Configuration Optimization and Channel Estimation in Hybrid Beamforming mmWave Systems With Channel Support Side Information

Lixiang Lian, *Member, IEEE*, and Vincent K. N. Lau, *Fellow, IEEE*

***Abstract*—Hybrid beamforming (HBF) is a promising technique in millimeter-wave (mmWave) systems with a limited number of radio frequency (RF) chains. Acquiring channel state information is essential for HBF design. Various compressive-sensing-based channel estimation (CE) schemes that exploit the sparse nature of mmWave channels have been proposed. The design of an analog beamforming matrix plays a key role in providing high CE quality with low pilot overhead. In this paper, a novel and flexible HBF architecture called adaptive sum product network (ASPN) is proposed, which can realize more flexible analog beamforming with less hardwares. Based on the ASPN architecture, a novel variational-Bayesian-inference-based compressive CE algorithm is proposed, which can fully exploit the channel support side information (CSSI) at the base station under a general uncertain measurement matrix induced by the off-grid parameters for power leakage elimination. To optimally configure the degrees of freedom in the ASPN according to the CSSI, a dynamic configuration optimization algorithm is proposed by formulating the CE performance metric in terms of the configuration parameters of ASPN. The importance and effectiveness of the flexibility introduced in ASPN HBF architecture, the compressive CE algorithm as well as the configuration algorithm are illustrated by the superb performance of the proposed scheme compared to the representative state-of-the-art techniques through extensive simulation results.**

***Index Terms*—Compressive channel estimation, channel support side information, configuration optimization, millimeter-wave systems.**

I. Introduction

MILLIMETER-WAVE (mmWave) massive multiple input multiple output (MIMO) is a promising technology for future 5G wireless communication systems, which can provide large spatial multiplexing gain and array gain to enhance both the capacity and energy efficiency of wireless systems [1]. However, the hardware cost and power consumption become unaffordable when a dedicated radio frequency (RF) chain is used for each antenna. To overcome this issue, hybrid beamforming (HBF) that splits the beamforming (BF) processing into the analog and digital domains has been proposed [2], [3] to reduce the number of RF chains. HBF designs rely on the availability of channel state information at the base station (CSIT). However, the limited number of RF chains pose a great challenge for the estimation of the high dimensional channel.

In a conventional time division duplex (TDD) MIMO system with a full set of RF chains, the downlink CSIT can be obtained via uplink training due to the channel reciprocity. However, in an HBF system with limited RF chain constraint in the base station (BS), only a small subspace of the mmWave channel can be measured at each time due to the reduced dimension of the measurement signal. Conventional uplink training requires more training pilots to accurately estimate the large scale channel. Existing literature has partially addressed the issues by designing the analog BF to enhance the signal noise ratio (SNR) of the received pilot symbols as well as exploiting sparsity in the channel matrix to be estimated [3]–[10]. Based on the hardware constraints considered in the design of analog BF, existing works on CE with HBF can be divided into following categories.

- **Constant Amplitude Constraint on Analog BF:** In practice, since the analog BF operates in the RF domain, it is desirable to exclude active elements in the realization of analog BF. There are some works that impose the constant modulus constraint in the entries of analog BF [3]–[8]. As a result, the solution requires only passive phase shifters (PSs) and passive adders only. In terms of the phase shifter resources required in the analog BF, these works can only work for static fully-connected HBF architecture, and hence, the number of RF PSs required is MS (M is the number of antennas, S is the number of RF chains), which is a huge number for massive MIMO systems.
- **Limited Phase Shifters in the Realization of Analog BF:** Since the analog BF operates in the RF frequency and hence, the PSs are expensive [9]–[11]. There are some works that considered the limited PSs in the realization of analog BF for CE with HBF. [9] introduced a PS reduced adaptive selection network for CE, in which a random part of the PS is disconnected from the antennas to reduce the power consumption. However, the authors did not design the adaptive selection network to enhance the CE performance. Moreover, a huge number of PSs are still required, i.e., $MS/10$, to achieve acceptable CE performance. The

Manuscript received September 5, 2019; revised June 4, 2020; accepted October 7, 2020. Date of publication October 15, 2020; date of current version October 29, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael B. Wakin. This work was supported in part by the Research Grants Council under Project 16209916 and in part by Huawei Technologies. *(Corresponding author: Lixiang Lian.)*

Lixiang Lian is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: lianlx@shanghaitech.edu.cn).

Vincent K. N. Lau is with the Department of ECE, The Hong Kong University of Science and Technology, Hong Kong (e-mail: eeknlau@ust.hk).

Digital Object Identifier 10.1109/TSP.2020.3031056

switches are used in [10] to replace the PSs to reduce the hardware cost. However, the switch-based HBF will cause significant performance degradation.

Due to the high price and power consumption of RF PSs [9]–[11], it is important to consider the design of analog BF for channel estimation under the constraints of limited number of RF PSs, which has not been fully addressed by the existing literature.

To accommodate to the limited PS resources, [11]–[14] proposed to design dynamic-connected HBF architecture to realize flexible analog BF matrix. However, the existing dynamic designs have the following drawbacks. First, the existing dynamic HBF architectures are designed to maximize the spectrum efficiency/sum rate or minimize the distance with a fully-digital design in [11]–[14] assuming perfect channel information without considering the CE performance. Second, the existing dynamic HBF structures have some restrictions. For example, the dynamic HBF structure in [13], [14] requires that the number of PSs is the same as the number of antennas, and one PS is used to connect each RF chain-antenna pair, which can only realize the restrictive analog BF with constant modulus entries; [11], [12] assume that the PSs are with fixed phases, which restricts the flexibility of providing a high HBF gain especially with limited PS resources; a large number of PSs are still required for each RF chain-antenna connection, i.e., ~ 40 PSs in [11] to achieve acceptable precision; the parallel processing RF PSs used in [12] can cause additional hardware costs and power consumptions.

In this paper, we consider channel estimation problem under hardware constrained HBF in the sense that the number of RF PSs and the number of RF chains are limited in the BS in the uplink direction and imperfect *channel support side information* (CSSI)[1] can be obtained at the BS. A joint CSSI-assisted configuration optimization and CCE algorithm is proposed to estimate the mmWave massive MIMO channels with low pilot overhead and a flexible HBF network. The main contributions are summarized as follows.

- **Adaptive Sum-Product Network Architecture for Analog BF:** One of the key contribution of the paper is to propose a novel and flexible hardware architecture to construct the analog BF in HBF solution. To overcome the restrictions of the existing HBF architectures, an *adaptive sum-product network* (ASPN) architecture is proposed, which is radically different from the existing architectures [2]–[14], [16] by enabling an arbitrary number of PSs with adjustable phases and a flexible mapping strategy between RF chains and antennas. For example, more than one PSs are allowed to connect one RF chain to the antenna. This allows synthesis of more general analog BF matrix where each entry can have different constant modulus. This enables more flexibility on the analog BF matrix. As a result of more flexible architecture, the proposed ASPN is more hardware efficient and can achieve much better CE performance under the same hardware constraint.
- **The Robust Channel Estimation Algorithm Based on the ASPN Solution:** Based on the proposed ASPN-based analog BF architecture, a robust and highly effective CE algorithm is proposed that can estimate the higher dimensional channel matrix from the low dimensional pilot measurements. Compared to the existing compressive sensing (CS) based CE methods, the proposed CSSI-assisted variational Bayesian inference (CAVBI) algorithm can fully exploit the imperfect CSSI under a general uncertain[2] measurement matrix induced by the flexible ASPN architecture and the off-grid parameters for power leakage elimination. The existing CS algorithms, such as orthogonal matching pursuit (OMP)-based algorithm [17], LASSO-based algorithm [6], [8], [18], approximate message passing (AMP)-based algorithm [19], [20] and Bayesian inference-based algorithm [21]–[23], which can only exploit the CSSI or handle the uncertain measurement matrix, cannot be effectively used to solve the problem considered in this paper.
- **Configuration Optimization of ASPN:** While the ASPN gives more freedom on the realization of analog BF matrix, it is highly non-trivial on how to configure the degrees of freedom dynamically based on CSSI. We pose the dynamic configuration as an optimization problem by minimizing the mean square error (MSE) of the compressive CE algorithm. The key challenge is on the development of the closed form objective function (MSE of CE algorithm) in terms of the configuration parameters. Furthermore, the problem is highly nonconvex. Using successive convex approximation (SCA) technique, a low complexity CSSI-assisted parallel successive convex approximation (CAP-SCA) configuration optimization algorithm is proposed.

The rest of the paper is organized as follows. In Section II, the system model for the mmWave CE is introduced. In Section III, we present the ASPN analog BF architecture. After this, we elaborate the CSSI-assisted CCE algorithm in Section IV for a given ASPN architecture and introduce the CSSI-assisted configuration optimization algorithm in Section V to optimally configure the degrees of freedom of ASPN architecture. Finally, the simulation results and the conclusion are given in Section VI and Section VII, respectively.

Notations: Upper case and lower case bold face letters denote matrices and vectors, respectively. $\mathbf{x}[m]$ means the m-th element of vector $\mathbf{x}$. $\mathcal{CN}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ stands for complex Gaussian vector $\mathbf{x}$ with expectation $\boldsymbol{\mu}$ and co-variance matrix $\boldsymbol{\Sigma}$. $\mathrm{tr}(\mathbf{X})$ stands for the trace of matrix $\mathbf{X}$. $\mathbf{X}=\mathrm{diag}(\mathbf{x})$ means matrix $\mathbf{X}$ is a diagonal matrix with $\mathbf{x}$ on the diagonal. $\mathbf{A}=\mathrm{Diag}(\mathbf{A}_1,\mathbf{A}_2)$ means matrix $\mathbf{A}$ is a diagonal block matrix with $\mathbf{A}_1$ and $\mathbf{A}_2$ on the diagonal. $\mathrm{vec}(\mathbf{X})$ means the vectorization of matrix $\mathbf{X}$. $a=\Theta(1)$ means that a is of order 1. $\delta(\cdot)$ stands for the Dirac delta function. $|\mathcal{H}|$ means the cardinality of the set $\mathcal{H}$. $\mathbf{X}^{-1}$, $\mathbf{X}^T$ and $\mathbf{X}^H$ denotes the inverse, transpose and conjugate transpose

[1]In practice, the channel support usually changes slowly [15] and the CSSI at the BS can be obtained from previously estimated channel supports [8].

[2]Here, the generality of the measurement matrix is induced by the flexible ASPN architecture. The uncertainty of the measurement matrix refers to the off-grid parameters introduced to eliminate the power leakage in the sparse representation of the mmWave channel.

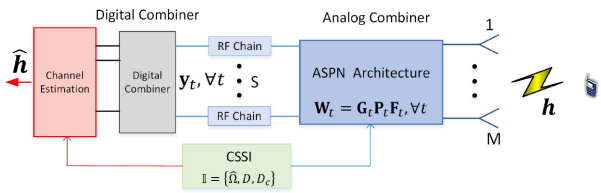

Fig. 1. Illustration of the uplink channel estimation with HBF reception at the BS.

of matrix $\mathbf{X}$, respectively. $\mathbb{N}_+$ stands for the positive integer set. Mod denotes the modulo operation. $(x)^*$ means the conjugate of x. $\mathbf{1}_{N\times M}$ stands for the all-one matrix with dimension $N \times M$. $\mathbf{I}_{N\times N}$ stands for the $N \times N$ dimensional identity matrix.

