# Message Passing-Based Sparse Channel Estimation Under Partially Coherent Wiener Phase Errors

Hamed Masoumi, *Student Member, IEEE*, and Nitin Jonathan Myers, *Member, IEEE*

***Abstract*—Compressive sensing (CS) is key to reduce the overhead in estimating sparse high dimensional channels at millimeter wave or terahertz frequencies. The channel measurements in CS are usually perturbed by random phase errors, commonly modeled as a Wiener process, at the oscillators. CS algorithms that ignore such phase errors fail to accurately estimate the channel. In practice, the phase errors are similar within a batch of measurements acquired in a short burst and the errors vary significantly across different batches, resulting in partially coherent measurements. We develop a message passing-based channel estimation algorithm that exploits the sparse structure of the channel together with the Wiener statistics of the phase errors. To this end, we absorb the phase errors into the sparse channel, and introduce three hidden variables to model its support, magnitude, and phase. We derive the message flows between these variables while incorporating Wiener phase noise statistics. Finally, we use alternating optimization to decouple the sparse channel and the phase errors from the vector estimated with our message-passing technique. Using simulations, we show that the proposed algorithm achieves better channel reconstruction than comparable benchmarks.**

***Index Terms*—Phase noise, mm-wave, THz, compressed sensing, belief propagation.**

## I. Introduction

COMPRESSIVE sensing (CS) [1] is a tool to recover sparse vectors from few linear measurements. CS can potentially reduce the overhead in estimating high-dimensional spatial channels at millimeter-wave (mmWave) or terahertz (THz) frequencies [2]. Due to the use of large antenna arrays at these frequencies, the dimension of the channel can be in the order of hundreds to hundreds of thousands [3]. Due to the sparse structure of spatial channels in the angle domain at mmWave or THz frequencies [2], CS allows for efficient channel estimation with fewer measurements than traditional methods, thereby reducing the overhead in link configuration.

The spatial channel measurements in mmWave or THz systems are acquired sequentially in time. In practice, the phase of these measurements is perturbed due to phase noise [4]. Phase noise occurs due to random fluctuations in the frequency and phase of an oscillator [5]. This noise, more prominent at higher frequencies [6], [7], significantly alters the received signal model compared to the standard CS model, which only accounts for additive noise in the linear measurements. On the one hand, standard CS algorithms that are agnostic to phase errors, including those based on message passing [8], [9], fail when the variance of the phase errors is large. On the other hand, classical phase noise tracking approaches [10], [11] do not exploit channel sparsity, making them unsuitable for mmWave or THz channel estimation.

Prior work [12], [13], [14] addresses the phase noise issue by considering only the magnitude of the acquired measurements for sparse channel estimation. Recently, it was shown in [15] that the sparse self-calibration technique from [4] can be used to jointly estimate the phase errors and the sparse channel. While these methods are robust to phase noise [4], [12], [13], [14], they overlook the temporal correlation of phase errors, which is common in practice [16]. One of the early works that exploit the temporal correlation in CS measurements uses an extended Kalman filter to track the phase variations [17]. The method developed in [17], however, only estimates a one-sparse approximation of the channel for single stream beamforming. This motivates the need to develop techniques that estimate all the sparse entries in the channel.

Recent work in [7], [15], [18], and [19] has exploited the partially coherent phase error structure in the CS measurements for channel estimation. Under this structure, the phase errors within short batches of CS measurements are nearly the same while they change considerably across different batches. The sparse Bayesian learning-based method in [7] considers phase errors, but it does not leverage the Wiener structure in the errors within the recovery algorithm. In [18], a two-stage partially coherent compressive phase retrieval (PC-CPR) was proposed to reconstruct the sparse channel from partially coherent CS measurements. The algorithm is initialized using the CPR method from [20], and the phase offsets and the sparse channel are iteratively estimated in a refinement stage. In [19], expectation-maximization generalized approximate message passing (EM-GAMP) [21] is used to first obtain a coarse estimation of the sparse channel, which is subsequently used to estimate the phase offsets. In [15], phase error-aware matching pursuit followed by alternating optimization was developed to recover the sparse channel and the phase errors. The methods in [7], [15], and [18] assume that the channel is exactly sparse, which is unrealistic. Further, [15], [18], [19] assume that the number of non-zero entries in the sparse

Received 20 February 2025; revised 21 June 2025 and 3 September 2025; accepted 4 September 2025. Date of publication 15 September 2025; date of current version 22 December 2025. An earlier version of this paper was presented in part at the 2025 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit) [DOI: 10.1109/EuCNC/6GSummit63408.2025.11037155]. The associate editor coordinating the review of this article and approving it for publication was A. A. D'Amico. *(Corresponding author: Nitin Jonathan Myers.)*

The authors are with Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: h.masoumi@tudelft.nl; n.j.myers@tudelft.nl).

Digital Object Identifier 10.1109/TWC.2025.3607275

channel is known, which is not practical. To the best of our knowledge, none of the existing methods exploit Wiener phase error statistics in partially coherent sparse recovery using measurements acquired over multiple batches.

We develop a message passing-based sparse recovery algorithm that exploits Wiener statistics of the phase errors in the channel measurements. In message passing-based methods, the unknown variables to be solved, i.e., the sparse channel entries or the phase errors, are represented as nodes in a factor graph. Information about these variables, such as sparsity or measurement-based likelihoods, is incorporated as factors. By iteratively exchanging messages (beliefs) between the factors and the variables, the marginal distributions of the variables are computed, which are subsequently used to estimate the unknowns. In this paper, we factorize the posterior joint probability density function (PDF) of the phase offsets and the sparse angle domain channel. Our factorization incorporates the temporal correlation in the phase-perturbed batches of CS measurements. Finally, we approximate the messages in our factor graph to develop a tractable channel estimation algorithm. Our main contributions are listed below.

- We develop a message passing-based algorithm to estimate a sparse spatial channel from phase-perturbed CS measurements. To capture the channel's sparsity in the angle domain, we define two random variables for each channel entry: one to model the support and another to model the magnitude. Additionally, to incorporate the temporal Wiener phase noise statistics, we introduce a third random process that models the combined effect of the channel's phase and the phase noise.
- We derive approximations for the messages exchanged between various nodes in our factor graph to enable a computationally tractable algorithm. To this end, we model the combination of the angle domain channel entries and the phase errors as a realization of a Gaussian mixture. The properties of this mixture are adapted to the statistics of the phase noise.
- We use alternating optimization to determine the angle domain channel and the phase-perturbations from our message passing-based estimate. Finally, we show by simulations that our algorithm achieves a lower channel reconstruction error than comparable benchmarks.

In our recent work [22], we developed a multiple measurement vector (MMV) AMP-based approach for sparse channel estimation from partially coherent measurements. While the idea of absorbing phase errors into the channel to form a collection of correlated channels was introduced in [22], the method in [22] leveraged only the shared support structure across this collection. In contrast, this paper not only exploits the common support structure but also incorporates the Wiener phase noise statistics and the shared magnitude structure across the collection. To achieve this, we develop a new message-passing algorithm distinct from the conventional MMV-AMP that exploits only the common support.

*Notation:* We use $a$, $\mathbf{a}$, and $\mathbf{A}$ to represent a scalar, vector, and matrix, respectively. The transpose and conjugate-transpose operators are indicated by $(\cdot)^{\mathrm{T}}$ and $(\cdot)^*$. The $i^{\mathrm{th}}$ column of $\mathbf{A}$ is $\mathbf{a}[i]$, $a[i]$ is the $i^{\mathrm{th}}$ entry of $\mathbf{a}$, and $A(i,j)$ is the $(i,j)^{\mathrm{th}}$ entry of $\mathbf{A}$. The Frobenius norm of $\mathbf{A}$ is $\|\mathbf{A}\|_{\mathrm{F}}$. $|\cdot|$ and $\angle\cdot$ give the element-wise magnitude and phase of a vector or matrix. For an integer $N$, $[N] = 1,\ldots,N$, and $\lfloor a \rfloor$ is the floor of $a$. The vectorization of $\mathbf{A}$, denoted by $\mathrm{vec}(\mathbf{A})$, is formed by stacking its columns. $\otimes$ denotes the Kronecker product. The inner product of matrices $\mathbf{A}$ and $\mathbf{B}$ is $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} A(i,j)B^*(i,j)$. We use $\mathsf{x}$ for a random variable and $x$ for its realization, with $p(x)$ as the PDF of $\mathsf{x}$ evaluated at $x$. The complex Gaussian PDF $\mathcal{CN}(x; a, b)$ has mean $a$ and variance $b$. Lastly, $\mathbb{R}_+^{N\times 1}$ is the set of $N \times 1$ vectors with non-negative real entries, and $\mathsf{j} = \sqrt{-1}$.

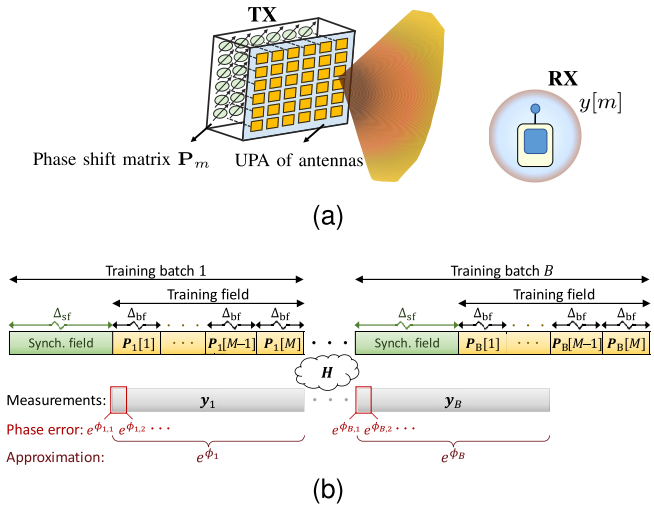



Fig. 1. A wireless system with a UPA at the TX in (a) and the frame structure for acquiring partially coherent channel measurements in (b). The phase error within each batch is assumed to be constant.

## II. Channel and System Model

In this section, we consider a point-to-point narrowband multiple-input single-output (MISO) system and discuss the partially coherent measurement model used for channel estimation. Extending our work to a multi-user scenario is beyond the scope of this paper, and we leave it for future work.

### A. Channel Model

We consider an $N \times N$ half-wavelength spaced uniform planar array (UPA) at the transmitter (TX) and a single antenna receiver (RX) as shown in Fig. 1(a). The $N^2$ dimensional narrowband MISO channel between the TX and RX is modeled as an $N \times N$ matrix $\mathbf{H}$. Let $L$ denote the propagation rays in the environment with the $\ell^{\mathrm{th}}$ ray having a complex gain of $\zeta_\ell$, an azimuth angle-of-departure (AoD) $\vartheta_{a,\ell}$ and an elevation AoD $\vartheta_{e,\ell}$. By defining the beamspace angles as $\omega_{a,\ell} = \pi \sin\vartheta_{e,\ell} \sin\vartheta_{a,\ell}$, $\omega_{e,\ell} = \pi \sin\vartheta_{e,\ell} \cos\vartheta_{a,\ell}$ [23] and the $N \times 1$ Vandermonde vector $\mathbf{a}_N(\omega)$ as

$$\mathbf{a}_N(\omega) = \left[1, e^{\mathsf{j}\omega}, e^{\mathsf{j}2\omega}, \cdots, e^{\mathsf{j}(N-1)\omega}\right]^{\mathrm{T}}, \tag{1}$$

the baseband channel matrix $\mathbf{H}$ is given by

$$\mathbf{H} = \sum_{\ell=1}^{L} \zeta_\ell \mathbf{a}_N(\omega_{e,\ell})\, \mathbf{a}_N^{\mathrm{T}}(\omega_{a,\ell}). \tag{2}$$

For typical mmWave or THz access points, the channel dimension $N^2$ can be in the order of hundreds to thousands.

High scattering at mmWave and THz carrier frequencies results in an approximately sparse channel in the angle domain [2]. By transforming $\mathbf{H}$ into the beamspace (angle domain), we can exploit this sparse structure in channel estimation. Since we assume a UPA at the TX, the 2D-Discrete Fourier Transform (DFT) of $\mathbf{H}$ is used for the beamspace representation. Let $\mathbf{U}_N$ denote the standard $N \times N$ unitary DFT matrix and $\mathbf{X}$ denote the beamspace representation of $\mathbf{H}$. Then, $\mathbf{X}$ and $\mathbf{H}$ are related as

$$\mathbf{H} = \mathbf{U}_N \mathbf{X} \mathbf{U}_N, \tag{3}$$

where $\mathbf{X}$ is approximately sparse. The channel $\mathbf{H}$ is unknown and the goal of this paper is to estimate $\mathbf{H}$ from its phase perturbed measurements acquired at the RX. To acquire one scalar-valued channel measurement $y[m]$, the TX applies a beam training matrix to its array as shown in Fig. 1(a).

### B. Frame Structure for Acquiring Measurements

We use Fig. 1(b) to explain the frame structure for acquiring measurements to estimate the channel. Our frame structure contains $B$ training batches, and is similar to the structure in IEEE 802.11ad/ay [24], [25]. Each batch consists of a synchronization field and a training field to obtain channel measurements. We define $\Delta_{\text{sf}}$ as the duration of the synchronization field and $\Delta_{\text{bf}}$ as the time needed to acquire one channel measurement within a batch. In the IEEE 802.11ad standard, $\Delta_{\text{sf}}$ is several $\mu$s while $\Delta_{\text{bf}}$ is in the order of tens of ns [26, Chapter 20]. We use $M$ to denote the number of distinct beamformers applied in each training batch, equivalently the RX acquires $M$ channel measurements per batch.

We assume a block fading channel that remains constant over $B$ training batches, which is reasonable when the channel coherence time is sufficiently long. For example, in a 60 GHz indoor environment, the coherence time is approximately 1 ms [27], significantly exceeding the training batch duration of about 44 $\mu$s with the IEEE 802.11ad/ay standards [15]. We use $\mathbf{P}_b[m] \in \mathbb{C}^{N \times N}$ to denote the beamformer applied at the TX to obtain the $m^{\text{th}}$ spatial channel measurement in batch $b$. This measurement is acquired at the RX and is denoted by $y_b[m]$. The measurements at the RX are perturbed by random phase noise, which is commonly modeled as a Wiener process [10]. Such a process is described using $\mathcal{L}(f_{\text{m}})$, the power spectral density of the phase noise process at an offset $f_{\text{m}}$ from the carrier frequency $f_{\text{c}}$ [16], i.e., $\mathcal{L}(f_{\text{m}}) = 10\log_{10}\left(f_{\text{c}}^2 c / \left(\pi^2 f_{\text{c}}^4 c^2 + f_{\text{m}}^2\right)\right)$ dBc/Hz. The phase noise parameter $c$ is an oscillator-dependent constant. We assume that this parameter does not change with time [16] as also considered in [7], [15], and [17]. An oscillator with a small $\mathcal{L}(f_{\text{m}})$ results in phase errors with low variance.

