf{\Sigma}_{\mathrm{cov}}\right) - \mathrm{Tr}\left(\mathbf{\Sigma}_{\mathrm{cov}}\left(\mathbf{\Sigma}_{\mathrm{cov}} + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{\Sigma}_{\mathrm{cov}}\right), \tag{19}$$

which is achieved if and only if the kernel is exactly the real covariance of $\mathbf{h}$, i.e., $\mathbf{\Sigma} = \mathbf{\Sigma}_{\mathrm{cov}}$, and the FAS channels are completely observed, i.e., $\Omega = \{1, \cdots, N\}$.

*Proof:* Constructive proof is given in Appendix B. ∎

TABLE I
COMPUTATIONAL COMPLEXITY OF DIFFERENT SCHEMES.

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

**Lemma 2** characterizes the lower bound of MSE while employing $\bar{S}$ 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 $\mathbf{\Sigma}_{\mathrm{cov}}$, the MSE of the Bayesian reconstructor will be exactly $\mathrm{Tr}((\mathbf{\Sigma}_{\mathrm{cov}}^{-1} + \sigma^{-2}\mathbf{I})^{-1})$, which coincides with the well-known posterior covariance formula in linear estimation theory [41]. From the perspective of statistical signal processing, **Lemma 2** 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** [30] and FAS-ML reconstructor in **Algorithm 2** should be employed online for channel estimation. Different from these schemes, the proposed S-BAR incorporates an hybrid offline and online implementation process. Specifically, the signal processing of S-BAR scheme is composed of two components, 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 $\mathbf{\Sigma}_\Omega$, which is updated $PM$ times. According to (14), the complexity of *Stage 1* is $\mathcal{O}\left(P^2M^2\left(P^2M^2+NPM+N^2\right)\right)$. At *Stage 2*, the computational complexity is from the weighted sum of received pilot $\mathbf{y}$, i.e., $\hat{\mathbf{h}} = \mathbf{w}^{\mathrm{H}}\mathbf{y}$. Thus, the computational complexity of *Stage 2* is $\mathcal{O}\left(N\right)$. For comparison, here we summarize the computational complexity of different schemes 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, the weighted sum $\hat{\mathbf{h}} = \mathbf{w}^{\mathrm{H}}\mathbf{y}$ in *Stage 2* can be realized 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.

### D. Kernel Selection of S-BAR

Selecting an appropriate kernel $\mathbf{\Sigma}$ for S-BAR is an essential step in building an effective regression model. The choice of

$\mathbf{\Sigma}$ 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 appropriate kernel should assign higher similarity to nearby ports and decrease influence rapidly with distance. Let $\mathbf{x}_n$ denote the position of the $n$-th port. After balancing the complexity and practicality, three kernel selections are recommended as follows.

*1) Exponential Kernel:* The exponential kernel $\mathbf{\Sigma}_{\exp}$, also known as the Laplacian kernel, is the most popular choice in Bayesian regression, which is given by

$$\mathbf{\Sigma}_{\exp}(n,n') = \alpha^2 e^{-\frac{\|\mathbf{x}_n - \mathbf{x}_{n'}\|^2}{\eta^2}} \tag{20}$$

for all $n, n' \in \{1, \cdots, N\}$, where $\alpha$ and $\eta$ are adjustable hyperparameters. Compared with the other kernels, the exponential kernel is less sensitive to outliers, and it is suitable to reconstruct channels without obvious regularity.

*2) Bessel Kernel:* The Bessel kernel $\mathbf{\Sigma}_{\text{bes}}$ is well-suited for capturing and modeling complex-valued data with oscillatory or periodic patterns. The kernel is given by

$$\mathbf{\Sigma}_{\text{bes}}(n,n') = \alpha^2 J_v\left(\frac{\|\mathbf{x}_n - \mathbf{x}_{n'}\|}{\eta}\right) \tag{21}$$

for all $n, n' \in \{1, \cdots, N\}$, $J_v$ is the $v$-order Bessel function of the first kind. $\mathbf{\Sigma}_{\text{bes}}$ has the flexibility to adapt to data that exhibits regular and repeating fluctuations, which may not be adequately modeled by other kernels. Thus, $\mathbf{\Sigma}_{\text{bes}}$ is suitable to reconstruct the channels with periodic patterns.

