

Iterative Channel Estimation Using LSE and Sparse Message Passing for MmWave MIMO Systems

Chongwen Huang, *Student Member, IEEE*, Lei Liu, *Student Member, IEEE*,
Chau Yuen, *Senior Member, IEEE*, Sumei Sun, *Fellow, IEEE*

Abstract—We propose an iterative channel estimation algorithm based on the Least Square Estimation (LSE) and Sparse Message Passing (SMP) algorithm for the Millimeter Wave (mmWave) MIMO systems. The channel coefficients of the mmWave MIMO are approximately modeled as a Bernoulli-Gaussian distribution and the channel matrix is sparse with only a few non-zero entries. By leveraging the advantage of sparseness, we propose an algorithm that iteratively detects the exact locations and values of non-zero entries of the sparse channel matrix. At each iteration, the locations are detected by the SMP, and values are estimated with the LSE. We also analyze the Cramér-Rao Lower Bound (CRLB), and show that the proposed algorithm is a minimum variance unbiased estimator under the assumption that we have the partial priori knowledge of the channel. Furthermore, we employ the Gaussian approximation for message densities under density evolution to simplify the analysis of the algorithm, which provides a simple method to predict the performance of the proposed algorithm. Numerical experiments show that the proposed algorithm has much better performance than the existing sparse estimators, especially when the channel is sparse. In addition, our proposed algorithm converges to the CRLB of the genie-aided estimation of sparse channels with only five turbo iterations.

Index Terms—Millimeter wave, iterative channel estimation, sparse message passing, Gaussian-Bernoulli distribution, Cramér-Rao lower bound, minimum variance unbiased estimator, Gaussian approximation.

I. INTRODUCTION

Millimeter wave (mmWave) has been receiving tremendous interests from the academia, industry, and government for future 5G cellular systems [1]–[10] due to the available spectrum from 30 GHz to 300 GHz. However, mmWave poses new challenges. One of them is the severe path-loss. Recent urban model experiments show that path losses are 40 dB worse at 28 GHz compared to 2.8 GHz [11], [12].

One way to overcome this severe path-loss of mmWave signal propagation is to increase the number of transmit and receive antennas [7], [13], [14]. With the large number of

antennas and relatively fewer channel paths, the mmWave channel is approximately sparse [7], [14]–[17]. The sparse feature is recently verified by measurements, for example, [11], [12], [18] reported that mmWave channels typically exhibit only 3–4 scattering clusters in dense-urban non-line-of-sight environments. Therefore, conventional MIMO iterative channel estimation methods [19]–[25] are not suitable for mmWave systems due to the different channel characteristic and system model (i.e., mmWave systems usually employ the hybrid analog/digital architecture for reducing hardware cost and power consumption [7], [11]). This prompts the need to design efficient channel estimation techniques for the mmWave systems.

For sparse channel estimation, several algorithms have been proposed in [26]–[29], [31]–[34]. They can be classified into three categories according to the required priori information (except the noise variance) of the channel. The algorithm in the first category requires to know the full knowledge of the channel's distribution, structure, etc., for example, the approximate message passing (AMP) algorithm [26] proposed by Donoho and Maleki. AMP is a low-complexity iterative Bayesian algorithm that can achieve approximately maximum a posteriori and minimum mean-squared error signal estimates. Iterative Detection/Estimation With Threshold (ITD-SE) [27], Adaptive Compressed Sensing (ACS) estimation algorithm proposed in [28] and Orthogonal Matching Pursuit (OMP) [29] are classified into the second category, which needs partial priori information of the channel, e.g., the degree of sparsity L . ITD-SE based Least Square Estimation (LSE) needs the fewer iterations, but its performance depends on the adaptive threshold selection scheme. ACS leverages the advanced compressed sensing theory and combines with the hybrid beamforming technique. Therefore, it is very suitable for the mmWave systems. The algorithms in the third category do not need any priori knowledge of the channel distribution except the noise variance, e.g., Sparse Bayesian Learning (SBL) [30], LASSO [32], Expectation-maximization Bernoulli-Gaussian Approximate Message Passing (EM-BG-AMP) [33], etc. However, most of these algorithms need to learn the channel in order to improve the estimation performance, for example, SBL that is implemented via the more robust T-MSBL [31]. In addition, although LASSO [32] and EM-BG-AMP [33] have the lower complexity, their solutions are generally not the sparsest, and EM-BG-AMP also requires independent and identically distributed (i.i.d.) zero-mean Gaussian training matrix. [33], [35]. Recently, [34] proposed a modified mean field (MF) message passing-based algorithm, which also belongs to the

The work of C. Yuen was supported by the MIT-SUTD International design center and NSFC 61750110529 Grant, and that of C. Huang by the PHC Merlion PhD program.

Chongwen Huang and Chau Yuen are with the Singapore University of Technology and Design, Singapore. Lei Liu is with the Singapore University of Technology and Design and City University of Hong Kong, Hong Kong, China. Sumei Sun is with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), 138632, Singapore. (e-mail: chongwen_huang@mymail.sutd.edu.sg, leiliuxidian@gmail.com, yuen-chau@sutd.edu.sg, sunsm@i2r.a-star.edu.sg).

The material in this paper was presented in part at the conference of the IEEE Globecom 2016 workshop on mobile communications in higher frequency bands, Washington D.C., USA, Dec. 2016 [1].

third category and can deliver even better performance with the lower complexity than the conventional vector-form MF SBL algorithm by introducing a few hard constraint factors. This also provides a promising method for the future mmWave channel estimation.

In this paper, we develop an iterative channel estimation algorithm based on the LSE, Expectation-Maximization (EM) and Sparse Message Passing (SMP) for mmWave MIMO systems with large antenna arrays at both the transmitter and receiver. A beamspace channel representation model is adopted which can capture the sparseness of physical mmWave channel and provides simple geometric interpretation of the scatter environment ([see [15], [16]]). Based on this representation, we further model the mmWave channel as a Bernoulli-Gaussian distribution. We summarize our main contributions of this paper as follows.

- We formulate a sparse channel estimation problem and propose a novel sparse channel estimation algorithm. Compared with existing sparse channel estimation methods, ours can yield a better performance since it not only can take full advantage of the inherent sparseness of the mmWave channel, but also can leverage both virtues of the LSE and SMP algorithms.
- We give the performance analysis of the proposed algorithm, and derive its upper bound and CRLB. Furthermore, we show that the proposed algorithm is the Minimum Variance Unbiased Estimator (MVUE) under the assumption that we have the partial priori knowledge of the channel.
- We employ the Extrinsic Information Transfer (EXIT) chart-based technique for the convergence analysis of the key part of the proposed algorithm, and provide insights on the iteration evolution of the proposed algorithm, based on which design parameters we optimize. This analysis use a Gaussian approximation for message densities under the density evolution, and adopt the Log-Likelihood Ratios (LLRs) of messages, which can reduce the complexity of the analysis.
- We evaluate the performance of the proposed estimation algorithm. Numerical simulations show that our algorithm exhibits far better performance than the classical LSE estimator, as well as existing sparse channel estimators (e.g., LASSO, ITD-SE, EM-BG-AMP, etc.). In addition, we also find that this algorithm can approximately achieve the CRLB with the fast convergence speed.

The rest of the paper is organized as follows. In Section II, we present the beamspace channel representation model for mmWave MIMO systems and formulate a sparse channel estimation problem. In Section III, we propose an iterative sparse channel estimation algorithm. The performance analysis of the proposed algorithm is given in Section IV. In Section V, simulation results demonstrating the performance of the proposed algorithms are given, before concluding the paper in Section VI.

Notation: a is a scalar, \mathbf{a} is a vector and \mathbf{A} is a matrix. \mathbf{A}^T , \mathbf{A}^H , \mathbf{A}^{-1} , \mathbf{A}^\dagger and $\|\mathbf{A}\|_F$ represent transpose, Hermitian (conjugate transpose), inverse, pseudo-inverse and Frobenius

norm of a matrix \mathbf{A} , respectively. $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of \mathbf{A} and \mathbf{B} , and $\text{vec}(\mathbf{A})$ is a vector stacking all the columns of \mathbf{A} . $\text{diag}(\mathbf{a})$ is a diagonal matrix with the entries of \mathbf{a} on its diagonal, and $\text{diag}(\mathbf{A})$ is a block diagonal matrix with the matrix \mathbf{A} as the block on its diagonal. $\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{V})$ is the Probability Distribution Function (PDF) of a complex Gaussian random vector \mathbf{x} with mean \mathbf{m} and covariance \mathbf{V} . We use the $E\{\cdot\}$, $\text{Var}\{\cdot\}$, $\exp(\cdot)$, and \mathbb{R} to denote the expectation, variance, nature exponential operation and the real field respectively. In addition, $E\{a|b\}$ denotes the conditional expectation of variable a given b , and $\text{Var}\{a|b\}$ denotes the conditional variance of the variable a given b .

II. SYSTEM MODEL

We consider a hybrid analog-digital mmWave communication system that has N_t transmit and N_r receive antennas at the transmitter and receiver respectively, and both of them have N_{RF} RF chains¹. The transmitter and receiver communicate via N_s data streams, such that $N_s \leq N_{RF} \leq N_r$ and $N_s \leq N_{RF} \leq N_t$ [28], [35]. Assuming frequency-flat fading channel, and that there is a $N_{RF} \times N_s$ baseband precoder \mathbf{F}_{BB} followed by an $N_t \times N_{RF}$ RF precoder \mathbf{F}_{RF} in the downlink transmission, we can denote $\mathbf{F} = \mathbf{F}_{RF}\mathbf{F}_{BB}$ as a $N_t \times N_s$ combined precoding matrix, and similarly, we denote \mathbf{C} as the $N_r \times N_s$ combining matrix, which is composed of the RF combiners \mathbf{C}_{RF} and baseband combiners \mathbf{C}_{BB} . For the traditional hybrid analog-digital model, the observed signal at the receiver can be written as [28], [35]–[38],

$$\mathbf{y} = \mathbf{C}^H \mathbf{H} \mathbf{F} \mathbf{s} + \mathbf{C}^H \mathbf{z}, \quad (1)$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is the transmitted signal, $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ is the received signal and $\mathbf{z} \in \mathbb{C}^{N_r \times 1}$ is the Gaussian noise with $\mathbf{z} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I})$.

Since mmWave channels are expected to have limited scattering, we adopt a geometric channel model with L scatterers. Each scatterer is further assumed to contribute a single propagation path between transmitters and receivers [11], [28], [35]. Under this model, the channel \mathbf{H} can be expressed as

$$\mathbf{H} = \sqrt{\frac{N_r N_t}{\rho}} \sum_{l=1}^L \alpha_l \mathbf{a}_r(\theta_l) \mathbf{a}_t^H(\phi_l), \quad (2)$$

where ρ denotes the average path-loss between the transmitter and receiver, α_l is the gain of the l th path, $\phi_l \in [0, 2\pi]$ and $\theta_l \in [0, 2\pi]$ denote the l th path's azimuth angles of departure and arrival of the transmitter and receiver respectively. Finally, $\mathbf{a}_t(\phi_l)$ and $\mathbf{a}_r(\theta_l)$ are the antenna array response vectors at the transmitter and receiver respectively [28], [36]–[38]. If a uniform linear arrays is used, $\mathbf{a}_t(\phi_l)$ can be written as

$$\mathbf{a}_t(\phi_l) = \frac{1}{\sqrt{N_t}} \left[1, e^{j \frac{2\pi}{\lambda} d \sin(\phi_l)}, \dots, e^{j (N_t-1) \frac{2\pi}{\lambda} d \sin(\phi_l)} \right]^T, \quad (3)$$

where λ is the signal wavelength, and d is the distance between antenna elements. The array response vectors at the receiver, $\mathbf{a}_r(\theta_l)$, can be written in a similar fashion. Then, the channel can be written in a more compact form as

$$\mathbf{H} = \mathbf{A}_r \text{diag}(\boldsymbol{\alpha}) \mathbf{A}_t^H, \quad (4)$$

¹For simplicity, we assume the same number of RF chains at the transmitter and receiver. The proposed architecture also can be extended to the case where has different numbers of RF chains.

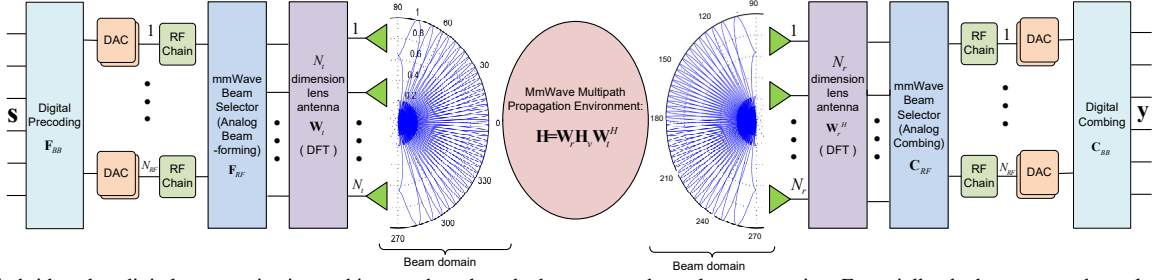


Fig. 1. The hybrid analog-digital communication architecture based on the beamspace channel representation. Essentially, the beamspace channel representation is to map the signal of the spatial domain to the signal of the beam domain by employing a carefully designed discrete lens antenna array instead of the electromagnetic antenna array. The Fourier transformation \mathbf{W}_t and \mathbf{W}_r can be seen as a mapping from the antenna domain onto a beam domain and the entries of the matrix \mathbf{H}_v can be interpreted as the channel gains between the N_t transmit and the N_r receive beams.

where $\alpha = \sqrt{\frac{N_r N_t}{\rho}} [\alpha_1, \alpha_2, \dots, \alpha_l]^T$. The matrices

$$\mathbf{A}_t = [\mathbf{a}_t(\phi_1), \mathbf{a}_t(\phi_2), \dots, \mathbf{a}_t(\phi_l)], \quad (5)$$

and

$$\mathbf{A}_r = [\mathbf{a}_r(\theta_1), \mathbf{a}_r(\theta_2), \dots, \mathbf{a}_r(\theta_l)], \quad (6)$$

contain the transmitter and receiver array response vectors. The mmWave multipath propagation channel usually consists of a few reflected path clusters [18], [35], [38], [39]. Large antenna arrays are deployed in the mmWave systems for combatting the high path loss. Hence, we usually have $L \ll \min\{N_r, N_t\}$.