For a Wiener process, the difference in phase errors is Gaussian with a variance that is directly proportional to the time interval between the sampling instants [16]. As a result, the phase error variations within a batch are smaller when compared to the variation across different batches. To develop our algorithm, we ignore phase error variations within a batch and model only the variations across different batches, while our simulations employ realistic phase errors that also vary within a batch. Under this assumption, we use $\phi_b$ to model the phase error in all the measurements acquired in batch $b$. We define $\sigma_{\text{pn}}^2 = 4\pi^2 f_{\text{c}}^2 c \Delta$ as the variance of the phase error differences, where $\Delta$ is the time duration of a batch. For a Wiener process, $\phi_b | \phi_{b-1} \sim \mathcal{N}\left(\phi_{b-1}, \sigma_{\text{pn}}^2\right)$. As a reference, we assume that the initial phase $\phi_1 = 0$. In Fig. 2, we show the actual phase noise and the $\phi_b$s for a particular realization. The $m^{\text{th}}$ channel measurement acquired in batch $b$ is

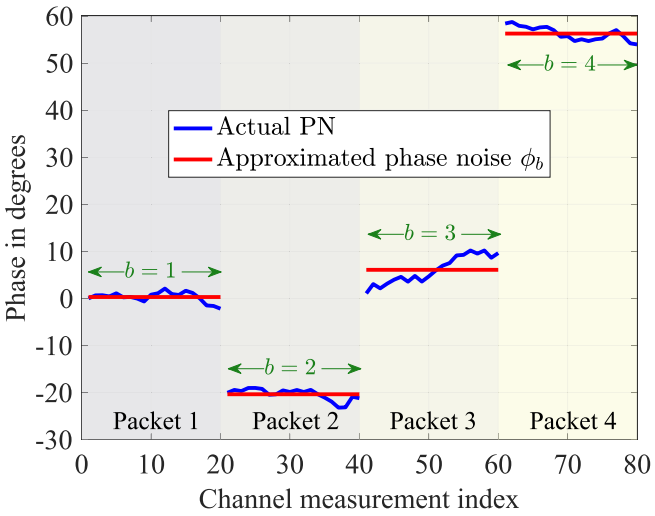



Fig. 2. A realization of the phase noise process and its approximation for $f_{\text{c}} = 60$ GHz, $B = 4$ batches, $M = 20$ measurements per batch, and phase noise power spectral density $\mathcal{L}(f_{\text{m}}) = -101.7$ dBc/Hz at $f_{\text{m}} = 1$ MHz [28]. Here, $\Delta_{\text{sf}} = 44\mu$s and $\Delta_{\text{bf}} = 128$ ns [15].

$$y_b[m] = e^{\mathrm{j}\phi_b} \langle \mathbf{H}, \mathbf{P}_b[m] \rangle + w_b[m], \qquad m \in [M], \tag{4}$$

where $w_b[m] \sim \mathcal{CN}(0, \sigma^2)$ is additive Gaussian noise. In this paper, the $MB$ phase-perturbed measurements in (4), acquired over $B$ batches, are used to estimate the channel $\mathbf{H}$.

Now, we discuss the partially coherent CS formulation [18] of the spatial measurements in (4), to exploit sparsity of the channel in the beamspace. We use $\mathbf{A}_b \in \mathbb{C}^{M \times N^2}$ to denote the CS matrix associated with the $M$ spatial measurements in (4) acquired in batch $b$. The $m^{\text{th}}$ row of $\mathbf{A}_b$ is $\text{vec}\left(\mathbf{P}_b^*[m]\right)^{\text{T}} \left(\mathbf{U}_N \otimes \mathbf{U}_N\right)$ [29], which depends on the beam training matrix $\mathbf{P}_b[m]$ applied at the TX. We define $\mathbf{y}_b$ as the vector of $M$ measurements $\{y_b[m]\}_{m=1}^M$ acquired in the $b^{\text{th}}$ batch, and $\mathbf{w}_b$ as the additive measurement noise vector containing $\{w_b[m]\}_{m=1}^M$. The vector version of the sparse beamspace $\mathbf{X}$ is denoted by the $N^2$-dimension vector $\mathbf{x}$, i.e., $\mathbf{x} = \text{vec}(\mathbf{X})$. Using $\langle \mathbf{H}, \mathbf{P}_b[m] \rangle = \mathbf{A}_b(m,:)\mathbf{x}$ [29] in (4) yields

$$\mathbf{y}_b = e^{\mathrm{j}\phi_b} \mathbf{A}_b \mathbf{x} + \mathbf{w}_b. \tag{5}$$

The phase errors $\{\phi_b\}_{b=2}^B$ and the sparse vector $\mathbf{x}$ are unknown and have to be estimated using the acquired measurements $\{\mathbf{y}_b\}_{b=1}^B$ and the known CS matrices $\{\mathbf{A}_b\}_{b=1}^B$. It is also known that $\phi_b | \phi_{b-1} \sim \mathcal{N}\left(\phi_{b-1}, \sigma_{\text{pn}}^2\right)$.

## III. Overview of Our Message Passing-Based Partially Coherent Sparse Recovery Algorithm

As estimating $\phi_b$s and $\mathbf{x}$ from the non-linear model in (5) is challenging, we first absorb the phase errors into the

sparse vector to create $B$ auxiliary vectors, each with the same dimension as $\mathbf{x}$. We then reformulate (5) as linear measurements of these auxiliary vectors. We show that these vectors are correlated, i.e., they share the same support and differ only by phase, which follows Wiener statistics. Our message-passing method leverages this correlation, sparsity in the beamspace, and the measurements in (5) to estimate $\mathbf{x}$.

### A. Sparse Prior and Correlated Structure of Auxiliary Vectors

A common sparsity-promoting prior is the Bernoulli-Gaussian (BG) PDF [30], [31], which was used to model sparse channels in [32]. Similar to [32], we model $\mathbf{x}$ as a realization of a random vector $\mathsf{x}$ that follows a BG prior. Specifically, $\mathsf{x}$ has independent and identically distributed (IID) entries which are zero with probability $1-\lambda$, i.e., $\Pr\{\mathsf{x}[n]=0\}=1-\lambda \forall n \in [N^2]$, and follow a Gaussian distribution otherwise. We consider a zero-mean complex Gaussian distribution with the same variance $\mu$ to write the BG prior as

$$p\left(x[n]\right)=(1-\lambda)\delta(x[n])+\lambda\mathcal{CN}\left(x[n];0,\mu\right), \tag{6}$$

where $\delta(\cdot)$ is the Dirac delta function.

In our method, we absorb the phase error terms $\{e^{\mathsf{j}\phi_b}\}_{b=1}^{B}$ into the sparse vector $\mathbf{x}$ to construct the auxiliary vectors

$$\mathbf{z}_b = e^{\mathsf{j}\phi_b}\mathbf{x}, \ \ \forall b \in [B]. \tag{7}$$

As $\mathbf{z}_b$ is just a scalar multiple of $\mathbf{x}$, $\mathbf{z}_b$ is sparse when $\mathbf{x}$ is sparse. Let $\mathcal{S}_{\mathbf{x}}$ denote the support of $\mathbf{x}$, i.e., $\mathcal{S}_{\mathbf{x}} \triangleq \{n : x[n] \neq 0\}$, and $\mathcal{S}_{\mathbf{z}_b}$ denote the support of $\mathbf{z}_b$. From (7), we observe that $\mathbf{z}_b$ has the same support as $\mathbf{x}$ for each $b$, i.e., $\mathcal{S}_{\mathbf{z}_1}=\mathcal{S}_{\mathbf{z}_2}=\cdots=\mathcal{S}_{\mathbf{z}_B}=\mathcal{S}_{\mathbf{x}}$. In addition, the entry-wise magnitudes of $\mathbf{x}$ and $\mathbf{z}_b$ are equal, i.e., $|\mathbf{z}_1|=|\mathbf{z}_2|=\cdots=|\mathbf{z}_B|=|\mathbf{x}|$. Furthermore, $\mathbf{z}_b$ and $\mathbf{z}_{b-1}$ differ only by a phase error which follows the Wiener statistics. Using (7), the measurements in (5) can be rewritten as

$$\mathbf{y}_b = \mathbf{A}_b\mathbf{z}_b + \mathbf{w}_b, \ \ \forall b \in [B]. \tag{8}$$

Our goal is to develop an algorithm to first estimate $\{\mathbf{z}_b\}_{b=1}^{B}$ from (8) while exploiting the correlation within this collection. Then, we use set of equations in (7) to recover $\mathbf{x}$ from $\{\hat{\mathbf{z}}_b\}_{b=1}^{B}$.

We discuss two methods to reconstruct the auxiliary vectors $\{\mathbf{z}_b\}_{b=1}^{B}$ from (8). The first method exploits sparsity and the common support across the auxiliary vectors, ignoring the shared magnitude and the Wiener statistics due to phase noise. We refer to this method as *Proposed SuppOnly*, which is solved using the MMV-AMP algorithm [33]. For more details on our this approach, we refer the reader to our conference version [22]. The emphasis of this paper is on our second method, which not only exploits sparsity and the shared support in $\{\mathbf{z}_b\}_{b=1}^{B}$ but also leverages the common magnitude structure and the Wiener phase noise statistics in $\{\mathbf{z}_b\}_{b=1}^{B}$.

In our message passing-based method, we model $\mathbf{z}_b$ as a realization of a random vector $\mathsf{z}_b$. The support, magnitude, and the phase of $\mathsf{z}_b[n]$ are modeled using random variables $\mathsf{s}[n] \in \{0,1\}$, $\mathsf{r}[n] \in \mathbb{R}_+$, and $\Theta_b[n] \in \mathbb{R}$ to express $\mathsf{z}_b[n]$ as

$$\mathsf{z}_b[n] = \mathsf{s}[n]\mathsf{r}[n]e^{\mathsf{j}\Theta_b[n]}. \tag{9}$$

Equivalently, (9) can be expressed as

$$p(z_b[n]|s[n],r[n],\theta_b[n]) = \begin{cases} \delta(z_b[n]-r[n]e^{\mathsf{j}\theta_b[n]}), & \text{if } s[n]=1 \\ \delta(z_b[n]), & \text{if } s[n]=0, \end{cases} \tag{10}$$

where $\theta_b[n]$ is a realization of $\Theta_b[n]$. From (10), we observe that $z_b[n]=0$ when $s[n]=0$, thereby aiding sparse priors. As $\mathsf{z}_b[n]$ has the same support as $\mathsf{x}_b[n]$, it follows from the BG prior on $\mathsf{x}_b[n]$ in (6) that [30], [31]

$$p\left(s[n]\right)=\lambda^{s[n]}\left(1-\lambda\right)^{1-s[n]}, s[n]\in\{0,1\}. \tag{11}$$

To determine the prior on $\mathsf{r}[n]$, we first observe that $|\mathsf{z}_b[n]| = |\mathsf{x}[n]|$ from (7). Next, it follows from [34, Appendix 5.B] that the magnitude of a random variable distributed as $\mathcal{CN}(x[n];0,\mu)$ has a Rayleigh distribution with parameter $\sqrt{\mu/2}$. Therefore, the PDF of $\mathsf{r}[n]$ is

$$p\left(r[n]\right)=\text{Rayleigh}\left(r[n];\sqrt{\mu/2}\right). \tag{12}$$

To find the prior on $\Theta_b[n]=\angle\mathsf{z}_b[n]$, we observe from (7) that $\angle\mathsf{z}_b[n]=\angle\mathsf{x}[n]+\varphi_b$, where $\varphi_b$ models the realization $\phi_b$. Since the non-zero entries of $\mathsf{x}$ are IID according to $\mathcal{CN}\left(x[n];0,\mu\right)$, their phases $\{\angle\mathsf{x}[n]\}_{n=1}^{N^2}$ have IID uniform distribution $\mathcal{U}([-\pi,\pi))$. As we assume $\phi_1=0$, we have $\angle\mathsf{z}_1=\angle\mathsf{x}$ from (7). Further, $\{\angle\mathsf{z}_b\}_{b=1}^{B}$ can be modeled as a new Wiener process with $\angle\mathsf{z}_1 \sim \mathcal{U}\left([-\pi,\pi)^{N^2}\right)$, i.e.,

$$\angle\mathsf{z}_b = \angle\mathsf{z}_{b-1} + \mathbf{1}_{N^2}\Phi_b, \tag{13}$$

where $\mathbf{1}_{N^2}$ is a vector of ones and $\Phi_b \sim \mathcal{N}\left(0,\sigma_{\text{pn}}^2\right)$. We observe from (7) that the same phase error is introduced in all entries of $\mathbf{z}_b$. Therefore, the phase noise innovation $\mathcal{N}\left(0,\sigma_{\text{pn}}^2\right)$ is same for all the elements of $\mathsf{z}_b$ in (13).

Incorporating a common phase noise innovation across all the elements of $\mathsf{z}_b$ is challenging as it leads to numerous loops in the corresponding factor graph. To aid a tractable message passing method, our model assumes that the phase error innovations within $\mathsf{z}_b$ are independent, i.e.,

$$\Theta_b[n]=\angle\mathsf{z}_b[n]=\angle\mathsf{z}_{b-1}[n]+\Phi_b[n], \tag{14}$$

where $\Phi_b[n]$ are IID as $\mathcal{N}\left(0,\sigma_{\text{pn}}^2\right)$. Therefore, (14) is a relaxation of the true model in (13). Under this approximation,

$$p\left(\theta_b[n]\middle|\theta_{b-1}[n]\right)=\mathcal{N}\left(\theta_b[n];\theta_{b-1}[n],\sigma_{\text{pn}}^2\right), \tag{15}$$

where $\theta_1[n] \sim \mathcal{U}([-\pi,\pi))$. Note that our message passing-based method to estimate the auxiliary vectors does not exploit the property that the true $\theta_b[n]$ is invariant with $n$. Our alternating optimization technique in Sec. IV, however, leverages this property to recover $\mathbf{x}$ from the estimates $\{\hat{\mathbf{z}}_b\}_{b=1}^{B}$.

As the supports of $\{\mathsf{z}_b\}_{b=1}^{B}$ are modeled by a single random vector $\mathsf{s}$ and their magnitudes are modeled by a single random vector $\mathsf{r}$, our method inherently incorporates the shared support and magnitude structure. The Wiener phase noise statistics is incorporated using (13), with the limitation that the phase errors within each batch are assumed to be IID.

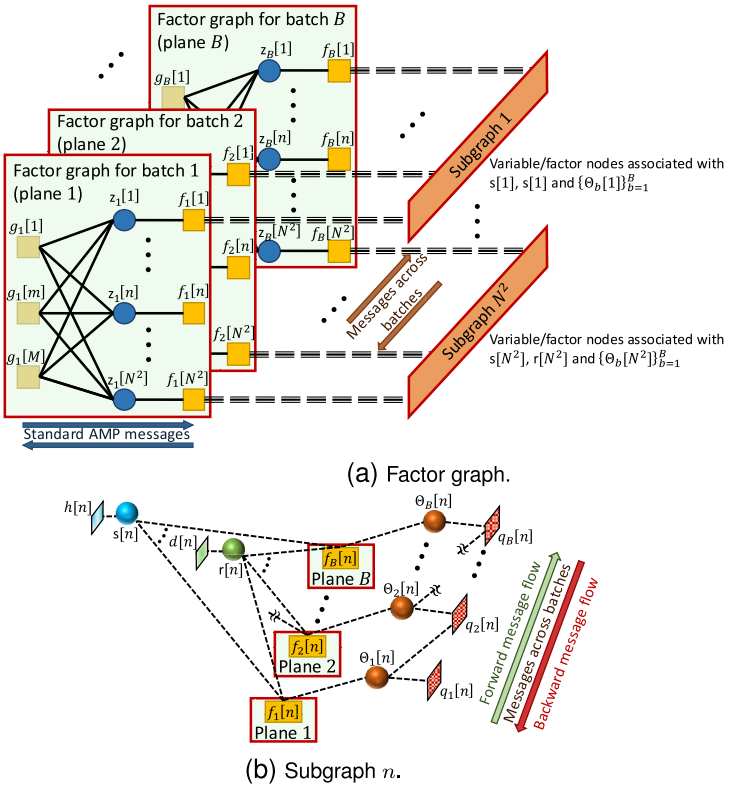



Fig. 3. The factor graph corresponding to the factorization of the posterior joint PDF in (16). Figure (b) shows the details of subgraph $n$ in the factor graph. In this paper, the MMSE estimate of $\{\mathbf{z}_b\}_{b=1}^B$ given the measurements $\{\mathbf{y}_b\}_{b=1}^B$ is computed, which is subsequently used to estimate $\mathbf{x}$. Our factor graph incorporates sparsity, shared support and magnitude structure, and Wiener phase noise statistics. The explicit form of the factors in our graph are listed in Table I.