*3) 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., $\mathbf{\Sigma}_{\text{cov}} = \mathsf{E}\left(\mathbf{h}\mathbf{h}^{\text{H}}\right)$. Since $\mathbf{\Sigma}_{\text{cov}}$ is unknown in practice, before employing S-BAR, we can train an approximated $\mathbf{\Sigma}_{\text{cov}}$ based on the statistical data. The kernel can be obtained by

$$\mathbf{\Sigma}_{\text{cov}} \approx \frac{1}{T}\sum_{t=1}^{T} \mathbf{h}_t \mathbf{h}_t^{\text{H}}, \tag{22}$$

where $\mathbf{h}_t$ is the channel at the $t$-th training timeslot and $T$ is the number of training timeslots. Since the second moment $\mathsf{E}\left(\mathbf{h}\mathbf{h}^{\text{H}}\right)$ does not change so frequently as channels, $\mathbf{\Sigma}_{\text{cov}}$ can be updated in a large timescale. In general, $\mathbf{\Sigma}_{\text{cov}}$ can achieve satisfactory estimation accuracy in most scenarios.

## 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 measurement noise on the estimation accuracy is investigated in Subsection V-C. Finally, the influence of pilot overhead on the estimation accuracy is studied in Subsection V-D.

### A. Simulation Setup and Benchmarks

Since we have assumed the normalized transmit power, the receiver signal-to-noise ratio (SNR) is defined as $\text{SNR} = \frac{\mathsf{E}\left(\|\mathbf{h}\|^2\right)}{\sigma^2}$, of which the default value is set to 20 dB. Let $\hat{\mathbf{h}}$ denote the estimated value of channel $\mathbf{h}$. The performance is evaluated by the normalized mean square error (NMSE), which is defined as $\text{NMSE} = \mathsf{E}\left(\frac{\|\mathbf{h}-\hat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2}\right)$.

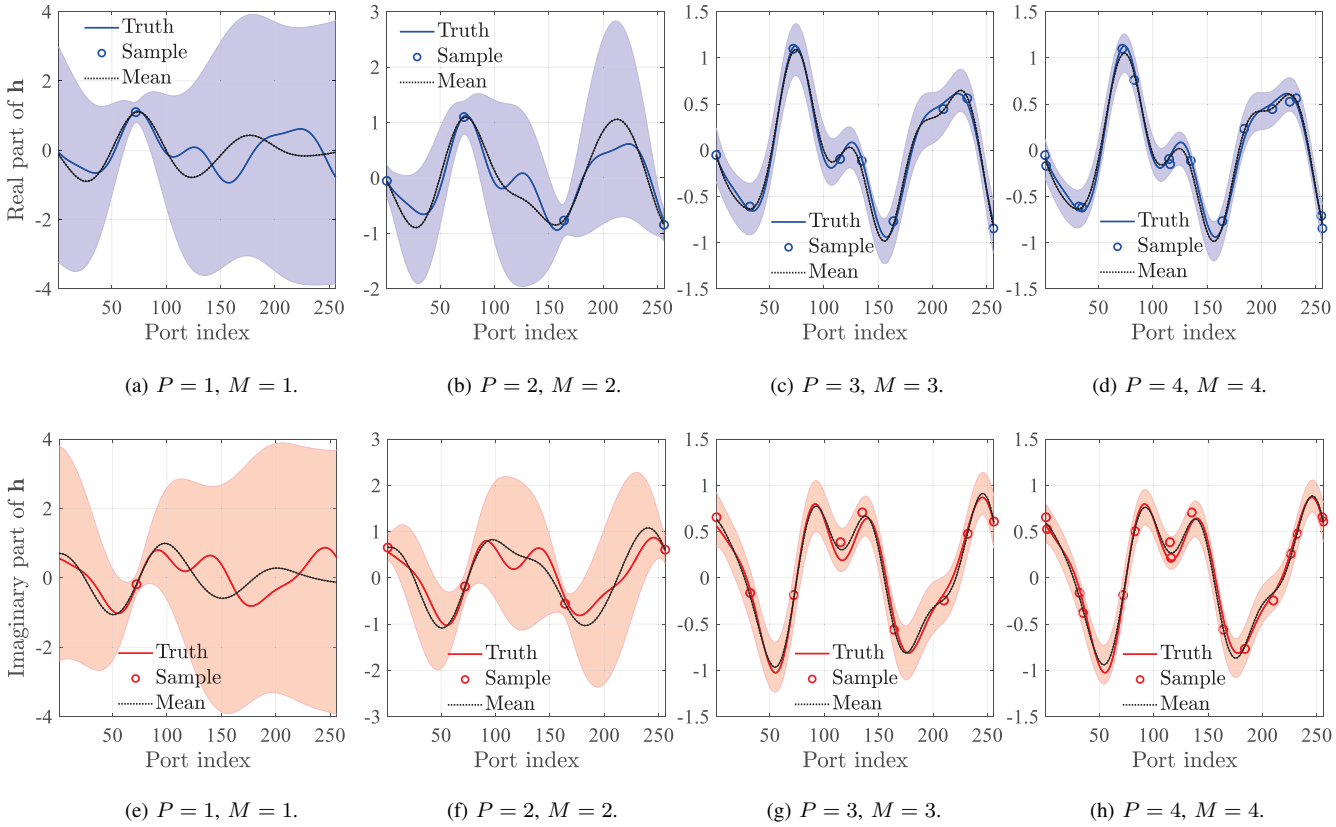
TABLE II
Simulation Parameters of FAS Channels

