Article | Received 25 January 2026; Revised 24 March 2026; Accepted 16 April 2026; Published 19 April 2026
https://doi.org/10.55092/aic20260006

Channel parameter estimation and localization based on structured sparse bayesian CP decomposition in mmWave massive MIMO-OFDM systems

Yizhuo Xia, Weijia Yu*, Yuyang Xu, Xin Luo and Jianhe Du

School of Information and Communication Engineering, Communication University of China, Beijing 100024, China

* Correspondence author; E-mail: yuweijia2019@cuc.edu.cn.

Highlights:

- Treats receiver-side hardware impairments as sparse noise, distinguishing it from Gaussian noise in mmWave massive MIMO-OFDM systems.
- Adopts the proposed SSBCP algorithm with latent binary variables to reduce the impact of sparse noise for robust and high-accuracy parameter extraction.
- Provides uniqueness analysis under Kruskal's condition and complexity analysis; simulations confirm robust estimation and localization under hardware impairments.

Abstract: In this paper, we propose a structured sparse Bayesian CANDECOMP/PARAFAC (SSBCP) algorithm for channel parameter estimation and localization in millimeter-wave (mmWave) massive multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) systems with receiver-side hardware impairments. Firstly, based on the physical mechanisms of receiver hardware impairments, the received signal containing both antenna-dependent sparse noise and additive Gaussian noise is constructed as a third-order parallel factor (PARAFAC) tensor. Secondly, by designing an equivalent hybrid precoding matrix, the original complex-valued tensor is transformed into a real-valued counterpart suitable for Bayesian processing. Thirdly, accurate estimation of the factor matrices is achieved through a structured sparse Bayesian tensor decomposition that incorporates binary latent variables to control the positions of sparse noise. Finally, the channel parameters are extracted from the estimated factor matrices and the localization is accomplished based on their geometric relationships. Simulation results show the proposed SSBCP algorithm outperforms existing algorithms across sparse noise ratios. Even under severe hardware distortion conditions, the proposed SSBCP algorithm maintains outstanding parameter estimation and localization performance in environments with multiple scattering paths.

Keywords: structured sparse Bayesian; parameter estimation and localization; mmWave massive MIMO-OFDM; hardware impairments; tensor decomposition

1. Introduction

Due to millimeter-wave (mmWave) massive multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) has emerged as a core technology for 5G and 6G wireless communication systems, leveraging large-scale antenna arrays and OFDM to significantly enhance spectral efficiency and mitigate multipath effects [1,2]. Accurate acquisition of channel state information (CSI) is crucial in these systems as it forms the foundation for reliable communication and high-precision localization [3]. However, the performance of massive MIMO-OFDM systems is constrained by several factors, including pilot overhead [4], hardware impairments [5,6], and electromagnetic interference [7–9], which contribute to hybrid noise effects such as sparse impulse noise and Gaussian noise. Among these, excessive pilot overhead reduces spectral efficiency, increases system complexity and computational costs, and elevates power consumption, ultimately impacting system throughput and real-time performance. Meanwhile, hardware impairments introduce various types of noise in communication systems, including phase noise, thermal noise, and distortions caused by factors such as in-phase/quadrature (IQ) imbalance, clock jitter-induced sampling errors, and intermodulation distortion [10]. More critically, these noise sources often coexist and are coupled, leading to nonlinear distortion, phase errors, and fluctuations in the radio frequency (RF) chain. These effects degrade signal quality and may result in additional bandwidth loss. Electromagnetic interference further exacerbates signal deterioration at the receiver, intensifying multipath effects and noise accumulation, which contribute to channel estimation deviations and increased positioning errors. Together, these limitations restrict spectral efficiency, system reliability, and the transmission capacity of large-scale antenna arrays.

Existing research has explored various strategies to mitigate pilot overhead in MIMO-OFDM systems, including pilot design [11–13], compressed sensing (CS) [14–18], and tensor-based approaches [19–22]. The work [11] presents a joint pilot design and channel estimation framework that further enhances pilot efficiency in OFDM systems. An efficient pilot design and gridless channel estimation strategy is developed in [12] for massive MIMO-OFDM systems, which exploits channel sparsity in the angle-delay domain to enhance accuracy and reduce overhead. Meanwhile, the work [13] introduces a gridless framework with optimized pilot design, achieving significant performance gains in estimation accuracy and pilot overhead reduction for massive MIMO-OFDM by leveraging the same sparsity property. In MIMO-OFDM systems, channel sparsity manifests as the concentration of the channel impulse response on a limited number of multipath components. CS algorithms such as orthogonal matching pursuit (OMP) [14,23], copressive sampling matching pursuit (CoSaMP) [24], sparsity adaptive matching pursuit (SAMP) [15], and simultaneous orthogonal matching pursuit (SOMP) [16] exploit this sparsity to significantly reduce pilot sequence length while maintaining estimation accuracy. In [18], CS is applied to eliminate velocity ambiguity in the pseudo-random time-division multiplexing-MIMO scheme, enhancing estimation precision. In [25], an algorithm based on CS is proposed that show promising results in lowering pilot overhead while preserving channel estimation performance in massive MIMO systems. Note that the algorithms mentioned above primarily rely on matrix representations, whereas tensor-based approaches can more effectively leverage their inherent multidimensional structure. In [20], channel parameter estimation is achieved through the classical alternating least squares (ALS) algorithm, which iteratively updates the

factor matrices until convergence. Likewise, the work [26] applies sparse recovery to localized channel modeling. Although these studies achieve notable progress in reducing pilot overhead, they are generally conducted under ideal assumptions without considering the impact of hardware impairments.

In practice, hardware impairments introduce additional distortions and mixed noise, which can severely degrade the performance of existing channel estimation algorithms and may even lead to estimation failure. To address these challenges, recent pilot allocation strategies have been developed to explicitly account for hardware impairments in cell-free massive MIMO systems [27]. Beyond pilot design, compensation techniques are also employed at both the transmitter and receiver sides. At the transmitter side of MIMO-OFDM systems, digital predistortion (DP) is widely employed to counteract nonlinear distortions caused by power amplifiers (PAs), thereby improving signal quality [28,29]. Adaptive precoding optimizes transmission signals to mitigate phase noise and IQ imbalance, while machine learning techniques can predict and compensate for RF impairments. At the receiver side, several approaches have been developed to handle hardware impairments. Blind signal recovery algorithms estimate and compensate for phase noise using statistical techniques, adaptive filtering suppresses mixed noise (e.g., sparse impulse noise and Gaussian noise), and error correction coding enhances resilience against hardware impairments. The work [30] investigates the application of Reconfigurable Intelligent Surface (RIS) in MIMO-OFDM systems to maximize spectral and energy efficiency while mitigating hardware impairments. The work [31] provides a survey on hardware impairments and compensation techniques in massive MIMO systems. Furthermore, the work [32] proposes a hardware impairment-aware transceiver design to optimize the performance of bidirectional full-duplex MIMO-OFDM systems.

Bayesian tensor decomposition provides a principled probabilistic model for channel estimation in noisy environments. By incorporating prior knowledge and modeling uncertainty, Bayesian algorithms enable robust inference of latent signal factors [33] and naturally accommodate mixed noise conditions. Bayesian CANDECOMP/PARAFAC Factorization (BCPF) in [21] leverages probabilistic modeling to handle missing data and quantify uncertainty in latent factors. For spatiotemporal traffic data, the Bayesian Gaussian CP (BGCP) tensor decomposition in [22] enables accurate imputation under temporally correlated corruption. In particular, the Bayesian model allows explicit treatment of sparse impulse noise alongside Gaussian noise, thereby enhancing estimation accuracy under hardware impairments. Recent advances such as the Bayesian sparse Kronecker product decomposition and streaming Bayesian CP decomposition algorithms [34] also demonstrate improved robustness against non-Gaussian and impulse-like noise.

However, existing algorithms often treat noise as a unified interference source, relying solely on the signal-to-noise ratio (SNR) as a performance metric without distinguishing the unique effects of different noise types. Moreover, many conventional approaches assume a Gaussian noise environment, resulting in insufficient robustness to sparse impulse noise. In this paper, we focus on receiver-side hardware impairments in mmWave massive MIMO-OFDM systems and propose a structured sparse Bayesian CP (SSBCP) algorithm for robust channel parameter estimation. By refining the noise modeling and exploiting tensor decomposition techniques, the proposed algorithm achieves accurate estimation under mixed noise conditions. The main contributions of this paper are summarized as follows:

- Considering receiver-side hardware impairments, we obtain the tensor representation of the received signal in mmWave massive MIMO-OFDM systems. These impairments induce sparse

noise that is present at certain locations within the received signal tensor. Unlike existing algorithms that treat sparse noise as part of Gaussian noise, the designed model can process sparse noise separately, thereby better leveraging its characteristics. Furthermore, tensor modeling fully utilizes the multidimensional structure of the received signal.