TABLE I
FACTORS OF THE JOINT PDF IN (16)

| Factor definition | Explicit form |
|---|---|
| $g_b[m] = p\left(y_b[m]\middle\vert \mathbf{z}_b\right)$ | $\mathcal{CN}\left(y_b[m]; \mathbf{a}_b^{\mathrm{T}}[m]\mathbf{z}_b, \sigma^2\right)$ |
| $f_b[n] = p\left(z_b[n]\middle\vert s[n], r[n], \theta_b[n]\right)$ | $\delta\left(z_b[n] - s[n]r[n]e^{j\theta_b[n]}\right)$ |
| $q_b[n] = p\left(\theta_b[n]\middle\vert \theta_{b-1}[n]\right)$ | $\mathcal{N}\left(\theta_b[n]; \theta_{b-1}[n], \sigma_{\mathrm{pn}}^2\right),\ \theta_1[n] \sim \mathcal{U}([-\pi, \pi))$ |
| $h[n] = p(s[n])$ | $\lambda^{s[n]}\left(1-\lambda\right)^{1-s[n]},\ s[n] \in \{0,1\}$ |
| $d[n] = p(r[n])$ | $\mathrm{Rayleigh}\left(r[n]; \sqrt{\mu/2}\right) = 2\frac{r[n]}{\mu}e^{-r^2[n]/\mu}$ |

## B. Factor Graph and Message Scheduling

We use the posterior joint PDF of $\{\mathsf{z}_b\}_{b=1}^B$, $\mathsf{s}$, $\mathsf{r}$ and $\{\Theta_b\}_{b=1}^B$ to develop our algorithm for estimating $\{\mathbf{z}_b\}_{b=1}^B$. By using the Bayes' rule and the dependencies between $\{\mathsf{z}_b\}_{b=1}^B$, $\mathsf{s}$, $\mathsf{r}$ and $\{\Theta_b\}_{b=1}^B$, we factorize the joint PDF as

$$
\begin{aligned}
&p\left(\{\mathbf{z}_b\}_{b=1}^B, \mathbf{s}, \mathbf{r}, \{\boldsymbol{\theta}_b\}_{b=1}^B \middle\vert \{\mathbf{y}_b\}_{b=1}^B\right) \\
&\propto \prod_{b=1}^B \left( \prod_{m=1}^M p\left(y_b[m]\middle\vert \mathbf{z}_b\right) \prod_{n=1}^N p\left(z_b[n]\middle\vert s[n], r[n], \theta_b[n]\right) \right. \\
&\left. p\left(\theta_b[n]\middle\vert \theta_{b-1}[n]\right) \right) \prod_{n=1}^N p\left(s[n]\right) \prod_{n=1}^N p\left(r[n]\right), \qquad (16)
\end{aligned}
$$

where $\propto$ denotes equality up to a constant scale factor, and $p\left(\theta_1[n]\middle\vert \theta_0[n]\right) = p\left(\theta_1[n]\right)$. See Table I for the factors in (16).

The dependencies among different variables in (16) is represented by a *factor graph* shown in Fig. 3. The factor graph contains two types of nodes: *factor nodes* shown as rectangles and *variable nodes* shown as circles. Each factor node represents a factor in (16) and is connected to the variable nodes corresponding to its arguments. The explicit form of all the factors in our graph are listed in Table I. Our factor graph, shown in Fig. 3 comprises $B$ planes, where each plane corresponds to a measurement batch in (8). The connections between different planes indicate the correlations across $\{\mathbf{z}_b\}_{b=1}^B$, which follow from (9) and (14). For example, the $n^{\mathrm{th}}$ factor nodes $\{f_b[n]\}_{b=1}^B$ in all planes are connected through $r[n]$ and $s[n]$ in Fig. 3(b). This is because $\{\mathsf{z}_b[n]\}_{b=1}^B$ have the same magnitude, i.e., $r[n]$, and are all jointly non-zero when $s[n] = 1$. The evolution of the phase noise process across different batches is modeled by inter-plane connections through variable nodes $\{\Theta_b[n]\}_{b=1}^B$ using the factors $\{q_b[n]\}_{b=1}^B$. The $q_b[n]$s, shown in Fig. 3(b), incorporate Wiener phase noise statistics from (15).

Our method aims to find the minimum mean squared error (MMSE) estimate of $\{\mathbf{z}_b\}_{b=1}^B$ given the measurements $\{\mathbf{y}_b\}_{b=1}^B$, using the factor graph in Fig. 3. To this end, the marginal posteriors $p\left(z_b[n]\middle\vert \{\mathbf{y}_b\}_{b=1}^B\right)$ are first estimated using the sum-product algorithm [35]. The means of these posteriors then yield the MMSE estimates. In each iteration of the sum-product algorithm, nodes in the factor graph exchange messages, also known as beliefs, which represent probability densities. There are two types of message flows: those sent from variables to neighboring factors, and those sent from factors to neighboring variables. A belief exchanged from a factor to a variable represents the probability distribution of

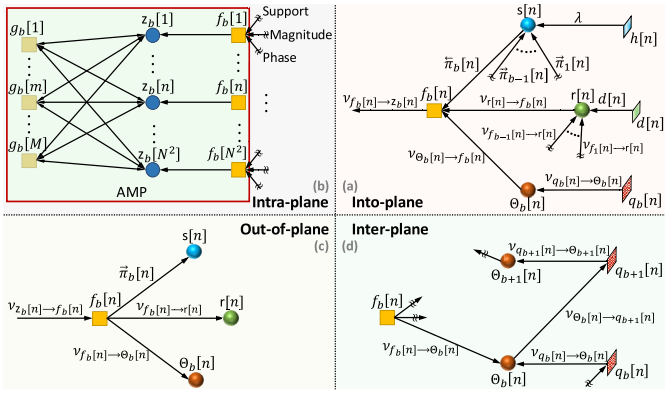



Fig. 4. Four sections of the factor graph illustrating the four key steps of the serial message schedule. It shows only the messages involved for nodes indexed by $n \in [N^2]$ and the nodes within plane $b \in [B]$.

the variable, as perceived from the factor's perspective. In contrast, the belief sent from a variable to a factor represents the variable's perception of its own probability distribution, informed by all the factor nodes it is connected to—except for the one it is messaging. In the sum-product algorithm [35], a message from a variable to a factor is obtained by multiplying all incoming messages from the other neighboring factors. A message from a factor to a variable is calculated by summing (or integrating) the product of all incoming messages over all variables, except for the one to which the message is sent.

Our algorithm propagates messages on a plane-by-plane basis to exploit the Wiener structure in the phase errors. In the first plane ($b = 1$) of Fig. 3, the linear regression problem in (8) is solved with a sparse prior on $\mathsf{z}_b$. This "intra-plane" step is solved using the standard AMP [9], which approximates the messages from a factor node $g_b[m]$ to a variable node $\mathsf{z}_b[n]$ as Gaussian PDF. This approximation is possible because the factor node $g_b[m]$ depends on $z_b[n]$ through the linear mix $\mathbf{a}_b^{\mathrm{T}}[m]\mathbf{z}_b$ where the elements of $\mathbf{a}_b^{\mathrm{T}}[m]$ have approximately similar norm [9]. After the intra-plane step converges, the marginal posterior of $\mathsf{z}_1[n]$ is estimated for each $n$. This posterior provides side-information about the support and the magnitude of $\mathsf{z}_2[n]$, a variable to be solved in the next plane ($b = 2$). The Wiener phase noise statistic, combined with the posterior, provides side-information on the phase of $\mathsf{z}_2[n]$. The "out-of-plane" step in our algorithm derives side-information from a plane, while the "inter-plane" step propagates this information to the next plane. In Fig. 4, the out-of-plane step shows $f_1[n]$ sending beliefs about the support $\mathsf{s}[n]$, magnitude $\mathsf{r}[n]$, and phase $\Theta_1[n]$, informed by the message from $\mathsf{z}_1[n]$. The support and magnitude messages are directly propagated to the next plane, while the Wiener statistics is used to compute the belief about $\Theta_2[n]$. Finally, the "into-plane" step propagates the support, magnitude, and phase side-information to the next plane. This information is used as a prior to solve for the posterior of $\mathsf{z}_2$ in (8) using the AMP. The procedure is continued until an estimate of the marginal posterior PDF for $\mathsf{z}_B[n]$ in the last plane, i.e., plane $B$, is obtained. Because the messages are passed in the direction of the increasing plane index $b$, we refer to this process as the *forward* message flow.

**Algorithm 1** Alternating Algorithm to Solve (17)

**Input:** $\{\hat{\mathbf{z}}_b\}_{b=1}^{B}$, Maximum number of iterations $T_{\max}$.
**Initialization:** $i = 1$, $\hat{\mathbf{x}}^{(1)} = \hat{\mathbf{z}}_1$.
**For** $i = 1, \cdots, T_{\max}$ **do**:
1. $\hat{\boldsymbol{\phi}}^{(i)} = -\angle\hat{\mathbf{Z}}^{*}\hat{\mathbf{x}}^{(i)}$. Note: $\angle\mathbf{c}$ returns the angle of $\mathbf{c}$
2. $\hat{\mathbf{x}}^{(i+1)} = \frac{1}{B}\hat{\mathbf{Z}}e^{-\mathrm{j}\hat{\boldsymbol{\phi}}^{(i)}}$.

**End For**
**Output:** $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(T_{\max})}$.

The forward flow can sequentially update the channel estimate with each new measurement batch.

When the measurements from all the $B$ batches are available for channel estimation, *backward* message flows can also be performed in addition to the forward message flows [30]. The backward flow propagates side-information from plane $b+1$ to plane $b$, i.e., measurements in the $b+1^{\text{th}}$ batch are used to derive priors needed to solve for $\mathsf{z}_b[n]$. After multiple forward-backward message flows, the marginal posterior $p(z_b[n]|\{\mathbf{y}_b\}_{b=1}^{B})$ is estimated by exploiting measurements from all the batches. The mean of these posteriors gives the auxiliary vector estimates $\{\hat{\mathbf{z}}_b\}_{b=1}^{B}$. To estimate $\mathbf{x}$ from these vectors, we stack $\{\hat{\mathbf{z}}_b\}_{b=1}^{B}$ in an $N^2 \times B$ matrix $\hat{\mathbf{Z}} = [\hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_B]$. We define $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_B]^{\mathrm{T}}$, $e^{\mathrm{j}\boldsymbol{\phi}} = \left[e^{\mathrm{j}\phi_1}, \ldots, e^{\mathrm{j}\phi_B}\right]^{\mathrm{T}}$ and $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_B]$. From (7), we observe that $\mathbf{Z}$ is a rank one matrix as $\mathbf{Z} = \mathbf{x}e^{\mathrm{j}\boldsymbol{\phi}^{\mathrm{T}}}$. Hence, we find a rank-one matrix of the form $\mathbf{x}e^{\mathrm{j}\boldsymbol{\phi}^{\mathrm{T}}}$ that is closest to $\hat{\mathbf{Z}}$ in terms of the Frobenius norm distance, i.e.,

$$\{\hat{\mathbf{x}}, \hat{\boldsymbol{\phi}}\} = \underset{\{\mathbf{x},\boldsymbol{\phi}\}}{\operatorname{argmin}} \ \|\hat{\mathbf{Z}} - \mathbf{x}e^{\mathrm{j}\boldsymbol{\phi}^{\mathrm{T}}}\|_{\mathrm{F}}, \tag{17}$$

where $\hat{\mathbf{x}}$ is our estimate for the sparse channel $\mathbf{x}$. To solve (17), we develop an alternating approach outlined in Algorithm 1.

## IV. Messages in the Proposed Algorithm

In this section, we derive the messages exchanged between the nodes within the factor graph in Fig. 3. While we primarily focus on forward message flows, the backward flows can be derived through a similar procedure.

We use $\nu_{i\to j}$ to denote the message sent from node $i$ to node $j$. Examples of these directional message flows are shown in Fig. 4. Note that the message exchanged between the factor $f_b[n]$ and the support variable $\mathsf{s}[n]$ is a Bernoulli probability mass function (PMF), e.g., $\nu_{f_b[n]\to\mathsf{s}[n]} = (\vec{\pi}_b[n], 1-\vec{\pi}_b[n])$ where $\vec{\pi}_b[n]$ is the belief that $\mathsf{s}[n]$ is 1. For ease of notation, we denote this Bernoulli PMF as $\nu_{f_b[n]\to\mathsf{s}[n]} = \vec{\pi}_b[n]$. Similarly, $\nu_{\mathsf{s}[n]\to f_b[n]} = \overleftarrow{\pi}_b[n]$. We use $\lambda = \nu_{h[n]\to\mathsf{s}[n]}$ to denote the Bernoulli PMF sent from the factor $h[n]$ to the variable $\mathsf{s}[n]$. Finally, as $d[n]$ itself is the belief about the magnitude $\mathsf{r}[n]$, we write $d[n] = \nu_{d[n]\to\mathsf{r}[n]}$.

### A. Messages Exchanged With the Amp Subgraph

The forward pass in our algorithm begins with the into-plane step for $b = 1$. As no side-information is available for $b = 1$, $\overleftarrow{\pi}_1[n] = h[n]$, $\nu_{\mathsf{r}[n]\to f_1[n]} = d[n]$, and $\nu_{\Theta_1[n]\to f_1[n]} = q_1[n]$. These distributions, listed in Table I, are based on the parameters $\lambda$ and $\mu$, which can either be estimated from a dataset

or learned using the expectation maximization algorithm [36]. Since, $h[n]$ is a Bernoulli PMF, $d[n]$ is a Rayleigh PDF and $q_1[n]$ is a uniform PDF over $[-\pi, \pi)$, it follows from (9) that into-plane message $\nu_{f_1[n]\to \mathsf{z}_1[n]}$ is the BG prior

$$\nu_{f_1[n]\to \mathsf{z}_1[n]} = (1-\lambda)\delta(z_1[n]) + \lambda \mathcal{CN}\left(z_1[n]; 0, \mu\right), \quad (18)$$

for each $n$. These into-plane messages provide beliefs on $\{\mathsf{z}_1[n]\}_{n=1}^{N^2}$, which are used in the intra-plane step for $b=1$.

The intra-plane step in plane $b$ performs standard AMP using the into-plane beliefs and the measurement-based likelihoods, i.e., the $g_b[m]$s. As the CS matrix $\mathbf{A}_b$ is usually dense, the factor graph in plane $b$ has several cycles, which require computing multi-dimensional integrals with the sum-product algorithm. To circumvent this high-complexity integration, AMP approximates $\nu_{g_b[m]\to \mathsf{z}_b[n]}$ as a Gaussian, by using the central limit theorem and the Taylor series expansion [9]. For the exact messages exchanged between $\{g_b[m]\}_{m=1}^{M}$ and $\mathsf{z}_b[n]$, we refer the reader to [9]. To apply the AMP, the message $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ is treated as a prior on $\mathsf{z}_b[n]$ for each $n$. For instance, these messages are the BG prior in (18) for $b=1$. After AMP iterations, the local posterior of $\mathsf{z}_b[n]$ based only on batch $b$ measurements can be obtained. We use $e_b[n]$ to denote the mean and $c_b[n]$ to denote the variance of this posterior. At the end of intra-plane AMP iterations, the messages $\{\nu_{\mathsf{z}_b[n]\to f_b[n]}\}_{n=1}^{N^2}$ given by

$$\nu_{\mathsf{z}_b[n]\to f_b[n]} = \mathcal{CN}(z_b[n]; e_b[n], c_b[n]), \quad (19)$$

are used to perform the out-of-plane step. Towards the end of this section, we derive the messages $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ for $b \geq 2$. These messages are deferred to a later point, as they depend on messages involving the phase variables.