| Parameter | QuaDRiGa [42] | SSC in (4) |
|---|---|---|
| Carrier frequency $f_c$ | 3.5 GHz | 3.5 GHz |
| Number of clusters | 23 | 9 |
| Number of rays | 20 | 100 |
| Path gains | [43, Table 7.7.1-2] | $\mathcal{CN}(0,1)$ |
| Incident angles | [43, Table 7.7.1-2] | $\mathcal{U}(-\pi,+\pi)$ |
| Max. Angle spread | 5° | 5° |
| Path delays | [43, Table 7.7.1-2] | \ |
| Max. Doppler shift | 10 Hz | \ |

*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 practical QuaDRiGa channel model in [42] and the spatially-sparse clustered (SSC) channel model in (4). The standard channel parameters in 3GPP TR 38.901 [43] is used to generate the QuaDRiGa channels, and the main values are set as shown in Table II. For the kernel settings of S-BAR, the hyperparameters are set to $\alpha = 1$ and $\eta = \frac{\lambda}{2\pi}$ to generate the exponential kernel $\mathbf{\Sigma}_{\exp}$ and the Bessel kernel $\mathbf{\Sigma}_{\text{bes}}$ [33]. Inspired by the covariance model in [4], [5], the order of Bessel function in $\mathbf{\Sigma}_{\text{bes}}$ is set as $\nu = 0$. To account for an ideal baseline, in both channel cases, we set the number of training timeslots as $T = 100$ to train a covariance kernel $\mathbf{\Sigma}_{\text{cov}}$ as the input of S-BAR.

*2) Simulation Schemes:* To show the effectiveness of our proposed S-BAR, here 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 [30], 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 [29] 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 ($\mathbf{\Sigma}$):** Given a kernel $\mathbf{\Sigma}$, **Algorithm 4**, i.e., the proposed S-BAR, is employed to estimate $\mathbf{h}$.




(a) $P = 1$, $M = 1$. (b) $P = 2$, $M = 2$. (c) $P = 3$, $M = 3$. (d) $P = 4$, $M = 4$.

(e) $P = 1$, $M = 1$. (f) $P = 2$, $M = 2$. (g) $P = 3$, $M = 3$. (h) $P = 4$, $M = 4$.

Fig. 4. 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.

### *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. 4 to intuitively show its behavior for different system parameters. The QuaDRiGa channel model is considered to generate channel $\mathbf{h}$. To show the ideal case, the covariance kernel $\mathbf{\Sigma}_{\text{cov}}$ is used as the input of S-BAR. Fig. 4 (a)-(d) show the real part of $\mathbf{h}$ as a function of the port index, and Fig. 4 (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 switch matrix $\mathbf{S}$ designed via S-BAR. The black dotted line "Mean" denotes the posterior mean of Bayesian regression $\boldsymbol{\mu}_\Omega$, i.e., the estimated channel $\hat{\mathbf{h}}$. The highlighted shadows in the figures represent the confidence interval of $\mathbf{h}$, defined as $[\boldsymbol{\mu}_\Omega(n) - 3\mathbf{\Sigma}_\Omega(n,n), \boldsymbol{\mu}_\Omega(n) + 3\mathbf{\Sigma}_\Omega(n,n)]$ for the $n$-th port channel. From this figure, we have the following three observations.