- We first design an equivalent hybrid precoding matrix to achieve complex-to-real tensor transformation while preserving the desired noise characteristics. Then, the structured sparse Bayesian CP decomposition of the resulting real tensor produces factor matrices that contain the channel parameters. To enhance the accuracy of this decomposition, the proposed SSBCP algorithm employs hierarchical prior and variational Bayesian principles, introducing latent binary variables to selectively activate sparse noise at partial positions, thereby achieving more accurate CP decomposition. Finally, the proposed SSBCP algorithm precisely extracts the angle of arrival (AoA), angle of departure (AoD), and delay parameters from the estimated factor matrices.
- We provide a uniqueness analysis for the constructed tensor model with theoretical guarantees under Kruskal's condition. Simulation results show that, by accounting for sparse noise characteristics, the proposed SSBCP algorithm performs posterior approximation of the received signal tensor within a variational Bayesian inference model, enabling robust estimation of channel parameters. Compared with ALS [20], BCPF [21], and BGCP [22], the proposed SSBCP algorithm achieves accurate scattering points (SPs) and mobile stations (MS) localization in non-line-of-sight (NLoS) environments, while maintaining robustness under severe hardware impairments.

The remainder of this paper is organized as follows: Section 2 presents the mmWave massive MIMO-OFDM system model with hardware impairments at the receiving antennas. Section 3 details the proposed SSBCP algorithm. Section 4 provides a proof of uniqueness for the CP decomposition, Section 5 analyzes the computational complexity, and Section 6 presents the simulation results. Finally, Section 7 concludes the paper.

Notations and Symbols: In this paper, we adopt the following notation conventions. Italic letters denote scalar values (e.g., a, A), while bold lowercase letters represent vectors (e.g., $\boldsymbol{a}$). Bold uppercase letters indicate matrices (e.g., $\boldsymbol{A}$), whereas uppercase calligraphic letters refer to tensors (e.g., $\mathscr{X}$). The sets $\mathbb{R}$ and $\mathbb{C}$ represent the real and complex number domains, respectively. Matrix operations are defined as follows: $(\cdot)^T$ represents the transpose, $(\cdot)^*$ denotes the conjugate, $(\cdot)^H$ denotes the Hermitian conjugate, and $(\cdot)^{-1}$ refers to the matrix inverse. Additionally, $\mathrm{tr}(\cdot)$ indicates the trace operation, $\|\cdot\|_2$ denotes the L_2 norm, and $\|\cdot\|_F$ stands for the Frobenius norm. $\mathrm{rank}(\boldsymbol{A})$ and k_A indicate the rank and Kruskal rank (k-rank) of $\boldsymbol{A}$, respectively. The symbol $\sqcup_n$ denotes the concatenation operation along the n-th dimension of a tensor. Furthermore, we define the following mathematical operations: $\circledast$ represents the Hadamard product, $\odot$ denotes the Khatri–Rao product, and $\circ$ corresponds to the outer product. The standard CP decomposition is formulated as $\mathscr{X} = \sum_{r=1}^{R} \boldsymbol{x}_{\cdot r}^{(1)} \circ \cdots \circ \boldsymbol{x}_{\cdot r}^{(N)} = [\![\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}]\!]$. We denote the multivariate Gaussian distribution as $\mathcal{N}(x \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the mean vector, and $\boldsymbol{\Sigma}$ is the covariance matrix. The Gamma distribution is denoted as $\mathrm{Ga}(x \mid a, b)$, where a is the shape parameter, and b is the rate parameter. The Bernoulli distribution is denoted as $\mathrm{Bernoulli}(x \mid \rho)$, where ρ is the success probability. The Beta distribution is denoted as $\mathrm{Beta}(x \mid a, b)$, where a and b are shape parameters. The notation $\mathbb{E}_q[\cdot]$ denotes the expectation with respect to the variational distribution $q(\cdot)$.

2. Channel and signal model

Consider a mmWave massive MIMO-OFDM downlink system as illustrated in Figure 1, which consists of a base station (BS) and an MS. The position of the BS, denoted by p_{BS}, is known. The MS position $p_{\mathrm{MS}} = \{x, y\} \in \mathbb{R}^2$, the clock bias B (expressed in meters), and the orientation σ are all unknown. The environment contains L SPs whose true positions $\{p_{\mathrm{SP},l}\}_{l=1}^{L}$ are uniformly distributed within a defined region. Time delay τ_l, measured in seconds, is associated with l-th path, and the corresponding physical distance is given by $d_l = c\tau_l - B$, where c is the speed of light. Moreover, we adopt the additive stochastic process model from [35] to characterize the hardware impairment at the receiver. The BS is equipped with N_{BS} antennas and M_{BS} RF chains, where $M_{\mathrm{BS}} < N_{\mathrm{BS}}$, while MS has N_{MS} antennas and N_{MS} RF chains. The channel comprises L scattering paths between the BS and MS. The mmWave Massive MIMO-OFDM system contains K training subcarriers among $\bar{K}$ total subcarriers. Over T consecutive time blocks, the BS employs a distinct beamforming vector for each subcarrier in every block. The beamforming vector for the t-th time block and the k-th subcarrier can be expressed as

$$x_k(t) = F_{\mathrm{RF}}(t) F_k(t) s_k(t), \forall k = 1, \ldots, K, \tag{1}$$

where $s_k(t) \in \mathbb{C}^{r \times 1}$ represents the transmitted data at the k-th subcarrier and t-th time block. The digital precoding matrix $F_k(t) \in \mathbb{C}^{M_{\mathrm{BS}} \times r}$ is used to adjust the transmitted signal, while the RF precoding matrix $F_{\mathrm{RF}}(t) \in \mathbb{C}^{N_{\mathrm{BS}} \times M_{\mathrm{BS}}}$ remains common across all subcarriers.

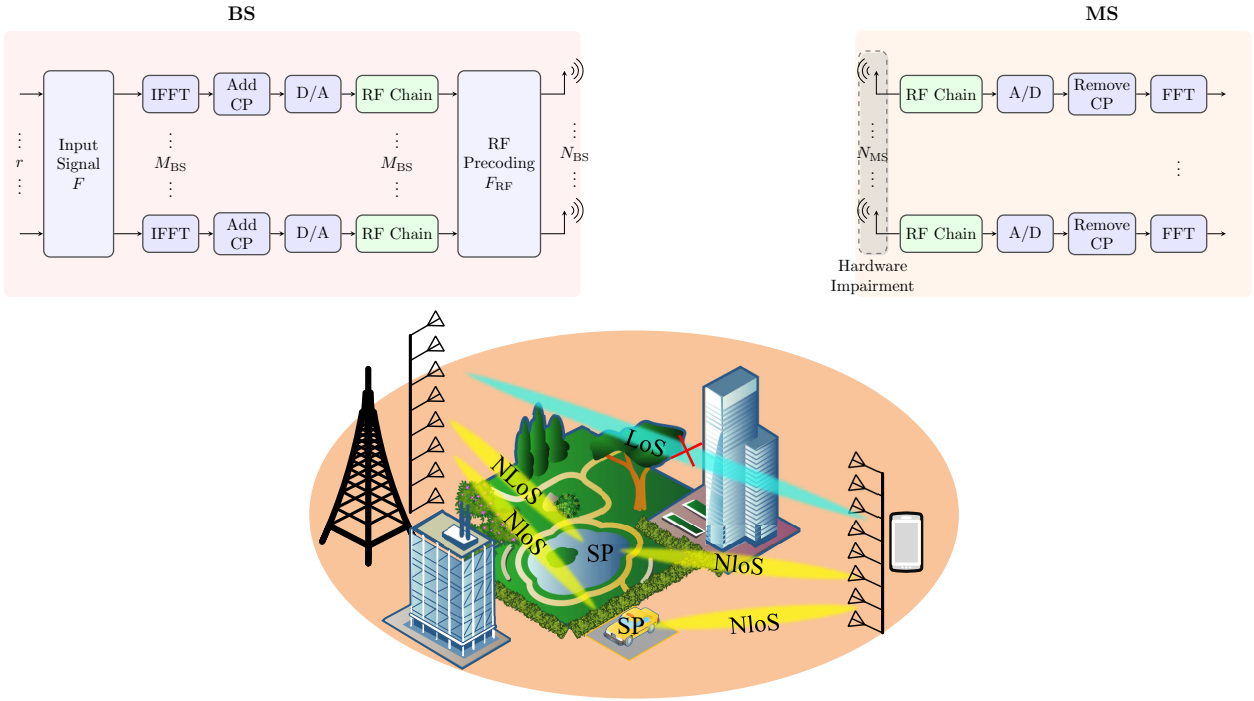

Figure 1. System model.