### B. Messages With the Magnitude Variables

In the out-of-plane step, the beliefs on $\mathsf{s}[n]$s, $\mathsf{r}[n]$s, and $\Theta_b[n]$s are updated using the posteriors in (19) from plane $b$. By the sum-product algorithm, $\nu_{f_b[n]\to \mathsf{r}[n]}$ is derived by first computing the product of $\nu_{\mathsf{s}[n]\to f_b[n]}$, the posterior from (19) and $\nu_{\Theta_b[n]\to f_b[n]}$. This product is then integrated over $\mathsf{s}[n]$ and $\Theta_b[n]$ to write

$$\begin{aligned}\nu_{f_b[n]\to \mathsf{r}[n]} =& (1-\overleftarrow{\pi}_b[n])\,\mathcal{CN}(0; e_b[n], c_b[n]) \\ &+\overleftarrow{\pi}_b[n]\int_{\theta[n]} \nu_{\Theta_b[n]\to f_b[n]} \mathcal{CN}\left(r[n]e^{\mathrm{j}\theta_b[n]}; e_b[n], c_b[n]\right).\end{aligned} \quad (20)$$

The first term in (20), corresponding to $s[n]=0$, is constant with respect to the variable $r[n]$ and this constant makes $\nu_{f_b[n]\to \mathsf{r}[n]}$ an improper distribution which does not integrate to 1 [33]. To address this issue, we use the idea from [33] that considers $\mathsf{s}[n] \in \{\epsilon, 1\}$ in the limit $\epsilon \to 0$, and define

$$\Omega\left(\overleftarrow{\pi}_b[n]\right) = \frac{\epsilon^2 \overleftarrow{\pi}_b[n]}{(1-\overleftarrow{\pi}_b[n]) + \epsilon^2 \overleftarrow{\pi}_b[n]}. \quad (21)$$

For a small $\epsilon$, the message $\nu_{f_b[n]\to \mathsf{r}[n]}$ in (20) is modified to

$$\begin{aligned}&\bar{\nu}_{f_b[n]\to \mathsf{r}[n]} \\ &= (1-\Omega(\overleftarrow{\pi}_b[n])) \int_{\theta_b[n]} \nu_{\Theta_b[n]\to f_b[n]} \mathcal{CN}\left(r[n]e^{\mathrm{j}\theta_b[n]}; \frac{1}{\epsilon}e_b[n], \frac{1}{\epsilon^2}c_b[n]\right) \\ &\quad + \Omega(\overleftarrow{\pi}_b[n]) \int_{\Theta_b[n]} \nu_{\Theta_b[n]\to f_b[n]} \mathcal{CN}(r[n]e^{\mathrm{j}\theta_b[n]}; e_b[n], c_b[n]),\end{aligned} \quad (22)$$

which can be shown to be a proper distribution. To obtain a closed form expression for (22), we set $\nu_{\Theta_b[n]\to f_b[n]}$ to a uniform distribution over $[-\pi, \pi)$. The uniform prior on $\Theta_b[n]$ is exact for $b=1$, as indicated in Table I. For $b>1$, however, the prior on $\Theta_b[n]$ may not be uniform over $[-\pi, \pi)$ due to the Wiener phase statistics. While the true prior on $\Theta_b[n]$ is used to compute beliefs on the phase variables, we assume that $\Theta_b[n]$ is uniform over $[-\pi, \pi)$ to compute the belief in (22). This assumption, although a limitation for $b>1$, is made to obtain a closed form expression for (22).

The closed form expression for $\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$, which approximates $\nu_{f_b[n]\to \mathsf{r}[n]}$, is given in (48) of Section A. This message $\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$, derived under the assumption that $\nu_{\Theta_b[n]\to f_b[n]} = 1/2\pi$, includes the modified Bessel function and propagating such a message is computationally expensive [30], [31]. To tackle this challenge, we approximate $\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$ as a Gaussian PDF, which can be fully characterized by its mean and variance. We define $\varrho_b[n] = |e_b[n]|^2/2c_b[n]$, $t_b[n]$ as the mean of $\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$ and $v_b[n]$ as its variance. The expressions for $t_b[n]$ and $v_b[n]$, derived in Section A, are

$$t_b[n] = \frac{\sqrt{c_b[n]/\pi}}{\epsilon(1-\Omega(\overleftarrow{\pi}_b[n])) + \Omega(\overleftarrow{\pi}_b[n])} \times \frac{\exp(\varrho_b[n])}{\mathrm{I}_0(\varrho_b[n])}, \quad (23)$$

$$\begin{aligned}v_b[n] =& \frac{\epsilon^{-1}(1-\Omega(\overleftarrow{\pi}_b[n])) + \Omega(\overleftarrow{\pi}_b[n])}{\epsilon(1-\Omega(\overleftarrow{\pi}_b[n])) + \Omega(\overleftarrow{\pi}_b[n])} \times \frac{c_b[n]}{2} \\ &\times \left(1 + 2\varrho_b[n]\left(1 + \frac{\mathrm{I}_1(\varrho_b[n])}{\mathrm{I}_0(\varrho_b[n])}\right)\right) - t_b^2[n].\end{aligned} \quad (24)$$

In summary, we assumed that $\nu_{\Theta_b[n]\to f_b[n]} = 1/2\pi$ to approximate the out-of-plane message $\nu_{f_b[n]\to \mathsf{r}[n]}$ as

$$\nu_{f_b[n]\to \mathsf{r}[n]} \propto \mathcal{N}(r[n]; t_b[n], v_b[n]). \quad (25)$$

We make use of (25) to compute the into-plane message $\nu_{\mathsf{r}[n]\to f_b[n]}$ for $b>1$. Using the sum-product rule, we can write from the factor graph in Fig. 3 that

$$\nu_{\mathsf{r}[n]\to f_b[n]} = p(r[n]) \prod_{\ell=1}^{b-1} \nu_{f_b[n]\to \mathsf{r}[n]}. \quad (26)$$

Substituting (25) and the Rayleigh PDF $d[n] = p(r[n])$ from Table I in (26), we can write

$$\nu_{\mathsf{r}[n]\to f_b[n]} \propto r[n]\mathcal{N}(r[n]; \eta_b[n], \gamma_b[n]), \quad \forall r[n] \in \mathbb{R}_+, \quad (27)$$

where

$$\eta_b[n] = \frac{\sum_{\ell=1}^{b-1} \frac{t_\ell[n]}{v_\ell[n]}}{\frac{2}{\mu} + \sum_{\ell=1}^{b-1} \frac{1}{v_\ell[n]}}, \quad \gamma_b[n] = \frac{1}{\frac{2}{\mu} + \sum_{\ell=1}^{b-1} \frac{1}{v_\ell[n]}}. \quad (28)$$

The simplification in (27) follows from the fact that the product of $b-1$ Gaussian PDFs is also Gaussian and that the Rayleigh PDF $p(r[n])$ from Table I can be rewritten as $p(r[n]) = 2\sqrt{\pi/\mu}\, r[n]\mathcal{N}(r[n]; 0, \mu/2)$ where $r[n] \in \mathbb{R}_+$.

### C. Messages With the Support Variables

Next, we derive the out-of-plane message $\nu_{f_b[n]\to \mathsf{s}[n]}$. By the sum-product rule,

$$\nu_{f_b[n]\to \mathsf{s}[n]} = \int_{z_b[n]} \int_{r[n]} \int_{\theta_b[n]} f_b[n] \nu_{\mathsf{z}_b[n]\to f_b[n]} \nu_{\mathsf{r}[n]\to f_b[n]} \nu_{\Theta_b[n]\to f_b[n]}. \tag{29}$$

For the Bernoulli PMF $\nu_{f_b[n]\to \mathsf{s}[n]}$, the belief that $\mathsf{s}[n]=1$ is

$$\vec{\pi}_b[n] = \frac{[\nu_{f_b[n]\to \mathsf{s}[n]}]_{\mathsf{s}[n]=1}}{[\nu_{f_b[n]\to \mathsf{s}[n]}]_{\mathsf{s}[n]=1} + [\nu_{f_b[n]\to \mathsf{s}[n]}]_{\mathsf{s}[n]=0}}. \tag{30}$$

To compute $\vec{\pi}_b[n]$, we derive $\nu_{f_b[n]\to \mathsf{s}[n]}$ under the assumption that $\nu_{\Theta_b[n]\to f_b[n]} = 1/2\pi$. The same assumption was made to approximate $\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$. To obtain a closed-form expression for $\nu_{f_b[n]\to \mathsf{s}[n]}$, we use a Rice distributed [37] approximation of $\nu_{\mathsf{r}[n]\to f_b[n]}$ in (27) with parameters $\rho_b[n]$, $\tau_b[n]$, i.e.,

$$\text{Rice}\left(r[n];\rho_b[n],\tau_b[n]\right) = \frac{r[n]}{\tau_b[n]} \exp\left(-\frac{r^2[n]+\rho_b^2[n]}{2\tau_b[n]}\right) \times \mathrm{I}_0\left(\frac{\rho_b[n]}{\tau_b[n]} r[n]\right), \tag{31}$$

with the same mean and variance as $\nu_{\mathsf{r}[n]\to f_b[n]}$. The Rice distribution is a reasonable choice as the non-zero entries of $\mathbf{x}$ have a complex Gaussian distribution. The parameters $\rho_b[n]$ and $\tau_b[n]$ can be computed from $\nu_{\mathsf{r}[n]\to f_b[n]}$ in (27), using the RiceFit function in MATLAB [38]. The inputs of this function are the mean and the variance of $\nu_{\mathsf{r}[n]\to f_b[n]}$, given by (61) and (62), in Section B. Finally, we define $\bar{\tau}_b[n] = (2\tau_b[n] + c_b[n])/2$ and approximate the probability in (30) as

$$\vec{\pi}_b[n] \approx \frac{\text{Rice}(|e_b[n]|;\rho_b[n],\bar{\tau}_b[n])}{\frac{2|e_b[n]|}{c_b[n]} \exp\left(-\frac{|e_b[n]|^2}{c_b[n]}\right) + \text{Rice}(|e_b[n]|;\rho_b[n],\bar{\tau}_b[n])}. \tag{32}$$

The derivation of (32) can be found in Section B.

Next, we compute $\nu_{\mathsf{s}[n]\to f_b[n]}$ for $b>1$. By the sum-product rule, the probability associated with this Bernoulli PMF is

$$\overleftarrow{\pi}_b[n] = \frac{\lambda \prod_{\ell=1}^{b-1} \vec{\pi}_\ell[n]}{(1-\lambda)\prod_{\ell=1}^{b-1}(1-\vec{\pi}_\ell[n]) + \lambda \prod_{\ell=1}^{b-1} \vec{\pi}_\ell[n]}, \tag{33}$$

where the probability derived in (32) is used for $\vec{\pi}_\ell[n]$.

### D. Messages With the Phase Variables

: We now derive $\nu_{f_b[n]\to\Theta_b[n]}$ and $\nu_{\Theta_b[n]\to q_{b+1}[n]}$. To obtain a tractable approximation for these messages, we assume that the phase of the intra-plane AMP estimate of $z_b[n]$ is free from errors. Under this assumption, the intra-plane estimate of $z_b[n]$ has a phase equal to that of its mean in (19), i.e., $\angle e_b[n]$. Specifically,

$$\nu_{f_b[n]\to\Theta_b[n]} = \delta\left(\theta_b[n] - \angle e_b[n]\right). \tag{34}$$

Due to the Dirac-delta PDF in (34), the inter-plane message $\nu_{\Theta_b[n]\to q_{b+1}[n]} = \nu_{f_b[n]\to\Theta_b[n]}\nu_{q_b[n]\to\Theta_b[n]}$ is also a Dirac impulse, i.e., $\nu_{\Theta_b[n]\to q_{b+1}[n]} \propto \delta(\theta_b[n] - \angle e_b[n])$. The message $\nu_{q_{b+1}[n]\to\Theta_{b+1}[n]}$ forwarded to the $b+1^{\text{th}}$ plane is then

$$\nu_{q_{b+1}[n]\to\Theta_{b+1}[n]} = \int_{\theta_b[n]} q_{b+1}[n]\nu_{\Theta_b[n]\to q_{b+1}[n]}$$

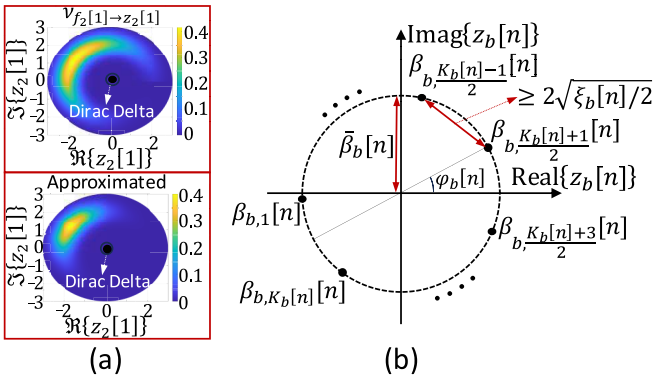



Fig. 5. In the forward pass, solving for auxiliary vector entries up to plane $b$ provides side information (belief) on the entry in plane $b$. We approximate this belief, denoted by $\nu_{f_b[n]\to z_b[n]}$, as a Bernoulli-Gaussian mixture, both of which are shown in (a). We assume that the Gaussian mixture components have the same variance $\xi_b[n]$, and that their means have the same magnitude denoted by the dots in (b).

$$\propto \mathcal{N}\left(\theta_{b+1}[n];\varphi_{b+1}[n],\sigma^2_{\text{pn}}\right). \tag{35}$$

where $\varphi_{b+1}[n] = \angle e_b[n]$. For $b>1$, we observe from Fig. 4 that the variable $\Theta_b[n]$ connects the factors $q_b[n]$ and $f_b[n]$ in the forward pass. Therefore, $\nu_{\Theta_b[n]\to f_b[n]} = \nu_{q_b[n]\to\Theta_b[n]}$, i.e.,

$$\nu_{\Theta_b[n]\to f_b[n]} \propto \mathcal{N}(\theta_b[n];\varphi_b[n],\sigma^2_{\text{pn}}). \tag{36}$$

### E. Into-Plane Message

$\nu_{f_b[n]\to \mathsf{z}_b[n]}$: For the first plane $b=1$, the message $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ is simply the BG prior in (18) which enforces sparsity. The belief on $\mathsf{z}_b[n]$ in the subsequent planes ($b\geq 2$) could incorporate more information than a simple BG prior. This additional information is due to the statistics of the support, magnitude, and phase learned from all preceding planes in the forward pass.