For capturing the inherent sparse characteristic of the physical mmWave modeling, we adopt a hybrid analog-digital communication architecture that is based on the beamspace channel representation as shown in Fig. 1. This beamspace representation also provides a tractable linear channel characterization, and offers a simple and transparent interpretation to the effects of scattering and array characteristics on channel capacity and diversity [15], [40]–[43]. Essentially, the beamspace channel representation is to map the signal of the spatial domain to the signal of the beam domain by employing a carefully designed discrete lens antenna array instead of the electromagnetic antenna array [10], [40]–[44]. In particular, the lens acts as a virtual passive phase shifter, focusing the incident electromagnetic wave to a certain region. This lens, when used jointly with antennas, exhibits two significant properties: (i) focused signal power at the front end achieving high directivity and gain, and (ii) concentrated signal power directed to a sub-region of the antenna array. These properties make the lens a practical and energy efficiency tool for implementing the RF frontend in beamforming systems [6], [45]–[47]. The finite dimensionality of the signal space allows the beamspace channel model that can be expressed as

$$\mathbf{H}_v = \mathbf{W}_r^H \mathbf{H} \mathbf{W}_t \quad (7)$$

where $\mathbf{W}_r \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{W}_t \in \mathbb{C}^{N_t \times N_t}$ are channel-invariant unitary DFT matrices [17], [40], and $\mathbf{W}_t \mathbf{W}_t^H = \mathbf{I}_{N_t}$, $\mathbf{W}_r \mathbf{W}_r^H = \mathbf{I}_{N_r}$. Note that $\mathbf{H}_v \in \mathbb{C}^{N_r \times N_t}$ is no longer diagonal. We recast (1) by the beamspace channel representation as

$$\mathbf{y} = \mathbf{C}^H \mathbf{W}_r \mathbf{H}_v \mathbf{W}_t^H \mathbf{F} \mathbf{s} + \mathbf{C}^H \mathbf{W}_r \mathbf{z}, \quad (8)$$

The Fourier transformation \mathbf{W}_t and \mathbf{W}_r can be seen as a mapping from the antenna domain onto a beam domain and the entries of the matrix \mathbf{H}_v can be interpreted as the channel gains between the N_t transmit and the N_r receive beams [40]. Assuming the channel is time-invariant in the blocks

$t \in \{1, \dots, T\}$. Then, $\mathbf{Y} \triangleq [\mathbf{y}_1, \dots, \mathbf{y}_T]$, $\mathbf{S} \triangleq [\mathbf{s}_1, \dots, \mathbf{s}_T]$, and $\mathbf{N} \triangleq [\mathbf{C}^H \mathbf{W}_r \mathbf{z}_1, \dots, \mathbf{C}^H \mathbf{W}_r \mathbf{z}_T] \triangleq [\mathbf{n}_1, \dots, \mathbf{n}_T]$. The channel model is rewritten as

$$\mathbf{Y} = \mathbf{C}^H \mathbf{W}_r \mathbf{H}_v \mathbf{W}_t^H \mathbf{F} \mathbf{S} + \mathbf{N}, \quad (9)$$

where $\mathbf{Y} \in \mathbb{C}^{N_s \times T}$, $\mathbf{S} \in \mathbb{C}^{N_s \times T}$, and $\mathbf{N} \in \mathbb{C}^{N_s \times T}$. We define $\mathbf{D} \triangleq \mathbf{C}^H \mathbf{W}_r$ and $\mathbf{X} \triangleq \mathbf{W}_t^H \mathbf{F} \mathbf{S}$. Then, we recast the (9) as

$$\mathbf{Y} = \mathbf{D} \mathbf{H}_v \mathbf{X} + \mathbf{N}. \quad (10)$$

Then, vectorizing (10) [28] yields

$$\begin{aligned} \text{vec}(\mathbf{Y}) &= \text{vec}(\mathbf{D} \mathbf{H}_v \mathbf{X}) + \text{vec}(\mathbf{N}) \\ &= (\mathbf{X}^T \otimes \mathbf{D}) \text{vec}(\mathbf{H}_v) + \text{vec}(\mathbf{N}). \end{aligned} \quad (11)$$

By defining the $\bar{\mathbf{y}} \triangleq \text{vec}(\mathbf{Y})$, $\bar{\mathbf{S}} \triangleq \mathbf{X}^T \otimes \mathbf{D}$, $\mathbf{h}_v \triangleq \text{vec}(\mathbf{H}_v)$ and $\bar{\mathbf{n}} \triangleq \text{vec}(\mathbf{N})$, (11) is equivalently rewritten as follows:

$$\bar{\mathbf{y}} = \bar{\mathbf{S}} \mathbf{h}_v + \bar{\mathbf{n}}, \quad (12)$$

where $\bar{\mathbf{y}} \in \mathbb{C}^{N_s T \times 1}$, $\bar{\mathbf{S}} \in \mathbb{C}^{N_s T \times N_r N_t}$, $\mathbf{h}_v \in \mathbb{C}^{N_r N_t \times 1}$ and $\bar{\mathbf{n}} \in \mathbb{C}^{N_s T \times 1}$. The mmWave channel estimation problem is simplified to estimate the beamspace channel vector \mathbf{h}_v by the equivalent training matrix $\bar{\mathbf{S}}$ and the observed vector $\bar{\mathbf{y}}$.

III. SPARSE CHANNEL ESTIMATION

In this section, we present an iterative channel estimation algorithm based on the LSE and SMP algorithm as shown in Fig. 2, which is named LSE-SMP. It consists of four phases: LSE Coarse Estimation (Step 1), Sparse Message Passing Detection (Step 2), Update for LSE Estimation and Sparsity Ratios (Step 3), and Decision and Output (Step 4). Since there is no priori knowledge of \mathbf{h}_v , we initially adopt the LSE method to obtain its coarse estimation. Then, in the Step 2, we consider the estimation of non-zero positions in the channel vector \mathbf{h}_v as a detection problem, and propose an SMP algorithm to find these non-zero positions under the estimated sparsity ratio. In the Step 3, we apply LSE method again by leveraging the estimated non-zero positions in the Step 2 to obtain the fine estimation of \mathbf{h}_v , and estimate the sparsity ratio by the Expectation-Maximization (EM) method. The fourth Step is to make the decision according to the performance requirements (MSE, number of iterations, etc.) and the estimation of the Step 2 and Step 3. If the MSE of LSE-SMP meets the requirement or the number of iterations reaches the limit, the final estimation of \mathbf{h}_v will be output, otherwise the Step 2 and Step 3 will repeat until we obtain a fine estimation.

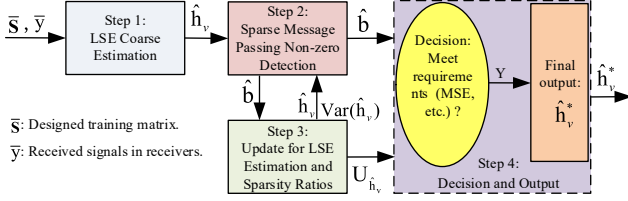


Fig. 2. The processing for the proposed LSE-SMP algorithm, which consists of four phases: LSE Coarse Estimation, Sparse Message Passing Detection, Update for LSE Estimation and Sparsity Ratios, and Decision and Output.

A. LSE Coarse Estimation

To find the Coarse Estimation $\hat{\mathbf{h}}_v$ based on the observed vector $\bar{\mathbf{y}}$ with a Mean Square Error (MSE) $E\{\|\hat{\mathbf{h}}_v - \mathbf{h}_v\|^2\}$, we can solve the following Least Square (LS) problem,

$$\begin{aligned}\hat{\mathbf{h}}_v &= \arg \min_{\mathbf{h}_v} \|\bar{\mathbf{y}} - \bar{\mathbf{S}}\mathbf{h}_v\|_2^2 \\ &= [\bar{\mathbf{S}}^H \bar{\mathbf{S}}]^{-1} \bar{\mathbf{S}}^H \bar{\mathbf{y}}.\end{aligned}\quad (13)$$

It is noted that it is MVUE in the sense of MSE for the deterministic signal when the estimator does not have any prior knowledge about either the sparsity structure of \mathbf{h}_v (i.e., the distribution and location of non-zero entries), or its degree of sparsity (i.e., L).

B. Sparse Message Passing Algorithm

After we get the Coarse Estimation of \mathbf{h}_v , we propose a fast iterative algorithm to find the positions of non-zero entries. This algorithm is named sparse message passing since it can take full advantage of the channel sparsity and message passing algorithm.

1) *Factor Graph Representation of the mmWave Channel:* In order to get better understanding of our proposed algorithm, we show the factor graph representation of the channel vector \mathbf{h}_v in the following. Firstly, we decompose the \mathbf{h}_v into a diagonal coefficient matrix $\mathbf{U}_{\mathbf{h}_v}$ and a column array \mathbf{b} . The column array $\mathbf{b} = [b_{ij}]_{N_r N_t \times 1} (i \in \{0, \dots, N_r\}, j \in \{0, \dots, N_t\})$ is called the position vector, and it represents the positions of non-zero in the coefficient matrix $\mathbf{U}_{\mathbf{h}_v}$. The $b_{ij} \in \{1, 0\}$ can be seen as a Bernoulli distribution. Then, the \mathbf{h}_v can be recast as

$$\mathbf{h}_v = \underbrace{\begin{bmatrix} h_{11} & & & 0 \\ & \ddots & & \\ & & h_{1N_t} & \\ 0 & & & h_{N_r N_t} \end{bmatrix}}_{=\mathbf{U}_{\mathbf{h}_v}} \underbrace{\begin{bmatrix} b_{11} \\ \vdots \\ b_{1N_t} \\ \vdots \\ b_{N_r N_t} \end{bmatrix}}_{=\mathbf{b}}. \quad (14)$$

Then, we rewrite (12) as

$$\underbrace{\begin{bmatrix} y_{11} & \dots & y_{1T} & \dots & y_{N_r T} \end{bmatrix}}_{\bar{\mathbf{y}}} = \bar{\mathbf{S}}\mathbf{h}_v + \bar{\mathbf{n}} \quad (15)$$

$$= \bar{\mathbf{S}}\mathbf{U}_{\mathbf{h}_v} \mathbf{b} + \bar{\mathbf{n}}.$$

According to factor graph analysis rules [19], [48]–[50], we can plot the factor graph to represent above equations, and it is shown in the Fig. 3. The nodes $(n_{11}, \dots, n_{N_r T})$ and $(h_{11}, \dots, h_{N_r N_t})$ are named the sum and variable nodes respectively.

The proposed SMP algorithm is considered for estimating positions of non-zero entries. It is similar to the belief propagation decoding process of the low density parity check code, in which the output message called extrinsic information on each edge is calculated by the messages on the other edges that are connected with the same node [19], [51]–[54].

2) *Message Update at Sum Nodes :* To analyze a sum node that is shown in the Fig. 4, we can obtain the k th ($k \in \{0, \dots, N_s\}$) data stream at the t th ($t \in \{0, \dots, T\}$) time block, and it can be expressed as [55]–[57]

$$y_{kt} = \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} s_{kt,ij} h_{ij} b_{ij} + n_{kt}. \quad (16)$$

As we mentioned before, there are L non-zero entries in the vector of \mathbf{b} . Then, we have the definition of the sparsity ratio $\eta = \frac{L}{N_r N_t}$. Assuming that the entries of $\mathbf{b} = [b_{11}, \dots, b_{N_t N_r}]$ are i.i.d., we can know the probability of the Bernoulli distribution, which can be denoted as

$$\begin{cases} p_0(b_{ij} = 1) = \hat{\eta}, \\ p_0(b_{ij} = 0) = 1 - \hat{\eta}, \end{cases} \quad (17)$$

where $\hat{\eta}$ is the estimation of η , and it will be updated by the EM method in the next phase. It should be pointed out that the initial value of η is set as 0.5. When N_t or N_r goes very large, the term $\sum_{i=1}^{N_r} \sum_{j=1}^{N_t} s_{kt,ij} h_{ij} b_{ij}$ can be approximated as the Gaussian distribution [50], [54], [58] according to the law of large numbers. When we compute the probability of $p(b_{ij} = 1)$ from the kt sum node to the ij variable node, we consider the messages from the other variable nodes ℓm ($\ell \neq i, m \neq j$ and $\ell \in \{1, 2, \dots, N_r\}, m \in \{1, 2, \dots, N_t\}$) to the sum node kt as the **equivalent Gaussian noise** n_{kt}^* . This can be expressed as

$$y_{kt} = \underbrace{s_{kt,ij} h_{ij} p(b_{ij} = 1)}_{\text{Desired Item}} + \underbrace{\sum_{\ell \neq i, m \neq j} \sum_{\ell=1}^{N_r} \sum_{m=1}^{N_t} s_{kt,\ell m} h_{\ell m} p(b_{\ell m} = 1)}_{\text{Equivalent Gaussian noise: } n_{kt}^*} + n_{kt}. \quad (19)$$