Firstly, as the number of $P$ and $M$ increases, the confidence interval, i.e., the vertical height of shadows, is gradually reduced for both real and imaginary parts of $\mathbf{h}$. 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 $\mathbf{\Sigma}_\Omega(n,n)$ becomes sufficiently small, the posterior mean $\mu_\Omega$ can well approximate $\mathbf{h}$. Secondly, Fig. 4 (c) and Fig. 4 (g) show that, nine samples are enough for the posterior mean $\boldsymbol{\mu}_\Omega$ to well approximate a 256-dimensional $\mathbf{h}$. This encouraging result suggests that, benefiting from the prior knowledge of strongly correlated channels, very few pilots and fluid antennas are sufficient to accurately reconstruct the high-dimensional channels. Thirdly, one can note that the sample spacing is usually large. The reason is that, for each sampling, the proposed S-BAR selects the channel with the largest posterior variance for measurements. When a port is selected and measured, the channel uncertainty of its nearby ports will decrease, which reduces the trend of selecting them as the subsequent samples. In fact, the channels selected by S-BAR should be weakly correlated so as to acquire as much information as possible. Consequently, when the number of pilots/antennas is small, one can hardly find many samples within a small region.

### *C. NMSE versus the Receiver SNR*

To evaluate the effect of noise on the estimation accuracy, we plot the NMSE as a function of the receiver SNR in Fig. 5 and Fig. 6, which follow the assumptions of QuaDRiGa model and SSC model in (4), respectively. Note that, the FAS-OMP and FAS-ML reconstructors have assumed the SSC model in (4). Thus, the QuaDRiGa channel case with many clusters can be viewed as their model-mismatched case, and the SSC channel case with few clusters can be viewed as their model-matched case. Particularly, due to the lack of obvious regularity for QuaDRiGa channels, the exponential kernel

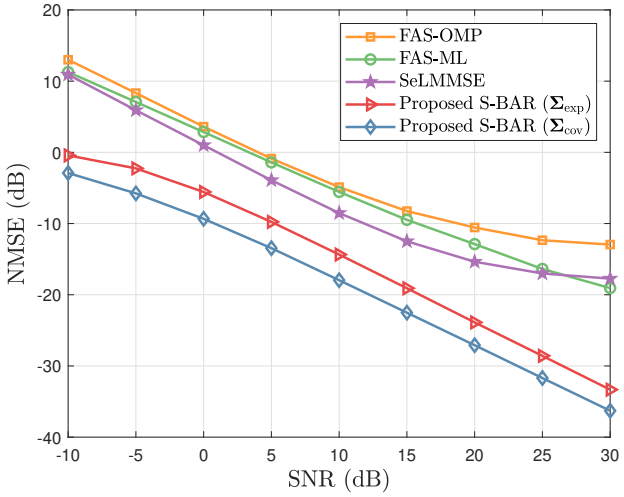



Fig. 5. The NMSE as a function of the receiver SNR under the assumption of QuaDRiGa channel model.

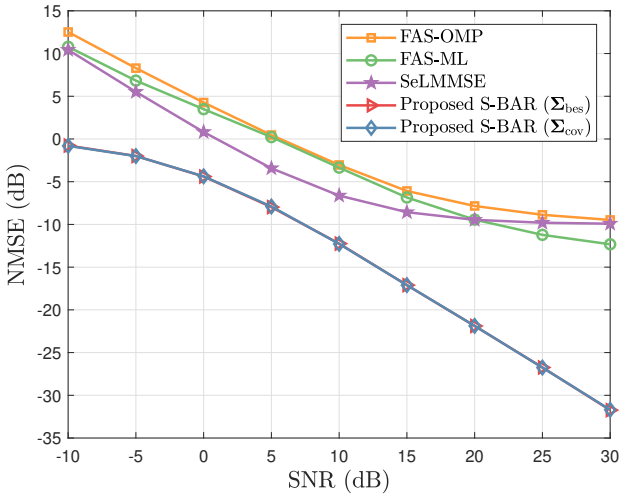



Fig. 6. The NMSE as a function of the receiver SNR under the assumption of SSC channel model.

$\mathbf{\Sigma}_{\text{exp}}$ is selected as the input of S-BAR in the QuaDRiGa case. Due to the periodic patterns of SSC channels in the spatial domain, the Bessel kernel $\mathbf{\Sigma}_{\text{bes}}$ is selected as the input of S-BAR in the SSC case [4], [5]. To show the performance of S-BAR in the ideal case, the pre-trained covariance kernel $\mathbf{\Sigma}_{\text{cov}}$ is also considered as the input of S-BAR 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, which implies the gradually higher estimation accuracy of all schemes. Particularly, one can find that the proposed S-BAR achieves the highest estimation accuracy for both SSC channels and QuaDRiGa channels. In most cases, the NMSE for S-BAR is of 5 dB lower than those for the other benchmark schemes. 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