The message $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ is computed by first evaluating the product $f_b[n]\nu_{\Theta_b[n]\to f_b[n]}\nu_{s[n]\to f_b[n]}\nu_{\mathsf{r}[n]\to f_b[n]}$, where the individual terms are listed in (36), (33) and (27). Then, the product is integrated over $\mathsf{r}[n]$, $\mathsf{s}[n]$, and $\Theta_b[n]$ to arrive at $\nu_{f_b[n]\to \mathsf{z}_b[n]}$. As this integration is challenging, we approximate it as a Bernoulli-Gaussian Mixture (BGM) distribution. Such an assumption aids computationally tractable messages for subsequent planes. An example of the true $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ and our BGM approximation is shown in Fig. 5(a). This belief, originating from the preceding plane, comprises non-zero probabilities concentrated on an annular segment as illustrated in Fig. 5(a). The radius of this segment is governed by the magnitude prior on $\mathsf{z}_2[1]$, while its angular spread is determined by the uncertainty in $\mathsf{z}_1[1]$ as well as the variance of the incremental phase error due to the Wiener process.

Our BGM approximation of $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ uniformly spaces the Gaussians on the annular segment shown in Fig. 5(a). We assume that the means of the complex Gaussians in our BGM are of the form

$$\beta_{b,k}[n] = \bar{\beta}_b[n] e^{\mathrm{j}\angle\beta_{b,k}[n]}, \tag{37}$$

to achieve a non-zero probability density over the annular segment. In (37), we use $\bar{\beta}_b[n]$ to denote the magnitude of

$\beta_{b,k}[n]$. The magnitude of means, i.e., $\bar{\beta}_b[n]$, is the same for all Gaussian components in our BGM. We set

$$\bar{\beta}_b[n] = \text{mean}\left(\nu_{\mathsf{r}[n]\to f_b[n]}\right), \tag{38}$$

and its expression is derived in (61). We assume the same variance of $\xi_b[n]$ for all Gaussian components in our BGM. We use $K_b[n]$ to denote the number of complex Gaussians in our BGM approximation and assume that $K_b[n]$ is odd. The weight of the $k^{\text{th}}$ GM component is denoted by $\omega_{b,k}[n] \in [0,1]$, with $\sum_{k=1}^{K_b[n]} \omega_{b,k}[n] = 1$. Our BGM approximation takes the form

$$\nu_{f_b[n]\to \mathsf{z}_b[n]} \approx (1 - \overleftarrow{\pi}_b[n])\delta(z_b[n]) + \overleftarrow{\pi}_b[n] \sum_{k=1}^{K_b[n]} \omega_{b,k}[n]\mathcal{CN}(z_b[n]; \beta_{b,k}[n], \xi_b[n]). \tag{39}$$

The parameters in our BGM, i.e., the $\omega_{b,k}[n]$s, $K_b[n]$, and $\xi_b[n]$ are learned from the true $\nu_{f_b[n]\to \mathsf{z}_b[n]}$.

Next, we observe that the Gaussian mixture in (39) approximates

$$\left[\nu_{f_b[n]\to \mathsf{z}_b[n]}\right]_{\mathsf{s}[n]\neq 0} = \overleftarrow{\pi}_b[n] \iint_{r[n]\theta_b[n]} \delta(z_b[n] - r[n]e^{\mathrm{j}\theta_b[n]})\nu_{\mathsf{r}[n]\to f_b[n]} \times \nu_{\Theta_b[n]\to f_b[n]}. \tag{40}$$

From (40), (27) and (36), it can be seen that $\left[\nu_{f_b[n]\to \mathsf{z}_b[n]}\right]_{\mathsf{s}[n]\neq 0}$ is exactly zero at $z_b[n] = 0$. To make sure that the mixture in (39) is nearly zero at $z_b[n] = 0$, the variance $\xi_b[n]$ is chosen such that the means, i.e., $\beta_{b,k}[n]$s, are at least $2\sqrt{\xi_b[n]/2}$ away from $z_b[n] = 0$. Also, the variance $\xi_b[n]$ should be large enough to encompass the annular region with a limited number of Gaussian mixture components, with this limit imposed for computational tractability. We use

$$\xi_b[n] = 2\min\left\{\left(\frac{\text{mean}\left(\nu_{\mathsf{r}[n]\to f_b[n]}\right)}{2}\right)^2, \text{var}\left(\nu_{\mathsf{r}[n]\to f_b[n]}\right)\right\}, \tag{41}$$

where $\text{var}\left(\nu_{\mathsf{r}[n]\to f_b[n]}\right)$, derived in (62), is the variance of the true belief $\nu_{\mathsf{r}[n]\to f_b[n]}$. When $\xi_b[n]$ from (41) is small, a substantial number of GM components may be needed to encompass the annular region. To avoid the high computational complexity in such a case, we use $\max\{\xi_b[n], \Xi_b[n]\}$ as the variance of the GM components in our approximation. Here, $\Xi_b[n]$ is chosen to make sure that $K_b[n]$ is smaller than 60.

The phase values of $\beta_{b,k}[n]$s are symmetrically placed around $\beta_{b,(K_b[n]+1)/2}[n]$, which is set to the mean of $\nu_{\Theta_b[n]\to \mathsf{z}_b[n]}$, i.e., $\angle\beta_{b,(K_b[n]+1)/2}[n] = \varphi_b[n]$. When $\sigma^2_{\text{pn}} \leq \xi_b[n]$, we use just $K_b[n] = 1$ complex Gaussian component to cover the annular region. When $\sigma^2_{\text{pn}} > \xi_b[n]$, the number of GM components $K_b[n]$ are chosen such that the Euclidean distance between any two neighboring $\beta_{b,k}[n]s$ is at least $2\sqrt{\xi_b[n]/2}$ as shown in Fig. 5(b). To satisfy this condition, we set

$$K_b[n] = 2\left\lfloor \frac{\pi - \sin^{-1}\left(\frac{\sqrt{\xi_b[n]/2}}{\bar{\beta}_b[n]}\right)}{2\sin^{-1}\left(\frac{\sqrt{\xi_b[n]/2}}{\bar{\beta}_b[n]}\right)} \right\rfloor + 1 \text{ and,} \tag{42}$$

$$\angle\beta_{b,k}[n] = \varphi_b[n] + \left(\frac{K_b[n]+1}{2} - k\right) \frac{2\pi - 2\sin^{-1}\left(\frac{\sqrt{\xi_b[n]/2}}{\bar{\beta}_b[n]}\right)}{K_b[n]-1}. \tag{43}$$

The derivations of (42) and (43) are provided in C.

Now, we explain how we determine the weights $\omega_{b,k}[n]$s that scale the GM components in (39). As seen from Fig. 5(a), the probability mass in the annular region tapers off around the central angle, due to Gaussian statistics of the phase error innovations. As a result, we assign the largest weight to the $(K_b[n]+1)/2$-th GM component with mean $\bar{\beta}_b[n]e^{\mathrm{j}\varphi_b[n]}$ and small weight to the components whose mean is farther away from $\beta_{b,(K_b[n]+1)/2}[n]$. We define $\{\tilde{\omega}_{b,k}[n]\}_{k=1}^{K_b[n]}$ as

$$\tilde{\omega}_{b,k}[n] = \mathcal{N}\left(\beta_{b,k}[n]; \beta_{b,(K_b[n]+1)/2}[n], \sigma^2_{\text{pn}}[n]\right), \tag{44}$$

to denote the unnormalized weights assigned to the GM components. The GM weights $\omega_{b,k}[n]$s are computed by normalizing the weights in (44) as

$$\omega_{b,k}[n] = \frac{\tilde{\omega}_{b,k}[n]}{\sum_{k=1}^{K_b[n]} \tilde{\omega}_{b,k}[n]}, \tag{45}$$

We notice from Fig. 5(a) that our approximated BGM belief in (39) well approximates the true $\nu_{f_b[n]\to \mathsf{z}_b[n]}$. We observed empirically that our message passing algorithm can successfully reconstruct the channel even with the approximated belief shown in Fig. 5(a).

After performing the "Intra-plane" step of plane $B$ in the first forward message flow, the backward message flow can be performed. To this end, the messages in the "Out-of-plane" step of plane $B$ are computed in the same way as in the forward message flow. These messages are then used to compute the messages for the "Inter-plane" and the "Into-plane" step of plane $B-1$. Next, the proposed GM approximation is used to obtain local priors for $\{\mathsf{z}_{B-1}[n]\}_{n=1}^{N^2}$ which are used to perform AMP iterations in the "Intra-plane" step. This procedure continues until it performs the "Intra-plane" step in plane 1 to obtain a belief about $\{\mathsf{z}_1[n]\}_{n=1}^{N^2}$. In the forward message flow, the messages related to plane $b$ are updated based on the information received from planes 1 up to plane $b-1$ while in the backward message flow the messages related to plane $b$ are updated based on the information received from planes $B$ down to plane $b+1$. Multiple rounds of forward and backward message flows can be used to refine the estimates $\{\hat{\mathsf{z}}_b\}_{b=1}^{B}$.

Now, we discuss the computational complexity of our developed message passing-based method outlined in Algorithm 2. As our algorithm performs AMP within each plane, and we limit the number of Gaussian mixture components similar to [21], the computational complexity of calculating parameters for the intra-plane messages, i.e., steps **1** and **7** in Algorithm 2, is dominated by matrix multiplication by $\mathbf{A}_b$ resulting in $\mathcal{O}\left(MN^2\right)$ complexity [21]. Furthermore, the parameters of the into-plane, out-of-plane, and inter-plane messages, i.e., steps **2–6** and **8–11**, are computed independently for each $n$ and do not depend on the number of measurements, resulting in an overall computational complexity of $\mathcal{O}\left(N^2\right)$. Therefore, assuming $T^{\text{amp}}$ AMP iterations per plane, the

**Algorithm 2** Channel Estimation With One Forward Pass**Input**: Parameters $\lambda$ and $\mu$ of the prior distribution (6) of $\mathbf{x}$, phase perturbed CS measurements $\{\hat{\mathbf{y}}_b\}_{b=1}^{B}$ (4), CS matrices $\{\mathbf{A}_b\}_{b=1}^{B}$, phase noise variance $\sigma_{\text{pn}}^2$ and the measurement noise $\sigma^2$. **While** $b \leq B$ **do**: **If** $b = 1$ **then**

1. Estimate $\hat{\mathbf{z}}_1[n]$ with the AMP [9] using prior in (18).
2. Determine $\vec{\pi}_1[n]$s assuming $\rho_b[n] = 0, \tau_b[n] = \mu/2$, $\forall n \in [N^2]$ using (32) for the first batch.
3. Find $\{t_1[n]\}_{n=1}^{N^2}$ using (23) and $\{v_1[n]\}_{n=1}^{N^2}$ using (24).
   **Else**
4. Calculate $\{\overleftarrow{\pi}_b[n]\}_{n=1}^{N^2}$ using (33).
5. Set $\varphi_b[n] = \angle e_{b-1}[n]$, $\forall n \in [N^2]$.
6. Compute parameters $\{\omega_{b,k}[n]\}_{k=1}^{K_b[n]}$ from (45), $\{\beta_{b,k}[n]\}_{k=1}^{K_b[n]}$, $\forall n \in [N^2]$ from (37), (38), (43), and $\{\xi_b[n]\}_{n=1}^{N^2}$ from (41) for the BGM approximation in (39).
7. Use learned BGM prior in (39) to estimate $\hat{\mathbf{z}}_b[n]$ with AMP.
8. Compute parameters $\{\rho_b[n]\}_{n=1}^{N^2}$ and $\{\tau_b[n]\}_{n=1}^{N^2}$ of the Rice distribution (31) fit to the magnitude-related messages in (27).
9. Find $\{\vec{\pi}_b[n]\}_{n=1}^{N^2}$ using (32).
10. Compute the parameters $\{\eta_b[n]\}_{n=1}^{N^2}$ and $\{\gamma_b[n]\}_{n=1}^{N^2}$ of the magnitude related message (27) going into the batch.
11. Compute $\{t_b[n]\}_{n=1}^{N^2}$ using (23) and $\{v_b[n]\}_{n=1}^{N^2}$ using (24).
    **End If**
12. $b \leftarrow b + 1$.
    **End While**
13. Determine $\hat{\mathbf{x}}$ using Algorithm 1 based on the estimates $\{\hat{\mathbf{z}}_b\}_{b=1}^{B}$ obtained from the above procedure.

**Output:** $\hat{\mathbf{x}}$.

overall complexity per plane is dominated by $\mathcal{O}\left(T^{\text{amp}}MN^2\right)$. Finally, for $B$ planes and $T^{\text{fb}}$ forward and backward message propagation, the overall computational complexity is $\mathcal{O}\left(T^{\text{fb}}BT^{\text{amp}}MN^2\right)$. The computational complexity of the standard AMP from [9] for sparse recovery using *BM* measurements is $\mathcal{O}\left(BT^{\text{amp}}MN^2\right)$. We observe that the complexity of our approach is higher than that of the AMP by a factor of $T^{\text{fb}}$. This is due to exploiting the Wiener statistics in the phase noise, which results in better channel reconstruction. The complexity of both methods, however, scales only linearly with the number of measurements $M$ and the array dimension $N^2$.

Due to the uniform prior $\mathcal{U}([-\pi, \pi])$ on the phase variable in plane 1, i.e., $\Theta_1$, the number of GM components used in the "Into-plane" step of plane 1 is higher than that of other planes. Fewer GM components are needed in the subsequent planes as the prior is concentrated (as shown in Fig. 5(a)). In Fig. 6, we investigate the impact of the number of GM components used to approximate $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ in (39). As can be seen, increasing the number of GM components in the approximation of $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ using (39) improves the NMSE of the estimate as it improves the approximation of $\nu_{f_b[n]\to \mathsf{z}_b[n]}$. It is observed, however, that the NMSE remains almost the same when the

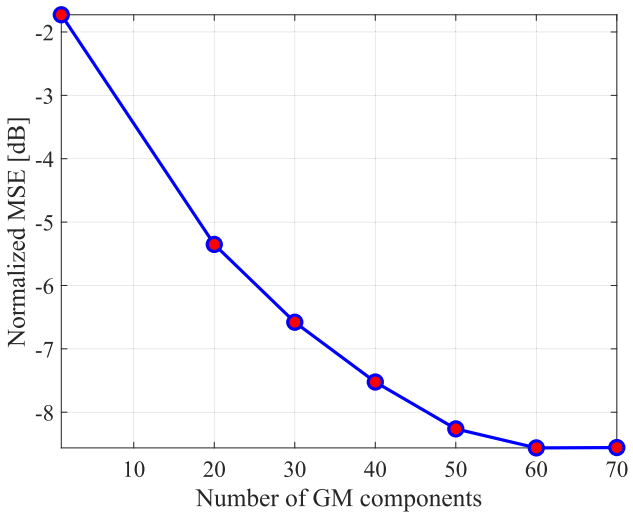



Fig. 6. NMSE with the number of Gaussian components used to approximate $\nu_{f_b[n]\to z_b[n]}$ in (39). Here, we use $M = 50$, $B = 4$, and SNR $= -15$ dB. We limit the maximum number of Gaussian components to 60 as the NMSE remains unchanged when more components are used".

number of GM components is greater than 60. We note that a large number of GM components in the approximation of $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ in (39) leads to a high computational complexity. To avoid this, we have limited the number of components in the GM approximation, which is 60 as per our study in Fig. 6.

An alternative to our GM approximation is to employ a variational inference approach [39], as done in [40], to approximate $\nu_{f_b[n]\to \mathsf{z}_b[n]}$ using tractable distributions. The approximation in variational inference is obtained by solving an optimization problem that reduces the discrepancy between the approximation and the exact distribution [39]. Incorporating this approach within our framework, however, is beyond the scope of this paper and we leave it for our future endeavors.