Furthermore, we can compute the mean value and variance of the equivalent Gaussian noise n_{kt}^* [58], [59]. These messages update at the sum nodes are given by

$$\begin{aligned}e_{kt \rightarrow ij}^s(\tau) &= E\{n_{kt}^* | s_{kt}, \hat{\mathbf{h}}(\tau), \mathbf{p}^v(\tau)\} \\ &= \sum_{\ell \neq i} \sum_{m \neq j} s_{kt,\ell m} \hat{h}_{\ell m}(\tau) p_{\ell m \rightarrow kt}^v(\tau), \\ v_{kt \rightarrow ij}^s(\tau) &= Var\{n_{kt}^* | s_{kt}, \hat{\mathbf{h}}(\tau), \mathbf{p}^v(\tau), \hat{\mathbf{v}}_h(\tau)\} \\ &= \sum_{\ell \neq i} \sum_{m \neq j} s_{kt,\ell m}^2 E\{\hat{h}_{\ell m}^2(\tau)\} E\{p_{\ell m \rightarrow kt}^{v^2}(\tau)\} \\ &\quad - s_{kt,\ell m}^2 \hat{h}_{\ell m}^2(\tau) p_{\ell m \rightarrow kt}^{v^2}(\tau) \\ &= \sum_{\ell \neq i} \sum_{m \neq j} s_{kt,\ell m}^2 p_{\ell m \rightarrow kt}^v(\tau) \hat{h}_{\ell m}^2(\tau) (1 - p_{\ell m \rightarrow kt}^v(\tau)) \\ &\quad + s_{kt,\ell m}^2 p_{\ell m \rightarrow kt}^v(\tau) v_{h_{\ell m}}(\tau) + \sigma_n^2.\end{aligned} \quad (20)$$

where τ denotes the iteration number, and σ_n^2 is the variance of the Gaussian noise, and $s_{kt} = [s_{kt,11}, \dots, s_{kt,ij}, \dots, s_{kt,N_r N_t}]^H$, $\hat{\mathbf{h}} = [\hat{h}_{11}, \dots, \hat{h}_{ij}, \dots, \hat{h}_{N_r N_t}]^H$, $\mathbf{p}^v = [p_{11 \rightarrow kt}^v, \dots, p_{ij \rightarrow kt}^v, \dots, p_{N_r N_t \rightarrow kt}^v]^T$ and $\hat{\mathbf{v}}_h = [v_{h_{11}}, \dots, v_{h_{ij}}, \dots, v_{h_{N_r N_t}}]^H$. In addition, $e_{kt \rightarrow ij}^s(\tau)$ and $v_{kt \rightarrow ij}^s(\tau)$ denote the mean and variance of the equivalent

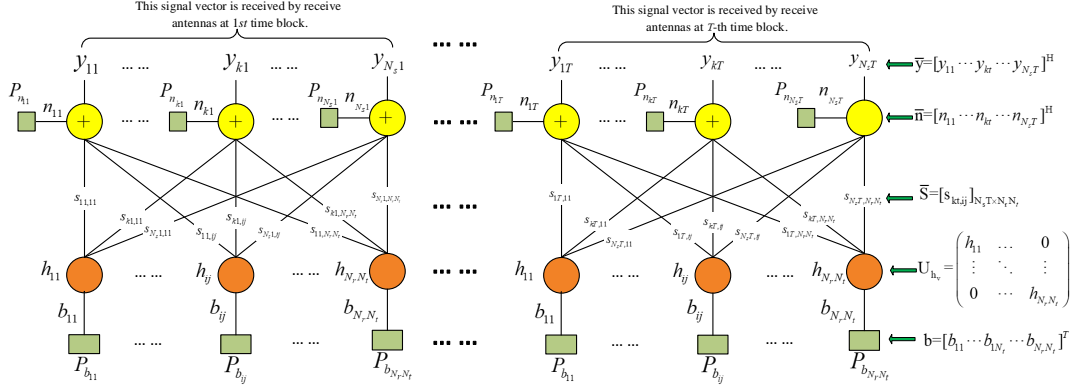


Fig. 3. The factor graph representation for the proposed sparse message passing detection algorithm. This factor graph is plotted on the basis of equations (14)-(16) and factor graph rules. The nodes \$(n_{11}, \dots, n_{N_r T})\$ and \$(h_{11}, \dots, h_{N_r N_t})\$ are named the sum and variable nodes respectively.

Gaussian noise \$n_{kt}^*\$ when the message \$s_{kt,ij}h_{ij}p(b_{ij} = 1)\$ passes from the \$kt\$ sum node to the \$ij\$ variable node at the \$\tau\$th iteration. Similarly, \$p_{\ell m \rightarrow kt}^v(\tau)\$ denotes the probability message of \$p(b_{\ell m} = 1)\$ passing from the \$\ell m\$ variable node to the \$kt\$ sum node at the \$\tau\$th iteration. \$\hat{h}_{\ell m}\$ and \$v_{h_{\ell m}}\$ that denote the mean and variance of \$h_{\ell m}\$ are estimated in the Step 3. Once we obtain the mean and variance of the equivalent Gaussian noise, we can compute the statistical probability of \$b_{ij} = 1\$ and \$b_{ij} = 0\$ as below equations in accordance to \$(y_{kt} - s_{kt,ij}h_{ij}p(b_{ij} = 1)) \sim \mathcal{N}(e_{kt \rightarrow ij}^s(\tau), v_{kt \rightarrow ij}^s(\tau))\$.

$$\begin{cases} P(b_{ij} = 1 | y_{kt}, s_{kt,ij}(\tau), \hat{h}_{ij}(\tau), v_{h_{ij}}(\tau)) \\ = \mathcal{N}(y_{kt}; e_{kt \rightarrow ij}^s(\tau) + s_{kt,ij}\hat{h}_{ij}(\tau), v_{kt \rightarrow ij}^s(\tau) + s_{ij}^2 v_{h_{ij}}(\tau)), \\ P(b_{ij} = 0 | y_{kt}, s_{kt,ij}(\tau), \hat{h}_{ij}(\tau), v_{h_{ij}}(\tau)) \\ = \mathcal{N}(y_{kt}; e_{kt \rightarrow ij}^s(\tau), v_{kt \rightarrow ij}^s(\tau)). \end{cases} \quad (22)$$

Then, we can give the probability message of \$b_{ij} = 1\$ passing from the \$kt\$ sum node to the \$ij\$ variable node as follows

$$\begin{aligned} p_{kt \rightarrow ij}^s(\tau) &= \left(1 + \frac{P(b_{ij} = 0 | y_{kt}, s_{kt,ij}(\tau), \hat{h}_{ij}(\tau), v_{h_{ij}}(\tau))}{P(b_{ij} = 1 | y_{kt}, s_{kt,ij}(\tau), \hat{h}_{ij}(\tau), v_{h_{ij}}(\tau))} \right)^{-1} \\ &= \left(1 + \sqrt{\frac{v_{kt \rightarrow ij}^s(\tau) + s_{ij}^2 v_{h_{ij}}(\tau)}{v_{kt \rightarrow ij}^s(\tau)}} \exp\left(\frac{(y_{kt} - e_{kt \rightarrow ij}^s(\tau))^2}{-2v_{kt \rightarrow ij}^s(\tau)} \right) \right. \\ &\quad \left. + \frac{(y_{kt} - e_{kt \rightarrow ij}^s(\tau) - s_{kt,ij}\hat{h}_{ij}(\tau))^2}{2(v_{kt \rightarrow ij}^s(\tau) + s_{ij}^2 v_{h_{ij}}(\tau))} \right)^{-1}. \end{aligned} \quad (23)$$

3) *Message Update at Variable Nodes* : In terms of the message update at variable nodes, we consider variable nodes as a broadcast process [60], [61] and the message update at the variable node is given by

$$p_{ij \rightarrow kt}^v(\tau+1) = \left(1 + \frac{\prod_{\kappa \neq k} p_{\kappa t \rightarrow ij}^s(\tau) \cdot p_0(b_{ij} = 0)}{\prod_{\kappa \neq k} p_{\kappa t \rightarrow ij}^s(\tau) \cdot p_0(b_{ij} = 1)} \right)^{-1}, \quad (24)$$

where \$\kappa \in \{1, \dots, N_s\}\$, and \$p_{\kappa t \rightarrow ij}^s(\tau+1)\$ denotes the probability message of \$b_{ij} = 1\$ passing from the \$it\$ sum node to the \$ij\$ variable node at the \$(\tau+1)\$th iteration. Furthermore, we can obtain the estimation of the Bernoulli variable \$b_{ij}\$ at the \$(\tau+1)\$th iteration as

$$\hat{b}_{ij}(\tau+1) = \left(1 + \frac{\prod_{k=1}^{N_s} p_{kt \rightarrow ij}^s(\tau) \cdot p_0(b_{ij} = 0)}{\prod_{k=1}^{N_s} p_{kt \rightarrow ij}^s(\tau) \cdot p_0(b_{ij} = 1)} \right)^{-1}. \quad (25)$$

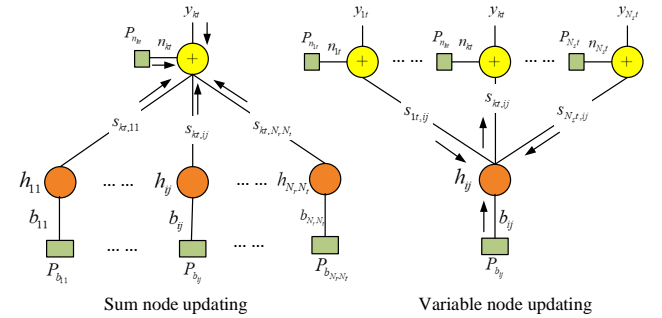


Fig. 4. Messages update at sum nodes and variable nodes. The output message called extrinsic information on each edge is calculated by the messages on the other edges that are connected with the same node. For the Gaussian-Bernoulli sparse signals, the messages passing on each edge are the probabilities of a Bernoulli distribution. The mean and variance of a Gaussian distribution are updated at the sum nodes, and they are used for computing the probability of a Bernoulli distribution.

Remark 1: It should be pointed out that \$p_{kt \rightarrow ij}^s(\tau+1)\$ is the extrinsic information and will be used to update messages of the sum nodes in the next iteration. On the other hand, \$\hat{b}_{ij}(\tau+1)\$ is updated based on the full information coming from all the sum nodes, and it will be used in the Step 3 for the estimation of \$\mathbf{h}_v\$.

Remark 2: The proposed SMP that is based on the message passing can obtain near optimal performance without heavy computational complexity by approximating the term \$\sum_{i=1}^{N_r} \sum_{j=1}^{N_t} s_{kt,ij}h_{ij}b_{ij}\$ as the Gaussian distribution. This is because the number of transmit antennas \$N_t\$ is large in mmWave communication systems as they operate at the higher communication spectrum, therefore, the approximation is accurate due to the law of large numbers. This also shows that our proposed algorithm is specialized for mmWave systems.

C. Update for LSE Estimation and Sparsity Ratio

1) *LSE Fine Estimation*: Once the positions of the non-zero entries have been estimated, the next step is to estimate the value of the coefficient matrix \$\mathbf{U}_{h_v}\$. For the problem, we propose a novel strategy based on the LSE method. This strategy is to **swap the positions** of \$h_{ij}\$ and \$b_{ij}\$ in the (14), so that we can get an accurate estimation by leveraging the sparsity of \$\mathbf{b}\$. Rewriting (15) as

$$\bar{\mathbf{y}} = \bar{\mathbf{S}}\mathbf{U}_{\mathbf{b}}\mathbf{h}_v + \bar{\mathbf{n}}, \quad (26)$$

where \$\mathbf{h}_v \in \mathbb{C}^{N_t N_r \times 1}\$, and \$\mathbf{b} = [\hat{b}_{ij}]_{N_r N_t \times 1}\$ is the vector estimated by the SMP algorithm. Similar with the LSE Coarse

Estimation, the estimation of \mathbf{h}_v can be obtained by solving the following LS problem

$$\hat{\mathbf{h}}_v = \arg \min_{\mathbf{h}_v} \{\|\bar{\mathbf{y}} - \bar{\mathbf{S}}\mathbf{U}_{\hat{\mathbf{b}}} \mathbf{h}_v\|_2^2\}. \quad (27)$$

Solving the above expression, we get the following estimator for \mathbf{h}_v as

$$\begin{cases} \hat{\mathbf{h}}_v(\tau) = \hat{\mathbf{Q}}^\dagger(\tau) (\bar{\mathbf{S}}\mathbf{U}_{\hat{\mathbf{b}}}(\tau))^H \bar{\mathbf{y}}, \\ \hat{\mathbf{v}}_h(\tau) = \sigma_n^2 \left((\bar{\mathbf{S}}\mathbf{U}_{\hat{\mathbf{b}}}(\tau))^H \bar{\mathbf{S}}\mathbf{U}_{\hat{\mathbf{b}}}(\tau) \right)^\dagger, \end{cases} \quad (28)$$

where $\hat{\mathbf{Q}}(\tau) = \mathbf{U}_{\hat{\mathbf{b}}}(\tau) \bar{\mathbf{S}}^H \bar{\mathbf{S}} \mathbf{U}_{\hat{\mathbf{b}}}(\tau)$, $\hat{\mathbf{v}}_h = \text{diag}[v_{h_{ij}}]_{N_r N_t \times N_r N_t}$. $\hat{\mathbf{h}}_v(\tau)$ and $\hat{\mathbf{v}}_h(\tau)$ denotes the estimated value and variance of \mathbf{h}_v at τ th iteration. After we obtain $\hat{\mathbf{h}}_v(\tau)$ and $\hat{\mathbf{v}}_h(\tau)$, these values will replace the $\hat{\mathbf{h}}_v(k-1)$ and $\hat{\mathbf{v}}_h(k-1)$ for calculating the mean and variance of the equivalent Gaussian noise n_{kt}^* in the iteration.