Then, the received signal $y_k(t) \in \mathbb{C}^{N_{\mathrm{MS}} \times 1}$ at the MS can be modeled as

$$y_k(t) = H_k x_k(t) + v_k(t) + w_k(t), \tag{2}$$

where $v_k(t)$ represents sparse noise, and $w_k(t)$ accounts for additive Gaussian noise, with the frequency-domain channel matrix H_k modeled as

$$H_k = \sum_{l=1}^{L} \alpha_l \exp(-j2\pi\tau_l f_s k/\bar{K}) a_{\mathrm{MS}}(\theta_l) a_{\mathrm{BS}}^T(\phi_l), \tag{3}$$

where α_l represents the complex gain of the l-th path, and f_s is the system sampling frequency. $a_{\mathrm{MS}}(\theta_l)$ and $a_{\mathrm{BS}}(\phi_l)$ denote the array steering vectors at the MS and BS, for the AoA θ_l and AoD ϕ_l, respectively. Substituting Equation (1) in Equation (2) gives

$$y_k(t) = H_k F_{\mathrm{RF}}(t) F_k(t) s_k(t) + v_k(t) + w_k(t). \tag{4}$$

Assuming that the digital precoding matrix and pilot symbols are identical for all subcarriers, the transmitted block at time t is defined as

$$p(t) \triangleq F_{\mathrm{RF}}(t) F(t) s(t). \tag{5}$$

Extending the system model over T time blocks yields the received signal at the k-th subcarrier as

$$Y_k = H_k P + V_k + W_k, \tag{6}$$

where

$$\begin{aligned} Y_k &\triangleq \big[y_k(1), \ldots, y_k(T)\big], \\ P &\triangleq \big[p(1), \ldots, p(T)\big], \\ V_k &\triangleq \big[v_k(1), \ldots, v_k(T)\big], \\ W_k &\triangleq \big[w_k(1), \ldots, w_k(T)\big]. \end{aligned}$$

Substituting $H_k(t)$ in Equation (6), we obtain

$$\begin{aligned} Y_k &= \sum_{l=1}^{L} \tilde{\alpha}_{l,k} a_{\mathrm{MS}}(\theta_l) a_{\mathrm{BS}}^T(\phi_l) P + V_k + W_k \\ &= \sum_{l=1}^{L} \tilde{\alpha}_{l,k} a_{\mathrm{MS}}(\theta_l) \tilde{a}_{\mathrm{BS}}(\phi_l) + V_k + W_k, \end{aligned} \tag{7}$$

where $\tilde{\alpha}_{l,k} \triangleq \alpha_l \exp(-j2\pi l f_s k/\bar{K})$, $\tilde{a}_{\mathrm{BS}}(\phi_l) \triangleq P^T a_{\mathrm{BS}}(\phi_l)$. Expanding along the subcarrier dimension K, the received signal can be expressed in tensor form as $\mathscr{Y} \in \mathbb{C}^{N_{\mathrm{MS}} \times T \times K}$

$$\mathscr{Y} = \sum_{l=1}^{L} a_{\mathrm{MS}}(\theta_l) \circ \tilde{a}_{\mathrm{BS}}(\phi_l) \circ (\alpha_l g(\tau_l)) + \mathscr{V} + \mathscr{W}, \tag{8}$$

where

$$\begin{aligned} a_{\mathrm{MS}}(\theta_l) &= \left[e^{j2\pi d\left(\frac{N_{\mathrm{MS}}-1}{2}\right)\frac{\theta_l}{\lambda}}, e^{j2\pi d\left(\frac{N_{\mathrm{MS}}-1}{2}-1\right)\frac{\theta_l}{\lambda}}, \ldots, \right. \\ &\qquad \left. e^{j2\pi d\left(-\frac{N_{\mathrm{MS}}-1}{2}\right)\frac{\theta_l}{\lambda}} \right]^T, \\ a_{\mathrm{BS}}(\phi_l) &= \left[e^{j2\pi d\left(\frac{N_{\mathrm{BS}}-1}{2}\right)\frac{\phi_l}{\lambda}}, e^{j2\pi d\left(\frac{N_{\mathrm{BS}}-1}{2}-1\right)\frac{\phi_l}{\lambda}}, \ldots, \right. \\ &\qquad \left. e^{j2\pi d\left(-\frac{N_{\mathrm{BS}}-1}{2}\right)\frac{\phi_l}{\lambda}} \right]^T, \\ g(\tau_l) &= \left[e^{(-j2\pi\tau_l f_s(1/\bar{K}))}, \ldots, e^{(-j2\pi\tau_l f_s(K/\bar{K}))} \right]^T. \end{aligned} \tag{9}$$

Furthermore, d denotes the antenna spacing, and λ is the carrier wavelength. $\mathscr{V} \in \mathbb{C}^{N_{\mathrm{MS}} \times T \times K}$ is the structured sparse noise tensor, and $\mathscr{W} \in \mathbb{C}^{N_{\mathrm{MS}} \times T \times K}$ is the additive Gaussian noise tensor. Define the following matrix as

$$\begin{aligned} A &\triangleq \left[a_{\mathrm{MS}}(\theta_1), \quad \ldots a_{\mathrm{MS}}(\theta_L)\right], \\ B &\triangleq P^T \left[a_{\mathrm{BS}}(\phi_1), \quad \ldots a_{\mathrm{BS}}(\phi_L)\right] = P^T A_{\mathrm{BS}}, \\ C &\triangleq \left[\alpha_1 g(\tau_1), \quad \ldots \alpha_L g(\tau_L)\right], \end{aligned} \tag{10}$$

then, we can rewrite Equation (8) in a more compact form as

$$\begin{aligned} \mathscr{Y} &= \mathscr{I}_L \times_1 A \times_2 P^T A_{\mathrm{BS}} \times_3 C + \mathscr{V} + \mathscr{W} \\ &= \mathscr{I}_L \times_1 A \times_2 B \times_3 C + \mathscr{V} + \mathscr{W}, \end{aligned} \tag{11}$$

where $\mathscr{I}_L \in \mathbb{R}^{L \times L \times L}$ is the identity tensor.

3. The proposed SSBCP algorithm

3.1. Real-valued representation

Due to the increased complexity of decomposing complex-valued tensors, the proposed SSBCP algorithm avoids direct processing. Instead, it transforms the complex tensor into a real-valued representation prior to structured sparse Bayesian CP decomposition, thereby enabling the extraction of factor matrices and channel parameters. Considering $a_{\mathrm{MS}}(\theta)$ and $a_{\mathrm{BS}}(\phi)$ are conjugate centrosymmetric [36], the matrices A and A_{BS} satisfy the following equation

$$\begin{cases} \Pi_{N_{\mathrm{MS}}} A^* = A \\ \Pi_{N_{\mathrm{BS}}} A_{\mathrm{BS}}^* = A_{\mathrm{BS}}. \end{cases} \tag{12}$$