## V. Simulation Results

We consider a narrowband system with a $16 \times 16$ UPA at the TX for $f_{\text{c}} = 60$ GHz. The TX-RX distance is 15 m. Unless otherwise specified, phase noise at $f_{\text{m}} = 1$ MHz offset from $f_{\text{c}}$ is $\mathcal{L}(f_{\text{m}}) = -101.7$ dBc/Hz [28]. The duration to obtain each measurement is $\Delta_{\text{bf}} = 128$ ns and the spacing between successive batches is $\Delta_{\text{sf}} = 44\mu$s [15], which results in $\sigma_{\text{pn}}^2 \approx 0.12$ [16]. We use $B = 4$ batches. The beam training matrices $\{\mathbf{P}_b[m]\}_{m=1}^{M} \forall b \in \{1, \cdots, B\}$ applied at the UPA are random circular shifts of a $16 \times 16$ perfect binary array [29]. For each beam training matrix at the TX, spread sequences of duration 128 ns are transmitted within the training fields in Fig. 1(b) [29], resulting in a spreading gain of about 20 dB. We use 100 urban micro line-of-sight channels from the NYU simulator [41]. Therefore, the AoDs can be off-grid resulting in approximately sparse beamspace channels. This channel model is more realistic than the model in [15] which assumes on-grid AoDs and known sparsity. Finally, Wiener phase noise is introduced considering time durations of $\Delta_{\text{sf}}$ and $\Delta_{\text{bf}}$. This induces phase errors in the CS measurements acquired both within and across batches. Although our algorithm models only one phase error per batch (see (5)), the proposed method

works even when the measurements within a batch are affected by varying phase errors.

The channel dataset is used to calculate the prior parameters $\lambda$ and $\mu$ in (6). To compute the sparsity rate $\lambda$, we first find the strongest entries in the angle domain channel that account for 95 percent of the energy in the channel. Denoting this number by $K$, we then calculate $\lambda$ as the mean of $K/N^2$ over all channels within the dataset. Next, the sample variance of the $K$ strongest entries from all the channels is used to compute $\mu$. This resulted in $\lambda \approx 0.07$ and $\mu \approx 13.3$ for our dataset. An expectation-maximization-based approach can be used in practice to learn these parameters; however, this extension is beyond the scope of our paper.

We use $\hat{\mathbf{H}}$ to denote the channel estimate, which is derived from the angle domain estimate $\hat{\mathbf{X}}$ as $\hat{\mathbf{H}} = \mathbf{U}_N \hat{\mathbf{X}} \mathbf{U}_N$. The normalized mean squared error (NMSE) is defined as $\mathbb{E}[\|\mathbf{H}-\hat{\mathbf{H}}\|_{\mathrm{F}}^2/\|\mathbf{H}\|_{\mathrm{F}}^2]$. The signal-to-noise ratio (SNR) is given by $\mathrm{SNR} = \sum_{b=1}^{B} \|\mathbf{A}_b\mathbf{x}\|^2/MB\sigma^2$. To avoid numerical issues at high SNRs, we use only one Gaussian mixture component, i.e., the one with the largest weight, to approximate $\nu_{f_b[n]\to \mathsf{z}_b[n]}$. We use $T_{\max} = 5$ in Algorithm 1 and $\epsilon = 10^{-6}$ in (21). To further limit the number of GM components, we remove the GM components whose corresponding non-normalized weight is $10^{-3}$ times smaller than the maximum non-normalized weight $\tilde{\omega}_{b,(K_b[n]+1)/2}[n]$ in (44).

We benchmark our method against our *SuppOnly* method [22], which solves (5) by exploiting only the common support across $\{\mathsf{z}_b\}_{b=1}^B$. Unlike [22], the technique proposed in this paper incorporates Wiener phase noise statistics and the common magnitude structure across $\{\mathsf{z}_b\}_{b=1}^B$, in addition to leveraging the support structure. We also benchmark against partially coherent sparse recovery based on *SparseLift* [4]. The application of *SparseLift* to solve (5) was discussed in [15], wherein the lifted matrix $\mathbf{\Gamma} = e^{\mathrm{j}\boldsymbol{\phi}}\mathbf{x}^{\mathrm{T}}$ is first estimated. For a fair comparison, we use AMP [9] to solve for $\mathbf{\Gamma}$. Then, the singular-value decomposition of the estimate $\hat{\mathbf{\Gamma}}$ is used to determine $\hat{\mathbf{x}}$ [15]. We also use the partially coherent matching pursuit (PCMP) algorithm from [15] to benchmark our method. PCMP is a greedy algorithm that iteratively detects the support of $\mathbf{x}$, similar to the matching pursuit algorithm. Then, the phase error and the coefficient corresponding to the detected support are estimated using an alternating optimization. PCMP requires knowledge of the sparsity level for $\mathbf{x}$. Therefore, we use PCMP algorithms with two different sparsity levels as benchmarks, denoted by PCMP1 and PCMP2. PCMP1 uses a sparsity level of $\lceil \lambda N^2 \rceil = 19$ and PCMP2 uses a sparsity level of $2\lceil \lambda N^2 \rceil = 38$ with $\lceil \cdot \rceil$ denoting the ceil operator. We also use a baseline, called *IgnoredPN*, that just stacks the CS matrices across all batches and employs standard AMP over the stacked phase perturbed measurements [9]. Finally, we compare our algorithm against ideal case, called *ZeroPN*, which uses phase error-free measurements for sparse recovery with the stacked CS matrices.

From Fig. 7, we observe that the NMSE with the proposed method and the method from [22] decrease until two forward message flows. This is because, to estimate $\mathsf{z}_b$ in the first message flow, only measurements until the current batch, i.e., $\{\mathbf{y}_\ell\}_{\ell=1}^b$, are used. By the end of the second forward message flow, two forward and one backward message flows are completed, ensuring that measurements from all the $B$ batches are used for estimation. Additional message flows beyond this point did not reduce the estimation error.

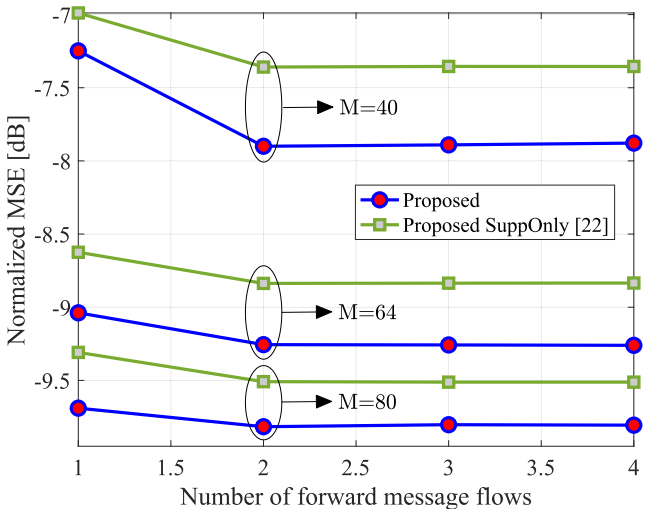



Fig. 7. NMSE with the number of forward passes for $B = 4$, SNR $= -15$ dB. NMSE stabilizes beyond the second pass, as measurements from all batches are used starting with the second pass.

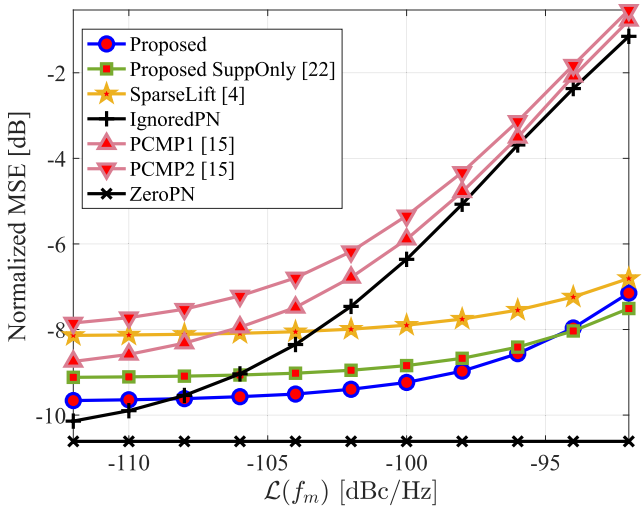



Fig. 8. NMSE with phase noise power spectral density $\mathcal{L}(f_{\mathrm{m}})$ for $B = 4$, $M = 64$ and SNR $= -15$ dB. Our method outperforms SparseLift [4] and SuppOnly [22] methods for $\mathcal{L}(f_{\mathrm{m}}) < -96$ dBc/Hz.

We notice from Fig. 8 that our method results in a lower NMSE than the benchmarks when $\mathcal{L}(f_{\mathrm{m}})$, the phase noise power spectral density, is smaller than $-96$ dBc/Hz. Note that *SparseLift* does not incorporate Wiener phase noise statistics. At high $\mathcal{L}(f_{\mathrm{m}})$, however, the phase errors within a batch become large enough to perturb the auxiliary estimates in each plane. We observe that our *proposed* method is severely impacted by these errors when compared to the *proposed SuppOnly* method. This is because our *proposed* method propagates not only the beliefs about support, but also beliefs about both magnitude and phase to subsequent planes, which are inaccurate under strong phase noise perturbations, i.e., $\mathcal{L}(f_{\mathrm{m}}) > -96$ dBc/Hz. These inaccuracies arise because our

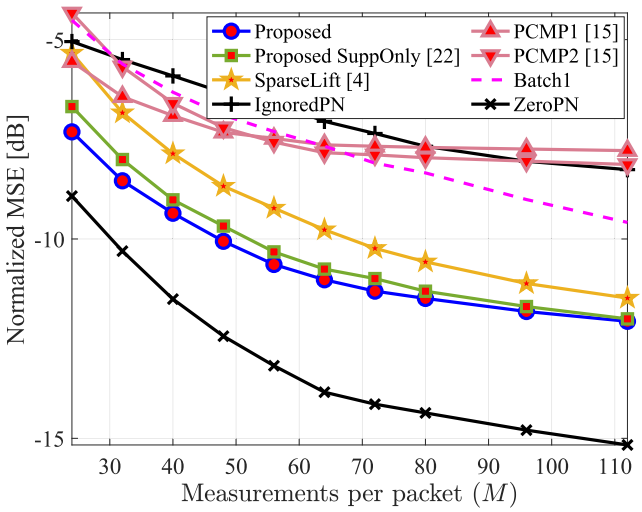
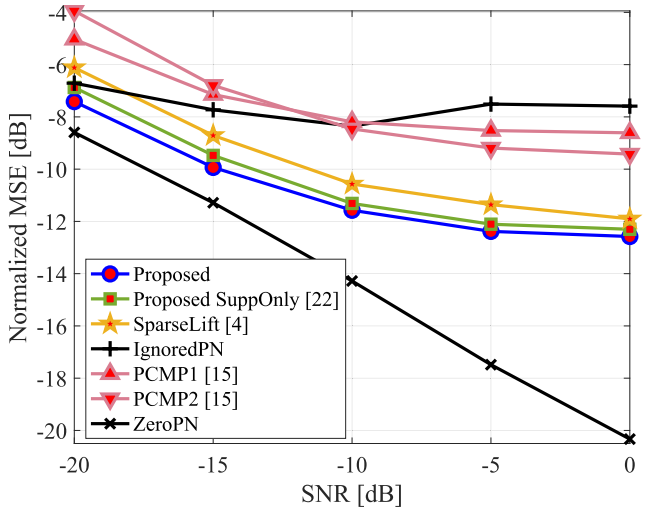
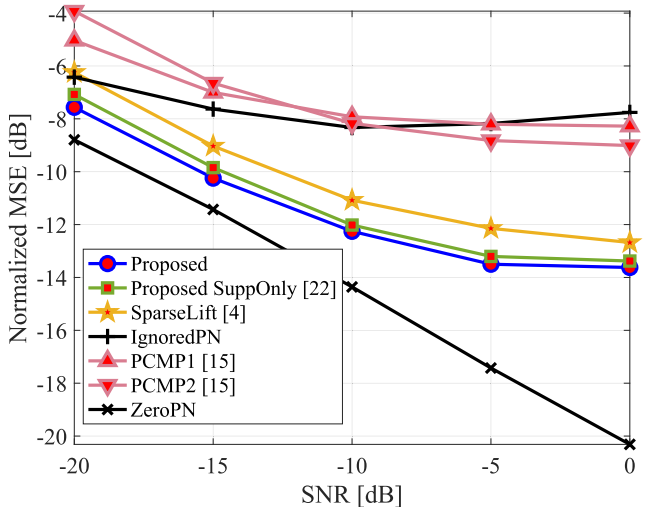



Fig. 9. NMSE with $M$ for $B = 4$, SNR $= -10$ dB. The proposed method results in a smaller NMSE than benchmarks with phase-perturbed measurements.




(a) Actual phase noise (varies within and across batches).




(b) Phase noise consistent with our model in (5).

Fig. 10. NMSE with SNR for $B = 4$, $M = 80$. Our method results in a smaller NMSE than the benchmarks for the phase noise in (a) and (b). In (a), the measurements are corrupted by the actual phase noise, which varies both within and across batches. In (b), consistent with our model in (5), the phase noise is constant over each batch, and it varies across batches. Approximating the actual phase noise with a constant phase noise within each batch results in a loss of about 0.9 dB in the NMSE at high SNR.

model assumes no phase noise variation within a batch, which is invalid at high $\mathcal{L}(f_{\rm m})$. These distorted phase and magnitude beliefs are propagated across the planes, ultimately leading to a poorer channel estimate with the *proposed* method than with the *proposed SuppOnly* method. This means that for low-quality oscillators with $\mathcal{L}(f_{\rm m}) > -96$ dBc/Hz, our *SuppOnly* method is suitable, while the *proposed* method is suitable when $\mathcal{L}(f_{\rm m}) < -96$ dBc/Hz. The derived method in Algorithm 2 is still valuable as the oscillators in [28], [42], and [43] have $\mathcal{L}(f_{\rm m}) < -96$ dBc/Hz. Finally, we observe that the *IgnoredPN* method that just stacks all the measurements performs better than our method when $\mathcal{L}(f_{\rm m}) < -108$ dBc/Hz. At 60 GHz, the oscillator in [28], [42], and [43] generates a phase noise between $-99.4$ dBc/Hz and $-104.6$ dBc/Hz, for which our method provides the lowest NMSE. Note that the range of $\mathcal{L}(f_{\rm m})$ for which our method performs best varies with the choice of $\Delta_{\rm sf}$ and $\Delta_{\rm bf}$.

We observe from Fig. 9 that the proposed method achieves about 0.6 dB improvement in the NMSE compared to our *SuppOnly* method. This is because our factor graph in Fig. 3 exploits additional structure in the auxiliary vectors, i.e., magnitude and phase correlation, while the *SuppOnly* method exploits only the common support structure. The gap between our proposed technique and the *SuppOnly*, however, reduces at a large $M$ as the beliefs enforcing common support structure have a stronger impact than those incorporating magnitude and phase correlation. Finally, we observe that there is almost no improvement in the NMSE with our method over *SuppOnly*, when $M \geq 95$. This is because our assumption that the phase noise is almost the same over a batch breaks down for a long batch, equivalently a large $M$. In Fig. 9, we also included *Batch1* method wherein standard AMP from [9] is used to estimate the beamspace using only measurements from batch 1. Specifically, the *Batch1* method solves for the sparse $\mathbf{x}$ from $\mathbf{y}_1 = \mathbf{A}_1\mathbf{x} + \mathbf{w}_1$ and ignores all the other measurements. Simply stacking the other measurements (*IgnoredPN*) does not work well compared to the *Batch1* method due to inter-batch phase errors. Our *proposed* method leverages useful information from other batches even under these phase errors, resulting in a lower NMSE than the *Batch1* method.