2) *EM update for the sparsity ratio* : We now present a classical EM algorithm [33], [62] to learn the sparsity ratio η . Since the channel vector \mathbf{h}_v can be modeled as i.i.d Bernoulli-Gaussian, then we have the marginal PDF as

$$p(h_{ij}; \eta, u_{h_v}, v_{h_v}) = (1-\eta)\delta(h_{ij}) + \eta \mathcal{N}(h_{ij}; u_{h_v}, v_{h_v}), \quad (30)$$

where $\delta(\cdot)$ is the Dirac delta, and u_{h_v} and v_{h_v} are the mean and variance of non-zero entries of h_{ij} respectively. Then, we can give the EM update for η by the estimated the parameters $\hat{\mathbf{r}}(\tau) \triangleq [\hat{\eta}(\tau), \hat{u}_{h_v}(\tau), \hat{v}_{h_v}(\tau)]$. In the sequel, we use the $\hat{\mathbf{r}}_{\setminus \eta}$ to denote the vector $\hat{\mathbf{r}}$ with the element η removed. Similar to the equation (29) in paper [33], the EM update for η can be written as

$$\hat{\eta}(\tau+1) = \arg \max_{\eta \in (0,1)} \sum_{j=1}^{N_t} \sum_{i=1}^{N_r} E\{\ln p(h_{ij}; \eta, \hat{\mathbf{r}}_{\setminus \eta}(\tau) | \bar{\mathbf{y}}; \hat{\mathbf{r}}(\tau))\}. \quad (31)$$

To maximize the value of η in the above equation, it is necessary to zero the derivative of the sum; i.e., that satisfies

$$\sum_{j=1}^{N_t} \sum_{i=1}^{N_r} \int_{h_{ij}} p(h_{ij} | \bar{\mathbf{y}}; \hat{\mathbf{r}}(\tau)) \frac{d \ln p(h_{ij}; \eta, \hat{\mathbf{r}}_{\setminus \eta}(\tau))}{d\eta} = 0. \quad (32)$$

Finally, we get the solution of (32) that is

$$\hat{\eta}(\tau+1) = \frac{1}{N_t N_r} \sum_{j=1}^{N_t} \sum_{i=1}^{N_r} \frac{1}{1 + \left(\frac{\eta(\tau) \mathcal{N}(r; \hat{h}_{ij}(\tau), \hat{v}_{h_{ij}}(\tau) + \mu^r)}{(1-\eta(\tau)) \mathcal{N}(r; 0, \mu^r)} \right)^{-1}}, \quad (33)$$

where $r = \frac{1}{N_s} \sum_{k=1}^{N_s} s_{kt,ij}^{-1} (y_{kt} - e_{kt \rightarrow ij}^s(\tau))$, $\mu^r = \frac{1}{N_s} \sum_{k=1}^{N_s} s_{kt,ij}^{-2} v_{kt \rightarrow ij}^s(\tau)$, and r can be interpreted as a μ^r -variance-AWGN corrupted observation of the true h_{ij} . Their detailed derivations are similar with the equation (9) and (15) of the paper [33].

D. Decision and Output of LSE-SMP

When the MSE of the LSE-SMP meets the requirement or the number of iterations reaches the limit, we output the final estimation of channel vector \mathbf{h}_v as

$$\begin{cases} \hat{\mathbf{h}}_v = \hat{\mathbf{Q}}^\dagger(\tau) (\bar{\mathbf{S}}\mathbf{U}_{\hat{\mathbf{b}}}(\tau))^H \bar{\mathbf{y}}, \\ \hat{\mathbf{h}}_v^* = \mathbf{U}_{\hat{\mathbf{b}}} \hat{\mathbf{b}}, \end{cases} \quad (34)$$

where $\hat{\mathbf{b}} = [\hat{b}_{ij}]_{N_r N_t \times 1}$. It should be pointed out that the final output is based on the SMP and LSE Fine Estimation.

Remark 3: The LSE fine estimation makes full use of the sparse information that is estimated by the SMP at each

iteration, which not only can significantly improve estimation performance of LSE, but also accelerate the whole algorithm to converge. In turn, the estimated result of LSE is used to improve the estimation accuracy of SMP in the next iteration. In other words, the LSE and SMP will help each other at each iteration for improving the performance of estimation until the MSE approaches the minimum or meets the system requirement. When the channel \mathbf{h}_v is more sparse, the advantage of the proposed algorithm becomes significant. Therefore, this is the another reason that the proposed algorithm is suitable for mmWave communication systems.

E. LLRs of LSE-SMP

From (24) and (25), we notice that the messages update for the variable nodes are easy to overflow in the simulations due to the multiplications of a large number of probabilities. Therefore, we use Log-Likelihood Ratios (LLRs) scheme [54], [63] to replace the computation of the non-zero probabilities during the message update process. This not only can prevent the overflow, but also can reduce the complexity of computation. The LLRs scheme can be written as follows

$$l_{kt \rightarrow ij}^s(\tau) = \log \frac{p_{kt \rightarrow ij}^s(\tau)}{1 - p_{kt \rightarrow ij}^s(\tau)}, \quad (36)$$

$$l_{ij \rightarrow kt}^v(\tau) = \log \frac{p_{ij \rightarrow kt}^v(\tau)}{1 - p_{ij \rightarrow kt}^v(\tau)}, \quad (37)$$

$$l_0 = \log \frac{P_0(b_{ij} = 1)}{1 - P_0(b_{ij} = 1)}, \quad (38)$$

for any $i \in \{1, \dots, N_r\}$, $j \in \{1, \dots, N_t\}$, $k \in \{1, \dots, N_s\}$ and τ (τ denotes the number of iterations). $l_{kt \rightarrow ij}^s(\tau)$ denotes the LLRs of the probability message of $b_{ij} = 1$ passing from the kt sum node to ij variable node. Similarly, $l_{ij \rightarrow kt}^v(\tau)$ denotes the LLRs of the probability message of $b_{ij} = 1$ passing from the ij variable node to kt sum node. Then, the LSE-SMP algorithm is updated by LLRs as follows.

1) *Message Update at Sum Nodes* : Based on (20), (21) and (23), the LLRs of Bernoulli-Gaussian messages updating at sum nodes are rewritten as

$$\begin{aligned} l_{kt \rightarrow ij}^s(\tau) = & -\log \left(\sqrt{\frac{v_{kt \rightarrow ij}^s(\tau) + s_{tj}^2 v_{h_{ij}}(\tau)}{v_{kt \rightarrow ij}^s(\tau)}} \right) \\ & - \frac{(y_{kt} - e_{kt \rightarrow ij}^s(\tau) - s_{kt,ij} \hat{h}_{ij}(\tau))^2}{2(v_{kt \rightarrow ij}^s(\tau) + s_{tj}^2 v_{h_{ij}}(\tau))} + \frac{(y_{kt} - e_{kt \rightarrow ij}^s(\tau))^2}{2v_{kt \rightarrow ij}^s(\tau)}. \end{aligned} \quad (39)$$

2) *Message Update at Variable Nodes* : Based on (24) and (25), the LLRs of messages updating at variable nodes are rewritten as

$$l_{ij \rightarrow kt}^v(\tau+1) = l_0 + \sum_{\kappa \neq k}^{N_s} l_{kt \rightarrow ij}^s(\tau), \quad (40)$$

$$l_{ij}^v(\tau+1) = l_0 + \sum_{k=1}^{N_s} l_{kt \rightarrow ij}^s(\tau), \quad (41)$$

$$\hat{b}_{ij}(\tau+1) = 1/(1 + e^{-l_{ij}^v(\tau+1)}). \quad (42)$$

where the message $l_{ij}^v(\tau+1)$ is updated based on the full information coming from all the sum nodes. It is used for calculating the $\hat{b}_{ij}(\tau+1)$.

From (39)-(42), it should be noticed that the LSE-SMP algorithm in the LLRs form will be more concise, which can reduce the complexity of computation and prevent the overflow of multiplications of a large number of probabilities, since LLRs transform the multiplication operations into the addition operations. In addition, we also can see that this will be convenient to analyze and predict the performance of the system in the following section IV.

F. LSE-SMP in the Matrix Form

In order to reduce the complexity of matrix inversions and multiplications, we can split the high dimension diagonal matrix and block diagonal matrix into some low dimension matrixs, since the transmit and receive antennas are independent each other as we mentioned before. As we defined before, we denote $i \in \{1, 2, \dots, N_r\}$, $j \in \{1, 2, \dots, N_t\}$, $k \in \{1, 2, \dots, N_s\}$ and $t \in \{1, 2, \dots, T\}$, and k denotes the τ th iteration in the following definitions. Then, we give some definitions as follows: $\mathbf{U}^s(\tau) = [e_{kt \rightarrow ij}^s(\tau)]_{N_s T \times N_r N_t}$, $\mathbf{V}^s(\tau) = [v_{kt \rightarrow ij}^s(\tau)]_{N_s T \times N_r N_t}$, $\mathbf{P}^s(\tau) = [p_{kt \rightarrow ij}^s(\tau)]_{N_s T \times N_r N_t}$, $\mathbf{L}^s(\tau) = [l_{kt \rightarrow ij}^s(\tau)]_{N_s T \times N_r N_t}$, $\mathbf{P}^v(\tau) = [p_{ij \rightarrow kt}^v(\tau)]_{N_r N_t \times N_s T}$, $\mathbf{L}^v(\tau) = [l_{ij \rightarrow kt}^v(\tau)]_{N_r N_t \times N_s T}$, $\mathbf{U}_{\hat{\mathbf{h}}_v}(\tau) = \text{diag}[\hat{h}_{ij}(\tau)]_{N_r N_t \times 1}$, $\mathbf{U}_{\hat{\mathbf{b}}}(\tau) = \text{diag}[\hat{b}_{ij}(\tau)]_{N_t N_s \times 1}$, $\hat{\mathbf{V}}_h(\tau) = \text{diag}[v_{h_{ij}}(\tau)]_{N_r N_t \times 1}$, $\bar{\mathbf{y}} = [y_{kt}]_{N_s T \times 1}$, $\hat{\mathbf{Q}}(\tau) = \mathbf{U}_{\hat{\mathbf{b}}}(\tau) \bar{\mathbf{S}} \bar{\mathbf{S}}^H \mathbf{U}_{\hat{\mathbf{b}}}(\tau)$.

In addition, we let $\mathbf{A}_{N_r \times N_t} * \mathbf{B}_{N_r \times N_t} = [a_{ij} b_{ij}]_{N_r \times N_t}$, $\mathbf{A}_{N_r \times N_t}^{(2)} = [a_{ij}^2]_{N_r \times N_t}$, $\mathbf{1}_{N_r \times N_t} = [1]_{N_r \times N_t}$, and $\mathbf{C}_{N_r \times N_r} = \mathbf{A}_{N_r \times N_t} \cdot \mathbf{B}_{N_t \times N_r}$ is the matrix product of \mathbf{A} and \mathbf{B} .

Then, the algorithm 1 shows the detailed process of the LSE-SMP in the LLRs matrix form.

IV. PERFORMANCE ANALYSIS

A. Cramér-Rao Low Bound Of LSE-SMP

In this section, we give the analysis of CRLB and show that our proposed LSE-SMP algorithm is unbiased under the assumption that we have the partial priori knowledge of the channel. Firstly, we consider the case that the channel vector \mathbf{h}_v is a deterministic and non-sparse. From [64], [65] and the signal model in (13), we can yield the CRLB of the conventional LSE as

$$\text{CRLB}_{\text{LSE}}^2 \geq \mathbf{C}_{\text{LSE}} = \sigma_n^2 (\bar{\mathbf{S}}^H \bar{\mathbf{S}})^{-1}, \quad (43)$$

where \mathbf{C}_{LSE} is the covariance matrix of \mathbf{h}_v for the LSE estimation. Note that LSE is the MVUE, and the detailed proof can be found in [64], [65]. Compared with the non-sparse case, the sparse case is slightly more complex. In particular, we are interested in the lower bound for the estimation accuracy,

² Two different CRLBs for MSE depends on the amount of knowledge of the estimators about the sparsity structure of the targeted deterministic signal [27]. Specifically, compared with the LSE, our proposed algorithm can estimate the sparse information (positions of non-zero entries) by employing the SMP and EM. Therefore, CRLBs are different for the LSE and LSE-SMP algorithm, which are denoted as CRLB_{LSE} and $\text{CRLB}_{\text{LSE-SMP}}$ respectively.