II. System Model

A. Uplink Channel Estimation With HBF At the BS

In this paper, we focus on the uplink mmWave channel estimation problem with a massive MIMO BS communicating with a single antenna user[3] and consider hybrid beamforming reception at the BS side where the number of RF chains S is much less than the number of antennas M. Note that channel estimation is performed at the BS during a training phase based on uplink pilot symbols. At the training phase, the user transmits T uplink training symbols to the BS. BS processes the received training symbols with an analog BF matrix $\mathbf{W}_t \in \mathbb{C}^{M\times S}$ at the t-th training slot to convert the M-dimensional received RF signals to the S-dimensional baseband signals. The baseband signal $\mathbf{y}_t \in \mathbb{C}^S$ at the S outputs of the RF chains will be used as baseband measurements to recover the original higher dimensional channel as illustrated in Fig. 1.[4]

Without loss of generality, we assume T identical pilot symbols are transmitted and the average transmit power is normalized to be one. After analog combining, the received baseband signal $\mathbf{y}_t \in \mathbb{C}^S, t = 1, \ldots, T$ at the BS can be written as

$$\mathbf{y}_t = \mathbf{W}_t^H (\mathbf{h} + \tilde{\mathbf{n}}_t) = \mathbf{W}_t^H \mathbf{h} + \mathbf{n}_t, t = 1, \ldots, T, \quad (1)$$

where $\mathbf{h} \in \mathbb{C}^M$ is the uplink channel. For clarity, we focus on a narrowband system with flat block fading channel, but the proposed algorithm can be easily modified to cover the wideband system as well. $\tilde{\mathbf{n}}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2\boldsymbol{I})$ is the measurement noise at the receiver antennas, and $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{W}_t^H\mathbf{W}_t)$ is the effective noise after combining using analog BF matrix $\mathbf{W}_t$. Let $\mathbf{y} = [\mathbf{y}_1; \cdots; \mathbf{y}_T]$ and $\mathbf{n} = [\mathbf{n}_1; \cdots; \mathbf{n}_T] \in \mathbb{C}^{ST}$. Then (1) can be expressed in a compact form as [5]–[9],

$$\mathbf{y} = \mathbf{W}^H\mathbf{h} + \mathbf{n}, \quad (2)$$

where $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}_n)$, with

$$\boldsymbol{\Sigma}_n = \mathrm{Diag}\left(\mathbf{W}_1^H\mathbf{W}_1, \ldots, \mathbf{W}_T^H\mathbf{W}_T\right). \quad (3)$$

Given $\mathbf{y}$ and $\mathbf{W}$, the goal of CE is to recover the uplink channel $\mathbf{h}$. Note that for each training slot, only an S dimensional signal can be measured due to the employment of HBF with S RF chains. If the least squares (LS) method is used to estimate the CSI $\mathbf{h}$ from $\mathbf{y}$, it requires at least M/S pilots, which will cause a huge CSI signaling overhead for the HBF system with $S \ll M$. Hence, it is necessary to design more effective CE methods to reduce the required training overhead under HBF system.

B. Off-Grid-based Sparse Channel Model

1) MmWave Channel Model: Since mmWave channels are expected to have limited scattering [24]–[26], we adopt a geometric channel model with L multipaths. Under this model, the channel $\mathbf{h}$ can be expressed as [3], [5], [10]

$$\mathbf{h} = \sum_{l=1}^{L} \alpha_l \mathbf{a}(\theta_l), \quad (4)$$

where α_l and θ_l represents the complex gain and angle of arrival (AoA) of the l-th multipath, respectively, $\mathbf{a}(\theta)$ is the array response vector at the BS for AoA θ. Let $\mathbf{p}_m = [p_m^x, p_m^y]^T$ be the position of the m-th antenna relative to the array's gravity center. For arrays without mutual antenna coupling and isotropic antennas, the array response vector $\mathbf{a}(\theta)$ for a given AoA θ has the following expression [27, Eq. (2)]:

$$[\mathbf{a}(\theta)]_m = \exp\left(\frac{2\pi i}{\lambda}\mathbf{p}_m^T\begin{bmatrix}\cos(\theta)\\ \sin(\theta)\end{bmatrix}\right), \quad (5)$$

where $[\mathbf{a}(\theta)]_m$ denotes the m-th element of $\mathbf{a}(\theta)$ and λ denotes the wavelength of the uplink propagation.

2) Off-Grid Basis for the Sparse Channel Model: Let $\{\hat{\vartheta}_1, \ldots, \hat{\vartheta}_M\}$ be a uniform sampling grid, which uniformly covers the angular domain $(0, 2\pi]$. In practice, the true AoAs $\{\theta_l\}_{l=1}^L$ usually do not lie exactly on the grid points. To handle the direction mismatch, we adopt an off-grid model for the angular domain channel representation. Specifically, if $\theta_l \notin \{\hat{\vartheta}_1, \ldots, \hat{\vartheta}_M\}$ and $\hat{\vartheta}_{m_l}$, $m_l \in \{1, \ldots, M\}$ is the nearest grid point to θ_l, θ_l can be written as

$$\theta_l = \hat{\vartheta}_{m_l} + \beta_{m_l}, \quad (6)$$

where β_{m_l} is the off-grid gap. Then we have $\mathbf{a}(\theta_l) = \mathbf{a}(\hat{\vartheta}_{m_l} + \beta_{m_l})$. The uplink channel $\mathbf{h}$ in (4) has a sparse representation with an off-grid basis as given by

$$\mathbf{h} = \mathbf{A}(\boldsymbol{\beta})\mathbf{x}, \quad (7)$$

where $\mathbf{x} \in \mathbb{C}^M$ is the sparse angular domain channel, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_M]^T$, $\mathbf{A}(\boldsymbol{\beta}) = [\mathbf{a}(\hat{\vartheta}_1 + \beta_1), \ldots, \mathbf{a}(\hat{\vartheta}_M + \beta_M)]$, and

$$\beta_{m_l} = \begin{cases}\theta_l - \hat{\vartheta}_{m_l}, & l = 1, \cdots L\\ 0, & \text{otherwise}\end{cases}. \quad (8)$$

[3]For a multi-user system, the users can send orthogonal pilot sequences to the BS for uplink training, there is no interference between different users. Therefore, we focus on the uplink CE problem for a reference user in this paper for clarity.

[4]Note that the analog BF matrix plays a key role in the channel estimation. On the one hand, with an HBF architecture, the mmWave channel can only be observed through the analog combining network, which projects the received RF signal to the baseband. On the other hand, the analog BF should also provide SNR gain to the received pilot symbols if properly designed. Therefore, the analog BF design has been the key focus in most of the existing papers studying CE problem with HBF architecture [4]–[10] and digital BF does not play any role in the CE based on the baseband measurements $\mathbf{y}_t$.

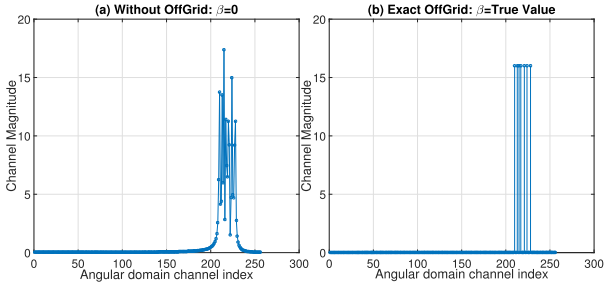

Fig. 2. Simulated angular domain channel using the mm-SSCM channel model. Consider a uniform linear array (ULA) with $M = 256$.

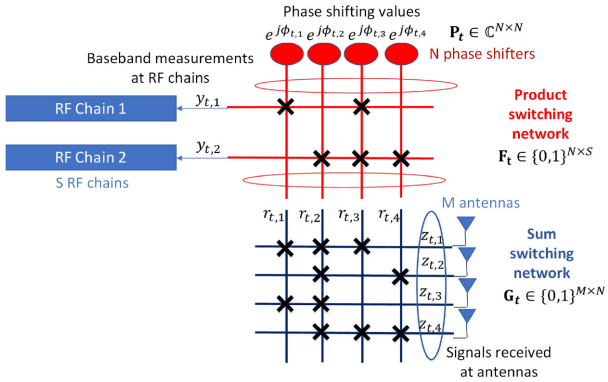

Fig. 3. The implementation of proposed adaptive HBF architecture. Assume $M = 4, S = 2, N = 4$, and t-th training slot. Cross on the switching fabric means the switch is on. $\mathbf{z}_t \in \mathbb{C}^4$ is the received signal at the antennas, $\mathbf{r}_t \in \mathbb{C}^4$ is the output signal of sum switching network, and the input signal of product switching network, $[e^{j\phi_{t,1}}, e^{j\phi_{t,2}}, e^{j\phi_{t,3}}, e^{j\phi_{t,4}}] \in \mathbb{C}^4$ stands for the phase shifting vector of 4 PSs.

Note that with the off-grid basis, the model could significantly alleviate the direction mismatch because there always exists some β_{m_l} making (6) hold exactly.

The sparse channel model in (4) and (7) is verified using the 28 GHz mmWave statistical spatial channel model (mm-SSCM) proposed in [26], which was developed based on the high frequency propagation measurements in New York City, and has been shown to faithfully reproduce realistic impulse responses of measured urban channels. Fig. 2 illustrates a simulated mmWave channel using the mm-SSCM. It can be seen that most of the channel energy concentrates on a few elements in the angular domain. Moreover, the channel tends to be more sparse if the off-grid parameter $\boldsymbol{\beta}$ can be accurately learned, and the power leakage induced by angle mismatch will be eliminated.

According to (7), the received signal $\mathbf{y}$ in (2) can be rewritten by

$$\mathbf{y} = \boldsymbol{\Phi}(\boldsymbol{\beta})\mathbf{x} + \mathbf{n}, \tag{9}$$

where $\boldsymbol{\Phi}(\boldsymbol{\beta}) = \mathbf{W}^H\mathbf{A}(\boldsymbol{\beta})$. To accurately estimate the mmWave channel, we need to jointly recover the sparse signal $\mathbf{x}$ and refine the off-grid vector $\boldsymbol{\beta}$ given the measurements $\mathbf{y}$ and analog BF matrix $\mathbf{W}$. Besides the sparse nature of the mmWave channel, the prior information available at the BS can also be exploited to enhance the CE performance.

C. Channel Support Side Information at the BS

Denote $\Omega \triangleq \{m : \mathbf{x}[m] \neq 0\}$ as the support of channel vector $\mathbf{x}$ in (9). In this paper, we assume that the BS knows an estimated channel support $\hat{\Omega}$, an upper bound D of the size of the true channel support (i.e., $|\Omega| \leq D$), and a lower bound D_c of the size of the intersection between the estimated and true channel support (i.e., $|\Omega \cap \hat{\Omega}| \geq D_c$). The *information set* $\mathbb{I} = \{\hat{\Omega}, D, D_c\}$ known at the BS is called *channel support side information* (CSSI). Denote the sparsity ratio lower bound within $\hat{\Omega}$ as $\kappa_1 = \frac{D_c}{D}$ and the sparsity ratio upper bound outside $\hat{\Omega}$ as $\kappa_2 = \frac{D-D_c}{M-D}$. One method to obtain CSSI is described as follows. As shown in [15], channel supports usually exhibit strong temporal correlation over consecutive time slots, and thereby, the CSSI $\mathbb{I}$ can be estimated from the previously estimated T_s channel supports $\{\hat{\Omega}(-T_s), \ldots, \hat{\Omega}(-1)\}$. Specifically, $\hat{\Omega}$ can be chosen as the latest estimated channel support, i.e., $\hat{\Omega} = \hat{\Omega}(-1)$.$D$ and D_c can be estimated as $D = \max_{t\in\{1,\ldots,T_S\}} |\hat{\Omega}(-t)|$ and $D_c = \min_{t\in\{2,\ldots,T_s\}} |\hat{\Omega}(-t) \cap \hat{\Omega}(-t+1)|$, respectively.

Similar CSSI has also been assumed in prior works on CS recovery algorithms [18], [28] and CCE algorithms [6], [8], [17]. As illustrated in [18], the CSSI quality depends on how much information it can provide about the position of the support. Specifically, when κ_1 and κ_2 deviate more from the mean $\frac{D}{M}$ (e.g., either close to 1 or close to 0), the CSSI quality is higher. But when $\kappa_1 = \kappa_2 = \frac{D}{M}$, we cannot obtain more prior information other than the sparsity level.

III. Adaptive HBF Architecture

The existing HBF architectures, such as the static fully-connected structure [2]–[8], the static partially-connected structure [16] and the dynamic fully-connected or partially-connected structure [9]–[14], have various restrictions. For example, the elements of analog BF matrix are restricted to have a constant modulus in [2]–[10], [13], [14], [16]; some only support a fixed number of PSs, e.g., MS PSs in [2]–[8] and M PSs in [13], [14], [16]; and dynamic HBF structures are designed assuming perfect CSI in [10]–[14]. To overcome these restrictions and consider the hardware constrained HBF with limited number of RF PSs, a flexible architecture, i.e.,*adaptive sum-product network* (ASPN) is proposed in this section, which enables an arbitrary number of PSs and a flexible mapping strategy between RF chains and antennas to realize a more flexible analog BF matrix with non-constant element modulus using a small number of PSs. To configure the degrees of freedom in the ASPN according to the CSSI, a dynamic configuration algorithm will be introduced in Section V.

A. Adaptive Sum-Product Hybrid Beamforming Network Architecture

For clarity, we focus on the t-th training slot ($1 \leq t \leq T$). The proposed ASPN contains $N \in \mathbb{N}_+$ PSs, a *product switching network* and a *sum switching network*, as illustrated in Fig. 3.