Fig. 10 shows that our technique achieves a lower NMSE than the *SuppOnly*, especially at a low SNR, as it exploits additional structure. As SNR increases, NMSE reduces with the proposed method and the *SuppOnly* method. The mismatch between the batch-based phase noise model considered in the measurement model (5) in our algorithm and the realistic model in our simulations results in the NMSE floor in Fig. 10(a). The error floor due to phase noise is also evident from the observation that the NMSE with *IgnoredPN* does not improve at high SNR. By comparing Fig. 10(b) and Fig. 10(a), we observe that approximating the actual phase noise with a constant phase noise within each batch results in a loss of

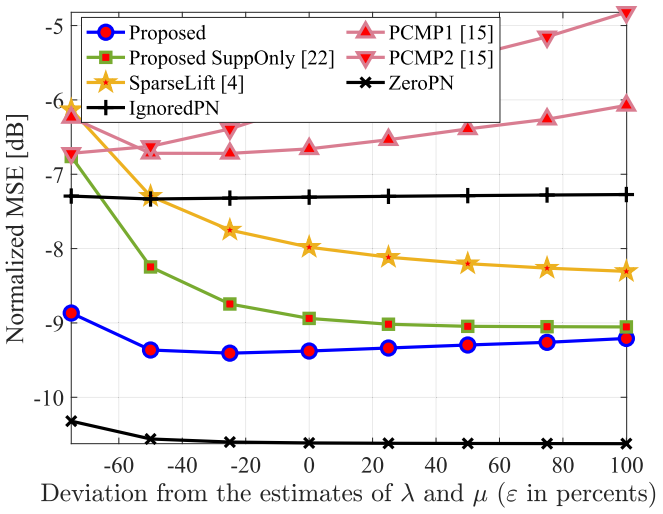



Fig. 11. NMSE with the deviation from the estimates of the prior parameters $\lambda$ and $\mu$ in percentages. Here, we use $M = 64$ for $B = 4$, SNR $= -15$ dB. Our proposed method is robust to inaccuracies in the prior parameters $\lambda$ and $\mu$, when compared to the benchmarks.

about 0.9 dB in the NMSE at high SNR. It is also seen that the NMSE saturates beyond a certain SNR, even with the same phase noise within each batch. The NMSE saturation is due to the use of the relaxed version of the true phase model in (13), given by (14), to construct our factor graph. Finally, we observe that PCMP1 and PCMP2 do not perform well compared to the proposed, the *SuppOnly* and the *SparseLift* methods. This is because PCMP is designed for exactly sparse channels with known sparsity levels, while the channels in our simulations are approximately sparse.

In Fig. 11, we investigate the robustness of our method to inaccuracies in prior parameters estimation. To this end, we add perturbations to the estimated priors, i.e., $\lambda$ and $\gamma$, as $(1+\varepsilon)\lambda$ and $(1+\varepsilon)\mu$ with $\varepsilon \in [-1, 1]$. As it is observed from Fig. 11, our *proposed* approach is more robust to mismatch in the estimated prior parameters than the benchmarks.

## VI. Potential Research Directions to Address Challenges With the Proposed Method

In this paper, a total of $BM$ spatial channel measurements are acquired using $B$ training batches with $M$ measurements per batch for channel estimation. The use of a large number of training batches results in a low spectral efficiency due to a high training overhead and a high computational complexity. One way to address this challenge is to use a single batch, i.e., $B = 1$. With only a single batch, however, as seen from Fig. 9, the channel estimation error is high due to the limited number of measurements acquired using a single batch. To reduce the training overhead while achieving a low estimation error, we suggest progressively decreasing the number of acquired measurements with the batch index $b$. The use of high measurements initially provides a good channel estimate. This estimate may be further refined using fewer phase-mismatched measurements. In our framework, this is equivalent to having CS matrices $\{\mathbf{A}_b\}_{b=1}^{B}$ with a decreasing number of rows as the batch index $b$ increases to $B$. Optimizing the number of rows across the $B$ batches is beyond the scope of this paper. In our work, we assume that the same number of measurements is acquired in any batch, although our method can also be applied to a more general setting with different numbers of measurements per batch.

In this work, we assume that the channel remains the same during the $B$ training batch transmissions. A potential future research direction is to extend the proposed method to the problem of channel tracking [44], [45], where the channel also changes across batches. In this case, the beliefs about the magnitude, support, and phase of the tracked channel from the previous batch can be used as side information to estimate the channel in the current batch. The channel tracking problem, however, is beyond the scope of this paper.

## VII. Conclusion

In this paper, we developed a message-passing-based technique that enables sparse channel estimation under partially coherent Wiener phase noise. Our approach absorbs the phase errors into the sparse channel to define a collection of phase-perturbed auxiliary sparse vectors. The proposed method operates in two stages. The first stage estimates the auxiliary vectors by exploiting their common support and magnitude structure, Wiener phase noise statistics, and sparsity. To this end, we developed a message passing-based approach to leverage these properties at inference. The second stage in our method uses alternating optimization to reconstruct the sparse channel from the auxiliary vector estimates. Numerical results demonstrate that our method achieves a lower normalized mean-squared error than comparable benchmarks.

## Appendix

### A. Derivation of $\bar{\nu}_{f_b[n]\to r[n]}$ and Its Mean and Variance

We first provide a closed form approximation for $\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$ assuming $\nu_{\Theta_b[n]\to f_b[n]} = 1/2\pi$, which is exact for $b = 1$. Then, we compute the mean and variance of $\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$ using this approximation. To this end, we replace the complex Gaussian PDFs in (22) with their explicit form to obtain

$$\begin{aligned}
&\bar{\nu}_{f_b[n]\to \mathsf{r}[n]} \\
&= \frac{\epsilon^2\left(1-\Omega(\overleftarrow{\pi}_b[n])\right)}{\pi c_b[n]} \exp\left(-\frac{|r[n]|^2+\epsilon^{-2}|e_b[n]|^2}{\epsilon^{-2}c_b[n]}\right) \\
&\quad \frac{1}{2\pi}\int_{\theta_b[n]} \exp\left(\Re\left(\frac{2r[n]e_b[n]}{\epsilon^{-1}c_b[n]}e^{-\mathrm{j}\theta_b[n]}\right)\right) \\
&\quad + \frac{\Omega(\overleftarrow{\pi}_b[n])}{\pi c_b[n]} \exp\left(-\frac{|r[n]|^2+\epsilon^{-2}|e_b[n]|^2}{\epsilon^{-2}c_b[n]}\right) \\
&\quad \frac{1}{2\pi}\int_{\theta_b[n]} \exp\left(\Re\left(\frac{2r[n]e_b[n]}{c_b[n]}e^{-\mathrm{j}\theta_b[n]}\right)\right). \qquad (46)
\end{aligned}$$

To compute the above integral, we use the identity

$$\int_{-\pi}^{\pi} \exp\left(\Re\left(\kappa e^{-\mathrm{j}\theta}\right)\right) d\theta = 2\pi \mathrm{I}_0(|\kappa|), \qquad (47)$$

where $\mathrm{I}_n(\cdot)$ is the modified Bessel function of the first kind of $n^{\text{th}}$ order. The identity in (47) holds for any constant $\kappa \in \mathbb{C}$ [46], with $\Re(\kappa)$ denoting the real part of $\kappa$. This results in

$\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$

$$\approx \frac{1-\Omega\left(\overleftarrow{\pi}_b[n]\right)}{\pi\epsilon^{-2}c_b[n]}\exp\left(-\frac{r^2[n]+\epsilon^{-2}|e_b[n]|^2}{\epsilon^{-2}c_b[n]}\right)\mathrm{I}_0\left(\frac{2\,|e_b[n]|}{\epsilon^{-1}c_b[n]}r[n]\right) + \frac{\Omega\left(\overleftarrow{\pi}_b[n]\right)}{\pi c_b[n]}\exp\left(-\frac{r^2[n]+|e_b[n]|^2}{c_b[n]}\right)\mathrm{I}_0\left(\frac{2\,|e_b[n]|}{c_b[n]}r[n]\right). \tag{48}$$

We use $t_b[n]$ to denote the mean and $v_b[n]$ to denote the variance of $\bar{\nu}_{f_b[n]\to \mathsf{r}[n]}$ in (48), which are given by

$$t_b[n] = \frac{\int_0^\infty r[n]\bar{\nu}_{f_b[n]\to\mathsf{r}[n]}dr[n]}{\int_0^\infty \bar{\nu}_{f_b[n]\to\mathsf{r}[n]}dr[n]}, \tag{49}$$

$$v_b[n] = \frac{\int_0^\infty r^2[n]\bar{\nu}_{f_b[n]\to\mathsf{r}[n]}dr[n]}{\int_0^\infty \bar{\nu}_{f_b[n]\to\mathsf{r}[n]}dr[n]} - t_b^2[n]. \tag{50}$$

To derive closed-form expressions for $t_b[n]$ and $v_b[n]$ in (49) and (50), we define $\varrho_b[n] = |e_b[n]|^2/2c_b[n]$ and use [47, Equation 10.43.24] and (48) to obtain

$$\int_0^\infty \bar{\nu}_{f_b[n]\to\mathsf{r}[n]}dr_b[n] \approx \frac{\epsilon\left(1-\Omega\left(\overleftarrow{\pi}_b[n]\right)\right)+\Omega\left(\overleftarrow{\pi}_b[n]\right)}{2\sqrt{\pi c_b[n]}}\exp(-\varrho_b[n])\,\mathrm{I}_0(\varrho_b[n])\,. \tag{51}$$

We simpify the numerator of $t_b[n]$ in (49) using the Rice distribution [37] defined in (31) and (48) as

$$r_b[n]\bar{\nu}_{f_b[n]\to\mathsf{r}[n]} \approx \frac{1-\Omega\left(\overleftarrow{\pi}_b[n]\right)}{2\pi}\mathrm{Rice}\left(r[n];\epsilon^{-1}|e_b[n]|\,,\frac{1}{2}\epsilon^{-2}c_b[n]\right) + \frac{\Omega\left(\overleftarrow{\pi}_b[n]\right)}{2\pi}\mathrm{Rice}\left(r[n];|e_b[n]|\,,\frac{1}{2}c_b[n]\right). \tag{52}$$

By using (52), the numerator of $t_b[n]$ can be expressed as

$$\int_0^\infty r[n]\bar{\nu}_{f_b[n]\to\mathsf{r}[n]}dr[n] \approx \frac{1}{2\pi}. \tag{53}$$

By substituting (53) and (51) in (49), we obtain the closed-form expression (23) for $t_b[n]$.

Next, we use (52) to rewrite the numerator in (50) as

$$\int_0^\infty r_b^2[n]\bar{\nu}_{f_b[n]\to\mathsf{r}[n]}dr[n] \approx \frac{1-\Omega\left(\overleftarrow{\pi}_b[n]\right)}{2\pi}\int_0^\infty r[n]\mathrm{Rice}\left(r[n];\epsilon^{-1}|e_b[n]|\,,\frac{1}{2}\epsilon^{-2}c_b[n]\right)dr[n] + \frac{\Omega\left(\overleftarrow{\pi}_b[n]\right)}{2\pi}\int_0^\infty r[n]\mathrm{Rice}\left(r[n];|e_b[n]|\,,\frac{1}{2}c_b[n]\right)dr[n]. \tag{54}$$

As the integrands in (54) are the Rice distribution means,

$$\int_0^\infty r_b^2[n]\bar{\nu}_{f_b[n]\to\mathsf{r}[n]}dr[n] \approx \frac{\epsilon^{-1}\left(1-\Omega\left(\overleftarrow{\pi}_b[n]\right)\right)+\Omega\left(\overleftarrow{\pi}_b[n]\right)}{4}\sqrt{\frac{c_b[n]}{\pi}}\mathrm{L}_{1/2}\left(-\frac{|e_b[n]|^2}{c_b[n]}\right). \tag{55}$$

where $\mathrm{L}_{1/2}(\cdot)$ denotes a Laguerre polynomial given by

$$\mathrm{L}_{1/2}(\alpha) = \exp\left(\frac{\alpha}{2}\right)\left[(1-\alpha)\mathrm{I}_0\left(-\frac{\alpha}{2}\right) - \alpha\mathrm{I}_1\left(-\frac{\alpha}{2}\right)\right]. \tag{56}$$

By using (56), (55) and (51) in (50), we obtain the closed-form expression (24) for $v_b[n]$.

### B. Derivation of $\vec{\pi}_b[n]$

We first compute the mean and the variance of $\nu_{\mathsf{r}[n]\to f_b[n]}$, which are then used to obtain $\rho_b[n]$ and $\tau_b[n]$ of the Rice distribution in (31). We define $\varsigma_b[n] = \eta_b[n]/\sqrt{2\gamma_b[n]}$, use $\iota_b[n]$ to denote the mean of $\nu_{\mathsf{r}[n]\to f_b[n]}$ and $\varkappa_b[n]$ to denote its variance. We have

$$\iota_b[n] = \frac{\int_0^\infty r[n]\nu_{\mathsf{r}[n]\to f_b[n]}}{\int_0^\infty \nu_{\mathsf{r}[n]\to f_b[n]}}, \quad \varkappa_b[n] = \frac{\int_0^\infty r^2[n]\nu_{\mathsf{r}[n]\to f_b[n]}}{\int_0^\infty \nu_{\mathsf{r}[n]\to f_b[n]}} - \iota_b^2[n]. \tag{57}$$

To derive closed-form expressions for (57), we denote the error function as $\mathrm{erf}(a) = (2/\sqrt{\pi})\int_0^a \exp\left(-t^2\right)dt$. Now, by substituting $\nu_{\mathsf{r}[n]\to f_b[n]}$ with its approximation from (27) and applying integral by parts, we obtain

$$\int_0^\infty \nu_{\mathsf{r}[n]\to f_b[n]}dr[n] \overset{(27)}{\approx} \frac{\eta_b[n]}{2}\left[1+\mathrm{erf}\left(\frac{\eta_b[n]}{\sqrt{2\gamma_b[n]}}\right)\right] + \sqrt{\frac{\gamma_b[n]}{2\pi}}\exp\left(-\frac{\eta_b^2[n]}{2\gamma_b[n]}\right), \tag{58}$$

$$\int_0^\infty r[n]\nu_{\mathsf{r}[n]\to f_b[n]}dr[n] \overset{(27)}{\approx} \frac{\eta_b^2[n]+\gamma_b[n]}{2}\left[1+\mathrm{erf}\left(\frac{\eta_b[n]}{\sqrt{2\gamma_b[n]}}\right)\right] + \eta_b[n]\sqrt{\frac{\gamma_b[n]}{2\pi}}\exp\left(-\frac{\eta_b^2[n]}{2\gamma_b[n]}\right), \tag{59}$$

$$\int_0^\infty r^2[n]\nu_{\mathsf{r}[n]\to f_b[n]}dr[n] \overset{(27)}{\approx} \frac{\eta_b[n](\eta_b^2[n]+3\gamma_b[n])}{2}\left[1+\mathrm{erf}\left(\frac{\eta_b[n]}{\sqrt{2\gamma_b[n]}}\right)\right] + (7\eta_b^2[n]+2\gamma_b[n])\sqrt{\frac{\gamma_b[n]}{2\pi}}\exp\left(-\frac{\eta_b^2[n]}{2\gamma_b[n]}\right). \tag{60}$$