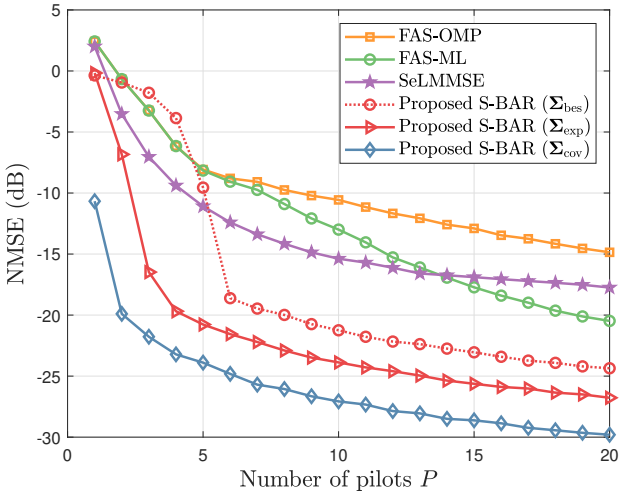



Fig. 7. The NMSE as a function of the number of pilots $P$ under the assumption of QuaDRiGa channel model.

provided by the measured channels may not be sufficient to capture all patterns of $\mathbf{h}$. In the QuaDRiGa channel case, these two estimators also suffer from the performance loss caused by the mismatched channel model. For SeLMMSE, although channel correlation is partially exploited, the unmeasured channels are directly obtained by zero-order interpolation. From the statistical perspective, the potential estimation errors of the unmeasured channels are not considered by SeLMMSE. In contrast, by building an experiential kernel, the proposed S-BAR has incorporated the effect of prior information into its estimator, which naturally considers the potential estimation errors of all channels. Through kernel-based port selection and regression, S-BAR can eliminate the uncertainty of massive channels simultaneously, even only with a small number of pilots. 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 $\mathbf{\Sigma}_{\text{exp}}$ and Bessel kernel $\mathbf{\Sigma}_{\text{bes}}$ can achieve similar performance as the covariance kernel $\mathbf{\Sigma}_{\text{cov}}$. In particular, the curves of the two S-BAR schemes are almost coincident in the SSC channel case. Recall that $\mathbf{\Sigma}_{\text{exp}}$ and $\mathbf{\Sigma}_{\text{bes}}$ are generated by experiential parameters, while $\mathbf{\Sigma}_{\text{cov}}$ is trained from real channel data. This observation indicates that, even if the real channel covariance $\mathbf{\Sigma}_{\text{cov}}$ is totally unknown, the experiential kernel $\mathbf{\Sigma}_{\text{exp}}$ and $\mathbf{\Sigma}_{\text{bes}}$ 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.

### *D. NMSE versus the Number of Pilots*

In order to study the required number of pilots to achieve a given estimation accuracy, we plot the NMSE as a function of the number of pilots $P$ in Fig. 7 and Fig. 8, which follow the assumptions of QuaDRiGa model and SSC model, respectively. In addition, to show the influence of non-preferred

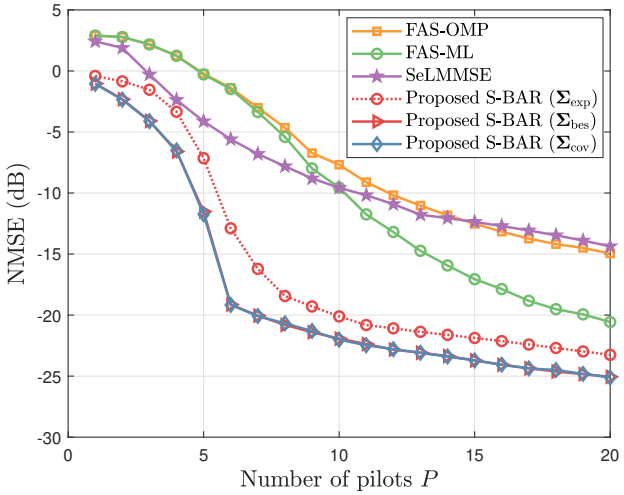



Fig. 8. The NMSE as a function of the number of pilots $P$ under the assumption of SSC channel model.

kernels on S-BAR in the two channel cases, we have also added the baseline that $\mathbf{\Sigma}_{\text{bes}}$ is selected as the input of S-BAR in Fig. 7 and the baseline that $\mathbf{\Sigma}_{\text{exp}}$ is selected as the input of S-BAR in Fig. 8. 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% when the kernel is appropriately selected. Furthermore, similar to Fig. 6 and Fig. 5, the preferred kernels $\mathbf{\Sigma}_{\text{exp}}$ and $\mathbf{\Sigma}_{\text{bes}}$ can realize similar performances to the pre-trained kernel $\mathbf{\Sigma}_{\text{cov}}$ in Fig. 7 and Fig. 8, respectively. This phenomenon further implies that, without additional training overhead to acquire $\mathbf{\Sigma}_{\text{cov}}$, the proposed S-BAR can still achieve satisfactory estimation accuracy.