Specifically, the received signals at the antennas are passed through the sum switching network to generate different sum terms. The sum terms are then passed through the product switching network, where the phases of different sum terms are shifted by available PSs to generate different product terms, and the product terms are adaptively combined at the RF chains to compose the baseband measurements.

Accordingly, the analog BF matrix has the following form:

$$\mathbf{W}_t = \mathbf{G}_t \mathbf{P}_t \mathbf{F}_t. \tag{10}$$

$\mathbf{G}_t \in \{0,1\}^{M\times N}$ describes the sum switching network, which maps the signal from M antennas to N PSs to generate different sum terms. $\mathbf{F}_t \in \{0,1\}^{N\times S}$ describes the product switching network, which is connected to the PS resources and maps the sum terms to the RF chains to generate different product terms. $\mathbf{P}_t \in \mathbb{C}^{N\times N}$ describes the phase shifting operation of the available N PSs, given by a diagonal matrix

$$\mathbf{P}_t = \mathrm{diag}\left(\left[e^{j\phi_{t,1}}, \ldots, e^{j\phi_{t,N}}\right]\right). \tag{11}$$

Considering B bits quantized PSs, the $\phi_{t,n}, \forall n \in \{1,\ldots,N\}$ can take values from the following quantized angle set:

$$\phi_{t,n} \in \mathcal{A} \triangleq \left\{0, \frac{2\pi}{2^B}, \ldots, \frac{\left(2^B-1\right)2\pi}{2^B}\right\}. \tag{12}$$

Remark 1: All the hardware components needed for the ASPN implementation are N quantized PSs and $NM + NS$ switches to compose the sum and product switching fabric. According to the detailed study in [10], the price and energy consumption of switches are insignificant compared to those of PSs [10], [11], [29]. Furthermore, the mmWave switches are capable of switching at a fraction of nanosecond speeds where smart switches have also been used for 60 GHz application at the receiver side in [30]. For these reasons, it's believed that the introduction of switches will incur negligible delay, cost and energy consumption in the proposed HBF design. The advantages of the proposed architecture in terms of hardware efficiency will be verified in Section VI.

B. Toy Example

Assume $M = 4$ antennas are coupled to $S = 2$ RF chains via an analog combining matrix $\mathbf{W}_t$ with $N = 4$ PSs. The configurations in Fig. 3 will be used as an example to illustrate the proposed ASPN architecture.

Denote the received RF signal at antennas as $\mathbf{z}_t \in \mathbb{C}^M$. After passing through the sum switching network, the output signal $\mathbf{r}_t \in \mathbb{C}^N$ can be written as

$$\mathbf{r}_t = \mathbf{G}_t^T \mathbf{z}_t, \mathbf{G}_t = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}. \tag{13}$$

Let $g_{t,m,n}$, $z_{t,m}$, $r_{t,n}$ denote the (m,n)-th element of $\mathbf{G}_t$, m-th element of $\mathbf{z}_t$ and n-th element of $\mathbf{r}_t$, respectively. $g_{t,m,n} = 1$ means $z_{t,m}$ is passed to the n-th output port to generate the sum terms with the other incoming signals to compose $r_{t,n}$; $g_{t,m,n} = 0$ means $z_{t,m}$ can not be passed to the n-th output port. After this, $\mathbf{r}_t$ will be passed through the product switching network, and the output signal $\mathbf{y}_t \in \mathbb{C}^S$ can be written as

$$\mathbf{y}_t = \mathbf{F}_t^T \mathbf{P}_t^H \mathbf{r}_t, \tag{14}$$

with

$$\mathbf{P}_t = \mathrm{diag}\left(\left[e^{j\phi_{t,1}}, e^{j\phi_{t,2}}, e^{j\phi_{t,3}}, e^{j\phi_{t,4}}\right]\right), \mathbf{F}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Specifically, let $f_{t,n,s}$, $e^{j\phi_{t,n}}$, $y_{t,s}$ denote the (n,s)-th element of $\mathbf{F}_t$, n-th diagonal value of $\mathbf{P}_t$ and s-th element of $\mathbf{y}_t$, respectively. $f_{t,n,s} = 1$ means $r_{t,n}$ is phase shifted by the n-th PS with $e^{-j\phi_{t,n}}$ to generate the product term, which is passed to the s-th RF chain to be combined with the other incoming phase-shifted signals to compose the final baseband measurement $y_{t,s}$; $f_{t,n,s} = 0$ means the signal $r_{t,n}$ doesn't contribute to the baseband measurement $y_{t,s}$.

Combining (13) and (14), we can get that

$$\mathbf{y}_t = \mathbf{F}_t^T \mathbf{P}_t^H \mathbf{G}_t^T \mathbf{z}_t = \mathbf{W}_t^H \mathbf{z}_t,$$

with $\mathbf{W}_t$ given by

$$\mathbf{W}_t = \begin{bmatrix} e^{j\phi_{t,1}} + e^{j\phi_{t,3}} & e^{j\phi_{t,2}} + e^{j\phi_{t,3}} \\ 0 & e^{j\phi_{t,2}} + e^{j\phi_{t,4}} \\ e^{j\phi_{t,1}} & e^{j\phi_{t,2}} \\ e^{j\phi_{t,3}} & e^{j\phi_{t,2}} + e^{j\phi_{t,3}} + e^{j\phi_{t,4}} \end{bmatrix}.$$

It can be seen that the proposed architecture eliminates the constant modulus constraints of the traditional HBF architectures by adaptively configuring each RF chain-antenna pair through the design of the ASPN, resulting in terms like 0, or the sum of several complex exponents. Hence, our architecture provides more flexibility to make sophisticated processing in analog BF and hence, higher BF gain to be realized.

IV. CSSI-Assisted VBI Algorithm for Compressive Channel Estimation

The proposed ASPN architecture can realize more flexible analog BF matrix subject to the limited number of PSs. Hence, we need to design a CE algorithm that can fully exploit the flexibility (and constraints) with the proposed ASPN architecture. Specifically, we need to propose a new channel estimation algorithm which can fully exploit the imperfect CSSI and works well under a general uncertain measurement matrix induced by the flexible ASPN architecture and the off-grid parameters for power leakage elimination. The existing prior information exploited CS algorithms in [6], [8], [17]–[20], [28] cannot be directly used, because the algorithms proposed therein can only work well under measurement matrix with good properties, such as independent and identically distributed (i.i.d.) Gaussian or partial orthogonal. Bayesian inference-based algorithms, such as variational Bayesian inference (VBI) [23] and sparse Bayesian learning (SBL) [21], [22] have been widely applied to solve the CS problem. However, the existing Bayesian methods only work for i.i.d. sparse prior or simple group-sparse prior, they cannot capture the sparsity structure induced by CSSI. Therefore, it's

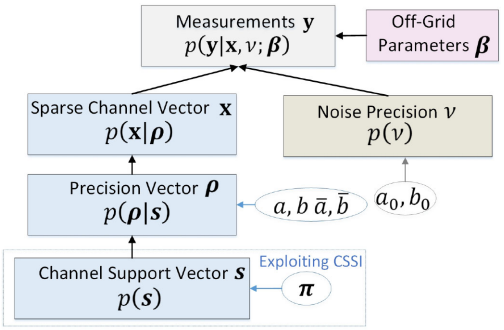

Fig. 4. Illustration of the three-layer probability model for CSSI-assisted CCE.

necessary to propose a novel CE algorithm to effectively estimate the mmWave channel under ASPN architecture.

A. Three-Layer Probability Model

To capture the sparsity structure of a mmWave channel induced by CSSI and facilitate the calculation of the proposed CAVBI algorithm, we propose a three-layer hierarchical probability model. The first layer of random variable is channel support vector $\boldsymbol{s} = [s_1, \ldots, s_M]^T \in \{0,1\}^M$, where s_m represents whether there is an active path arriving at the BS from the m-th angle grid. The second layer of random variable is precision vector $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_M] \in \mathbb{R}^M$, where ρ_m represents the precision (inverse of the variance) of x_m. The third layer of random variable is the sparse channel vector $\mathbf{x}$. Then the joint prior distribution of $\mathbf{x}, \boldsymbol{\rho}, \boldsymbol{s}$ is given by

$$p(\mathbf{x}, \boldsymbol{\rho}, \boldsymbol{s}) = \underbrace{p(\mathbf{x}|\boldsymbol{\rho})}_{\text{Sparse Channel}} \underbrace{p(\boldsymbol{\rho}|\boldsymbol{s})}_{\text{Precision}} \underbrace{p(\boldsymbol{s})}_{\text{Channel Support}}. \quad (15)$$

Fig. 4 illustrates the structure of the proposed probability model, and the details are elaborated below.

1) Probability Model for the Channel Support Vector: Based on CSSI $\mathbb{I}$, the prior distribution of $\boldsymbol{s}$can be obtained as

$$p(\boldsymbol{s}) = \prod_{m=1}^{M} \pi_m^{s_m} (1-\pi_m)^{1-s_m}, \quad (16)$$

where π_m is the prior probability of $p(s_m = 1)$, which can be calculated based on the CSSI as

$$\pi_m = \begin{cases} \kappa_1, & \text{if } m \in \hat{\Omega} \\ \kappa_2, & \text{if } m \notin \hat{\Omega} \end{cases}. \quad (17)$$

2) Probability Model for the Precision Vector and Sparse Channel: To allow the flexibility to model local characteristics of the signal, a non-stationary Gaussian prior with a distinct precision ρ_m for each element x_m of $\mathbf{x}$ is considered, i.e.,

$$p(\mathbf{x}|\boldsymbol{\rho}) = \prod_{m=1}^{M} \mathcal{CN}\left(x_m; 0, \rho_m^{-1}\right). \quad (18)$$

The precisions ρ_m are further constrained by treating them as random variables and imposing a Gamma prior distribution to them. The Gamma prior is selected because it is conjugate to the Gaussian, hence the associated Bayesian inference can be performed in closed form [22], [23]. The conditional prior of precision vector $\boldsymbol{\rho}$ is given by

$$p(\boldsymbol{\rho}|\boldsymbol{s}) = \prod_{m=1}^{M} \Gamma(\rho_m; a, b)^{s_m} \Gamma\left(\rho_m; \bar{a}, \bar{b}\right)^{1-s_m}, \quad (19)$$

where $\Gamma(\rho; a, b)$ is a Gamma hyper-prior with shape parameter a and rate parameter b.

The priors in (18) and (19) can be used to capture the sparse nature of mmWave channels, as explained below. When $s_m = 1$, there is an active path from the m-th angle grid to the BS. In this case, the shape and rate parameter a, b of the precision ρ_m should be chosen such that $\frac{a}{b} = \mathbb{E}[\rho_m] = \Theta(1)$ since the variance ρ_m^{-1} of x_m is $\Theta(1)$ when it is active. When $s_m = 0$, there is no active path from the m-th angle grid to the BS. In this case, the shape and rate parameter $\bar{a}, \bar{b}$ of the precision ρ_m should be chosen such that $\frac{\bar{a}}{\bar{b}} = \mathbb{E}[\rho_m] \gg 1$ since the variance ρ_m^{-1} of x_m is close to zero when it is inactive.

B. Problem Formulation

Based on the measurement model (9) and the assumption of complex Gaussian noise, we have

$$p(\mathbf{y}|\mathbf{x}, \nu; \boldsymbol{\beta}) = \mathcal{CN}\left(\mathbf{y}; \boldsymbol{\Phi}(\boldsymbol{\beta})\mathbf{x}, \nu^{-1}\boldsymbol{\Sigma}_n\right), \quad (20)$$

where $\nu = \sigma^{-2}$ represents the noise precision. Since ν is usually unknown, it can be modeled as a Gamma hyperprior

$$p(\nu) = \Gamma(\nu; a_0, b_0), \quad (21)$$

where we set $a_0, b_0 \to 0$ as in [21]–[23].