To construct a real-valued tensor $\mathscr{Z} \in \mathbb{R}^{N_{\mathrm{MS}} \times T \times 2K}$, P needs to be set as a centrally Hermitian matrix [37], *i.e.*,

$$P = \Pi_{N_{\mathrm{BS}}} P^* \Pi_T. \tag{13}$$

Thus, we obtain

$$\begin{aligned} \overline{\mathscr{Y}} =& \mathscr{Y}^* \times_1 \Pi_{N_{\mathrm{MS}}} \times_2 \Pi_T \times_3 \Pi_K \\ =& \mathscr{I}_L \times_1 \Pi_{N_{\mathrm{MS}}} A^* \times_2 \Pi_T P^H A_{\mathrm{BS}}^* \times_3 \Pi_K C^* \\ &+ \overline{\mathscr{V}} + \overline{\mathscr{W}} \\ =& \mathscr{I}_L \times_1 A \times_2 P^T A_{\mathrm{BS}} \times_3 \Pi_K C^* + \overline{\mathscr{M}}, \end{aligned} \tag{14}$$

where $\mathscr{M} = \mathscr{V} + \mathscr{W}$, $\overline{\mathscr{M}} = \mathscr{M}^* \times_1 \Pi_{N_{\mathrm{MS}}} \times_2 \Pi_T \times_3 \Pi_K$, $\overline{\mathscr{V}} = \mathscr{V}^* \times_1 \Pi_{N_{\mathrm{MS}}} \times_2 \Pi_T \times_3 \Pi_K$, and $\overline{\mathscr{W}} = \mathscr{W}^* \times_1 \Pi_{N_{\mathrm{MS}}} \times_2 \Pi_T \times_3 \Pi_K$. Further, the expanded tensor $\widetilde{\mathscr{Z}} \in \mathbb{R}^{N_{\mathrm{MS}} \times T \times 2K}$ is obtained:

$$\begin{aligned} \widetilde{\mathscr{Z}} =& \mathscr{Y} \sqcup_3 \overline{\mathscr{Y}} \\ =& \left(\mathscr{I}_L \times_1 A \times_2 P^T A_{\mathrm{BS}} \times_3 C + \mathscr{M}\right) \sqcup_3 \\ & \left(\mathscr{I}_L \times_1 A \times_2 P^T A_{\mathrm{BS}} \times_3 \Pi_K C^* + \overline{\mathscr{M}}\right) \\ =& \mathscr{I}_L \times_1 A \times_2 P^T A_{\mathrm{BS}} \times_3 \begin{bmatrix} C \\ \Pi_K C^* \end{bmatrix} + \widetilde{\mathscr{M}}, \end{aligned} \tag{15}$$

where $\widetilde{\mathscr{M}} = \mathscr{M} \sqcup_3 \overline{\mathscr{M}}$. Combining Equations (13) and (14), the complex tensor $\widetilde{\mathscr{Z}}$ can be transformed into a real-valued tensor $\mathscr{Z}$ as

$$\begin{aligned}\mathscr{Z} =& \widetilde{\mathscr{Z}} \times_1 Q_{N_{\mathrm{MS}}}^H \times_2 Q_T^H \times_3 Q_{2K}^H \\ =& \left(\mathscr{I}_L \times_1 A \times_2 P^T A_{\mathrm{BS}} \times_3 \begin{bmatrix} C \\ \Pi_K C^* \end{bmatrix} \right) \\ & \times_1 Q_{N_{\mathrm{MS}}}^H \times_2 Q_T^H \times_3 Q_{2K}^H + \widehat{\mathscr{M}} \\ =& \left(\mathscr{I}_L \times_1 Q_{N_{\mathrm{MS}}}^H A \times_2 Q_T^H B \times_3 Q_{2K}^H \begin{bmatrix} C \\ \Pi_K C^* \end{bmatrix} \right) + \widehat{\mathscr{M}},\end{aligned} \tag{16}$$

where $\widehat{\mathscr{M}} = \widetilde{\mathscr{M}} \times_1 Q_{N_{\mathrm{MS}}}^H \times_2 Q_T^H \times_3 Q_{2K}^H$, and the transformation matrix Q is defined as

$$\begin{aligned} Q_{2n} &= \frac{1}{\sqrt{2}} \begin{bmatrix} I_n & jI_n \\ \Pi_n & -j\Pi_n \end{bmatrix}, \\ Q_{2n+1} &= \frac{1}{\sqrt{2}} \begin{bmatrix} I_n & 0 & jI_n \\ 0 & \sqrt{2} & 0 \\ \Pi_n & 0 & -j\Pi_n \end{bmatrix}, \end{aligned} \tag{17}$$

which n is a positive integer. For notational convenience, we define

$$X^{(1)} \triangleq Q_{N_{\mathrm{MS}}}^H A, \quad X^{(2)} \triangleq Q_T^H B, \quad X^{(3)} \triangleq Q_{2K}^H \begin{bmatrix} C \\ \Pi_K C^* \end{bmatrix}, \tag{18}$$

so that $\mathscr{Z}$ can be compactly written as

$$\mathscr{Z} = \mathscr{I}_L \times_1 X^{(1)} \times_2 X^{(2)} \times_3 X^{(3)} + \widehat{\mathscr{M}}. \tag{19}$$

3.2. Bayesian CP decomposition

Considering Equation (19), $\mathscr{Z}$ can be constructed as the following Bayesian tensor model

$$\mathscr{Z} = \mathscr{X} + \mathscr{Q} \circledast \mathscr{S} + \mathscr{E}, \tag{20}$$

where $\mathscr{X} = [\![X^{(1)}, X^{(2)}, X^{(3)}]\!]$ denotes the noise-free received signal tensor; $\mathscr{S} \in \mathbb{R}^{N_{\mathrm{MS}} \times T \times 2K}$ is the sparse noise tensor; $\mathscr{E} \sim \mathscr{N}(0, \tau^{-1})$ is an additive Gaussian noise tensor; and $\mathscr{Q} \in \mathbb{R}^{N_{\mathrm{MS}} \times T \times 2K}$ is the latent variable tensor, defined as

$$\mathscr{Q} = q \circ 1, \tag{21}$$

where $q \in \mathbb{R}^{N_{\mathrm{MS}}}$ is a latent variable vector defined along the first dimension, with each element q_{i_1} being a binary variable indicating whether an anomaly exists for the corresponding antenna

$$q_{i_1} \in \{0, 1\}, \quad i_1 = 1, \ldots, N_{\mathrm{MS}}, \tag{22}$$

where $1 \in \mathbb{R}^{T \times 2K}$ is an all-ones matrix, used to broadcast the effect of q_{i_1} across the second and third dimensions, *i.e.*, $\mathscr{Q}_{i_1 i_2 i_3} = q_{i_1}$.

The sparse noise $\mathscr{S}$ is modulated by the latent variable tensor $\mathscr{Q}$ via the Hadamard product $\circledast$, ensuring that sparse noise only affects selected antennas. Thus, the contribution of sparse noise $\mathscr{S}$ is regulated as $\mathscr{Q} \circledast \mathscr{S}$, ensuring that for any index (i_1, i_2, i_3), if $q_{i_1} = 0$, the corresponding noise component $\mathscr{S}_{i_1, i_2, i_3}$ is completely suppressed, and does not affect the observed data. However, when $q_{i_1} = 1$, the corresponding sparse noise component contributes to the data generation process. Equation (20) thereby

enforces a structured sparsity pattern where anomalous noise is activated only on a subset of antennas, while preserving the analytical tractability of the structured sparse Bayesian CP decomposition algorithm. The probabilistic graphical model of the structured sparse Bayesian CP decomposition is shown in Figure 2. For any observed data point (i_1, i_2, i_3), the model expresses the observed value as

$$\mathscr{Z}_{i_1,i_2,i_3} = \langle x_{i_1}^{(1)}, x_{i_2}^{(2)}, x_{i_3}^{(3)} \rangle + q_{i_1} \mathscr{S}_{i_1,i_2,i_3} + \mathscr{E}_{i_1,i_2,i_3}, \tag{23}$$

where $x_{i_1}^{(1)}, x_{i_2}^{(2)}, x_{i_3}^{(3)}$ are the i_1-th, i_2-th, i_3-th row vectors of the CP factor matrices $X^{(1)}, X^{(2)}, X^{(3)}$ respectively, and their inner product $\langle x_{i_1}^{(1)}, x_{i_2}^{(2)}, x_{i_3}^{(3)} \rangle$ denotes $\mathscr{X}_{i_1,i_2,i_3}$. The sparse noise component $\mathscr{S}_{i_1,i_2,i_3}$ only influences the observed data when $q_{i_1} = 1$; otherwise, its contribution is nullified.