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 $\mathbf{\Sigma}_{\text{bes}}$ 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 $\mathbf{\Sigma}_{\text{exp}}$ 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, one can find that, 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 influenced by both the incomplete regression and measurement noise. 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, leading to a slowly decreased NMSE. On the other hand, in the QuaDRiGa channel case, this interesting phenomenon is not observed. The reason is that, unlike the SSC channels, the property of spatial periodicity is not obvious for the QuaDRiGa channels. Thereby, the irregularity makes it difficulty for S-BAR to fully capture all patterns of channels via a few samples. It also explains why the bessel kernel $\mathbf{\Sigma}_{\text{bes}}$ is not the preferred kernel for the QuaDRiGa channel case.

## VI. Conclusions

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, since reconfigurable intelligent surfaces (RISs) can provide additional control degrees of freedom (DoFs) for FAS channels [44], the cooperation between RISs and FASs may be able to improve the accuracy of CSI acquisitions [45].

## Appendix A
## Proof of **Lemma 1**

When $\dim(\Omega) = PM$, according to the definition of $\boldsymbol{\mu}_\Omega$ in (13), the square error of employing S-BAR can be rewritten as

$$\begin{aligned}\|\boldsymbol{\mu}_\Omega - \mathbf{h}\|^2 &= \left\|\left(\mathbf{\Sigma}(\Omega,:)\right)^{\mathrm{H}}\left(\mathbf{\Sigma}(\Omega,\Omega)+\sigma^2\mathbf{I}_{PM}\right)^{-1}\mathbf{y}-\mathbf{h}\right\|^2 \\ &\overset{(a)}{=} \Big\|\Big(\left(\mathbf{S\Sigma}\right)^{\mathrm{H}}\left(\mathbf{S\Sigma S}^{\mathrm{H}}+\sigma^2\mathbf{I}_{PM}\right)^{-1}\mathbf{S}-\mathbf{I}_N\Big)\mathbf{h} \\ &\qquad + \left(\mathbf{S\Sigma}\right)^{\mathrm{H}}\left(\mathbf{S\Sigma S}^{\mathrm{H}}+\sigma^2\mathbf{I}_{PM}\right)^{-1}\mathbf{z}\Big\|^2 \\ &\overset{(b)}{=} \left\|\left(\mathbf{\Pi}^{\mathrm{H}}\mathbf{S}-\mathbf{I}_N\right)\mathbf{h}+\mathbf{\Pi}^{\mathrm{H}}\mathbf{z}\right\|^2 \end{aligned} \tag{23}$$

where $(a)$ holds since $\mathbf{\Sigma}(\Omega,:) = \mathbf{S\Sigma}$, $\mathbf{h}(\Omega) = \mathbf{Sh}$, and $\mathbf{\Sigma}(\Omega,\Omega) = \mathbf{S\Sigma S}^{\mathrm{H}}$; $(b)$ holds according to (18). Next, recalling the properties $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}_{PM}, \sigma^2\mathbf{I}_{PM}\right)$ and $\mathsf{E}\left(\mathbf{hh}^{\mathrm{H}}\right) = \mathbf{\Sigma}_{\text{cov}}$, the MSE can be derived as

$$\begin{aligned} E &= \mathsf{E}\left(\|\boldsymbol{\mu}_\Omega - \mathbf{h}\|^2\right) \\ &\overset{(c)}{=} \mathrm{Tr}\left(\left(\mathbf{\Pi}^{\mathrm{H}}\mathbf{S}-\mathbf{I}_N\right)\mathbf{\Sigma}_{\text{cov}}\left(\mathbf{\Pi}^{\mathrm{H}}\mathbf{S}-\mathbf{I}_N\right)^{\mathrm{H}}\right)+\sigma^2\mathrm{Tr}\left(\mathbf{\Pi}^{\mathrm{H}}\mathbf{\Pi}\right) \end{aligned}$$

$$= \mathrm{Tr}\left(\mathbf{\Pi}^{\mathrm{H}}\left(\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}}\mathbf{S}^{\mathrm{H}}+\sigma^2\mathbf{I}_{PM}\right)\mathbf{\Pi}\right) - 2\Re\left(\mathrm{Tr}\left(\mathbf{\Pi}^{\mathrm{H}}\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}}\right)\right) + \mathrm{Tr}\left(\mathbf{\Sigma}_{\mathrm{cov}}\right), \quad (24)$$