Algorithm 1 LSE-SMP Algorithm

- 1: **Input:** \mathbf{S} , T , N_s , N_r , N_t , N_{kte} , σ_n^2 , $\epsilon > 0$, $\tau = 0$, initial value $\eta = 0.5$, and calculate $\bar{\mathbf{S}}^{(2)}$ and $(\bar{\mathbf{S}}^H \bar{\mathbf{S}})^{-1}$,
- 2: **Initialized Coarse LSE Estimation:** $\hat{\mathbf{V}}_h(0) = \sigma_n^2 [\bar{\mathbf{S}} \bar{\mathbf{S}}^H]^{-1}$, $\mathbf{P}^v(0) = \mathbf{0}$, $\mathbf{U}_{\hat{\mathbf{h}}_v}(0)$ and $\mathbf{L}^v(0) = \mathbf{0}$.
- 3: **Do**
- 4: $\tilde{\mathbf{V}}^*(\tau) = (\mathbf{U}_{\hat{\mathbf{h}}_v}^{(2)}(\tau) \cdot \mathbf{P}^v(\tau)) * (\mathbf{1}_{N_s T \times N_r N_t} - \mathbf{P}^v(\tau))$
- 5: $+\hat{\mathbf{V}}_h(\tau) \cdot \mathbf{P}^v(\tau) * \bar{\mathbf{S}}^{(2)H}$, $\tilde{\mathbf{U}}^*(\tau) = \mathbf{U}_{\hat{\mathbf{h}}_v}(\tau) \cdot \bar{\mathbf{S}}^H * \mathbf{P}^v(\tau)$,
- 6: and $\mathbf{P}^v(\tau) = (\mathbf{1}_{N_r N_t \times N_s T} + e^{-\mathbf{L}^v(\tau)})^{-1}$,
- 7:
$$\begin{bmatrix} \mathbf{U}^s(\tau) \\ \mathbf{V}^s(\tau) \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{U}}^*(\tau) \cdot \mathbf{1}_{N_s T \times 1} \\ \sigma_n^2 \cdot \mathbf{1}_{N_r N_t \times 1} + \tilde{\mathbf{V}}^*(\tau) \cdot \mathbf{1}_{N_s T \times 1} \end{bmatrix} \cdot \mathbf{1}_{1 \times N_s T}$$

$$- \begin{bmatrix} \tilde{\mathbf{U}}^{*H}(\tau) \\ \tilde{\mathbf{V}}^{*H}(\tau) \end{bmatrix},$$
- 8:
$$\mathbf{L}^s(\tau) = \frac{1}{2} \log \frac{\mathbf{V}^s(\tau) + \bar{\mathbf{S}}^{(2)} \cdot \hat{\mathbf{V}}_h(\tau) + (\bar{\mathbf{y}} \cdot \mathbf{1}_{1 \times N_r N_t} - \mathbf{U}^s(\tau))^2}{\mathbf{V}^s(\tau)} - \frac{(\bar{\mathbf{y}} \cdot \mathbf{1}_{1 \times N_r N_t} - \mathbf{U}^s(\tau) - \bar{\mathbf{S}} \cdot \mathbf{U}_{\hat{\mathbf{h}}_v}(\tau))^2}{2(\mathbf{V}^s(\tau) + \bar{\mathbf{S}}^{(2)} \cdot \hat{\mathbf{V}}_h(\tau))},$$
- 9:
$$\begin{bmatrix} \mathbf{L}^v(\tau+1) \\ \mathbf{L}(\tau+1) \\ \hat{\mathbf{b}}(\tau+1) \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{N_s T \times 1} \cdot [\mathbf{1}_{1 \times N_s T} \cdot \mathbf{L}^s(\tau)] - \mathbf{L}^s(\tau) \\ \mathbf{1}_{N_s T \times 1} \cdot [\mathbf{1}_{1 \times N_s T} \cdot \mathbf{L}^s(\tau)] \\ \mathbf{1}_{N_r N_t \times N_s T} / (\mathbf{1}_{N_r N_t \times N_s T} + e^{-\mathbf{L}(\tau+1)}) \end{bmatrix}$$

$$+ \begin{bmatrix} (\mathbf{L}^s(0))^H \\ (\mathbf{L}^s(0))^H \\ 0 \end{bmatrix}.$$
- 10:
$$\begin{bmatrix} \hat{\mathbf{h}}_v(\tau) \\ \hat{\mathbf{V}}_h(\tau) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{Q}}^\dagger(\tau) (\bar{\mathbf{S}} \mathbf{U}_{\hat{\mathbf{b}}}(\tau))^H \bar{\mathbf{y}} \\ \sigma_n^2 (\hat{\mathbf{Q}}^\dagger(\tau))^\dagger \end{bmatrix},$$
- 11:
$$\hat{\eta}(\tau+1) = \frac{1}{N_t N_r} \sum_{j=1}^{N_t} \sum_{i=1}^{N_r} \frac{1}{1 + \left(\frac{\eta(\tau) \mathcal{N}(\tau; \hat{h}_{ij}(\tau), \hat{v}_{h_{ij}}(\tau) + \mu^r)}{(1 - \eta(\tau)) \mathcal{N}(\tau; 0, \mu^r)} \right)^{-1}},$$
- 12: $\tau = \tau + 1$,
- 13: **While** ($(\|\mathbf{U}_{\hat{\mathbf{h}}_v}(\tau+1) - \mathbf{U}_{\hat{\mathbf{h}}_v}(\tau)\|_2 < \epsilon$ & $\|\mathbf{L}(\tau+1) - \mathbf{L}(\tau)\|_2 < \epsilon$) or $\tau \leq N_{kte}$)
- 14:
$$\begin{bmatrix} \hat{\mathbf{h}}_v \\ \hat{\mathbf{V}}_h \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{Q}}^\dagger(\bar{\mathbf{S}} \mathbf{U}_{\hat{\mathbf{b}}})^H \bar{\mathbf{y}} \\ \mathbf{U}_{\hat{\mathbf{h}}_v} \hat{\mathbf{b}} \end{bmatrix},$$
- 15: **Output:** $\hat{\mathbf{h}}_v$ and $\hat{\mathbf{h}}_v^*$.

however, the proposed LSE-SMP algorithm involves a big loop that contains the LSE estimation, solving Gaussian functions, SMP estimation and EM learning, hence it is extremely difficult to establish an analytical model. Recalling the in (14), we know that \mathbf{h}_v can be decomposed into the two parts ($\mathbf{U}_{\hat{\mathbf{h}}_v}$ and \mathbf{b}). Therefore, we can take the way like alternating minimization method [66]–[68] to analyze these two parts independently. In other words, when we analyze the estimated performance of $\mathbf{U}_{\hat{\mathbf{h}}_v}$, we assume that the \mathbf{b} is given, and vice versa. Therefore, we have the following assumption 1.

Assumption 1: The channel \mathbf{h}_v is deterministic and we have the priori knowledge of non-zero positions \mathbf{b} .

The paper [50], [69] had already shown that the message passing-based algorithm can converge to the LMMSE under the Gaussian distribution from the theoretical perspective. In addition, we also perform a few simulations under the typical settings, and simulations show that the estimation for non-zero positions is accurate, especially with small number of non-zero entries. These simulation results was attached as the supplemental materials. Under the Assumption 1, we have

Theorem 1: The proposed LSE-SMP is the MVUE, and

can achieve the CRLB that is given by

$$\text{CRLB}_{\text{LSE-SMP}}^3 \geq \sigma_n^2 \left((\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{S}}\mathbf{U}_b \right)^\dagger. \quad (44)$$

Proof: The first step is to verify that the our proposed estimator is unbiased under the Assumption 1. Recalling the signal model in (26) and the definition of the unbiased estimator, we get

$$\begin{aligned} E(\hat{\mathbf{h}}_v) &= E \left\{ \left((\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{S}}\mathbf{U}_b \right)^\dagger (\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{y}} \right\} \\ &= E \left\{ \left((\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{S}}\mathbf{U}_b \right)^\dagger (\bar{\mathbf{S}}\mathbf{U}_b)^H (\bar{\mathbf{S}}\mathbf{U}_b \mathbf{h}_v + \bar{\mathbf{n}}) \right\} \\ &= \mathbf{U}_b \mathbf{h}_v = \mathbf{h}_v. \end{aligned} \quad (45)$$

So, it is a unbiased estimator. The next step is to compute its CRLB and verify that our proposed LSE-SMP algorithm can achieve the CRLB under the Assumption 1. As previously mentioned, the additive white Gaussian noise is modeled as $\mathcal{N}(0, \sigma_n^2 \mathbf{I})$ and the channel \mathbf{h}_v is a deterministic vector. Recalling the signal model in (12), we can get

$$\bar{\mathbf{y}} \sim p(\bar{\mathbf{y}}|\mathbf{h}_v) = \mathcal{N}(\bar{\mathbf{S}}\mathbf{h}_v, \sigma_n^2 \mathbf{I}), \quad (46)$$

where $p(\bar{\mathbf{y}}|\mathbf{h}_v)$ is the probability density function of $\bar{\mathbf{y}}$ under the condition of known \mathbf{h}_v . Then, we can compute the

$$\frac{\partial \ln p(\bar{\mathbf{y}}|\mathbf{h}_v)}{\partial \mathbf{h}_v} = \frac{1}{\sigma_n^2} \left[(\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{y}} - (\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{S}}\mathbf{U}_b \mathbf{h}_v \right]. \quad (47)$$

Then, we obtain the following expression for the Fisher Information Matrix (FIM):

$$\mathbf{I}(\mathbf{h}_v) = -E \left\{ \frac{\partial^2 \ln p(\bar{\mathbf{y}}|\mathbf{h}_v)}{\partial \mathbf{h}_v^2} \right\} = \frac{1}{\sigma_n^2} (\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{S}}\mathbf{U}_b. \quad (48)$$

We note that $\bar{\mathbf{S}}\mathbf{U}_b$ has the rank no larger than L due to the multiplication of $\bar{\mathbf{S}}$ by \mathbf{U}_b . The matrices $\bar{\mathbf{S}}\mathbf{U}_b$ and $(\bar{\mathbf{S}}\mathbf{U}_b)^H$ have some all zero columns (and rows), so it is singular. For this type singular matrix, it need to meet a constraint [71], otherwise our proposed estimator (35) has infinite variance that renders the CRLB useless. Before we analyze this constraint, we firstly compute the following key identity

$$\begin{aligned} \mathbf{G} &= \left((\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{S}}\mathbf{U}_b \right)^\dagger (\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{S}}\mathbf{U}_b \\ &= \text{diag}(\mathbf{b}) \neq \mathbf{I}_{N_r N_t}. \end{aligned} \quad (49)$$

Then, the constraint is given [71] by

$$\mathbf{G} = \mathbf{G}\mathbf{I}(\mathbf{h}_v)\mathbf{I}(\mathbf{h}_v)^\dagger. \quad (50)$$

Plugging the (48) and (49) into (50), we obtain $\mathbf{G} = \mathbf{G}\mathbf{I}(\mathbf{h}_v)\mathbf{I}(\mathbf{h}_v)^\dagger = \mathbf{U}_b$ that holds. This means that the variance of our proposed estimator is finite. Since the FIM $\mathbf{I}(\mathbf{h}_v)$ in (48) is singular, the expression for the CRLB can be computed following [27], [71], which yields,

$$\begin{aligned} \text{CRLB}_{\text{LSE-SMP}} &\geq \mathbf{C}_{\text{LSE-SMP}} = \mathbf{G}\mathbf{I}(\mathbf{h}_v)^\dagger \mathbf{G}^H \\ &= \sigma_n^2 \left((\bar{\mathbf{S}}\mathbf{U}_b)^H \bar{\mathbf{S}}\mathbf{U}_b \right)^\dagger, \end{aligned} \quad (51)$$

where $\mathbf{C}_{\text{LSE-SMP}}$ is the covariance matrix of $\hat{\mathbf{h}}_v^*$ for the LSE-SMP estimation [70]. From the CRLB theorem [65], we know that the unbiased estimator attain the CRLB if and only if

$$\frac{\partial \ln p(\bar{\mathbf{y}}|\mathbf{h}_v)}{\partial \mathbf{h}_v} = \mathbf{I}(\mathbf{h}_v) \left(\hat{\mathbf{h}}_v - \mathbf{h}_v \right), \quad (52)$$

always holds.

Plugging (34), (47) and (48) into (52), this verifies that our proposed LSE-SMP estimators is the MVUE, and can achieve the CRLB under the Assumption 1. ■

Corollary 1: For the deterministic and sparse channel, we have that the MSE of LSE estimator is the upper bound of that of the proposed LSE-SMP estimator. This also be denoted by $\text{MSE}_{\text{LSE-SMP}} \leq \text{MSE}_{\text{LSE}}$.

Proof: Firstly, we compute the MSE of the LSE estimator. It can be denoted by

$$\begin{aligned} \text{MSE}_{\text{LSE}} &= E\{\|\hat{\mathbf{h}}_v - \mathbf{h}_v\|^2\} \\ &= \text{trace}\{\mathbf{C}_{\text{LSE}}\} = \sum_{l=1}^{N_r N_t} [\mathbf{C}_{\text{LSE}}]_{l,l}, \end{aligned} \quad (53)$$

where $l \in \{1, 2, \dots, N_r N_t\}$. Similarly, we can obtain the MSE of the LSE-SMP estimator as follows,

$$\begin{aligned} \text{MSE}_{\text{LSE-SMP}} &= E\{\|\hat{\mathbf{h}}_v^* - \mathbf{h}_v\|^2\} \\ &= \text{trace}\{\mathbf{C}_{\text{LSE-SMP}}\} = \sum_{l=1}^{N_r N_t} [\mathbf{C}_{\text{LSE-SMP}}]_{l,l}. \end{aligned} \quad (54)$$

Since the channel vector is L sparse, $\mathbf{C}_{\text{LSE-SMP}}$ has no more than L eigenvalues. Furthermore, we notice that $\mathbf{C}_{\text{LSE-SMP}}$ is obtained from the full rank matrix \mathbf{C}_{LSE} by replacing $N_r N_t - L$ rows and corresponding columns with all zero entries at each index l for which $b_l = 0$. Since the $\bar{\mathbf{S}}$ is a non-singular matrix, it is easy to prove that both \mathbf{C}_{LSE} and $\mathbf{C}_{\text{LSE-SMP}}$ are the symmetric positive definite matrices, thus their all eigenvalues are greater than zero. We denote $0 < \lambda_{N_r N_t} \leq \lambda_{N_r N_t-1} \leq \dots \leq \lambda_1$ and $0 < \lambda_L^r \leq \lambda_{L-1}^r \leq \dots \leq \lambda_1^r$ as the eigenvalues of \mathbf{C}_{LSE} and $\mathbf{C}_{\text{LSE-SMP}}$ respectively. By applying the theorem 4.3.17 in [72] obtains

$$\lambda_1 \geq \lambda_1^r \geq \lambda_2 \geq \lambda_2^r \geq \dots \geq \lambda_L \geq \lambda_L^r \dots, \quad (55)$$

and therefore

$$\text{trace}\{\mathbf{C}_{\text{LSE}}\} = \sum_{l=1}^{N_r N_t} \lambda_l \geq \sum_{l=1}^L \lambda_l^r = \text{trace}\{\mathbf{C}_{\text{LSE-SMP}}\}. \quad (56)$$

Combining (56), and recalling (53) and (54), we know that $\text{MSE}_{\text{LSE-SMP}} \leq \text{MSE}_{\text{LSE}}$. ■

B. Analysis of Iterative Evolution of LSE-SMP

In this section, we will analyze the iterative evolution performance of the LSE-SMP. As we mentioned before, it is not easy to analyze the proposed algorithm directly by using the existing methods, i.e., density evolution algorithm [54], [73]. Actually, the core of the proposed algorithm is SMP algorithm, and the convergence behavior of the proposed algorithm is determined by the convergence behavior of the SMP algorithm. Therefore, following the last subsection, we also expect to leverage the alternating minimization method for ruling out the influence of LSE and EM algorithm, and focus on the SMP only. Then, we have the following assumption.