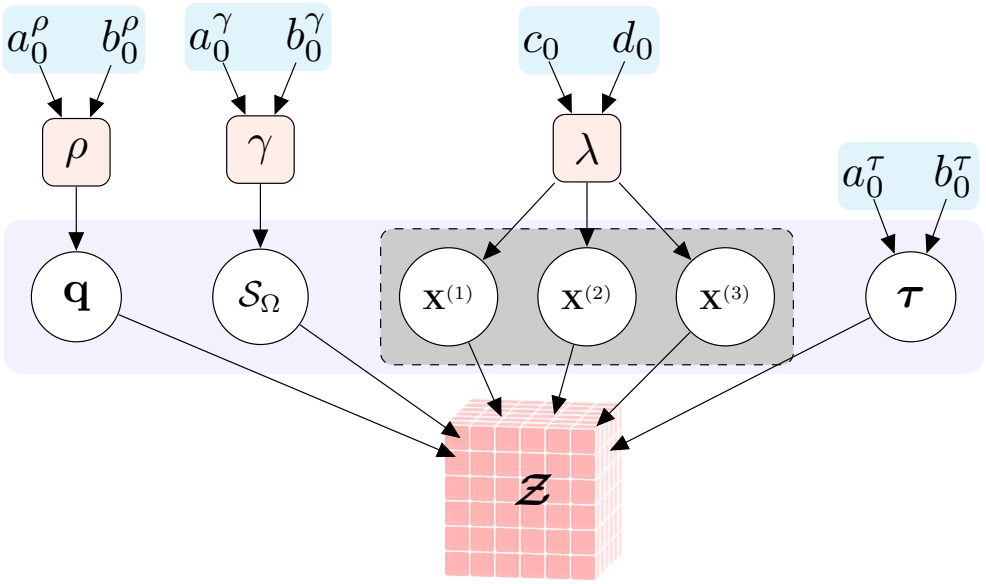

Figure 2. Probabilistic graphical model of the structured sparse bayesian CP decomposition.

(1) Prior distribution for each parameter

Prior distributions are typically determined using conjugate distributions, Jeffreys' principle, or the maximum entropy principle. To ensure closed-form posterior updates, we assume all parameters in the model follow conjugate priors. Since the noise $\mathscr{E}$ follows a Gaussian distribution $\mathscr{E}_{i_1,i_2,i_3} \sim \mathcal{N}\left(0, \tau^{-1}\right)$, the precision parameter τ is assigned a Gamma prior $p(\tau) = \mathrm{Ga}\left(\tau \mid a_0^\tau, b_0^\tau\right)$, where a_0^τ and b_0^τ denote the shape and rate hyperparameters of the prior, respectively.

For the sparse component $\mathscr{S}$, a hierarchical Gaussian prior is adopted, leading to a marginal Student-t distribution

$$p(\mathscr{S}_\Omega|\gamma) = \prod_{i_1,i_2,i_3} \mathcal{N}(\mathscr{S}_{i_1,i_2,i_3}|0, \gamma_{i_1,i_2,i_3}^{-1})^{O_{i_1,i_2,i_3}}. \tag{24}$$

Each element in the sparse noise tensor $\mathscr{S}$ has an individual precision parameter γ_{i_1,i_2,i_3}, which follows a Gamma prior as

$$p(\gamma_{i_1,i_2,i_3}) = \mathrm{Ga}\left(\gamma_{i_1,i_2,i_3} \mid a_0^\gamma, b_0^\gamma\right), \tag{25}$$

where a_0^γ and b_0^γ are shape and rate parameters, respectively. Notably, as γ_{i_1,i_2,i_3} approaches infinity, the corresponding sparse noise component $\mathscr{S}_{i_1,i_2,i_3}$ is forced to zero.

For factor matrices $X^{(n)}$, a Gaussian prior is assumed to maintain closed-form posterior updates as

$$
\begin{aligned}
p\left(X^{(1)} \mid \lambda\right) &= \prod_{i_1=1}^{N_{\mathrm{MS}}} \mathcal{N}\left(x_{i_1}^{(1)} \mid 0, \lambda^{-1}\right), \\
p\left(X^{(2)} \mid \lambda\right) &= \prod_{i_2=1}^{T} \mathcal{N}\left(x_{i_2}^{(2)} \mid 0, \lambda^{-1}\right), \\
p\left(X^{(3)} \mid \lambda\right) &= \prod_{i_n=3}^{2K} \mathcal{N}\left(x_{i_3}^{(3)} \mid 0, \lambda^{-1}\right),
\end{aligned}
\tag{26}
$$

where $\lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_R\}$ is a diagonal matrix of precision parameters. Each column precision parameter λ_r follows a Gamma prior

$$
p(\lambda) = \prod_{r=1}^{R} \mathrm{Ga}\left(\lambda_r \mid c_M^r, d_M^r\right). \tag{27}
$$

Thus, at the M-th iteration, λ follows a Gamma prior, where c_M^r and d_M^r are the prior shape and rate parameters for the r-th column, respectively.

To allow the model to automatically identify antennas affected by noise, the latent variable q_{i_1} is assigned a sparse prior, incorporating a global mixing coefficient ρ. Specifically, for each antenna i_1, we set

$$
p\left(q_{i_1} \mid \rho\right) = \mathrm{Bernoulli}\left(q_{i_1} \mid \rho\right), \tag{28}
$$

where q_{i_1} takes values 0 or 1, indicating whether an anomaly exists on the antenna. To model the prior knowledge of the global proportion of anomalous antennas, we assign a Beta hyperprior to ρ

$$
p(\rho) = \mathrm{Beta}(a_0^\rho, b_0^\rho), \tag{29}
$$

where a_0^ρ and b_0^ρ are shape parameters for the Beta distribution respectively.

Assuming all parameters and latent variables are independent in the prior, the joint probability density function is expressed as

$$
\begin{aligned}
p(\mathcal{Z}_\Omega, \Theta) = p\Big(\mathcal{Z}_\Omega, \{X^{(n)}\}_{n=1}^3, \mathcal{S}_\Omega, q, \lambda, \gamma, \tau\Big) \\
p\Big(\mathcal{Z}_\Omega \mid \{X^{(n)}\}_{n=1}^3, \mathcal{S}_\Omega, q, \tau\Big) \\
\times \prod_{n=1}^{3} p\Big(X^{(n)} \mid \lambda\Big)\, p\Big(\mathcal{S}_\Omega \mid \gamma\Big) \\
\times p(\lambda)\, p(\gamma)\, p(\tau)\, p\left(q \mid \rho\right)\, p(\rho),
\end{aligned}
\tag{30}
$$

where Ω represents the set of observed entries (*i.e.*, only locations where data is available), and $\Theta = \{X^{(1)}, X^{(2)}, X^{(3)}, \lambda, \mathcal{S}_\Omega, \gamma, \tau, q, \rho\}$ is the set of unknown parameters and latent variables. The likelihood function for observed entries (i_1, i_2, i_3) is

$$
\begin{aligned}
&p\Big(\mathcal{Z}_{i_1,i_2,i_3} \mid \{X^{(n)}\}, \mathcal{S}_{i_1,i_2,i_3}, q_{i_1}, \tau\Big) = \\
&\mathcal{N}\Big(\mathcal{Z}_{i_1,i_2,i_3} \mid \langle x_{i_1}^{(1)}, x_{i_2}^{(2)}, x_{i_3}^{(3)} \rangle + q_{i_1} \mathcal{S}_{i_1,i_2,i_3}, \tau^{-1}\Big)^{\mathcal{O}_{i_1,i_2,i_3}},
\end{aligned}
\tag{31}
$$

where $\mathcal{O}_{i_1,i_2,i_3}$ is an indicator function, which equals 1 if the corresponding entry is observed, and 0 otherwise.

(2) Posterior inference and parameter updates

Since exact Bayesian inference for our model is intractable, approximate inference methods must be

used. We employ mean-field variational inference [21,38], which seeks a factorized posterior distribution that minimizes the Kullback-Leibler (KL) divergence [39,40]:

$$\mathrm{KL}\Big(q(\Theta)\,\Big\|\,p(\Theta|\mathscr{Z}_\Omega)\Big) = \ln p(\mathscr{Z}_\Omega) - \mathscr{L}(q), \tag{32}$$

where the variational lower bound $\mathscr{L}(q)$ is defined as

$$\mathscr{L}(q) = \int q(\Theta)\ln\frac{p(\mathscr{Z}_\Omega,\Theta)}{q(\Theta)}d\Theta. \tag{33}$$

Using a general variational update formula, each parameter is sequentially updated as

$$\ln q_j(\Theta_j) = \mathbb{E}_{q_{(\Theta\setminus\Theta_j)}}\Big[\ln p\left(\mathscr{Z}_\Omega,\Theta\right)\Big] + \text{const.} \tag{34}$$

Below, we describe the specific update rules for each parameter.