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

## Appendix B
## Proof of **Lemma 2**

Observing (17), one can find that the MSE $E$ is a quadratic form with respect to matrix $\mathbf{\Pi}$. Thus, the minimum MSE can be achieved by finding an $\mathbf{\Sigma}$ such that $\frac{\partial E}{\partial \mathbf{\Pi}} = \mathbf{0}_{PM\times N}$ holds. According to (17), the partial derivative of $E$ with respect to $\mathbf{\Pi}$ can be derived as

$$\frac{\partial E}{\partial \mathbf{\Pi}} = \left(\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}}\mathbf{S}^{\mathrm{H}}+\sigma^2\mathbf{I}_{PM}\right)^{*}\mathbf{\Pi}^{*} - \left(\mathbf{\Sigma}_{\mathrm{cov}}\mathbf{S}^{\mathrm{H}}\right)^{\mathrm{T}}. \quad (25)$$

By letting $\frac{\partial E}{\partial \mathbf{\Pi}} = \mathbf{0}_{PM\times N}$ and substituting (18) into (25), the original proof can be reduced to prove that

$$\left(\mathbf{S}\mathbf{\Sigma}\mathbf{S}^{\mathrm{H}}+\sigma^2\mathbf{I}_{PM}\right)^{-1}\mathbf{S}\mathbf{\Sigma} = \left(\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}}\mathbf{S}^{\mathrm{H}}+\sigma^2\mathbf{I}_{PM}\right)^{-1}\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}}. \quad (26)$$

Since the proposed S-BAR does not depend on the value of noise power, (26) should hold for arbitrary $\sigma^2$. By considering the special case when $\sigma^2 \to \infty$, (26) is equivalent to

$$\mathbf{S}\left(\mathbf{\Sigma} - \mathbf{\Sigma}_{\mathrm{cov}}\right) = \mathbf{0}_{PM\times N} \quad (27)$$

Furthermore, since our derivation also does not impose the specific form of $\mathbf{S}$, $\mathbf{\Sigma} = \mathbf{\Sigma}_{\mathrm{cov}}$ is necessary to make the equation always hold. By substituting $\mathbf{\Sigma} = \mathbf{\Sigma}_{\mathrm{cov}}$, $\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}} = \mathbf{\Sigma}_{\mathrm{cov}}(\Omega,:)$, $\mathbf{S}\mathbf{h} = \mathbf{h}(\Omega)$, and $\mathbf{S}\mathbf{\Sigma}_{\mathrm{cov}}\mathbf{S}^{\mathrm{H}} = \mathbf{\Sigma}_{\mathrm{cov}}(\Omega,\Omega)$ into (17), the MSE $E$ can be rewritten as

$$E = \mathrm{Tr}\left(\mathbf{\Sigma}_{\mathrm{cov}}\right) - \mathrm{Tr}\left(\mathbf{\Sigma}_{\mathrm{cov}}\left(:,\Omega\right)\left(\mathbf{\Sigma}_{\mathrm{cov}}\left(\Omega,\Omega\right)+\sigma^2\mathbf{I}_{PM}\right)^{-1}\mathbf{\Sigma}_{\mathrm{cov}}\left(\Omega,:\right)\right), \quad (28)$$

which is exactly the trace of the posterior covariance $\mathbf{\Sigma}_{\Omega}$ in (14). In Bayesian linear regression, the posterior variance of sample path, i.e., each diagonal entry of $\mathbf{\Sigma}_{\Omega}$, monotonically decreases as the number of samples increases [40]. Therefore, the MSE $E$ achieves its minimum when all port channels are fully observed (although it may be impractical), i.e., $\Omega = \{1,\cdots,N\}$. By substituting $\Omega = \{1,\cdots,N\}$ into (28), we obtain $E_{\min}$ in (19), which completes the proof.

## References


[1] Z. Zhang, J. Zhu, L. Dai, and R. Heath, "Successive bayesian reconstructor for FAS channel estimation," in *Proc. IEEE Wireless Commun. Netw. Conf. (IEEE WCNC'24)*, Dubai, United Arab Emirates, Apr. 2024, pp. 1–5.

[2] K. Zheng, L. Zhao, J. Mei, B. Shao, W. Xiang, and L. Hanzo, "Survey of large-scale MIMO systems," *IEEE Commun. Surv. Tutorials*, vol. 17, no. 3, pp. 1738–1760, Thirdquarter 2015.