Denote the complete hidden variables as $\boldsymbol{\Theta} = \{\nu, \mathbf{x}, \boldsymbol{\rho}, \boldsymbol{s}\}$. Given measurements $\mathbf{y}$ and analog BF matrix $\mathbf{W}$, our goal is to estimate the angular domain channel $\mathbf{x}$ and refine the off-grid parameters $\boldsymbol{\beta}$ exploiting the imperfect CSSI. In particular, the off-grid parameters $\boldsymbol{\beta}$ are obtained by maximizing the likelihood function as follows:

$$\begin{aligned} \boldsymbol{\beta}^\star &= \arg\max_{\boldsymbol{\beta}} \ln p(\mathbf{y}; \boldsymbol{\beta}) \\ &= \arg\max_{\boldsymbol{\beta}} \int_{\boldsymbol{\Theta}} \ln p(\mathbf{y}, \boldsymbol{\Theta}; \boldsymbol{\beta})\, d\boldsymbol{\Theta}. \end{aligned} \quad (22)$$

For a given $\boldsymbol{\beta}$, we aim at computing the conditional marginal posterior $p(\mathbf{x}|\mathbf{y}; \boldsymbol{\beta})$, where

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}; \boldsymbol{\beta}) &\propto \int_{\boldsymbol{\Theta}_{-\mathbf{x}}} p(\mathbf{y}, \boldsymbol{\Theta}; \boldsymbol{\beta})\, d_{-\mathbf{x}} \\ &= \int_{\boldsymbol{\Theta}_{-\mathbf{x}}} p(\mathbf{x}, \boldsymbol{\rho}, \boldsymbol{s})\, p(\mathbf{y}|\mathbf{x}, \nu; \boldsymbol{\beta})\, p(\nu) d_{-\mathbf{x}}. \end{aligned} \quad (23)$$

$\propto$ denotes equality after scaling and $\int_{\boldsymbol{\Theta}_{-\mathbf{x}}}(\cdot)d_{-\mathbf{x}}$ denotes integration over $\boldsymbol{\rho}, \boldsymbol{s}, \nu$. Once the estimate of $\boldsymbol{\beta}$, i.e., $\boldsymbol{\beta}^\star$, and the associated posterior $p(\mathbf{x}|\mathbf{y}; \boldsymbol{\beta}^\star)$ are obtained, the maximum a posteriori (MAP) estimation of $\mathbf{x}$ can also be obtained through $\mathbf{x}^\star = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}; \boldsymbol{\beta}^\star)$.

However, it is challenging to solve the problem in (22) and calculate the posterior in (23) directly, because it is hard to calculate the closed-form multi-dimensional integration over

$\boldsymbol{\Theta}$ in (22) and (23). Moreover, the objective in (22) is a high-dimensional non-convex function. It is difficult to directly use the optimization method (i.e., gradient method) to solve it. To overcome these challenges, we propose a CAVBI algorithm to find the approximation of the posterior distribution in (23), and the approximate stationary point of problem (22).

C. Overview of the CAVBI Algorithm

The basic idea of the CAVBI algorithm is to simultaneously approximate the intractable posterior $p(\mathbf{x}|\mathbf{y};\boldsymbol{\beta})$ with a tractable variational distribution $q(\mathbf{x};\boldsymbol{\beta})$ and maximize the log-likelihood $\ln p(\mathbf{y};\boldsymbol{\beta})$ with respect to (w.r.t.) $\boldsymbol{\beta}$ as in (22). In summary, the CAVBI algorithm performs iterations between the following two major steps until convergence.

- **CAVBI-E Step:** Given $\boldsymbol{\beta}$, calculate the approximate posterior of $p(\mathbf{x}|\mathbf{y};\boldsymbol{\beta})$, denoted as $q(\mathbf{x};\boldsymbol{\beta})$ by incorporating the CSSI into the inference process, as elaborated in Section IV-C2. An approximate posterior of $p(\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\beta})$ can also be obtained, denoted as $q(\boldsymbol{\Theta};\boldsymbol{\beta})$.
- **CAVBI-M Step:** Given $q(\boldsymbol{\Theta};\boldsymbol{\beta}) \approx p(\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\beta})$, construct a surrogate function (lower bound) for the objective function $\ln p(\mathbf{y};\boldsymbol{\beta})$, then maximize the surrogate function w.r.t. $\boldsymbol{\beta}$, as elaborated in Section IV-C1.

1) Cavbi-M Step: It is difficult to directly maximize the log-likelihood function $\ln p(\mathbf{y};\boldsymbol{\beta})$ due to the lack of closed-form expression. To make the problem tractable, in the CAVBI-M step, an inexact majorization-minimization (MM) method in [21] is adopted, which maximizes a surrogate function of $\ln p(\mathbf{y};\boldsymbol{\beta})$ w.r.t. $\boldsymbol{\beta}$. Specifically, let $u(\boldsymbol{\beta};\dot{\boldsymbol{\beta}})$ be the surrogate function constructed at some fixed point $\dot{\boldsymbol{\beta}}$, which satisfies the following properties:

$$u(\dot{\boldsymbol{\beta}};\dot{\boldsymbol{\beta}}) = \ln p\left(\mathbf{y};\dot{\boldsymbol{\beta}}\right);\ u(\boldsymbol{\beta};\dot{\boldsymbol{\beta}}) \le \ln p\left(\mathbf{y};\boldsymbol{\beta}\right), \forall \boldsymbol{\beta}; \tag{24}$$

$$\left.\frac{\partial u(\boldsymbol{\beta};\dot{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\dot{\boldsymbol{\beta}}} = \left.\frac{\partial \ln p\left(\mathbf{y};\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\dot{\boldsymbol{\beta}}}. \tag{25}$$

Inspired by the expectation maximization (EM) algorithm [31], we use the following surrogate function:

$$u(\boldsymbol{\beta};\dot{\boldsymbol{\beta}}) = \int p\left(\boldsymbol{\Theta}|\mathbf{y};\dot{\boldsymbol{\beta}}\right) \ln \frac{p\left(\boldsymbol{\Theta},\mathbf{y};\boldsymbol{\beta}\right)}{p\left(\boldsymbol{\Theta}|\mathbf{y};\dot{\boldsymbol{\beta}}\right)} d\boldsymbol{\Theta}. \tag{26}$$

It can be shown that the surrogate function in (26) satisfies (24)-(25). Then in the CAVBI-M step of the i-th iteration, $\boldsymbol{\beta}$ is updated as

$$\boldsymbol{\beta}^{i+1} = \arg\max_{\boldsymbol{\beta}} u(\boldsymbol{\beta};\boldsymbol{\beta}^i), \tag{27}$$

where $\boldsymbol{\beta}^i$ stands for the value of $\boldsymbol{\beta}$ at the i-th iteration. However, the maximization problem in (27) is non-convex and it is difficult to find its optimal solution. Therefore, we use an inexact MM algorithm, where $\boldsymbol{\beta}^{i+1}$ is obtained by applying a gradient update as follows:

$$\boldsymbol{\beta}^{i+1} = \boldsymbol{\beta}^i + \Delta^i \cdot \left.\frac{\partial u(\boldsymbol{\beta};\boldsymbol{\beta}^i)}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^i}, \tag{28}$$

where Δ^i is the stepsize determined by the Armijo rule [21]. Alternatively, a fixed stepsize can be used as mentioned in [21] to reduce the computational complexity. The detailed expression of the gradient in (28) is given in Appendix .

This approach, to simply increase and not necessarily maximize $u(\boldsymbol{\beta};\boldsymbol{\beta}^i)$ is known as the Generalized EM (GEM). Based on the convergence of the GEM in [31], we have the following theorem for the convergence of the CAVBI algorithm.

Theorem 1 (Convergence of the CAVBI Algorithm): Suppose the surrogate function $u(\boldsymbol{\beta};\boldsymbol{\beta}^i)$ satisfies (24)-(25). If at each iteration, the inexact (gradient) update as in (28) for the off-grid parameter $\boldsymbol{\beta}$ is performed, the iterates generated by the CAVBI algorithm converge to a stationary point of Problem (22).

Therefore, if the exact posterior $p(\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\beta}^i)$ for given $\boldsymbol{\beta}^i$ can be calculated, we can construct the surrogate function in (26), and the corresponding CAVBI algorithm converges to a stationary point of (22). Unfortunately, in our case, the exact posterior $p(\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\beta}^i)$ is intractable due to the loops in the factor graph. Thus, in the CAVBI-E step, we propose to find an alternative probability density function $q(\boldsymbol{\Theta};\boldsymbol{\beta}^i)$ to approximate the posterior $p(\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\beta}^i)$, which has a factorized form as

$$q\left(\boldsymbol{\Theta};\boldsymbol{\beta}^i\right) = q\left(\nu;\boldsymbol{\beta}^i\right) q\left(\mathbf{x};\boldsymbol{\beta}^i\right) q\left(\boldsymbol{\rho};\boldsymbol{\beta}^i\right) q\left(\boldsymbol{s};\boldsymbol{\beta}^i\right), \tag{29}$$

where $q(\mathbf{x};\boldsymbol{\beta}^i) \approx p(\boldsymbol{x}|\mathbf{y};\boldsymbol{\beta}^i)$, $q(\boldsymbol{\rho};\boldsymbol{\beta}^i) \approx p(\boldsymbol{\rho}|\mathbf{y};\boldsymbol{\beta}^i)$, etc. Based on the posterior approximation $q(\boldsymbol{\Theta};\boldsymbol{\beta}^i)$, the tractable surrogate function can be constructed as

$$\hat{u}(\boldsymbol{\beta};\dot{\boldsymbol{\beta}}) = \int q\left(\boldsymbol{\Theta};\boldsymbol{\beta}^i\right) \ln \frac{p\left(\boldsymbol{\Theta},\mathbf{y};\boldsymbol{\beta}\right)}{q\left(\boldsymbol{\Theta};\boldsymbol{\beta}^i\right)} d\boldsymbol{\Theta}, \tag{30}$$

which is expected to approximately satisfy (24)-(25). Therefore, after the convergence of the CAVBI algorithm with a tractable surrogate function, we can obtain an approximate stationary solution $\hat{\boldsymbol{\beta}}$ of (22) and the associated (approximate) conditional marginal posterior $p(\boldsymbol{x}|\mathbf{y};\hat{\boldsymbol{\beta}}) \approx q(\mathbf{x};\hat{\boldsymbol{\beta}})$.

2) CAVBI-E Step: In the CAVBI-E step, the off-grid parameter $\boldsymbol{\beta}$ is fixed, therefore it will be omitted in the probability expressions for simplicity. $\boldsymbol{\Theta}_l$ is used to denote the l-th variable in $\boldsymbol{\Theta}$, such as $\nu, \mathbf{x}, \boldsymbol{\rho}, \boldsymbol{s}$. Let $\mathcal{L} = \{l|\forall \boldsymbol{\Theta}_l \in \boldsymbol{\Theta}\}$. The approximate conditional marginal posteriors could be calculated by minimizing the Kullback-Leibler divergence (KLD) between $p(\boldsymbol{\Theta}|\mathbf{y})$ and $q(\boldsymbol{\Theta})$ subject to a factorized form constraint on $q(\boldsymbol{\Theta})$ as follows:

$$\begin{aligned} \mathscr{A}_{\mathrm{VBI}}: q^\star\left(\boldsymbol{\Theta}\right) &= \arg\min_{q(\boldsymbol{\Theta})} \int q(\boldsymbol{\Theta}) \ln \frac{q(\boldsymbol{\Theta})}{p\left(\boldsymbol{\Theta}|\mathbf{y}\right)} d\boldsymbol{\Theta} \\ &\text{s.t. } (29), \int q\left(\boldsymbol{\Theta}_l\right) d\boldsymbol{\Theta}_l = 1, \forall l \in \mathcal{L}. \end{aligned} \tag{31}$$

Problem $\mathscr{A}_{\mathrm{VBI}}$ is non-convex and thus we aim at finding a stationary solution (denoted by $q^*(\boldsymbol{\Theta})$), as defined below.

Definition 1 (Stationary Solution): $q^*(\boldsymbol{\Theta}) = \prod_{l\in\mathcal{L}} q^*(\boldsymbol{\Theta}_l)$ is called a stationary solution of Problem $\mathscr{A}_{\mathrm{VBI}}$ if it satisfies all the constraints in $\mathscr{A}_{\mathrm{VBI}}$ and $\forall l \in \mathcal{L}$,

$$\begin{aligned} &q^*\left(\boldsymbol{\Theta}_l\right) \\ &= \arg\min_{q(\boldsymbol{\Theta}_l)} \int \prod_{j\neq l} q^*\left(\boldsymbol{\Theta}_j\right) q\left(\boldsymbol{\Theta}_l\right) \ln \frac{\prod_{j\neq l} q^*\left(\boldsymbol{\Theta}_j\right) q\left(\boldsymbol{\Theta}_l\right)}{p\left(\boldsymbol{\Theta}|\mathbf{y}\right)} d\boldsymbol{\Theta}. \end{aligned}$$

By finding a stationary solution $q^*(\boldsymbol{\Theta})$ of $\mathscr{A}_{\mathrm{VBI}}$, the approximate posterior $q^*(\boldsymbol{\Theta}_l) \approx p(\boldsymbol{\Theta}_l|\mathbf{y}), \forall l \in \mathcal{L}$ can be obtained. A

stationary solution of $\mathscr{A}_{\rm VBI}$ can be obtained via alternately optimizing each individual $q(\mathbf{\Theta}_l), l \in \mathcal{L}$, as will be proved by Lemma 1. Specifically, for given $q(\mathbf{\Theta}_j), \forall j \neq l$, the optimal $q(\mathbf{\Theta}_l)$ that minimizes the KLD in $\mathscr{A}_{\rm VBI}$ is given by [23]

$$q\left(\mathbf{\Theta}_l\right) \propto \exp\left(\left\langle \ln p\left(\mathbf{y}, \mathbf{\Theta}\right)\right\rangle_{\prod_{j\neq l} q(\mathbf{\Theta}_j)}\right), \tag{32}$$

where $\langle f(x)\rangle_{\psi(x)} = \int f(x)\psi(x)dx$. Based on (32), the update equations of all variables are given in the subsequent subsections. Note that the operator $\langle\cdot\rangle_{\mathbf{\Theta}_l}$ is equivalent to $\langle\cdot\rangle_{q(\mathbf{\Theta}_l)}$, and $\langle f(\mathbf{\Theta}_l)\rangle$ is equivalent to $\langle f(\mathbf{\Theta}_l)\rangle_{q(\mathbf{\Theta}_l)}$.