(a) Posterior updates for factor matrices $X^{(n)}$: Since each row of the factor matrices is assumed independent in the prior and follows a Gaussian distribution, the posterior also remains Gaussian. Thus, for the n-th factor matrix, we use the update

$$q_n\big(X^{(n)}\big) = \prod_{i_n=1}^{I_n}\mathscr{N}\Big(x_{i_n}^{(n)} \mid \tilde{x}_{i_n}^{(n)}, V_{i_n}^{(n)}\Big). \tag{35}$$

The posterior parameters are updated as

$$\begin{aligned}\tilde{x}_{i_n}^{(n)} =& \mathbb{E}_q[\tau]V_{i_n}^{(n)}\mathbb{E}_q\Big[X_{i_n}^{(\setminus n)T}\Big] \\ & \mathrm{vec}\Big((\mathscr{Z}-\mathbb{E}_q[\mathscr{Q}\circledast\mathscr{S}])_{\mathbb{I}(\mathscr{O}_{i_n}=1)}\Big), \\ V_{i_n}^{(n)} =& \Big(\mathbb{E}_q[\tau]\mathbb{E}_q\Big[X_{i_n}^{(\setminus n)T}X_{i_n}^{(\setminus n)}\Big] + \mathbb{E}_q[\lambda]\Big)^{-1},\end{aligned} \tag{36}$$

where $X_{i_n}^{(\setminus n)}$ is the Khatri–Rao product of all factor matrices except for the n-th one, and $(\cdot)_{\mathbb{I}(\mathscr{O}_{i_n}=1)}$ denotes the observation element of the i_n-th slice. Since the random variables $\mathscr{Q}$ and $\mathscr{S}$ are independent, the expectation of their Hadamard product satisfies $\mathbb{E}[\mathscr{Q}\circledast\mathscr{S}] = \mathbb{E}[\mathscr{Q}]\circledast\mathbb{E}[\mathscr{S}]$. To improve computational efficiency, precomputed expectations such as Khatri–Rao products are utilized in the updates.

(b) Posterior Updates for Low-Rank Tensor $\mathscr{X}$: Once the factor matrices are updated, the expectation of the low-rank tensor is computed as

$$\mathbb{E}_q\big[\mathscr{X}\big] = [\![\mathbb{E}_q\big[X^{(1)}\big],\mathbb{E}_q\big[X^{(2)}\big],\mathbb{E}_q\big[X^{(3)}\big]]\!]. \tag{37}$$

(c) Posterior Updates for Precision Parameter λ: Using a Gamma prior, the posterior for each column-wise precision parameter is updated as

$$q(\lambda) = \prod_{r=1}^{R}\mathrm{Ga}\Big(\lambda_r \mid c_M^r, d_M^r\Big), \tag{38}$$

where

$$\begin{aligned}c_M^r &= c_0^r + \frac{1}{2}(N_{\mathrm{MS}}+T+2K), \\ d_M^r &= d_0^r + \frac{1}{2}\sum_{n=1}^{3}\mathbb{E}_q\Big[x_{\cdot r}^{(n)T}x_{\cdot r}^{(n)}\Big].\end{aligned} \tag{39}$$

(d) Posterior Updates for Noise Precision τ: Similarly, using a Gamma prior, the posterior update for

the noise precision τ follows

$$q(\tau) = \mathrm{Ga}(\tau \mid a_M^\tau, b_M^\tau), \tag{40}$$

where

$$\begin{aligned} a_M^\tau =& a_0^\tau + \frac{1}{2}\sum_{i_1,i_2,i_3} \mathscr{O}_{i_1,i_2,i_3}, \\ b_M^\tau =& b_0^\tau + \frac{1}{2}\mathbb{E}_q\Big[\big\|\mathscr{O} \circledast \big(\mathscr{Z} - [\![X^{(1)}, X^{(2)}, X^{(3)}]\!]\big) \\ & - \mathscr{Q} \circledast \mathscr{S}\big)\big\|_F^2\Big]. \end{aligned} \tag{41}$$

(e) Posterior Updates for Latent Variable q: Since q_{i_1} is modeled as a Bernoulli random variable, its posterior is updated based on its prior and the likelihood of observing anomalies in the data:

$$\ln q(q_{i_1}) \propto \mathbb{E}_q\Big[\ln p\big(\mathscr{Z}_{i_1,:,:} \mid q_{i_1}, \cdots\big)\Big] + \mathbb{E}_q\Big[\ln p\big(q_{i_1} \mid \rho\big)\Big]. \tag{42}$$

For each antenna i_1, we compare the likelihood ratios with and without sparse noise activation

$$\ln \frac{q(q_{i_1}=1)}{q(q_{i_1}=0)} = \mathbb{E}_q\Big[\ln \frac{\rho}{1-\rho}\Big] + \Delta_{i_1}, \tag{43}$$

where Δ_{i_1} measures the improvement in fitting the data when including sparse noise as

$$\begin{aligned} \Delta_{i_1} =& \mathbb{E}_q[\tau] \sum_{(i_2,i_3)\in\Omega_{i_1}} \Bigg(\mathbb{E}_q[\mathscr{S}_{i_1,i_2,i_3}]\big(\mathscr{Z}_{i_1,i_2,i_3} \\ & - \mathbb{E}_q[\langle x_{i_1}^{(1)}, x_{i_2}^{(2)}, x_{i_3}^{(3)}\rangle]\big) - \frac{1}{2}\big(\mathbb{E}_q[\mathscr{S}_{i_1,i_2,i_3}^2] + \sigma_{i_1,i_2,i_3}^2\big)\Bigg), \end{aligned} \tag{44}$$

with $\Omega_{i_1} = \{(i_2,i_3) : (i_1,i_2,i_3) \in \Omega\}$ denoting the observed positions for antenna i_1. Thus, the posterior probability is

$$\mathbb{E}_q[q_{i_1}] = \frac{1}{1+\exp(-\Delta_{i_1}^*)}, \tag{45}$$

where $\Delta_{i_1}^* = \mathbb{E}_q\Big[\ln \frac{\rho}{1-\rho}\Big] + \Delta_{i_1}$.

(f) Posterior Update for Global Mixing Coefficient ρ: Since ρ follows a Beta prior, its posterior remains Beta-distributed:

$$q(\rho) = \mathrm{Beta}(a_M^\rho, b_M^\rho), \tag{46}$$

where

$$\begin{aligned} a_M^\rho &= a_0^\rho + \sum_{i_1=1}^{N_{\mathrm{MS}}} \mathbb{E}_q[q_{i_1}], \\ b_M^\rho &= b_0^\rho + N_{\mathrm{MS}} - \sum_{i_1=1}^{N_{\mathrm{MS}}} \mathbb{E}_q[q_{i_1}]. \end{aligned} \tag{47}$$

(g) Posterior Updates for Sparse Noise $\mathscr{S}$: For every $(i_1,i_2,i_3) \in \Omega$, the sparse noise posterior is updated as

$$q\big(\mathscr{S}_{i_1i_2i_3}\big) = \mathscr{N}\Big(\mathscr{S}_{i_1i_2i_3} \mid \tilde{\mathscr{S}}_{i_1i_2i_3}, \sigma_{i_1i_2i_3}^2\Big). \tag{48}$$

To correctly incorporate the antenna-dependent activation of sparse noise, the posterior mean

update is adjusted as

$$\begin{aligned}\tilde{\mathscr{S}}_{i_1i_2i_3} &= \sigma^2_{i_1i_2i_3}\mathbb{E}_q[\tau]\Big(\mathscr{Z}_{i_1i_2i_3} - \mathbb{E}_q\big[\langle x^{(1)}_{i_1}, x^{(2)}_{i_2}, x^{(3)}_{i_3}\rangle\big]\Big)\mathbb{E}_q[q_{i_1}],\\ \sigma^2_{i_1i_2i_3} &= \Big(\mathbb{E}_q[\gamma_{i_1i_2i_3}] + \mathbb{E}_q[\tau]\Big)^{-1}.\end{aligned} \tag{49}$$

This ensures that if $q_{i_1} = 0$, the sparse noise component is fully suppressed, while when $q_{i_1} = 1$, the update remains unchanged.