Assumption 2: We have the priori knowledge of non-zero entries of \mathbf{h}_v , and its the mean and variance are denoted as u_h and σ_h^2 respectively.

We design the training signal $\bar{\mathbf{S}}$ that is the Gaussian distribution with the zero mean and variance σ_s^2 . Many previous research results show that (40) can be well approximated by Gaussian densities (see [54]). we denote the mean and variance of message $l_{ij \rightarrow kt}^v$ at the variable node as u_v and σ_v^2 . There is an important condition, called the symmetry condition. For a Gaussian signal with the mean u_v and variance σ_v^2 , this

condition reduces to $\sigma_v^2 = 2u_v$ for the message updating at variable nodes [54], [55], [74], which means that we only need to keep the mean u_v of the message $l_{ij \rightarrow kt}^v$. Then, we have

Theorem 2: Under the assumption 2, we can obtain a closed-form update rule for variable nodes after any a few iterations. It denotes by

$$u_v(\tau+1) = l_0 + \frac{T-1}{\sqrt{4\pi u_v(\tau)}} \int_{\mathbb{R}} \zeta(l^v(\tau)) e^{-\frac{(l^v(\tau)-u_v(\tau))^2}{4u_v(\tau)}} dl^v(\tau), \quad (57)$$

with

$$\zeta(l^v(\tau)) = \frac{1}{2} \left(\frac{-\beta - a_1 snr^{-1}}{a_2 + a_1 (snr^{-1} + 1)} + \frac{\beta e^{2l^v(\tau)} + a_1 snr^{-1}}{a_2 + a_1 snr^{-1}} \right) - \log \sqrt{1 + \frac{a_1}{a_2 + a_1 snr^{-1}}}, \quad (58)$$

where k denotes the number of iterations, $snr = \frac{\sigma_h^2 \sigma_v^2}{\sigma_h^2}$, $\beta = u_h^2 \sigma_h^{-2}$, $a_1 = (1 + e^{l^v(\tau)})^2$, $a_2 = (N_t - 1) e^{l^v(\tau)} (1 + e^{l^v(\tau)} + \beta)$, and $l_{ij \rightarrow kt}^v(\tau)$ is denoted by $l^v(\tau)$.

Proof: Proof of the Theorem 2: see the Appendix. ■

Although it is difficult to obtain the exact solution of the mean u_v and variance σ_v^2 of $l^v(\tau)$, we can study and analyze the performance of SMP algorithm by simulating its Extrinsic Information Transfer (EXIT) chart [63], [69] based on the Theorem 2. This allows us to take full advantage of properties of the EXIT chart to predict the performance of our proposed algorithm. It is noted that the EXIT chart in this paper is a little difference with the EXIT chart in the paper [63] and [69]. The main difference is that we take the variances of variable nodes as the extrinsic information, while the paper [63] uses the mutual information as the extrinsic information, and the paper [69] uses the MSE as the extrinsic information. In fact, the LLRs of variances message can be transformed into mutual information and MSE directly. Furthermore, we can obtain the Corollary 2 that provides a significant criterion to judge the performance of the estimator.

Corollary 2: Assuming that the LSE-SMP has converged, if the variable nodes have the larger convergence variance σ_v^2 , then the LSE-SMP algorithm has the better estimation performance.

Proof: Based on the EXIT chart analysis technique, [63] has proved that there is a connection between the variance σ_v^2 of the variable node and the bit error rate (BER) after an arbitrary number of iterations. According to the equation (26) in the [63], the estimation of bit error probability can be approximated by the following

$$P_b \approx \frac{1}{2} \operatorname{erfc} \left(\frac{\sigma_v}{2\sqrt{2}} \right) = \frac{1}{2} \operatorname{erfc} \left(\frac{1}{2} \sqrt{u_v} \right), \quad (59)$$

where $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$ is the complementary error function, and is a monotonically decreasing and continuous on $(-\infty, +\infty)$. This means that the BER will decrease with the increase of the convergence variance σ_v^2 . Therefore, the Corollary 2 is proved. ■

Remark 4: The Theorem 2 shows the basic relationship between $u_v(\tau+1)$, $u_v(\tau)$ and some parameters, i.e., training sequence length T , snr , coefficient of variation of the channel β , etc. Moreover, the Corollary 2 provides a remarkable connection between the variance σ_v^2 of the variable node and

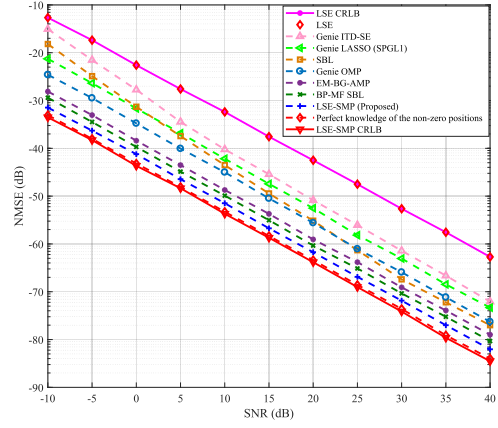


Fig. 5. This figure shows the SNR versus the average NMSE performance of various estimate algorithms. In the figure, $N_t = 32$, $N_r = 64$, $\eta = 0.007$, $\beta = 10$ with the $\sigma_h^2 = 10$, and the training sequence length $T = 64$. The result shows the proposed LSE-SMP estimator exhibits the best NMSE performance among the tested algorithms.

the performance of the algorithm, which provides a significant criterion for our proposed algorithm. Based on the Theorem 2 and Corollary 2, we take the advantage of the EXIT chart analysis to obtain a few insights on the proposed algorithm in the numerical results section.

V. NUMERICAL RESULTS

In this section, we report the results of a detailed numerical study on the performance of our proposed LSE-SMP algorithm using the Monte-Carlo simulations. For all numerical study, we considered the channel estimation problem in a 32×64 mmWave MIMO system. The channel vector \mathbf{h}_v was randomly generated based on (3) and (8), and the non-zero entries follow a Gaussian distribution. Throughout, we considered $\text{SNR} \triangleq E\{\|\tilde{\mathbf{S}}\|_F^2 / \|\mathbf{n}\|_F^2\}$ in the interval $[-10, 40] \text{ dB}$; defined the coefficients of variation of \mathbf{h}_v as $\beta = u_h^2 / \sigma_h^2$ with the $\sigma_h^2 = 10$; and the performance metric was the Normalized Mean Square Error (NMSE), given by $E\{\frac{\|\hat{\mathbf{h}}_v - \mathbf{h}_v\|_F^2}{\|\mathbf{h}_v\|_F^2}\}$. There are 500 different channel realizations and the average results are reported.

A. Performance Comparison

Fig. 5 shows the average channel estimation NMSE performance of the proposed LSE-SMP algorithm, LSE, Genie aided ITD-SE [27], genie-tuned LASSO (via SPGL1 [75]), SBL [30] (via the T-MSBL [31]), genie-tuned OMP [29], EM-BG-AMP [33] (in sparse mode) and BP-MF SBL [34]. All algorithms were run under the suggested defaults to obtain their best performance by the varied maximum number of iterations. Additionally, we also compute the CRLB for the classical LSE and the proposed LSE-SMP estimator. The result shows that our proposed LSE-SMP estimator exhibits the best NMSE performance among the tested algorithms, and reduces the NMSE by 3.5 dB relative to EM-BG-AMP and 1.5 dB relative to BP-MF SBL. As expected, the CRLB for the proposed LSE-SMP is the lowest, and this result is consistent with that of classical LSE estimator with the perfect knowledge of the non-zero positions. It is also seen from the Fig. 5 that the gap between LSE-SMP and LSE-SMP CRLB is much

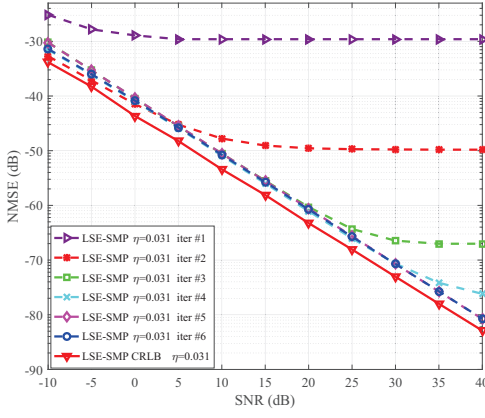


Fig. 6. This figure shows the average NMSE performance of the LSE-SMP estimator and its CRLB versus SNR under different turbo iterations. In the figure, $N_t = 32$, $N_r = 64$, $\eta = 0.031$, $\beta = 10$ with the $\sigma_h^2 = 10$, and the training sequence length $T = 64$. This result shows that the LSE-SMP algorithm reaches the convergence just need five iterations.

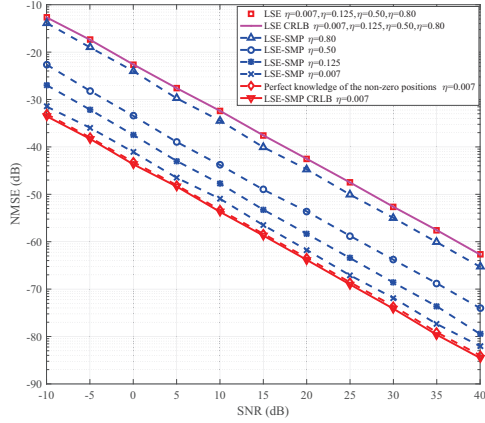


Fig. 7. This figure shows SNR versus the average NMSE performance comparison of LSE-SMP, LSE channel estimates and their CRLBs for different sparsity ratios. In the figure, $N_t = 32$, $N_r = 64$, $\eta \in \{0.007, 0.125, 0.50, 0.80\}$, $T = 64$, $\beta = 10$ with the $\sigma_h^2 = 10$, and the iterations = 6. It can be seen that the LSE-SMP CRLB is the lower bound of the proposed LSE-SMP and the LSE CRLB is the upper bound.

smaller, about 1.8dB. This gap is partly due to the errors in the detection of the non-zero positions in sparse message passing phase and partly to the fact that all our detection strategies rely on a coarse initial estimate of the channel.

B. Effect of Iterations

Fig. 6 shows the average channel estimation NMSE performance for the LSE-SMP algorithm under several turbo iterations. The result shows that the NMSE performance of the LSE-SMP algorithm will be lower with the increasing of iterations, and we also find that the gap of the NMSE performance between the adjacent iterations for the LSE-SMP algorithm will be smaller with more iterations. The main reason is that the parameters of sparsity ratio η and non-zeroes position vector \mathbf{b} are estimated more and more accurate. On the other hand, after the fifth turbo iteration, the NMSE performance have no significant improvement and it is very close to our analyzed LSE-SMP CRLB. This demonstrates that the convergence speed of the LSE-SMP algorithm is fast (just need five iterations).

C. Effect of Sparsity Ratios

For further investigating the effect of sparsity ratio to our proposed algorithm, we ran the algorithm to obtain its best performance and changed the sparsity ratio η from 0.007 to 0.80. Simulation results in Fig. 7 show that CRLB of the LSE-SMP is the lower bound and CRLB of the LSE is the upper bound. When the channel become more sparse, the performance of the proposed LSE-SMP will be better and approaches the lower bound. These results verified the analysis of Theorem 1 and Corollary 1. In addition, these results also show that the LSE-SMP is able to exploit the sparsity of the channel. To be specific, The NMSE performance of the LSE keeps unchanged under different sparsity ratios, while the NMSE of the LSE-SMP will decrease with the decrease of sparsity ratios. This indicates that the LSE-SMP scheme will be very suitable for channel estimation for mmwave systems since the channel of mmwave systems is sparse.

D. Effect of Training Sequence Length

In Fig. 8(a) and (b), we investigate the effect of different training sequence lengths by tracking the input and output LLRs of variance messages (σ_{in}^2 and σ_{out}^2) of the variable nodes of the LSE-SMP algorithm. From the results of Theorem 2 and Corollary 2, we know that the performance of the LSE-SMP algorithm can be measured by the convergence variances of the variable nodes. Therefore, we can leverage the EXIT chart based technique to analyse the system performance and design optimal parameters for some specific application scenarios. More specifically, Fig. 8(a) shows that the variance messages always converges to only one fixed point, and the longer the training sequence length, the better the estimation performance (BER). On the other hand, Fig. 8(b) shows that the increase in training sequence length will reduce the number of iteration needed, but at the cost of the overall computational time as the computational complexity of each iteration is exponential to the demission of matrix [19], [48]–[50]. Therefore, through the EXIT chart analysis, we can optimize the training sequence length to achieve a balance between the estimation performance and computational time. In addition, we also find that the shorter training sequence length will result in the narrower space between the input and output variance traces. With the decrease of training sequence length, two traces of σ_{in}^2 and σ_{out}^2 will approach to intersect. If two traces have more one intersections, this means that the proposed algorithm fails to converge. To be specific, $T = 16$ is the shortest training sequence that can make the proposed LSE-SMP work and converge under the setting of figures. Similarly, another an important function of the EXIT chart analysis is to predict the training sequence length under given the BER according to the Theorem 2 and Corollary 2.