Substituting (58), (59), and (60) in (57), we have

$$\iota_b[n] = \frac{\int_0^\infty r[n]\nu_{\mathsf{r}[n]\to f_b[n]}}{\int_0^\infty \nu_{\mathsf{r}[n]\to f_b[n]}} \overset{(27)}{\approx} \eta_b[n] + \frac{\gamma_b[n]\left(1+\mathrm{erf}\left(\varsigma_b[n]\right)\right)}{\eta_b[n]\left(1+\mathrm{erf}\left(\varsigma_b[n]\right)\right)+\sqrt{\frac{2\gamma_b[n]}{\pi}}\exp\left(-\varsigma_b^2[n]\right)}, \tag{61}$$

$$\varkappa_b[n] = \frac{\int_0^\infty r^2[n]\nu_{\mathsf{r}[n]\to f_b[n]}}{\int_0^\infty \nu_{\mathsf{r}[n]\to f_b[n]}} - \iota_b^2[n] \overset{(27)}{\approx} \eta_b^2[n]+2\gamma_b[n]-\iota_b^2[n] + \gamma_b[n]\eta_b[n]\frac{1+\mathrm{erf}\left(\varsigma_b[n]\right)+\frac{12}{\sqrt{\pi}}\varsigma_b[n]\exp\left(-\varsigma_b^2[n]\right)}{\eta_b[n]\left(1+\mathrm{erf}(\varsigma_b[n])\right)+\sqrt{\frac{2\gamma_b[n]}{\pi}}\exp(-\varsigma_b^2[n])}, \tag{62}$$

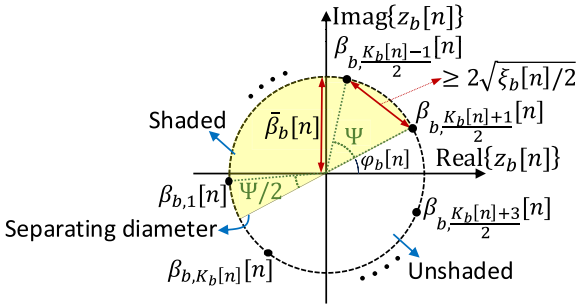



Fig. 12. We use mixture of $K_b[n]$ complex Gaussian PDFs to approximate $\nu_{f_b[n]\to z_b[n]}$. We assume that these Gaussian PDFs share a common variance, denoted by $\xi_b[n]$, and that their means have the same magnitude, denoted by $\bar{\beta}_b[n]$.

Now, we use the RiceFit function in MATLAB [48] to find $\rho_b[n]$ and $\tau_b[n]$ of the Rice distribution in (31). These parameters are tuned such that the Rice distribution's mean is equal to $\iota_b[n]$ (61) and its variance is equal to $\varkappa_b[n]$ (62).

Next, we derive an approximation of $[\nu_{f_b[n]\to \mathsf{s}[n]}]_{\mathsf{s}[n]=1}$ assuming $\nu_{\Theta_b[n]\to f_b[n]} = 1/2\pi$. To this end, we use (47) to compute the integral with respect to $\theta_b[n]$ in (29) when $s[n]=1$. This results in

$$[\nu_{f_b[n]\to \mathsf{s}[n]}]_{\mathsf{s}[n]=1} \overset{(29)}{=} \frac{1}{\pi c_b[n]} \int_{r[n]} \exp\left(-\frac{r^2[n]+|e_b[n]|^2}{c_b[n]}\right) \mathrm{I}_0\left(\frac{2\,|e_b[n]|}{c_b[n]} r[n]\right) \nu_{\mathsf{r}[n]\to f_b[n]}. \tag{63}$$

By substituting $\nu_{\mathsf{r}[n]\to f_b[n]}$ in (29) with the Rice distribution in (31) and using [47, Equation 10.43.28] in computing the integral with respect to $r[n]$, we have

$$[\nu_{f_b[n]\to \mathsf{s}[n]}]_{\mathsf{s}[n]=1} \propto \frac{1}{2\pi\,|e_b[n]|} \mathrm{Rice}\left(|e_b[n]|\,;\ \rho_b[n], \frac{2\tau_b[n]+c_b[n]}{2}\right). \tag{64}$$

Similarly, $[\nu_{f_b[n]\to \mathsf{s}[n]}]_{s[n]=0}$ is

$$\nu_{f_b[n]\to \mathsf{s}[n]}\big|_{s[n]=0} \propto \mathcal{CN}(0; e_b[n], c_b[n])\,. \tag{65}$$

By substituting (64) and (65) in (30), we can derive (32).

### C. Derivation of $K_b[n]$ in (42) and the Angles in (43)

In this section, we find the number of Gaussian components for the GM model (39) such that the $\ell_2$-norm distance between the means of any two neighboring components is at least $2\sqrt{\xi_b[n]/2}$ as shown in Fig. 12. We place the components with means $\{\beta_{b,k}\}_{b=1}^{(K_b[n]-1)/2}$ in one half of the circle indicated by the shaded region. Also, components with means $\{\beta_{b,k}\}_{b=(K_b[n]-1)/2}^{K_b[n]}$ are in unshaded half of the circle, and the component with mean $\beta_{b,(K_b[n]+1)/2} = \bar{\beta}_b[n]e^{\mathrm{j}\varphi_b[n]}$ lies on a diameter of the circle that separates these two half circles. We set the means of the components in the unshaded part to be the mirror of those in the shaded part with respect to the separating diameter. Hence, it is sufficient to identify the means in the shaded part to determine the rest of the means.

Let $\Psi \in (0,\pi)$ denote the angle between the line segments connecting the center of the circle to $\beta_{b,(K_b[n]+1)/2}$ and $\beta_{b,(K_b[n]-1)/2}$ as shown in Fig. 12. From geometry, we can infer from Fig. 12 that for $\Psi \geq 2\sin^{-1}\left(\sqrt{\xi_b[n]/2}/\bar{\beta}_b[n]\right)$, we have $\|\beta_{b,(K_b[n]-1)/2} - \beta_{b,(K_b[n]+1)/2}\| \geq 2\sqrt{\xi_b[n]/2}$. To meet this condition on $\Psi$, we can find that the maximum number of components in the shaded half is

$$\frac{K_b[n]-1}{2} = \left\lfloor \frac{\pi - \Psi/2}{\Psi} \right\rfloor. \tag{66}$$

Now, (42) immediately follows from (66).

To find the phase of the means $\beta_{b,k}$s, we set $\angle\beta_{b,1}[n] = \varphi_b[n] + \left(\pi - \frac{\Psi}{2}\right)$, without loss of generality. Next, we distribute the remaining components evenly by an angular distance of

$$\angle\beta_{b,k}[n] - \angle\beta_{b,k+1}[n] = \frac{2\pi-\Psi}{K_b[n]-1}, \quad k = 1, \cdots, K_b[n]-1, \tag{67}$$

which results in (43).

## REFERENCES


[1] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[2] H. Sarieddeen, M.-S. Alouini, and T. Y. Al-Naffouri, "An overview of signal processing techniques for terahertz communications," *Proc. IEEE*, vol. 109, no. 10, pp. 1628–1665, Oct. 2021.

[3] I. F. Akyildiz, C. Han, and S. Nie, "Combating the distance problem in the millimeter wave and terahertz frequency bands," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 102–108, Jun. 2018.

[4] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, Nov. 2015, Art. no. 115002.

[5] G. Hueber and R. B. Staszewski, *Multi-Mode/Multi-Band RF Transceivers for Wireless Communications: Advanced Techniques, Architectures, and Trends*. Hoboken, NJ, USA: Wiley, 2011, pp. 451–496.

[6] Z. Sha and Z. Wang, "Channel estimation and equalization for terahertz receiver with RF impairments," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1621–1635, Jun. 2021.

[7] R. Zhang, B. Shim, and H. Zhao, "Downlink compressive channel estimation with phase noise in massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5534–5548, Sep. 2020.

[8] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[9] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 2168–2172.

[10] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sep. 2005.

[11] S. Shayovitz and D. Raphaeli, "Message passing algorithms for phase noise tracking using Tikhonov mixtures," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 387–401, Jan. 2016.

[12] M. E. Rasekh and U. Madhow, "Noncoherent compressive channel estimation for mm-Wave massive MIMO," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 889–894.

[13] X. Li, J. Fang, H. Duan, Z. Chen, and H. Li, "Fast beam alignment for millimeter wave communications: A sparse encoding and phaseless decoding approach," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4402–4417, Sep. 2019.

[14] K.-H. Liu, X. Li, H. Zhao, and G. Fan, "Structured phase retrieval-aided channel estimation for millimeter-wave/sub-terahertz MIMO systems," in *Proc. IEEE 96th Veh. Technol. Conf.*, Sep. 2022, pp. 1–5.

[15] W. Yi, N. J. Myers, and G. Joseph, "Sparse millimeter wave channel estimation from partially coherent measurements," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2024, pp. 1897–1902.

[16] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterization," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 47, no. 5, pp. 655–674, May 2000.

[17] H. Yan and D. Cabria, "Compressive sensing based initial beamforming training for massive MIMO millimeter-wave systems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 620–624.
[18] C. Hu, X. Wang, L. Dai, and J. Ma, "Partially coherent compressive phase retrieval for millimeter-wave massive MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 68, pp. 1673–1687, 2020.
[19] C. Wei, H. Jiang, J. Dang, L. Wu, and H. Zhang, "Accurate channel estimation for mmWave massive MIMO with partially coherent phase offsets," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2170–2174, Sep. 2022.
[20] L. Zhang, G. Wang, G. B. Giannakis, and J. Chen, "Compressive phase retrieval via reweighted amplitude flow," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5029–5040, Oct. 2018.
[21] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.
[22] H. Masoumi and N. J. Myers, "Message passing-based sparse channel estimation from partially coherent measurement vectors," in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G Summit)*, Jun. 2025, pp. 637–642.
[23] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[24] T. Nitsche, C. Cordeiro, A. Flores, E. Knightly, E. Perahia, and J. Widmer, "IEEE 802.11ad: Directional 60 GHz communication for multi-gigabit-per-second Wi-Fi," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 132–141, Dec. 2014.
[25] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 186–192, Dec. 2017.
[26] *IEEE Standard for Information Technology–Telecommunications and Information Exchange Between Systems–Local and Metropolitan Area Networks–Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, 2021, pp. 1–4379.
[27] P. F. M. Smulders, "Statistical characterization of 60-GHz indoor radio channels," *IEEE Trans. Antennas Propag.*, vol. 57, no. 10, pp. 2820–2829, Oct. 2009.
[28] J. Rimmelspacher, R. Weigel, A. Hagelauer, and V. Issakov, "A quad-core 60 GHz push-push 45 nm SOI CMOS VCO with -101.7 dBc/Hz phase noise at 1 MHz offset, 19% continuous FTR and -187 dBc/Hz FoMT," in *Proc. IEEE 44th Eur. Solid State Circuits Conf. (ESSCIRC)*, Sep. 2018, pp. 138–141.
[29] N. J. Myers, A. Mezghani, and R. W. Heath Jr., "FALP: Fast beam alignment in mmWave systems with low-resolution phase shifters," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8739–8753, Dec. 2019.
[30] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5270–5284, Nov. 2013.
[31] S. Rangan, A. K. Fletcher, V. K. Goyal, E. Byrne, and P. Schniter, "Hybrid approximate message passing," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4577–4592, Sep. 2017.
[32] J. Mo, P. Schniter, and R. W. Heath Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.
[33] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 340–354, Jan. 2013.
[34] A. F. Molisch, *Wireless Communications*, vol. 34. Hoboken, NJ, USA: Wiley, 2012.
[35] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[36] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 39, no. 1, pp. 1–22, Sep. 1977.
[37] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 24, no. 1, pp. 46–156, Jan. 1945.
[38] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM J. Optim.*, vol. 9, no. 1, pp. 112–147, Jan. 1998.
[39] E. Angelino, M. J. Johnson, and R. P. Adams, "Patterns of scalable Bayesian inference," *Found. Trends Mach. Learn.*, vol. 9, nos. 2–3, pp. 119–247, 2016.
[40] N. Wu, H. Li, D. He, A. Nallanathan, and T. Q. S. Quek, "Integrated sensing and communication receiver design for OTFS-based MIMO system: A unified variational inference framework," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 4, pp. 1339–1353, Apr. 2025.
[41] S. Sun, G. R. MacCartney, and T. S. Rappaport, "A novel millimeter-wave channel simulator and applications for 5G wireless communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
[42] J. Rimmelspacher, R. Ciocoveanu, G. Steffan, M. Bassi, and V. Issakov, "Low power low phase noise 60 GHz multichannel transceiver in 28 nm CMOS for radar applications," in *Proc. IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, Los Angeles, CA, USA, Aug. 2020, pp. 19–22.
[43] H. Jia, P. Guan, W. Deng, Z. Wang, and B. Chi, "A low-phase-noise quad-core millimeter-wave fundamental VCO using circular triple-coupled transformer in 65-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 58, no. 2, pp. 371–385, Feb. 2023.
[44] L. Lian, A. Liu, and V. K. N. Lau, "Exploiting dynamic sparsity for downlink FDD-massive MIMO channel tracking," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2007–2021, Apr. 2019.
[45] Y. Wei, M.-M. Zhao, A. Liu, and M.-J. Zhao, "Channel tracking and prediction for IRS-aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 563–579, Jan. 2023.
[46] K. V. Mardia and P. E. Jupp, *Directional Statistics*. Hoboken, NJ, USA: Wiley, 2009.
[47] F. W. J. Olver et al. (2024). *NIST Digital Library of Mathematical Functions*. [Online]. Available: https://dlmf.nist.gov/10.43
[48] G. Ridgway. (2024). *Rice/Rician Distribution*. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/14237-rice-rician-distribution

**Hamed Masoumi** (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree with Delft Center for Systems and Control (DCSC), Delft University of Technology (TU Delft), The Netherlands. His research interests include the intersection of wireless communications, signal processing, and optimization. He was a recipient of the Best Student Paper Award at the IEEE SPAWC 2022.

**Nitin Jonathan Myers** (Member, IEEE) received the B.Tech. and M.Tech. degrees in electrical engineering from Indian Institute of Technology (IIT) Madras in 2016 and the Ph.D. degree in electrical and computer engineering (ECE) from The University of Texas at Austin (UT Austin) in 2020.

He is currently a tenured Assistant Professor with Delft Center for Systems and Control, Delft University of Technology (TU Delft). Prior to joining TU Delft, he was a Senior Engineer with the 5G Modem Research and Development Team, Samsung Semiconductor Inc., San Diego. His research interests include optimization and multi-dimensional signal processing, with applications to communications and sensing.

Dr. Myers received the DAAD WISE Scholarship in 2014 and a Silver Medal, Institute Merit Prize in 2016 during his undergraduate days at IIT Madras. At UT Austin, he received the University Graduate Continuing Fellowship 2019–2020, and the 2018 and 2019 ECE Research Awards from the Cockrell School of Engineering. He was recognized as an Exemplary Reviewer by IEEE TRANSACTIONS ON COMMUNICATIONS in 2019 and IEEE WIRELESS COMMUNICATIONS LETTERS in 2018. His research won the IEEE ICASSP 2020 Video Contest Runner-Up, the Best Student Paper Award at IEEE SPAWC 2022, and the Best Live Demo Award Finalist at IEEE Sensors 2024. At TU Delft, he was recognized as the Best Lecturer in the 2nd year of the Bachelor's Program in Mechanical Engineering (2021–2022), and he received the IDEA League Fellowship, in 2025.