[3] K. K. Wong, A. Shojaeifard, K.-F. Tong, and Y. Zhang, "Performance limits of fluid antenna systems," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2469–2472, Nov. 2020.

[4] K.-K. Wong, A. Shojaeifard, K.-F. Tong, and Y. Zhang, "Fluid antenna systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1950–1962, Mar. 2020.

[5] K.-K. Wong and K.-F. Tong, "Fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4801–4815, Jul. 2021.

[6] L. Zhu, W. Ma, and R. Zhang, "Movable-antenna array enhanced beamforming: Achieving full array gain with null steering," *arXiv preprint arXiv:2308.08787*, Aug. 2023.

[7] L. Zhu, W. Ma, B. Ning, and R. Zhang, "Movable-antenna enhanced multiuser communication via antenna position optimization," *arXiv preprint arXiv:2302.06978*, Feb. 2023.

[8] W. Ma, L. Zhu, and R. Zhang, "MIMO capacity characterization for movable antenna systems," *IEEE Trans. Wireless Commun. (early access)*, Sep. 2023.

[9] L. Zhu, W. Ma, and R. Zhang, "Movable antennas for wireless communication: Opportunities and challenges," *arXiv preprint arXiv:2306.02331*, Jun. 2023.

[10] 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.

[11] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," *IEEE Microw. Mag.*, vol. 5, no. 1, pp. 46–56, Jan. 2004.

[12] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[13] 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.

[14] 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.

[15] 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. 95 058–95 064, Jul. 2019.

[16] 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 in Communications and Networks*, vol. 3, p. 853416, Mar. 2022.

[17] 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.

[18] 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, 2012.

[19] 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.

[20] K. Wong, K. 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.

[21] 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. (IEEE SPAWC'22)*, Oulu, Finland, Jul. 2022, pp. 1–5.

[22] 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. (IEEE WCNC'22)*, Austin, TX, USA, Apr. 2022, pp. 78–83.

[23] Z. Chai, K.-K. Wong, K.-F. Tong, Y. Chen, and Y. Zhang, "Port selection for fluid antenna systems," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1180–1184, May 2022.

[24] 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. 2022.

[25] N. Waqar, K.-K. Wong, K.-F. 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.

[26] Z. Cheng, N. Li, J. Zhu, X. She, C. Ouyang, and P. Chen, "Sum-rate maximization for movable antenna enabled multiuser communications," *arXiv preprint arXiv:2309.11135*, Sep. 2023.

[27] Y. Wu, D. Xu, D. W. K. Ng, W. Gerstacker, and R. Schober, "Movable antenna-enhanced multiuser communication: Optimal discrete antenna positioning and beamforming," *arXiv preprint arXiv:2308.02304*, Aug. 2023.

[28] 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," *arXiv preprint arXiv:2308.09512*, Aug. 2023.

[29] 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.

[30] W. Ma, L. Zhu, and R. Zhang, "Compressed sensing based channel estimation for movable antenna communications," *arXiv preprint arXiv:2306.04333*, Jun. 2023.

[31] 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. China (IEEE/CIC ICCC Workshops'23)*, Chengdu, China, Aug. 2023, pp. 1–6.

[32] 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.

[33] C. Williams and C. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems*, vol. 8. MIT Press, 1995. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/1995/file/7cce53cf90577442771720a370c3c723-Paper.pdf

[34] 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.

[35] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.*, vol. 85, pp. 1–16, Aug. 2018.

[36] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "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.

[37] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[38] 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.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Nov. 2014, [Online] Available: https://stanford.edu/~boyd/papers/pdf/admm_distr_stats.pdf.

[40] M. A. Oliver and R. Webster, "Kriging: A method of interpolation for geographical information systems," *Int. J. Geographical Inf. Sys.*, vol. 4, no. 3, pp. 313–332, Apr. 1990.

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

[42] S. Jaeckel, L. Raschkowski, K. Börner, 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.

[43] 3GPP TR, "Study on channel model for frequencies from 0.5 to 100 GHz," *3GPP TR 38.901 version 14.0.0 Release*, Dec. 2019.

[44] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. passive RIS: Which will prevail in 6G?" *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1707–1725, Mar. 2022.

[45] A. Shojaeifard, K.-K. Wong, K.-F. Tong, Z. Chu, A. Mourad, A. Haghighat, I. Hemadeh, N. T. Nguyen, V. Tapio, and M. Juntti, "MIMO evolution beyond 5G through reconfigurable intelligent surfaces and fluid antenna systems," *Proc. IEEE*, vol. 110, no. 9, pp. 1244–1265, Sep. 2022.