(h) Posterior Updates for Gamma-Distributed Precision Parameter γ: For each sparse noise precision parameter γ_{i_1,i_2,i_3}, the posterior follows a Gamma distribution

$$q(\gamma_{i_1,i_2,i_3}) = \mathrm{Ga}\Big(\gamma_{i_1,i_2,i_3} \mid a_M^{\gamma_{i_1,i_2,i_3}}, b_M^{\gamma_{i_1,i_2,i_3}}\Big). \tag{50}$$

The parameters are updated as

$$\begin{aligned}a_M^{\gamma_{i_1,i_2,i_3}} &= a_0^{\gamma} + \frac{1}{2},\\ b_M^{\gamma_{i_1,i_2,i_3}} &= b_0^{\gamma} + \frac{1}{2}\Big(\tilde{\mathscr{S}}^2_{i_1,i_2,i_3} + \sigma^2_{i_1,i_2,i_3}\Big).\end{aligned} \tag{51}$$

(3) Variational Lower Bound Computation

To simultaneously monitor the changes in the approximate lower bound $\mathscr{L}(q)$ and reduce redundant computations, we calculate the variational lower bound at the end of each iteration:

$$\mathscr{L}(q) = \mathbb{E}_q\Big[\ln p\big(\mathscr{Z}_\Omega, \Theta\big)\Big] - \mathbb{E}_q\Big[\ln q(\Theta)\Big]. \tag{52}$$

We predefine a tolerance threshold tol: If $\mathscr{L}(q) < tol$ or $\Delta\mathscr{L}(q) < tol_1$, it indicates convergence of the lower bound, and the iteration is terminated. Additionally, the effective model rank R is dynamically updated based on posterior energy estimates of individual columns in the factor matrices. Specifically, for each column r of the factor matrix $X^{(n)}$: Compute its energy contribution, such as the expectation $\mathbb{E}_q[\|x^{(n)}_{:,r}\|^2]$. If the contribution falls below a predefined pruning threshold, the column is considered irrelevant, and its corresponding precision parameter λ_r is forced toward infinity. The effective rank R_{eff} is defined as the count of remaining unpruned components and is updated throughout the iteration process.

Once R_{eff} remains unchanged compared to the previous iteration and the variational lower bound satisfies the convergence criterion, the iterative process terminates. If the number of iterations exceeds a preset limit, and the rate of change in $\mathscr{L}(q)$ falls below tol or the relative change is below tol_1, while the effective rank remains less than or equal to the target rank, the iteration is forcefully terminated to prevent unnecessary computations. The initialization of the Bayesian model involves setting the factor matrices $X^{(n)}$ using either random values or singular value decomposition (SVD), where SVD-based initialization enhances convergence and estimation accuracy in tensor-based Bayesian models and mmWave massive MIMO systems [34,37]. Noise and precision hyperparameters are assigned non-informative small values, and the latent binary variable q is initialized to zero. The model follows a strict convergence criterion, requiring the variational lower bound improvement $\Delta\mathscr{L}$ to be less than 10^{-6}, maintaining a stable effective rank R_{eff} over multiple iterations, and limiting the maximum number of iterations to 2000. This empirical value was determined through experiments based on the hyperparameter settings and convergence threshold adopted from prior work [21], ensuring sufficient accuracy while maintaining reasonable runtime.

Hyperparameters play a crucial role in defining prior distributions, influencing the model's convergence behavior and estimation accuracy. The Gaussian noise precision parameter, denoteed by τ, is assigned a Gamma prior $\tau \sim \text{Gamma}(a_0^\tau = 10^{-6}, b_0^\tau = 10^{-6})$, ensuring a weakly informative prior that allows for flexible adaptation to noise levels. Similarly, each element of the sparse noise tensor $\mathscr{S}$ has an individual precision parameter γ_{i_1,i_2,i_3}, following a Gamma prior $\gamma_{i_1,i_2,i_3} \sim \text{Gamma}(a_0^\gamma = 10^{-6}, b_0^\gamma = 10^{-6})$, enabling adaptive sparsity control across different tensor entries. For the factor matrices $X^{(n)}$, the column precision parameters λ_r are assigned Gamma priors with shape parameters $c_0^r = 10^{-6}$ and rate parameters $d_0^r = 10^{-6}$, maintaining model flexibility in low-rank representation. To enhance the identification of antennas affected by sparse noise, the latent variable q_{i_1} follows a Bernoulli distribution with a global mixing parameter ρ, which itself follows a Beta prior with shape parameters $a_0^\rho = 1$ and $b_0^\rho = N_{\text{MS}}$. These hyperparameter choices collectively ensure model flexibility, robustness against varying noise conditions, and stable numerical behavior during variational inference updates, while maintaining consistency with the conjugate prior framework for efficient posterior computation.

For numerical stability, a regularization term $+\varepsilon I$ (with $\varepsilon = 10^{-8}$) is introduced in covariance matrix computations to prevent singularity issues. Additionally, during sparse noise computation, a stochastic perturbation with noise strength 0.1 is applied to prevent deterministic biases in the estimation process. These hyperparameter choices collectively ensure model flexibility, robustness against varying noise conditions, and stable numerical behavior during iterative updates.

The proposed SSBCP algorithm leverages conjugate priors and variational Bayesian inference, ensuring closed-form parameter updates. Furthermore, real-time lower-bound computation provides continuous monitoring of convergence. Automatic rank updates remove redundant factor components, improving model efficiency. Structured sparsity enforcement ensures that sparse noise appears only on selected antennas, thereby reflecting physical constraints in real-world applications.

3.3. Channel estimation

As discussed in the next subsection, the structured sparse Bayesian CP decomposition exhibits uniqueness up to scaling and permutation ambiguities under mild conditions. Specifically, the estimated factor matrices and the true factor matrices are related as follows. Assume that the real-valued tensor undergoes structured sparse Bayesian CP decomposition, yielding the estimated real-valued matrices $\hat{X}^{(1)}, \hat{X}^{(2)}, \hat{X}^{(3)}$. According to [20], we have

$$\begin{aligned} \hat{X}^{(1)} &= Q_{N_{\text{MS}}}^H A \Lambda_1 \Gamma + E_1, \\ \hat{X}^{(2)} &= Q_T^H B \Lambda_2 \Gamma + E_2, \\ \hat{X}^{(3)} &= Q_{2K}^H \begin{bmatrix} C \\ \Pi_K C^* \end{bmatrix} \Lambda_3 \Gamma + E_3, \end{aligned} \tag{53}$$

where $\Lambda_1, \Lambda_2, \Lambda_3$ are diagonal matrices constrained by $\Lambda_1 \Lambda_2 \Lambda_3 = I$, while Γ represents an unknown permutation matrix. The terms E_1, E_2, E_3 correspond to the estimation errors of the three factor matrices.

To extract the optimal angles and delay parameters from the estimated factor matrices, we evaluate the similarity between the estimated factors and predefined array beam directions. Let $\hat{a}_l$ denote the l-th column of $Q_{N_{\text{MS}}} \hat{X}^{(1)}$, $\hat{b}_l$ denote the l-th column of $Q_T \hat{X}^{(2)}$, and $\hat{c}_l$ denote the l-th column of the first K

rows of $Q_{2K}\hat{X}^{(3)}$. The AoA θ_l is estimated by maximizing the correlation between the estimated vector $\hat{a}_l$ and the MS array steering vector $a_{\mathrm{MS}}(\theta_l)$:

$$\hat{\theta}_l = \arg\max_{\theta_l} \frac{|\hat{a}_l^H a_{\mathrm{MS}}(\theta_l)|}{\|\hat{a}_l\|_2 \|a_{\mathrm{MS}}(\theta_l)\|_2}. \tag{54}$$

The AoD ϕ_l is inferred by maximizing the correlation between the estimated vector $\hat{b}_l$ and the transformed BS steering vector $\tilde{a}_{\mathrm{BS}}(\phi_l)$:

$$\hat{\phi}_l = \arg\max_{\phi_l} \frac{|\hat{b}_l^H \tilde{a}_{\mathrm{BS}}(\phi_l)|}{\|\hat{b}_l\|_2 \|\tilde{a}_{\mathrm{BS}}(\phi_l)\|_2}. \tag{55}$$

The delay τ_l is obtained by minimizing the error between the estimated delay vector $\hat{c}_l$ and the delay vector $g(\tau_l)$:

$$\hat{\tau}_l = \arg\min_{\tau_l} \frac{|\hat{c}_l^H g(\tau_l)|}{\|\hat{c}_l\|_2 \|g(\tau_l)\|_2}. \tag{56}$$

3.4. Localization

Each NLoS path provides geometric constraints that can be interpreted as two directional lines in space. One line originates from the BS and extends in the direction of the AoD, the other originates from the MS and extends in the direction of the AoA, adjusted by the MS orientation σ. These lines do not represent full trajectories but rather candidate positions for the SPs along each path.