D. Detailed Implementations of the CAVBI-E Step

1) Initialization: In order to trigger the alternating optimization (AO), the following initialization for $q(\mathbf{x}), q(\boldsymbol{\rho}), q(\boldsymbol{s})$ will be used in the first iteration.

- Initialize $q(\boldsymbol{s})$ as $q(\boldsymbol{s}) = p(\boldsymbol{s})$ using (16).
- Initialize a gamma distribution for $\boldsymbol{\rho}$: $q(\boldsymbol{\rho}) = \prod_m \Gamma(\rho_m; \tilde{a}_m, \tilde{b}_m)$, where $\tilde{a}_m = \pi_m a + (1-\pi_m)\bar{a}$, $\tilde{b}_m = \pi_m b + (1-\pi_m)\bar{b}$.
- Initialize a Gaussian distribution for $\mathbf{x}$: $q(\boldsymbol{x}) = \mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma})$, where $\boldsymbol{\mu} = \mathbf{\Sigma}\mathbf{\Phi}(\boldsymbol{\beta})^H \Sigma_n^{-1}\mathbf{y}$ and $\mathbf{\Sigma} = (\operatorname{diag}\langle\boldsymbol{\rho}\rangle + \mathbf{\Phi}(\boldsymbol{\beta})^H\Sigma_n^{-1}\mathbf{\Phi}(\boldsymbol{\beta}))^{-1}$.

2) Update for ν: For given $q(\mathbf{x})$, $q(\nu)$ can be derived as

$$q(\nu) = \Gamma(\nu; \tilde{a}_0, \tilde{b}_0), \tag{33}$$

where $\tilde{a}_0 = a_0 + ST$ and $\tilde{b}_0 = b_0 + \langle(\mathbf{y} - \mathbf{\Phi}(\boldsymbol{\beta})\mathbf{x})^H\mathbf{\Sigma}_n^{-1}(\mathbf{y} - \mathbf{\Phi}(\boldsymbol{\beta})\mathbf{x})\rangle$.

3) Update for $\mathbf{x}$: For given $q(\nu)$ and $q(\boldsymbol{\rho})$, $q(\mathbf{x})$ can be derived as

$$q(\mathbf{x}) = \mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma}), \tag{34}$$

where $\mathbf{\Sigma} = (\langle\nu\rangle\,\mathbf{\Phi}(\boldsymbol{\beta})^H\Sigma_n^{-1}\mathbf{\Phi}(\boldsymbol{\beta}) + \operatorname{diag}\langle\boldsymbol{\rho}\rangle)^{-1}$ and $\boldsymbol{\mu} = \langle\nu\rangle\,\mathbf{\Sigma}\mathbf{\Phi}(\boldsymbol{\beta})^H\Sigma_n^{-1}\mathbf{y}$.

4) Update for $\boldsymbol{\rho}$: For given $q(\boldsymbol{s})$ and $q(\mathbf{x})$, $q(\boldsymbol{\rho})$ can be derived as

$$q(\boldsymbol{\rho}) = \prod_{m=1}^{M} \Gamma\left(\rho_m, \tilde{a}_m, \tilde{b}_m\right), \tag{35}$$

where $\tilde{a}_m = \langle s_m\rangle\, a + \langle 1 - s_m\rangle\,\bar{a} + 1$ and $\tilde{b}_m = \langle s_m\rangle\, b + \langle 1 - s_m\rangle\,\bar{b} + \langle|x_m|^2\rangle$.

5) Update for $\boldsymbol{s}$: For given $q(\boldsymbol{\rho})$, $q(\boldsymbol{s})$ can be derived as

$$q(\boldsymbol{s}) = \prod_{m=1}^{M} \tilde{\pi}_m^{s_m}\left(1-\tilde{\pi}_m\right)^{1-s_m}, \tag{36}$$

where $\tilde{\pi}_m = \frac{1}{C}\pi_m\frac{b^a}{\Gamma(a)}\exp((a-1)\langle\ln\rho_m\rangle - b\langle\rho_m\rangle)$ and C is the normalization constant, given by $C = \pi_m\frac{b^a}{\Gamma(a)}\exp((a-1)\langle\ln\rho_m\rangle - b\langle\rho_m\rangle) + (1-\pi_m)\frac{\bar{b}^{\bar{a}}}{\Gamma(\bar{a})}\exp((\bar{a}-1)\langle\ln\rho_m\rangle - \bar{b}\langle\rho_m\rangle)$.

The involved expectations are given as follows for $\forall m \in \{1,\ldots,M\}$:

$$\begin{aligned}&\left\langle(\mathbf{y} - \mathbf{\Phi}(\boldsymbol{\beta})\mathbf{x})^H\,\mathbf{\Sigma}_n^{-1}\,(\mathbf{y} - \mathbf{\Phi}(\boldsymbol{\beta})\mathbf{x})\right\rangle\\ &= (\mathbf{y} - \mathbf{\Phi}(\boldsymbol{\beta})\boldsymbol{\mu})^H\,\mathbf{\Sigma}_n^{-1}\,(\mathbf{y} - \mathbf{\Phi}(\boldsymbol{\beta})\boldsymbol{\mu})\\ &\quad + \operatorname{Tr}\left(\mathbf{\Phi}(\boldsymbol{\beta})^H\mathbf{\Sigma}_n^{-1}\mathbf{\Phi}(\boldsymbol{\beta})\mathbf{\Sigma}\right),\end{aligned} \tag{37}$$

Algorithm 1: CAVBI Algorithm for CCE.

1: **Input:** $\mathbf{y}, \mathbf{W}, \mathbf{A}(\mathbf{0})$, CSSI $\mathbb{I} = \{\hat{\Omega}, D, D_c\}$, hyperparameters: $a = b = 1, \bar{a} = 1, \bar{b} = 0.001, a_0 = b_0 = 0.001$.
2: **Output:** Channel estimate: $\hat{\mathbf{h}}$
3: Set $\boldsymbol{\beta} = \mathbf{0}$; Initialize the distribution functions $q(\mathbf{x})$, $q(\boldsymbol{\rho})$ and $q(\boldsymbol{s})$, and the related expectations.
4: **while** not converge **do**
5: **% CAVBI-E Step:**
6: Update $q(\nu)$ using (33) and the expectations in (38).
7: Update $q(\mathbf{x})$ using (34) and the expectations in (37) and (40).
8: Update $q(\boldsymbol{\rho})$ using (35) and the expectations in (39).
9: Update $q(\boldsymbol{s})$ using (36) and the expectations in (38).
10: **% CAVBI-M Step:**
11: Construct the surrogate function in (30) using the approximate posterior output of CAVBI-E Step, i.e., $q(\mathbf{\Theta})$, and update $\boldsymbol{\beta}$ using (28).
12: **end while**
13: $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$, $p(\boldsymbol{x}|\mathbf{y}; \hat{\boldsymbol{\beta}}) \approx q(\mathbf{x}; \hat{\boldsymbol{\beta}})$, then we have $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}; \hat{\boldsymbol{\beta}}) = \boldsymbol{\mu}$, $\hat{\mathbf{h}} = \mathbf{A}(\hat{\boldsymbol{\beta}})\hat{\mathbf{x}}$.

$$\langle\nu\rangle = \frac{\tilde{a}_0}{\tilde{b}_0}, \langle s_m\rangle = \tilde{\pi}_m, \langle 1 - s_m\rangle = 1 - \tilde{\pi}_m, \tag{38}$$

$$\langle\rho_m\rangle = \frac{\tilde{a}_m}{\tilde{b}_m}, \langle\ln\rho_m\rangle = \psi\left(\tilde{a}_m\right) - \ln\left(\tilde{b}_m\right), \tag{39}$$

$$\langle\boldsymbol{x}\rangle = \boldsymbol{\mu}, \left\langle|x_m|^2\right\rangle = |\mu_m|^2 + \chi_{m,m}, \tag{40}$$

where $\psi(x) = \frac{d}{dx}\ln(\Gamma(x))$ is the digamma function, μ_m is the m-th element of $\boldsymbol{\mu}$, and $\chi_{m,m}$ is the m-th diagonal element of $\mathbf{\Sigma}$.

6) Convergence of the CAVBI-E Step: We have the following convergence lemma for the CAVBI-E step.

Lemma 1 (Convergence of the CAVBI-E Step): Every limiting point generated by the CAVBI-E Step using (33), (34), (35), and (36) with the initialization in Section IV-D1 is a stationary solution of Problem $\mathscr{A}_{\rm VBI}$.

The proof can be found in Appendix B.

Finally, the overall CAVBI algorithm is summarized in Algorithm 1.

V. CSSI-Assisted Configuration Optimization

In this section, we formulate the configuration optimization problem and develop closed form MSE in terms of configuration parameters of ASPN and apply optimization techniques (SCA to be specific) to come up with an efficient algorithm to dynamically configure the ASPN based on CSSI, which in conjunction with the proposed channel estimation technique yields improved performance with less hardwares.

A. Configuration Optimization Formulation

According to the posterior approximation of channel vector $\mathbf{x}$, i.e., $q(\mathbf{x}) = \mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma})$, given by (34) in the CAVBI

algorithm, the MSE of MAP estimation can be calculated as

$$\mathrm{MSE}\left(\mathbf{W}\right) = \frac{1}{M}\mathbb{E}\left\|\mathbf{x}-\hat{\mathbf{x}}\right\|^2 = \frac{1}{M}\mathrm{Tr}\left(\mathbf{\Sigma}\right). \tag{41}$$

Recall that in the CAVBI algorithm, the posterior approximation is updated iteratively until the convergence to the stationary point, and the final MSE of CAVBI estimation depends on the stationary solution $q^*(\mathbf{x})$, whose closed form is hard to obtain. Alternatively, the initial MSE is chosen as the optimization metric to seek the optimal configuration of $\mathbf{W}$.

Based on the initial posterior of $\mathbf{x}$ in Section IV-D1, we can get the initial MSE of $\mathbf{x}$ as

$$\varphi\left(\mathbf{W}\right) = \frac{1}{M}\mathrm{Tr}\left(\mathbf{A}^H\mathbf{W}\mathbf{\Sigma}_n^{-1}\mathbf{W}^H\mathbf{A}+\mathbf{R}^{-1}\right)^{-1}, \tag{42}$$

where $\mathbf{A} \triangleq \mathbf{A}(\mathbf{0})$. $\mathbf{\Sigma}_n$ is given by (3), which is also a function of $\mathbf{W}$. $\mathbf{R}$ is given by $\mathbf{R} = \mathrm{diag}\left\langle\boldsymbol{\rho}\right\rangle^{-1}$, where the m-th diagonal element is $\left\langle\rho_m\right\rangle^{-1} = \tilde{b}_m/\tilde{a}_m$. Note that $\mathbf{R}$ can be used to incorporate the CSSI, as explained below. If the CSSI quality is high, e.g., $\pi_m = 1, \forall m \in \hat{\Omega}; \pi_m = 0, \forall m \notin \hat{\Omega}$, the m-th diagonal element of $\mathbf{R}$ is $\Theta(1)$ for $m \in \hat{\Omega}$ and close to 0 for $m \notin \hat{\Omega}$. If the CSSI quality is very poor, e.g., $\pi_m = \frac{D}{M}, \forall m$, the diagonal elements of $\mathbf{R}$ are all equal.