E. Effect of the Coefficient of Variation of the Channel

Fig. 9 presents the effect of the coefficient of variation β of the channel vector \mathbf{h}_v for the proposed LSE-SMP algorithm. Then, we plotted the EXIT chart for the coefficient of variation. We can see that there is only one convergent

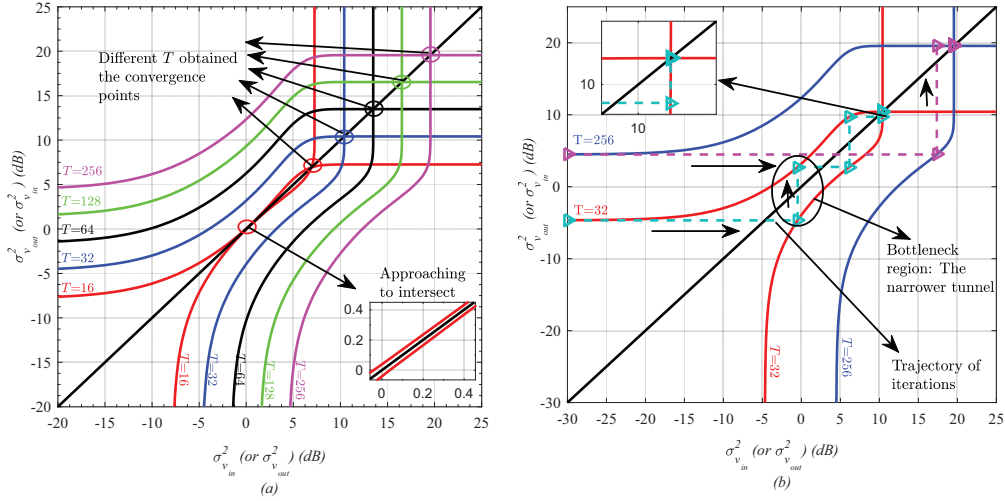


Fig. 8. This figure shows the EXIT chart analysis of the LSE-SMP algorithm, considering the case that $N_t = 32, N_r = 64, SNR = 10dB, \beta = 10$ with $\sigma_h^2 = 10$, and $\eta = 0.125$. σ_{in}^2 and σ_{out}^2 are the input and output variance messages of the variable nodes in the LLRs form. (a), $T \in \{16, 32, 64, 128, 256\}$ (b), $T \in \{32, 256\}$.

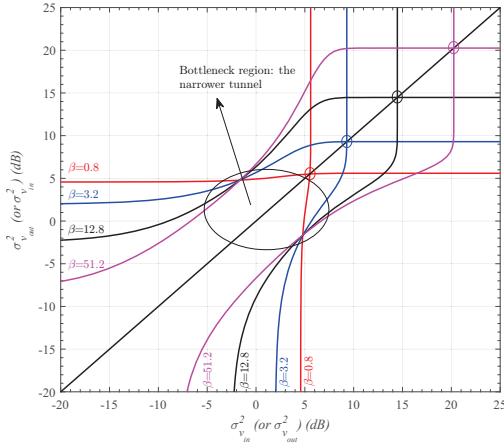


Fig. 9. This figure shows the EXIT chart analysis of the LSE-SMP algorithm with $T = 64, N_t = 32, N_r = 64$ and $SNR = 10dB$, where we set the coefficients of variation $\beta \in \{0.8, 3.2, 12.8, 51.2\}$ under the $\sigma_h^2 = 10$. The result shows that the more dispersion channel condition will leads to the lower system performance, while needs the less time to achieve the convergence.

variance point for the different coefficients of variation, and the convergent variance point will be higher with the increase of the coefficient of variation, while the *bottleneck region* that denotes the narrower region between the σ_{in}^2 and σ_{out}^2 traces keeps unchanged. Therefore, this also means that the higher variance point needs more iterations to achieve. In other words, these results indicate that the more dispersion channel condition will leads to the lower system performance, but needs the less time to achieve the convergence.

F. Complexity of the Algorithms

Fig. 10 shows the NMSE versus the runtime of various algorithms. We evaluate the runtime of each algorithm on a typical personal computer. By varying the maximum number of iterations in ITD-SE, SBL, OMP, EM-BG-AMP, BP-MF SBL and LSE-SMP algorithms, we obtained their NMSE-runtime frontier. The other two algorithms are represented by two points. We notice that the ITD-SE obtains the best performance-complexity trade-off when the NMSE is larger than $-66dB$, LASSO gives the best trade-off when the NMSE is between $-66dB$ and $-73dB$, EM-BG-AMP gives the best

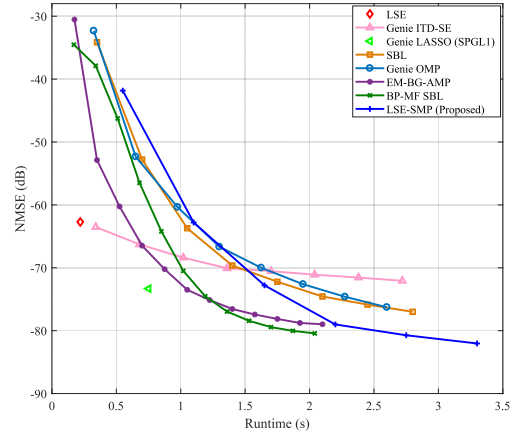


Fig. 10. This figure shows the NMSE versus the runtime of various algorithms. In the figure, $N_t = 32, N_r = 64, SNR = 40dB, \beta = 10$ with $\sigma_h^2 = 10, \eta = 0.007$, and $T = 64$. This result indicates that the proposed LSE-SMP obtains the best performance-complexity trade-off than the tested algorithms when the NMSE is less than $-80dB$.

trade-off when the NMSE is between $-73dB$ and $-76dB$, BP-MF SBL gives the best trade-off when the NMSE is between $-76dB$ and $-80dB$, and LSE-SMP is the best when the NMSE is less than $-80dB$. In other words, the proposed LSE-SMP algorithm obtains the best performance than other tested algorithms, although it may need to spend more time.

VI. CONCLUSION

We have presented a sparse channel estimation algorithm (LSE-SMP) for mmWave MIMO systems, which leverages both virtues of the SMP and LSE algorithms. We have analyzed the CRLB of the proposed LSE-SMP algorithm, and showed that the algorithm is MVUE under the assumption that we have the partial priori knowledge of the channel. Next, we have also shown an EXIT chart-based analysis technique for the convergence analysis of the key part of the proposed algorithm, and for the selection of optimal design parameters. Simulation experiments have verified that the proposed algorithm reduces the NMSE by about $2dB$ relative to the best of existing algorithms. In addition, it has been shown that the proposed algorithm typically needed only five turbo iterations

to approximately achieve CRLB. Since the correlation among the adjacent entries in the beam domain channel was ignored in this paper, a possible research direction is to take it into consideration in the future work.

APPENDIX

Then, (40) simply becomes Plugging (22) into (60), and we obtain the update as following,

$$u_v(\tau+1) = l_0 + E \left\{ \sum_{\kappa \neq k}^{N_s} \frac{(y_{\kappa t} - e_{\kappa t \rightarrow ij}^s(\tau) - s_{\kappa t, ij} \hat{h}_{ij}(\tau))^2}{-2(v_{\kappa t \rightarrow ij}^s(\tau) + s_{\kappa t, ij}^2 v_{h_{ij}}(\tau))} + \frac{(y_{\kappa t} - e_{\kappa t \rightarrow ij}^s(\tau))^2}{2(v_{\kappa t \rightarrow ij}^s(\tau))} - \log \sqrt{\frac{v_{\kappa t \rightarrow ij}^s(\tau) + s_{\kappa t, ij}^2 v_{h_{ij}}(\tau)}{v_{\kappa t \rightarrow ij}^s(\tau)}} \right\}. \quad (61)$$

In order to be convenient for analysis, we define equation (62) as follow,

Since transmit and receive antennas are independent and symmetric, we omit the subscript of $l_{ij \rightarrow kt}^v(\tau)$, and use $l^v(\tau)$ to denote the $l_{ij \rightarrow kt}^v(\tau)$ in the following paper. Due to $s_{kt, ij} \sim \mathcal{N}(0, \sigma_s^2)$, we can obtain $E(s_{kt, ij}^2) = \text{Var}(s_{kt, ij}) + E(s_{kt, ij})^2 = \sigma_s^2$. Similarly, we can get $E(n_{kt}^2) = \text{Var}(n_{kt}) + E(n_{kt})^2 = \sigma_n^2$. Plugging (20), (21) and (39) into the above expression, we obtain

$$\begin{aligned} \zeta(l^v(\tau)) = & -\log \left(\sqrt{\frac{(N_t-1)\sigma_s^2}{1+e^{-l^v(\tau)}}(\sigma_h^2 + \frac{u_h^2}{1+e^{l^v(\tau)}}) + \sigma_s^2\sigma_h^2 + \sigma_n^2} \right. \\ & \left. \frac{(N_t-1)\sigma_s^2}{1+e^{-l^v(\tau)}}(\sigma_h^2 + \frac{u_h^2}{1+e^{l^v(\tau)}}) + \sigma_n^2 \right. \\ & \left. \frac{-\sigma_s^2 u_h^2}{(1+e^{l^v(\tau)})^2} - \sigma_n^2 \right) \\ & + \frac{2 \left(\frac{(N_t-1)\sigma_s^2}{1+e^{-l^v(\tau)}}(\sigma_h^2 + \frac{u_h^2}{1+e^{l^v(\tau)}}) + \sigma_s^2\sigma_h^2 + \sigma_n^2 \right)}{\frac{\sigma_s^2 u_h^2}{(1+e^{-l^v(\tau)})^2} + \sigma_n^2} \\ & + \frac{2 \left(\frac{(N_t-1)\sigma_s^2}{1+e^{-l^v(\tau)}}(\sigma_h^2 + \frac{u_h^2}{1+e^{l^v(\tau)}}) + \sigma_n^2 \right)}{\frac{\sigma_s^2 u_h^2}{(1+e^{-l^v(\tau)})^2} + \sigma_n^2}. \end{aligned} \quad (63)$$

To simplify the expression, we yield equation (64).

where $snr = \frac{\sigma_s^2 \sigma_h^2}{\sigma_n^2}$.

The above expression looks like complex, however, we find that it is symmetric and has a few common terms. Then, we can define $\beta = u_h^2 \sigma_h^{-2}$, $a_1 = (1 + e^{l^v(\tau)})^2$, and $a_2 = (N_t - 1)e^{l^v(\tau)}(1 + e^{l^v(\tau)} + \beta)$. The $\phi(l^v(\tau))$ can be denoted by

$$\begin{aligned} \zeta(l^v(\tau)) = & \frac{1}{2} \left(\frac{-\beta - a_1 snr^{-1}}{a_2 + a_1 (snr^{-1} + 1)} + \frac{\beta e^{2l^v(\tau)} + a_1 snr^{-1}}{a_2 + a_1 snr^{-1}} \right) \\ & - \log \sqrt{1 + \frac{a_1}{a_2 + a_1 snr^{-1}}}, \end{aligned} \quad (65)$$

Since l^v is Gaussian with the mean u_v and variance $2u_v$, we obtain the following update expression by the definition of the expectation,

$$u_v(\tau+1) = l_0 + \frac{T-1}{\sqrt{4\pi u_v(\tau)}} \int_{\mathbb{R}} \zeta(l^v(\tau)) e^{-\frac{(l^v(\tau) - u_v(\tau))^2}{4u_v(\tau)}} dl^v(\tau). \quad (66)$$

Therefore, we have the Theorem 2.

ACKNOWLEDGEMENT

The work of C. Yuen was supported by the MIT-SUTD International design center and NSFC 61750110529 Grant, and that of C. Huang by the PHC Merlion PhD program. The authors would like to thank the associate editor and the reviewers for their valuable comments and suggestions.

REFERENCES

- [1] C. Huang, L. Liu, C. Yuen and S. Sun, "A LSE and Sparse Message Passing-Based Channel Estimation for mmWave MIMO Systems," *Proc. IEEE Globecom Workshops*, Washington, D.C., USA, 2016, pp. 1-6.
- [2] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work," *IEEE Access*, vol. 1, pp. 335-349, 2013.
- [3] M. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Rangan, T. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164-1179, Jun. 2014.
- [4] T. S. Rappaport, R. W. Heath Jr., R. C. Daniels, and J. Murdock, "Millimeter Wave Wireless Communications". *Prentice Hall*, Sep. 2014.
- [5] X. Gao, L. Dai and A. M. Sayeed, "Low RF-Complexity Technologies to Enable Millimeter-Wave MIMO with Large Antenna Array for 5G Wireless Communications," in *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211-217, Apr. 2018.
- [6] C. Huang, A. Zappone, M. Debbah and C. Yuen, "Achievable Rate Maximization by Passive Intelligent Mirrors," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Calgary, AB, 2018, pp. 3714-3718, Apr. 2018.
- [7] R. W. Heath Jr., N. G. Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, pp. 436-453, 2016.
- [8] K. Venugopal, A. Alkhateeb, R. W. Heath and N. G. Prelcic, "Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, New Orleans, LA, pp. 6493-6497, 2017.
- [9] J. Rodríguez-Fernández, K. Venugopal, N. González-Prelcic and R. W. Heath, "A frequency-domain approach to wideband channel estimation in millimeter wave systems," *2017 IEEE Int. Conf. on Commun. (ICC)*, Paris, 2017, pp. 1-7.
- [10] K. Venugopal, A. Alkhateeb, N. González Prelcic and R. W. Heath, "Channel Estimation for Hybrid Architecture-Based Wideband Millimeter Wave Systems," in *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996-2009, Sep. 2017.
- [11] T. S. Rappaport, G. R. MacCartney, Jr., M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design" *IEEE Trans. on Commun.*, vol. 63, no. 9, pp. 3029-3056, Sep. 2015.
- [12] S. Sun, G. R. MacCartney, Jr., and T. S. Rappaport, "Millimeter-Wave Distance-Dependent Large-Scale Propagation Measurements and Path Loss Models for Outdoor and Indoor 5G Systems," in *the 10th European Conference on Antennas and Propagation (EuCAP 2016)*, Apr. 2016.
- [13] Q. Xue, X. Fang, M. Xiao and L. Yan, "Multi-user Millimeter Wave Communications with Nonorthogonal Beams," in *IEEE Trans. on Veh. Technol.*, vol. 66, no. 7, pp. 5675-5688, Jul. 2017.
- [14] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.
- [15] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. on Signal Process.*, pp. 2563-2579, Oct. 2002.
- [16] P. Schniter and A. M. Sayeed, "A Sparseness-Preserving Virtual MIMO Channel Model," in *Proc. Conf. on Information Sciences and Systems*, (Princeton, NJ), pp. 36-41, Mar. 2004.
- [17] P. Schniter and A. Sayeed, "Channel Estimation and Precoder Design for Millimeter-Wave Communications: The Sparse Way," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, pp. 273-277, Nov. 2014.
- [18] A. Ghosh, T. Thomas, M. Cudak, R. Ratasuk, P. Moorut, F. Vook, T. Rappaport, G. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, , vol. 32, no. 6, pp.1152-1163, Jun. 2014
- [19] H. Niu, M. Shen, J. A. Ritcey and H. Liu, "A factor graph approach to iterative channel estimation and LDPC decoding over fading channels," *IEEE Trans. on Wireless Commun.*, vol. 4, pp.1345-1350, Jul. 2005.
- [20] M. F. Flanagan and A. D. Fagan, "Iterative Channel Estimation, Equalization, and Decoding for Pilot-Symbol Assisted Modulation Over Frequency Selective Fast Fading Channels," in *IEEE Trans. on Veh. Technol.*, vol. 56, no. 4, pp. 1661-1670, Jul. 2007.
- [21] Q. H. Guo, L. Ping and D. Huang, "A low-complexity iterative channel estimation and detection technique for doubly selective channels" in *IEEE Trans. on Wireless Commun.*, vol. 8, no. 8, pp. 4340-4349, August 2009.