Remark 2: Even though the ASPN architecture proposed in Section III enables an arbitrary number of PSs, the configuration optimization metric $\varphi(\mathbf{W})$ in (42) requires that $N \geq S$, so that the noise covariance matrix $\mathbf{\Sigma}_n$ is full rank and invertible. Therefore, we consider $N \geq S$ for the ASPN configuration design in the following sections.

The optimal $\mathbf{W}$ is given by solving the following optimization problem:

$$\mathscr{A}: \min_{\mathbf{W}} \ \varphi\left(\mathbf{W}\right) \tag{43}$$

$$\text{s.t. } \mathbf{W} = \left[\mathbf{W}_1, \ldots, \mathbf{W}_T\right], \tag{44}$$

$$\mathbf{W}_t = \mathbf{G}_t\mathbf{P}_t\mathbf{F}_t, \forall t, \tag{45}$$

$$\mathbf{P}_t = \mathrm{diag}\left(\left[e^{j\phi_{t,1}}, \ldots, e^{j\phi_{t,N}}\right]\right), \phi_{t,n} \in \mathcal{A}, \forall t, n, \tag{46}$$

$$\mathbf{G}_t \in \{0,1\}^{M\times N}, \mathbf{F}_t \in \{0,1\}^{N\times S}, \forall t. \tag{47}$$

Due to the element-wise discrete constraints on the matrices $\mathbf{G}_t, \mathbf{P}_t, \mathbf{F}_t$, problem $\mathscr{A}$ is non-convex, which is hard to solve. To make the problem tractable, we first ignore the discrete constraints and treat the variables in $\mathscr{A}$ as continuous variables. After finding the optimal $\mathbf{W}^*$, the resulting solution will be projected onto the feasible discrete set.

According to (44), (45) and (46), $\mathbf{W}$ can be represented by the vector $\mathbf{w} \in \mathbb{R}^{(MN+N+NS)T}$, given by

$$\mathbf{w} = [\underbrace{\mathbf{g}_1}_{\mathbf{w}_1}; \cdots; \underbrace{\mathbf{g}_T}_{\mathbf{w}_T}; \underbrace{\mathbf{f}_1}_{\mathbf{w}_{T+1}}; \cdots; \underbrace{\mathbf{f}_T}_{\mathbf{w}_{2T}}; \underbrace{\boldsymbol{\phi}_1}_{\mathbf{w}_{2T+1}}; \cdots; \underbrace{\boldsymbol{\phi}_T}_{\mathbf{w}_{3T}}], \tag{48}$$

where $\mathbf{g}_t = \mathrm{vec}(\mathbf{G}_t) \in \{0,1\}^{MN}$, $\mathbf{f}_t = \mathrm{vec}(\mathbf{F}_t) \in \{0,1\}^{NS}$ and $\boldsymbol{\phi}_t = [\phi_{t,1}; \cdots; \phi_{t,N}] \in \mathcal{A}^N$. Then the original problem $\mathscr{A}$ can be relaxed as

$$\tilde{\mathscr{A}}: \quad \min_{\mathbf{w}\in\mathcal{W}} \varphi(\mathbf{w}), \tag{49}$$

Algorithm 2: CAPSCA Algorithm for Analog BF Matrix Design.

1: **Input:** $\mathbf{A}, \mathbf{R}, \boldsymbol{\tau} \triangleq (\tau_i)_{i=1}^{3T} > 0, \{\gamma^k\} > 0, \mathbf{w}^0 \in \mathcal{W}$; set $k = 0$.
2: If $\mathbf{w}^k$ satisfies a suitable termination criterion: STOP.
3: **% Parallel Updating**
4: Compute $\hat{\mathbf{w}}_i, \forall i \in \{1, \ldots, 3T\}$ using (56).
5: Update $\mathbf{w}_i^{k+1}, \forall i \in \{1, \ldots, 3T\}$ according to (57).
6: $k \leftarrow k+1$, and go to Line 1.
7: $\mathbf{w}^* = \mathbf{w}^k$, project $\mathbf{w}^*$ back to the discrete feasible set using (51)-(53) to obtain the final solution $\hat{\mathbf{w}}$.

where $\mathcal{W}$ is a convex set, given by

$$\mathcal{W} = \left\{\mathbf{w} : \mathbf{g}_t \in [0,1]^{MN}; \mathbf{f}_t \in [0,1]^{NS}; \boldsymbol{\phi}_t \in \mathbb{R}^N, \forall t\right\}. \tag{50}$$

Therefore, $\mathcal{W}$ has a decoupled form: $\mathcal{W} = \{\mathbf{w} : \mathbf{w}_i \in \mathcal{W}_i, i = 1, \ldots, 3T\}$, where $\mathcal{W}_i$ is a convex set.

After finding the optimized $\mathbf{w}^* = [\mathbf{g}_1^*, \ldots, \boldsymbol{\phi}_T^*]$ for problem $\tilde{\mathscr{A}}$, we project it onto the discrete feasible set to obtain the final $\mathbf{w}$ as follows:

$$\hat{\mathbf{g}}_t = \arg\min_{\mathbf{g}_t} \left\|\mathbf{g}_t - \mathbf{g}_t^*\right\|, \text{s.t. } \mathbf{g}_t \in \{0,1\}^{MN}, \forall t, \tag{51}$$

$$\hat{\mathbf{f}}_t = \arg\min_{\mathbf{f}_t} \left\|\mathbf{f}_t - \mathbf{f}_t^*\right\|, \text{s.t. } \mathbf{f}_t \in \{0,1\}^{NS}, \forall t, \tag{52}$$

$$\hat{\phi}_{t,n} = \arg\min_{\phi\in\mathcal{A}} \left|\left(\phi_{t,n}^* \mod 2\pi\right) - \phi\right|, \forall n, t. \tag{53}$$

B. CSSI-Assisted Parallel Successive Convex Approximation

The availability of high performance multi-core computing platforms makes it increasingly desirable to develop parallel optimization methods. Therefore, we aim at developing parallel solutions for problem $\tilde{\mathscr{A}}$, whereby operations can be carried out on all block variables $\mathbf{w}_i$ at the same time.

Algorithm 2 summarizes the key steps of the proposed CAPSCA algorithm.[5] Specifically, given $\mathbf{w}^k$ at iteration k, the block variables $\{\mathbf{w}_i\}_{i=1}^{3T}$ can be updated in parallel by solving the following subproblem for each block variable $\mathbf{w}_i, i \in \{1, \ldots, 3T\}$:

$$\hat{\mathbf{w}}_i = \arg\min_{\mathbf{w}_i\in\mathcal{W}_i} \hat{\varphi}\left(\mathbf{w}_i; \mathbf{w}^k\right), \tag{54}$$

where $\hat{\varphi}(\mathbf{w}_i; \mathbf{w}^k)$ is the convex approximation of $\varphi(\cdot)$ at the point $\mathbf{w}^k$ w.r.t. the block variable $\mathbf{w}_i$ that preserves the first order properties of $\varphi(\cdot)$ w.r.t. $\mathbf{w}_i$. We use the quadratic approximation of $\varphi(\cdot)$, given by

$$\hat{\varphi}\left(\mathbf{w}_i; \mathbf{w}^k\right) \triangleq \varphi\left(\mathbf{w}^k\right) + \left(\nabla_{\mathbf{w}_i}\varphi\left(\mathbf{w}^k\right)\right)^T\left(\mathbf{w}_i - \mathbf{w}_i^k\right) + \tau_i\left\|\mathbf{w}_i - \mathbf{w}_i^k\right\|^2, \tag{55}$$

[5]The parallel computation feature refers to the algorithm to determine the dynamic configuration of the ASPN. In 5G systems, there is Cloud-Radio Access Network (C-RAN) [32] which provides parallel computation clusters. As such, the proposed configuration algorithm can leverage the parallel computation clusters in such C-RAN platform.

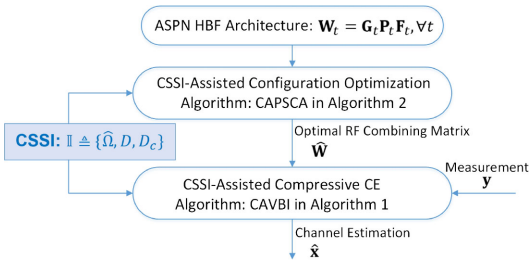

Fig. 5. Overall flow of CSSI-assisted configuration optimization and CCE algorithm.

where $\tau_i > 0$ is a constant. Since the subproblems corresponding to different block variables are strictly convex quadratic optimization problems, the solution for (54) is unique and has a closed form:

$$\hat{\mathbf{w}}_i = \mathbb{P}_{\mathcal{W}_i}\left[-\frac{1}{2\tau_i}\nabla_{\mathbf{w}_i}\varphi\left(\mathbf{w}^k\right) + \mathbf{w}_i^k\right], \tag{56}$$

where $\mathbb{P}_{\mathcal{W}_i}[\cdot]$ denotes the projection onto the convex region $\mathcal{W}_i$. Given $\hat{\mathbf{w}}_i$, $\mathbf{w}_i$ is updated according to

$$\mathbf{w}_i^{k+1} = \mathbf{w}_i^k + \gamma^k\left(\hat{\mathbf{w}}_i - \mathbf{w}_i^k\right), \tag{57}$$

where γ^k is a sequence to be properly chosen. Then the above iteration is carried out until convergence.

Based on the convergence proof of the parallel SCA algorithm in [33, Theorem 1], we have the following similar convergence statement for the proposed CAPSCA algorithm.

Theorem 2 (Convergence of CAPSCA Algorithm): Let $\{\mathbf{w}^k\}$ be the sequence generated by Algorithm 2. Suppose that $\{\gamma^k\}$ satisfies the following conditions: i) $\gamma^k \in (0, 1]$; ii) $\gamma^k \to 0$; iii) $\sum_k \gamma^k = +\infty$; iv) $\sum_k (\gamma^k)^2 < +\infty$. Then every limit point of $\{\mathbf{w}^k\}$ is a stationary solution of problem $\tilde{\mathscr{A}}$.

C. Overall CE Algorithm and Complexity Analysis

The overall flow of CSSI-assisted configuration optimization and CCE algorithm is summarized in Fig. 5. The computational complexity of the proposed algorithm is analyzed as follows. We focus on the number of complicated operations, such as multiplications and divisions[6] to show the main computational burden of the proposed algorithm.

1) Complexity of the CAPSCA Algorithm: The computational complexities of updating $\mathbf{g}_t$, $\mathbf{f}_t$ and $\boldsymbol{\phi}_t$ are the same, which are $\mathcal{O}(M^3 + MSN)$. Due to the parallel updating, the total computational requirement of CAPSCA algorithm is $\mathcal{O}(M^3 + MSN)$ per iteration.

2) Complexity of the CAVBI Algorithm: The computational complexity of the CAVBI algorithm in each iteration is given as follows.

- The complexity in updating $q(\nu)$ is $\mathcal{O}(M^2ST)$.
- The complexity in calculating $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ is $\mathcal{O}(M^2ST)$ and $\mathcal{O}(M^2)$, respectively.

[6]There are other mathematical operations involved in the updating of $q(\boldsymbol{s})$, such as exponential/logarithm/digamma functions, and the number of these operations is $\mathcal{O}(M)$, which is insignificant compared to the number of multiplications and divisions (which is a cubic order).

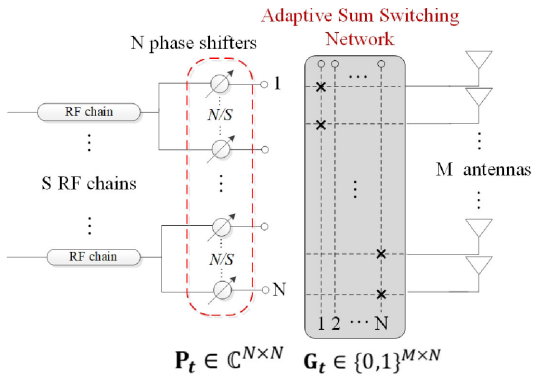

Fig. 6. Symmetric adaptive HBF architecture with fixed loadings.

- The complexity in updating $q(\boldsymbol{\rho})$ and $q(\boldsymbol{s})$ is $\mathcal{O}(M)$.
- The complexity in updating $\boldsymbol{\beta}$ is $\mathcal{O}(M^2ST)$ if the fixed stepsize is used.

Therefore, the total computational requirement of the CAVBI algorithm is $\mathcal{O}(M^2ST)$ per iteration.

D. Symmetric Adaptive HBF Architecture Design

The ASPN architecture in Section III is general and enables asymmetric signal flow in the architecture, as shown in Fig. 3, where the loading for each hardware component (i.e., RF chain, PS and antenna) can be dynamically changing based on the configuration optimization result. In such case, there could be a scenario where one RF chain is connected to one antenna only, and another RF chain is connected to all the antennas. To simplify the design of RF matching, we consider the following symmetric adaptive HBF (SA-HBF) design.

As illustrated in Fig. 6, we assume N is divisible by S, and the connection between RF chains and PSs is fixed, where each RF chain is connected to N/S PSs and each PS is connected to one RF chain. The phase shifting values $\mathbf{P}_t \in \mathbb{C}^{N\times N}$ and the sum switching network $\mathbf{G}_t \in \{0,1\}^{M\times N}$ between PSs and antennas are adaptively designed under the fixed loading constraint[7] by solving the following optimization problem:

$$\mathscr{A}^s: \min_{\mathbf{W}} \ \varphi(\mathbf{W}) \tag{58}$$

$$\text{s.t.} \ \ \mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_T], \tag{59}$$

$$\mathbf{W}_t = \mathbf{G}_t\mathbf{P}_t\mathbf{F}_t, \mathbf{G}_t \in \{0,1\}^{M\times N}, \forall t \tag{60}$$

$$\sum_{n=1}^{N} g_{t,m,n} = \iota, \sum_{m=1}^{M} g_{t,m,n} = \varrho, \forall m, n, t, \tag{61}$$

$$\mathbf{P}_t = \text{diag}\left(\left[e^{j\phi_{t,1}}, \ldots, e^{j\phi_{t,N}}\right]\right), \phi_{t,n} \in \mathcal{A}, \forall t, n, \tag{62}$$

where $\mathbf{F}_t = \text{Diag}(\mathbf{1}_{N/S\times 1}, \ldots, \mathbf{1}_{N/S\times 1}) \in \{0,1\}^{N\times S}$ is a block diagonal matrix with S all-one vectors on the diagonal, $1 \le \iota \le N$ is a constant integer representing that each antenna is

[7]Such a fixed loading constraint guarantees an RF chain is always connected to a fixed number of N/S PSs. This substantially simplifies the impedance of RF matching.

connected to ι PSs, $1 \leq \varrho \leq M$ is a constant integer representing that each PS is connected to ϱ antennas.

Similar to the method adopted to solve problem $\mathscr{A}$, $\mathscr{A}^s$ is first relaxed as

$$\tilde{\mathscr{A}}^s: \min_{\mathbf{w}^s \in \mathcal{W}^s} \varphi(\mathbf{w}^s), \tag{63}$$

where

$$\mathbf{w}^s = [\underbrace{\mathbf{g}_1}_{\mathbf{w}_1^s}; \cdots; \underbrace{\mathbf{g}_T}_{\mathbf{w}_T^s}; \underbrace{\boldsymbol{\phi}_1}_{\mathbf{w}_{T+1}^s}; \cdots; \underbrace{\boldsymbol{\phi}_T}_{\mathbf{w}_{2T}^s}],$$

$$\mathcal{W}^s = \left\{\mathbf{w}^s : \mathbf{g}_t \in [0,1]^{MN}, \mathbf{L}\mathbf{g}_t = \iota \mathbf{1}_{M\times 1}, \mathbf{D}\mathbf{g}_t = \varrho \mathbf{1}_{N\times 1}, \forall t\right\},$$

$\mathbf{L} = [\mathbf{I}_{M\times M}, \ldots, \mathbf{I}_{M\times M}] \in \{0,1\}^{M\times MN}$ and $\mathbf{D} = \text{Diag}(\mathbf{1}_{1\times M}, \ldots, \mathbf{1}_{1\times M}) \in \{0,1\}^{N\times MN}$. Then the CAPSCA in Algorithm 2 is adopted to solve problem $\tilde{\mathscr{A}}^s$. After this, the optimized $(\mathbf{w}^s)^*$ can be obtained. Then we project it onto the discrete feasible set to obtain the final $\mathbf{w}^s$ which satisfies the discrete constraint in (60)-(62).

VI. Simulation Results

In this section, we verify the importance and the effectiveness of the flexibility introduced in ASPN architecture, the compressive CE algorithm as well as the configuration algorithm under the realistic mmWave channel model, i.e., mm-SSCM, which is proposed in [26] for high frequency bands (28-73 GHz). The following baselines are considered.

- **Baseline 1** (VBI [23]): This is the standard VBI algorithm with the randomly generated RF training vectors.
- **Baseline 2** (CA-Weighted LASSO [18]): The CSSI-assisted weighted LASSO algorithm in [18] can optimally tune the regularization parameters according to the CSSI accuracy. The RF training vectors are randomly generated.
- **Baseline 3** (CAVBI-Fixed): The CAVBI algorithm proposed in Section IV is used to estimate mmWave channel. A fixed HBF architecture is considered and the RF training vectors are randomly generated.
- **Baseline 4** (Fully-Digital LS): Consider a traditional fully-digital system where $S = M$, the user sends one pilot for uplink training. The baseband received signal is $\mathbf{y} = \mathbf{h} + \tilde{\mathbf{n}}$. Then LS is used to estimate $\mathbf{h}$ from $\mathbf{y}$.

Consider the BS is equipped with $M = 128$ antennas and $S = 8$ RF chains, and $S \leq N \leq M$. We focus on the simulations for the ULA with $\lambda/2$ inter antenna spacing. In the following simulation results, CAVBI-PSCA and CAVBI-Symmetric stands for the proposed scheme with general ASPN architecture and SA-HBF architecture, respectively. For the SA-HBF design, we consider $\iota = 1$ and $\varrho = M/N$.[8] For Baseline 1-3, we consider a fixed symmetric HBF architecture.[9] Note that the simulation results of Baseline 2 for $N < 64$ are not shown due to its bad performance under non-i.i.d. measurement matrix. 1-bit resolution PSs are considered. The normalized MSE (NMSE) is used as the performance metric, defined as $\text{NMSE} = \frac{1}{K}\sum_{k=1}^{K} \frac{\|\mathbf{h}_k - \hat{\mathbf{h}}_k\|^2}{\|\mathbf{h}_k\|^2}$, where $\hat{\mathbf{h}}_k$ is the estimate of $\mathbf{h}_k$ at the k-th frame, and $K = 200$ in all schemes.

[8]For Baseline 1-3 and CAVBI-Symmetric schemes, we consider the case when N and M is divisible by S and N, respectively.

[9]which means that each RF chain is connected to N/S PSs and each PS is connected to M/N antennas uniformly without overlapping.

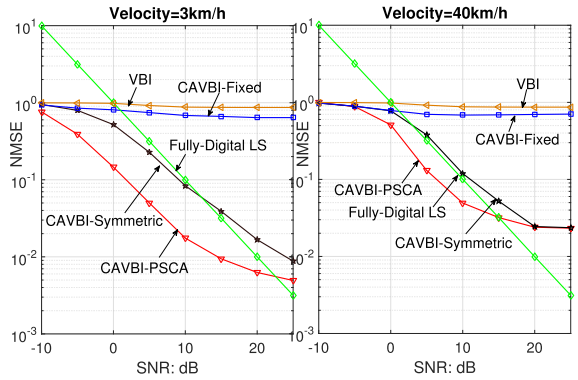

Fig. 7. NMSE vs SNR for different user velocities with $N = 32, T = 4$.

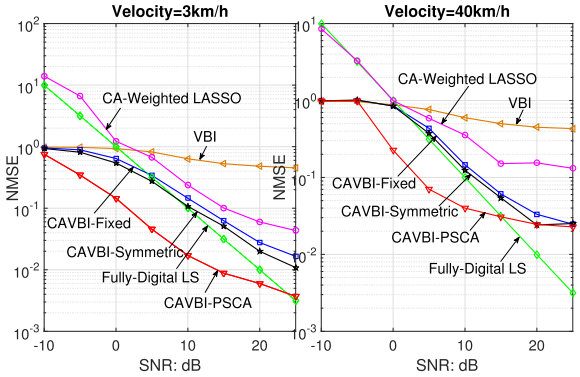

Fig. 8. NMSE versus SNR for different user velocities with $N = 128, T = 4$.

A. Impact of SNR

In Fig. 7 and Fig. 8, we compare the NMSE performance of various schemes versus SNR under different user velocities. Note that the velocity parameter will affect the quality of the CSSI provided by the previous channel estimates. When the velocity is relatively low, the previous estimated channels can provide more information about the current channel to be estimated, as such, higher quality CSSI can be obtained. The proposed algorithm, Baseline 2 and 3 can exploit the higher quality CSSI to improve the CE performance. However, the Fully-Digital LS estimates channels independently over time, without exploiting the CSSI. Therefore, the velocity parameter should not affect the performance of Fully-Digital LS algorithm.

Compared to standard VBI, the CAVBI achieves a better CE performance especially for a large number of PSs due to the exploitation of CSSI. Compared to the CA-Weighted LASSO, the CAVBI works well even for a small number of PSs and low SNR due to its robustness to the measurement matrix and noise. Compared to the fixed HBF architecture, the SA-HBF architecture achieves a significant performance gain when the number of PSs is small, and the performance gap decreases

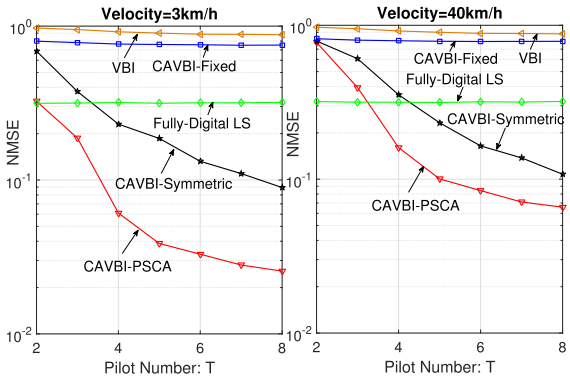

Fig. 9. NMSE versus pilot number T for different user velocities with $N = 32$, SNR $= 5$ dB.

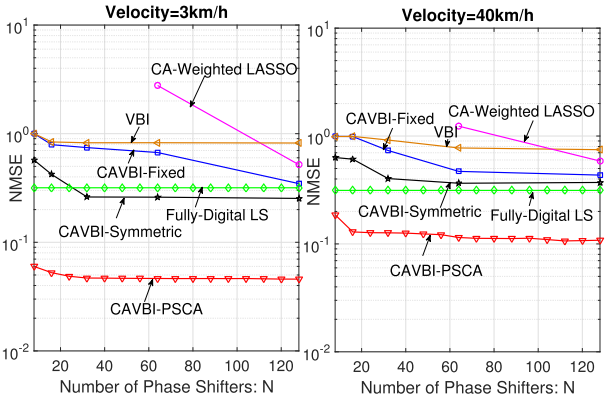

Fig. 10. NMSE versus the number of PSs N for different user velocities with $T = 4$, SNR $= 5$ dB.

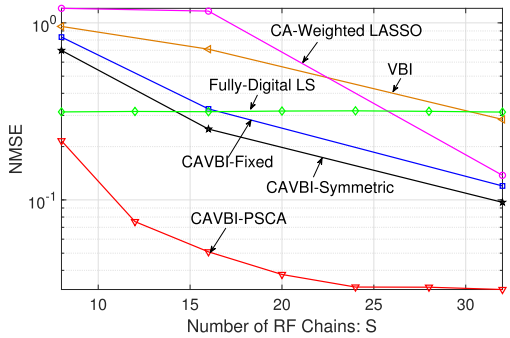

Fig. 11. NMSE versus the number of RF chains S with $N = 64$, $T = 2$, SNR $= 5$ dB, user velocity is 3 km/h.

when the number of PSs increases. This attributes to the adaptive design of the sum switching network. Compared to the SA-HBF architecture, the proposed general ASPN architecture achieves a big performance gain for various number of PSs. Therefore, optimizing the dynamic HBF architecture based on the CE performance metric can indeed enhance the final channel recovery performance, and more design feasibility can lead to larger performance gain. Even though the SA-HBF architecture could lead to performance degradation compared to the general ASPN architecture when we simplify the hardware implementation, SA-HBF still provides a flexible architecture to balance the achievable performance and implementation complexity, especially when PS resource is very limited.

Compared to the fully-digital system, the proposed CAVBI-PSCA performs better in a low SNR regime. For example, when the SNR is less than 20 dB for 3 km/h user velocity (15 dB for 40 km/h user velocity), the CAVBI-PSCA in the HBF system performs better than the LS in the fully-digital system, with only $T = 4$ pilots and $S = 8$ RF chains. Therefore, joint configuration optimization and CCE with CSSI can greatly improve the CE performance in HBF mmWave systems.