$$E\{l_{ij \rightarrow kt}^v(\tau + 1)\} = u_v(\tau + 1) = E\left\{l_0 + \sum_{\kappa \neq k}^{N_s} l_{\kappa t \rightarrow ij}^s(\tau)\right\} \quad (60)$$

$$= E\left\{l_0 + \sum_{\kappa \neq k}^{N_s} \log \frac{\mathcal{N}(y_{kt}|e_{\kappa t \rightarrow ij}^s(\tau) + s_{\kappa t, ij} \hat{h}_{ij}(\tau), v_{\kappa t \rightarrow ij}^s(\tau) + s_{\kappa t, ij}^2 v_{h_{ij}}(\tau))}{\mathcal{N}(y_{kt}|e_{\kappa t \rightarrow ij}^s(\tau), v^s(\tau)_{\kappa t \rightarrow ij}(\tau))}\right\}.$$

$$\zeta(l_{ij \rightarrow kt}^v(\tau)) = \frac{(y_{kt} - e_{\kappa t \rightarrow ij}^s(\tau))^2}{2(v_{\kappa t \rightarrow ij}^s(\tau))} - \frac{(y_{kt} - e_{\kappa t \rightarrow ij}^s(\tau) - s_{\kappa t, ij} \hat{h}_{ij}(\tau))^2}{2(v_{\kappa t \rightarrow ij}^s(\tau) + s_{\kappa t, ij}^2 v_{h_{ij}}(\tau))} - \log \sqrt{\frac{v_{\kappa t \rightarrow ij}^s(\tau) + s_{\kappa t, ij}^2 v_{h_{ij}}(\tau)}{v_{\kappa t \rightarrow ij}^s(\tau)}}. \quad (62)$$

$$\zeta(l^v(\tau)) = -\log \left(\sqrt{1 + \frac{(1+e^{l^v(\tau)})^2}{(N_t-1)e^{l^v(\tau)}(1+e^{l^v(\tau)}+u_h^2\sigma_h^{-2})+(1+e^{l^v(\tau)})^2snr^{-1}}} \right) + \frac{1}{2} \left(\frac{-u_h^2\sigma_h^{-2} - (1+e^{l^v(\tau)})^2snr^{-1}}{(N_t-1)e^{l^v(\tau)}(1+e^{l^v(\tau)}+u_h^2\sigma_h^{-2})+(1+e^{l^v(\tau)})^2(sn r^{-1}+1)} + \frac{u_h^2\sigma_h^{-2}e^{2l^v(\tau)} + (1+e^{l^v(\tau)})^2snr^{-1}}{(N_t-1)e^{l^v(\tau)}(1+e^{l^v(\tau)}+u_h^2\sigma_h^{-2})+(1+e^{l^v(\tau)})^2snr^{-1}} \right), \quad (64)$$

- [22] Y. Zhu, D. Guo and M. L. Honig, "A message-passing approach for joint channel estimation, interference mitigation, and decoding," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 12, pp. 6008-6018, Dec. 2009.
- [23] S. Park, B. Shim and J. W. Choi, "Iterative Channel Estimation Using Virtual Pilot Signals for MIMO-OFDM Systems," in *IEEE Trans. on Signal Process.*, vol. 63, no. 12, pp. 3032-3045, Jun. 2015.
- [24] S. Wu, L. L. Kuang, Z. Y. Ni, D. Huang, Q. H. Guo and J. H. Lu, "Message-Passing Receiver for Joint Channel Estimation and Decoding in 3D Massive MIMO-OFDM Systems," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 12, pp. 8122-8138, Dec. 2016.
- [25] Q. Guo and J. Xi, "Approximate Message Passing with Unitary Transformation", Apr. 2015. [Online]. Available: <http://arxiv.org/abs/1504.04799>.
- [26] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction" in *Proc. of Inform. Theory Workshop*, (Cairo, Egypt), Jan. 2010.
- [27] C. Carbonelli, S. Vedantam and U. Mitra, "Sparse Channel Estimation with Zero Tap Detection," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 5, pp. 1743-1763, May 2007.
- [28] A. Alkhateeb, O. El Ayach, G. Leus and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," in *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831-846, Oct. 2014.
- [29] J. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 53, pp. 4655-4666, Dec. 2007.
- [30] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, pp. 2153-2164, Aug. 2004.
- [31] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, pp. 912-926, Sep. 2011.
- [32] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267-288, 1996.
- [33] J. P. Vila and P. Schniter, "Expectation-Maximization Gaussian-Mixture Approximate Message Passing," in *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658-4672, Oct. 1, 2013.
- [34] C. Zhang, Z. Yuan, Z. Wang, and Q. Guo, "Low complexity sparse Bayesian learning using combined belief propagation and mean field with a stretched factor graph," *Signal Processing*, vol. 131, pp. 344-349, 2017.
- [35] J. Mo, P. Schniter, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2014.
- [36] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge Univ. Press, 2005.
- [37] V. Raghavan and A. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 345-364, Jan. 2011.
- [38] A. Alkhateeb and R. W. Heath, "Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems," in *IEEE Trans. on Com.*, vol. 64, no. 5, pp. 1801-1818, May 2016.
- [39] C. Gustafson, K. Haneda, S. Wyne, F. Tufvesson, "On mm-wave multipath clustering and channel modeling," *IEEE Trans. Antennas and Propag.*, vol. 62, no. 3, pp. 1445-1455, Mar. 2014.
- [40] J. Brady, N. Behdad, and A. M. Sayeed, "BeamSpace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas and Propag.*, vol. 61, pp. 3814-3827, Jul. 2013.
- [41] A. Sayeed and J. Brady, "High Frequency Differential MIMO: Basic Theory and Transceiver Architectures", in *Proc. of IEEE ICC*, London, pp. 1-6, Jun. 2015.
- [42] L. Dai, X. Gao, S. Han, I. Chih-Lin and X. Wang, "BeamSpace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Proc. Int. Conf. on Commun. in China (ICCC)*, Chengdu, 2016, pp. 1-6.
- [43] A. Sayeed and J. Brady, "BeamSpace MIMO Channel Modeling and Measurement: Methodology and Results at 28GHz," *IEEE Globecom Workshops*, Washington, DC, 2016, pp. 1-6.
- [44] J. Lee, G. T. Gil and Y. H. Lee, "Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications," in *IEEE Trans. on comm.*, vol. 64, no. 6, pp. 2370-2386, Jun. 2016.
- [45] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557-1571, Apr. 2016.
- [46] Y. J. Cho, G. Suk, B. Kim, D. K. Kim and C. Chae, "RF Lens-Embedded Antenna Array for mmWave MIMO: Design and Performance," in *IEEE Comm. Mag.*, vol. 56, no. 7, pp. 42-48, Jul. 2018.
- [47] C. Huang, G. Alexandropoulos, A. Zaponne, M. Debbah, and C. Yuen, "Energy efficient multi-user MISO communication using low resolution large intelligent surfaces," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018, pp. 1-6.
- [48] H. A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping and F. R. Kschischang, "The Factor Graph Approach to Model-Based Signal Processing" in *IEEE Proc.*, vol. 95, no. 6, pp. 1295-1322, Jun. 2007.
- [49] M. Nassar, P. Schniter, and B. Evans, "A Factor-Graph Approach to Joint OFDM Channel Estimation and Decoding in Impulsive Noise Environments," *IEEE Trans. on Signal Process.*, vol. 62, no. 6, pp. 1576-1589, Mar. 2014.
- [50] L. Liu, C. Yuen, Y. L. Guan, Y. Li and Y. Su, "Convergence Analysis and Assurance for Gaussian Message Passing Iterative Detector in Massive MU-MIMO Systems," in *IEEE Trans. on Wireless Commun.*, vol. 15, no. 9, pp. 6487-6501, Sept. 2016.
- [51] Q. Guo and L. Ping, "LMMSE turbo equalization based on factor graphs" *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 311-319, 2008.
- [52] T. L. Narasimhan and A. Chockalingam, "Channel Hardening-Exploiting Message Passing (CHEMP) Receiver in Large-Scale MIMO Systems," in *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 847-860, Oct. 2014.
- [53] L. Liu, C. Yuen, Y. L. Guan, Y. Li and Y. P. Su, "A Low-Complexity Gaussian Message Passing Iterative Detection for Massive MU-MIMO Systems," in *Proc IEEE ICICS*, Singapore, Dec. 2015.
- [54] S.-Y. Chung, T. J. Richardson and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," in *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 657-670, Feb. 2001.
- [55] T. J. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, pp. 599-618, Feb. 2001.
- [56] L. Liu, C. Yuen, Y. L. Guan, Y. Li and C. Huang, "Gaussian Message Passing Iterative Detection for MIMO-NOMA Systems with Massive Access," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, 2016, pp. 1-6.
- [57] C. Huang, L. Liu and C. Yuen, "Asymptotically Optimal Estimation Algorithm for the Sparse Signal With Arbitrary Distributions," in *Trans. on Veh. Technol.*, vol. 67, no. 10, pp. 10070-10075, Oct. 2018.
- [58] Q. Su and Y. C. Wu, "Convergence analysis of the variance in Gaussian belief propagation" *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5119-5131, Oct. 2014.

- [59] L. Liu, C. Huang, Y. Chi, C. Yuen, Y. L. Guan and Y. Li, "Sparse Vector Recovery: Bernoulli-Gaussian Message Passing", *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.
- [60] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave Division Multiple-Access," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 938-947, Apr. 2006.
- [61] L. Liu, Y. Li, Y. Su, and Y. Sun, "Quantize-and-Forward Strategy for Interleave-Division Multiple-Access Relay Channel," *IEEE Trans. on Veh. Technol.* 65(3): pp.1808-1814, 2016.
- [62] T. K. Moon, "The expectation-maximization algorithm," in *IEEE Signal. Process. Mag.*, vol. 13, no. 6, pp. 47-60, Nov 1996.
- [63] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp.1727-1737, Oct. 2001.
- [64] A. Van den Bos, *Parameter estimation for scientists and engineers*. John Wiley and Sons, 2007.
- [65] Steven M. Kay., *Fundamentals of statistical signal processing: Estimation Theory*. 1st edition. Upper Saddle River, NJ: Prentice Hall PTR, 1993.
- [66] U. Niesen, D. Shah and G. W. Wornell, "Adaptive Alternating Minimization Algorithms," in *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1423-1429, Mar. 2009.
- [67] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "An alternating minimization method for sparse channel estimation," in *Proc. 9th Int. Conf. Latent Variable Anal. Signal Separation*, 2010, pp. 319–327.
- [68] X. Lu, T. Gong, P. Yan, Y. Yuan and X. Li, "Robust Alternative Minimization for Matrix Completion," in *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 42, no. 3, pp. 939-949, Jun. 2012.
- [69] K. Bhattad and K. R. Narayanan, "An MSE-based transfer chart for analyzing iterative decoding schemes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 22-38, Jan. 2007.
- [70] Couillet, R., F. Pascal, and J. W. Silverstein, "Robust Estimates of Covariance Matrices in the Large Dimensional Regim", *IEEE Trans. Inf. Theory*, vol. 60, issue 11, pp. 7269-7278, 2014.
- [71] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87-89, Jan. 2001.
- [72] R. A. Horn, C. R. Johnson, *Matrix Analysis*, 2nd ed, 2013, Cambridge Univ. Press.
- [73] A. Maleki, "Coherence analysis of iterative thresholding algorithms," , *47th Annu. Allerton Conf. on Comm., Control., and Comp.*, Monticello, IL, 2009, pp. 236-243.
- [74] T. J. Richardson, M. A. Shokrollahi and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," in *IEEE Trans. on Inf. Theory*, vol. 47, no. 2, pp. 619-637, Feb 2001.
- [75] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Scientif. Comput.*, vol. 31, no. 2, pp. 890-912, 2008.