B. Impact of Pilot Number

In Fig. 9, we compare the NMSE performance of various schemes versus pilot number T under different user velocities. It shows that with dynamic HBF architecture, the proposed CAVBI-PSCA and CAVBI-Symmetric achieve sufficient performance gain compared to the CS-based CE baselines with static HBF architecture. By optimally configuring the flexibility of the ASPN architecture, the CAVBI-PSCA can achieve better performance compared to the SA-HBF design, which can only provide limited flexibility for the analog BF. Moreover, the proposed CAVBI-PSCA can outperform the full-digital system with $T > 2$ for 3 km/h user velocity and $T > 3$ for 40 km/h user velocity.

C. Impact of the Number of Phase Shifters

In Fig. 10, we compare the NMSE performance of various schemes versus the number of PSs N under different user velocities. It shows that as long as $N \geq S = 8$, CAVBI-PSCA can outperform all CE baselines for any number of PSs, and $N = 24$ PSs are sufficient for the CAVBI-PSCA algorithm as when $N \geq 24$, the NMSE performance saturates. Contrarily, $N = 32$ PSs are sufficient for the CAVBI-Symmetric algorithm to achieve a steady performance. CA-Weighted LASSO and CAVBI-Fixed requires $N = M = 128$ PSs to achieve satisfactory performance. And Fully-Digital LS requires $S = 128$ RF chains to be able to work. This illustrates that the proposed CAVBI-PSCA is much more hardware-efficient than existing CE algorithms in HBF systems, with the best CE performance under the same hardware constraint.

D. Impact of the Number of RF Chains

In Fig. 11, we compare the NMSE performance versus the number of RF chains S. It shows that the proposed CAVBI-PSCA algorithm outperforms all the baselines under the same hardware constraints. Compared to the Fully-Digital LS, the proposed algorithm can achieve better performance with much less number of RF chains.

E. Impact of Different Antenna Configurations

In Fig. 12, we compare the NMSE performance of various schemes for uniform circular array (UCA). It shows that the proposed algorithm achieves the best CE performance compared to the CS-based baseline algorithms. Compared to fully-digital

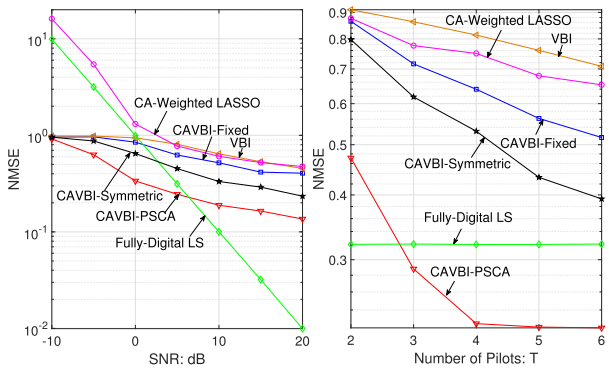

Fig. 12. NMSE performance comparison for UCA with $S = 8$, $N = 64$, user velocity is 3 km/h. Left: NMSE versus SNR with $T = 4$; Right: NMSE versus pilot number T with SNR $= 5$ dB.

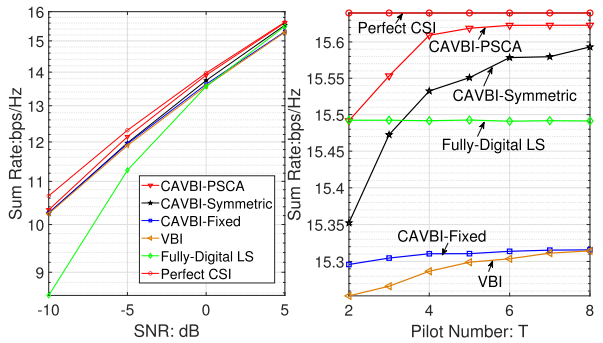

Fig. 13. Sum rate performance comparison with ULA, user velocity is 3 km/h, $S = 8$, $N = 32$. Left: Sum rate versus SNR with $T = 4$; Right: Sum rate versus pilot number T with SNR $= 5$ dB.

LS baseline, the proposed algorithm can achieve better CE performance when SNR is low (e.g., less than 5 dB) or more pilots are available (e.g., larger than 3). It also shows that the proposed scheme can be applied to different antenna array configurations.

F. Comparison of the Sum Rate Performance

In Fig. 13, the sum rates of the different schemes are compared. For comparison, we also plot the sum rate when perfect CSI is available at the BS. The BS transmits single data stream to the user using the fairness-aware greedy analog precoding algorithm in [34] together with the singular vector digital BF. The analog and digital precoding matrices are calculated from the estimated channels. It can be observed that the proposed algorithm achieves the highest sum rate among all baselines. When the SNR becomes higher and number of pilots becomes larger, the performance gap between proposed scheme and perfect CSI becomes smaller.

VII. Conclusion

In this paper, a joint configuration optimization and CCE algorithm in HBF system with hardware constraints is proposed. Specifically, we first propose a novel and flexible ASPN architecture subject to a limited number of PSs. The proposed ASPN analog BF can be applied to any number of RF PSs. Moreover, the ASPN allows more than one PSs to connect one RF chain with the antenna, which can realize more flexible analog BF matrix with non-constant modulus elements. Then, we propose a CSSI-assisted CCE algorithm based on the ASPN architecture, which can fully exploit the CSSI under a general uncertain measurement matrix. After this, we propose a dynamic configuration algorithm for the novel ASPN architecture. The configuration parameters are chosen to minimize the MSE of the channel estimation. The proposed configuration algorithm has a highly parallel structure and can be implemented using a multi-core platform to reduce the computational complexity. Finally, the performance advantage of the proposed overall CE algorithm is verified using a practical mmWave channel model.

Appendix

A. Gradient Update for Off-Grid Parameters

The derivative $\boldsymbol{\xi}_\beta = \frac{\partial u(\boldsymbol{\beta};\boldsymbol{\beta}^i)}{\partial \boldsymbol{\beta}}$ can be calculated as $\boldsymbol{\xi}_\beta = [\xi(\beta_1), \ldots, \xi(\beta_M)]^T$, with $\xi(\beta_m) = -2c_0 \mathrm{Re}(\mathbf{a}_m'^H \mathbf{W} \boldsymbol{\Sigma}_n^{-1} (\mathbf{W}^H \mathbf{a}_m c_1 + \mathbf{c}_2))$, where $c_0 = \tilde{a}_0/\tilde{b}_0$, $\mathbf{a}_m = \mathbf{a}(\hat{\vartheta}_m + \beta_m)$, $\mathbf{a}_m' = d\mathbf{a}(\hat{\vartheta}_m + \beta_m)/d\beta_m$, $c_1 = |\mu_m|^2 + \chi_{m,m}$, $\mathbf{c}_2 = \mathbf{W}^H \sum_{l \neq m} \chi_{l,m} \mathbf{a}_l - \mu_m^* \mathbf{y}_{-m}$, $\mathbf{y}_{-m} = \mathbf{y} - \mathbf{W}^H \sum_{l \neq m} \mu_l \mathbf{a}(\hat{\vartheta}_l + \beta_l)$, μ_m and $\chi_{l,m}$ denote the m-th and (l,m)-th element of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

B. Proof of Lemma 1

In our problem, the factorized form of $q(\boldsymbol{\Theta})$ in (29) can be considered as parameterized functions, e.g., a gamma distribution function with parameters $\tilde{\alpha}_m$ and $\tilde{\beta}_m$ for $q(\boldsymbol{\rho})$, a Gaussian distribution function with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for $q(\mathbf{x})$, etc. As a result, the optimization problem $\mathscr{A}_{\mathrm{VBI}}$ can be converted into a conventional parameter optimization problem. The sequence of the objective function in (31) generated by the alternating updating is non-increasing. Together with the fact that the objective function (31) has a lower bound, the sequence converges to a limit. Let the surrogate function be chosen as the objective function itself. Since for any variable in problem $\mathscr{A}_{\mathrm{VBI}}$ we have a unique solution, given by (32), according to Theorem 5 in [21], the CAVBI-E step converges to a stationary solution of Problem $\mathscr{A}_{\mathrm{VBI}}$.

References

[1] S. Mumtaz, J. Rodriguez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*. New York, NY, USA: Academic Press, 2016.
[2] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[3] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[4] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
[5] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015, pp. 2909–2913.

[6] L. Lian, A. Liu, and V. K. Lau, "Optimal-tuned weighted LASSO for massive MIMO channel estimation with limited RF chains," in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–6.
[7] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. IEEE Int. Symp. Inf. Theory*, 2016, pp. 860–864.
[8] A. Liu, V. Lau, and M. L. Honig, "Compressive RF training for massive MIMO with channel support side information," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3628–3641, Jul. 2019.
[9] J. Yang, C.-K. Wen, S. Jin, and F. Gao, "Beamspace channel estimation in mmWave systems via cosparse image reconstruction technique," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4767–4782, Oct. 2018.
[10] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?," in *Proc. Inf. Theory Appl. Workshop*, 2015, pp. 90–97.
[11] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog–digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, 2016.
[12] X. Yu, J. Zhang, and K. B. Letaief, "A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 282–297, May 2018.
[13] X. Zhu, Z. Wang, L. Dai, and Q. Wang, "Adaptive hybrid precoding for multiuser massive MIMO," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 776–779, Apr. 2016.
[14] Y. Teng, M. Wei, A. Liu, V. Lau, and Y. Zhang, "Mixed-timescale per-group hybrid precoding for multiuser massive MIMO systems," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 675–679, May 2018.
[15] K. Huang, R. W. Heath Jr, and J. G. Andrews, "Limited feedback beamforming over temporally-correlated channels," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1959–1975, May 2009.
[16] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.
[17] X. Rao and V. K. Lau, "Compressive sensing with prior support quality information and application to massive mimo channel estimation with temporal correlation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4914–4924, Sep. 2015.
[18] L. Lian, A. Liu, and V. K. Lau, "Weighted LASSO for sparse recovery with statistical prior support information," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1607–1618, Mar. 2018.
[19] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. 44th Annu. Conf. Inf. Sci. Syst.*, Mar. 2010, pp. 1–6.
[20] L. Lian, A. Liu, and V. K. Lau, "Exploiting dynamic sparsity for downlink FDD-massive MIMO channel tracking," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2007–2021, Apr. 2019.
[21] J. Dai, A. Liu, and V. K. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Trans. Signal Process.*, vol. 66, no. 10, 2018.
[22] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
[23] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[24] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.
[25] H. Zhang, S. Venkateswaran, and U. Madhow, "Channel modeling and MIMO capacity for outdoor millimeter wave links," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2010, pp. 1–6.
[26] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Tech.*, vol. 64, no. 7, pp. 2207–2225, Jul. 2016.
[27] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10–22, Feb. 1998.
[28] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, Sep. 2010.
[29] R. L. Schmid, P. Song, C. T. Coen, A. Ç. Ulusoy, and J. D. Cressler, "On the analysis and design of low-loss single-pole double-throw W-band switches utilizing saturated SiGe HBTs," *IEEE Trans. Microw. Theory Tech.*, vol. 62, no. 11, pp. 2755–2767, Nov. 2014.
[30] N. Celik and M. F. Iskander, "Design, optimization, and verification of an antenna array for the 60 GHz hybrid smart antenna system," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, 2010, pp. 1–4.
[31] C. J. Wu, *et al.*, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.
[32] C. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.
[33] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1874–1889, Apr. 2015.
[34] A. Liu and V. K. N. Lau, "Impact of csi knowledge on the codebook-based hybrid beamforming in massive MIMO," *IEEE Trans. Signal Process.*, vol. 64, no. 24, pp. 6545–6556, Dec. 2016.

Lixiang Lian (Member, IEEE) received the B.Eng. degree in information and communication engineering from Zhejiang University, Hangzhou, China, in 2014, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2020. Since April 2020, she has been an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. Her research interests include wireless communication, machine learning, and compressive sensing.

Vincent K. N. Lau (Fellow, IEEE) received the B.Eng. degree (Distinction First Hons.) from the University of Hong Kong, Hong Kong, in 1992 and the Ph.D. degree from Cambridge University, Cambridge, U.K., in 1997. He joined Bell Labs from 1997 to 2004 and the Department of ECE, Hong Kong University of Science and Technology (HKUST) in 2004. He is currently a Chair Professor and the Founding Director of Huawei-HKUST Joint Innovation Lab, HKUST. His current research interests include robust and delay-optimal cross layer optimization for MIMO/OFDM wireless systems, interference mitigation techniques for wireless networks, massive MIMO, M2M and network control systems.