Let p_{BS} denote the known BS position. For the l-th path, the directional point from the BS is defined as

$$s_l^{\mathrm{BS}}(\rho_1) = p_{\mathrm{BS}} + \rho_1 u(\phi_l), \tag{57}$$

where $u(\cdot) = [\cos\cdot,\ \sin\cdot]^T$ is the unit direction vector. The directional point from the MS is defined as

$$s_l^{\mathrm{MS}}(\rho_2) = p_{\mathrm{MS}} + \rho_2 u(\pi - \theta_l + \sigma). \tag{58}$$

The total propagation distance along the path is constrained by τ_l, such that

$$\rho_1 + \rho_2 = d_l = c\tau_l - B. \tag{59}$$

Due to measurement noise, the two directional lines for each path do not intersect exactly. To resolve this, we adopt a weighted least-squares formulation based on geometric consistency. For each path l, we define

$$\begin{aligned} i_l &= p_{\mathrm{BS}} - (d_l - B)\, u(\pi - \theta_l + \sigma), \\ j_l &= (d_l - B)\big(u(\phi_l) + u(\pi - \theta_l + \sigma)\big), \end{aligned} \tag{60}$$

so that the MS position satisfies $p_{\mathrm{MS}} = i_l + v_l j_l$ with $v_l \in [0,1]$. The estimate of the MS position is obtained by minimizing the sum of weighted orthogonal distances from p_{MS} to each line $i_l + v_l j_l$:

$$\widehat{p}_{\mathrm{MS}} = \Big(\sum_{l=1}^{L} \xi_l \big(I - \bar{j}_l \bar{j}_l^T\big)\Big)^{-1} \Big(\sum_{l=1}^{L} \xi_l \big(I - \bar{j}_l \bar{j}_l^T\big) i_l\Big), \tag{61}$$

where $\bar{j}_l = j_l / \|j_l\|$ and $\xi_l > 0$ is a confidence weight assigned to the l-th path, which can be chosen according to the estimated reliability of the extracted parameters.

Since the solution depends on the unknown clock bias B and orientation σ, we determine their optimal values by a two-dimensional (2D) search that minimizes the following residual function:

$$R(\sigma,B)=\frac{2}{N_sL(L-1)}\sum_{n=1}^{N_s}\sum_{l=1}^{L}\sum_{l'>l}d_{l,l'}^{(n)}, \tag{62}$$

where N_s is the number of Monte Carlo samples and $d_{l,l'}^{(n)}$ denotes the shortest Euclidean distance between the lines corresponding to paths l and l' in the n-th sample. The pair $(\widehat{\sigma},\widehat{B})$ that minimizes $R(\sigma,B)$ is used to compute the final MS position estimate via the weighted least-squares solution above.

Once the estimated MS position $\widehat{p}_{\mathrm{MS}}$, estimated clock bias $\widehat{B}$, and estimated orientation $\widehat{\sigma}$ are obtained, the position of the l-th SP can be estimated as the point that lies closest to both the BS-originating ray and the MS-originating ray. The closed-form estimate is given by

$$\widehat{p}_{\mathrm{SP},l}=\left(C_{\mathrm{BS},l}+C_{\mathrm{MS},l}\right)^{-1}\left(C_{\mathrm{BS},l}\,p_{\mathrm{BS}}+C_{\mathrm{MS},l}\,\widehat{p}_{\mathrm{MS}}\right), \tag{63}$$

where $C_{\mathrm{BS},l}=I-u(\phi_l)u(\phi_l)^T$ and $C_{\mathrm{MS},l}=I-u(\pi-\theta_l+\sigma)u(\pi-\theta_l+\sigma)^T$.

Algorithm 1 summarizes the complete procedure of the proposed SSBCP algorithm.

Algorithm 1 The Proposed SSBCP Algorithm

1: Construct the received tensor signal $\mathscr{Y}\in\mathbb{C}^{N_{\mathrm{MS}}\times T\times K}$ using Equation (8);
2: Obtain real-valued tensor $\mathscr{Z}\in\mathbb{C}^{N_{\mathrm{MS}}\times T\times 2K}$ using Equations (12)–(19);
3: **structured sparse Bayesian CP decomposition:**
  3.1: **Initialization**: Initialize factor matrices $\{X^{(n)}\}_{n=1}^{3}$,
    sparse noise term $\mathscr{S}$; and hyperparameters λ, γ, τ, ρ;
    initialize latent variable q;
  Repeat;
  3.2: **For** $n=1$ to 3
      Update factor matrix $X^{(n)}$ using Equation (36);
    EndFor
  3.3: Compute $\mathbb{E}_q[\mathscr{X}]$ using Equation (35);
  3.4: Update λ using Equation (38);
  3.5: Update τ using Equation (40);
  3.6: Update q and ρ using Equations (45) and (46);
  3.7: Update $\mathscr{S}$ and its precision γ using Equations (48) and (50);
  3.8: Recalculate the variational lower bound $\mathscr{L}(q)$ using Equation (52);
  3.9: Prune inactive components and dynamically adjust model rank R;
  Until the lower bound converges and the effective rank stabilizes;
  3.10: Extract posterior mean estimates of factor matrices $\hat{X}^{(n)}$
    and latent variables q; infer missing data based on the posterior distribution of $\mathscr{S}$;
4: Estimate channel parameters $\{\hat{\theta}_l,\hat{\phi}_l,\hat{\tau}_l\}_{l=1}^{L}$ from estimated factor matrices with Equations (54)–(56);
5: Estimate MS position $\widehat{p}_{\mathrm{MS}}$ and SP positions $\{\widehat{p}_{\mathrm{SP},l}\}_{l=1}^{L}$ using Equations (61) and (63).

4. Uniqueness

Kruskal's condition is widely recognised as a sufficient condition for the essential uniqueness of the CP decomposition. For the real-valued tensor $\mathscr{Z}$, Kruskal's condition [41] is satisfied

$$k_{X^{(1)}} + k_{X^{(2)}} + k_{X^{(3)}} \geq 2L + 2, \tag{64}$$

where $k_{X^{(1)}}, k_{X^{(2)}}, k_{X^{(3)}}$ denote the k-ranks of factor matrices $X^{(1)}, X^{(2)}, X^{(3)}$, respectively, the structured sparse Bayesian CP decomposition is essentially unique up to column permutation and scaling ambiguities, which do not affect the parameter extraction from factor matrices. When θ_l is distinct, based on the full k-rank characteristic of $Q_{N_{\mathrm{MS}}}$, the k-rank of $X^{(1)}$ satisfies [37]: $k_{X^{(1)}} = k_{Q_{N_{\mathrm{MS}}}^H A} = k_A \geq \min(N_{\mathrm{MS}}, L)$. Similarly, when ϕ_l is distinct and rank $(P) \geq L$, the k-rank of $X^{(2)}$ satisfies: $k_{X^{(2)}} = k_{Q_T^H (P^T A_{\mathrm{BS}})} = k_{P^T A_{\mathrm{BS}}} \geq \min(T, L)$. When τ_l is distinct, the k-rank of $X^{(3)}$ satisfies $k_{X^{(3)}} \geq \min(2K, L)$. If the system parameters satisfy

$$N_{\mathrm{MS}} \geq L, \quad T \geq L, \quad 2K \geq L, \tag{65}$$

then Equation (64) can be rewritten as

$$\min(N_{\mathrm{MS}}, T, 2K, L) \geq 2. \tag{66}$$

5. Computational complexity analysis

This section analyzes the computational complexity per variational iteration of the SSBCP algorithm. The dominant cost stems from factor updates and sparse-noise updates. Updating each factor involves applying the Khatri–Rao product of the other two factors to the unfolded residual; a direct implementation incurs $\mathcal{O}\big(LN_{\mathrm{MS}}T(2K)\big)$. By leveraging Gram precomputation and Hadamard products, the explicit Khatri–Rao construction is replaced by a Gram computation of $\mathcal{O}\big(L^2(N_{\mathrm{MS}} + T + 2K)\big)$, accompanied by the mean-update term of $\mathcal{O}\big(LN_{\mathrm{MS}}T(2K)\big)$. Each factor update also requires inverting an $L \times L$ posterior covariance matrix. If a distinct inverse is computed for every row, the cost accumulates to $\mathcal{O}\big(L^3(N_{\mathrm{MS}} + T + 2K)\big)$; this reduces to $\mathcal{O}(L^3)$ when a common covariance per mode is reused or approximated. Sparse noise and binary activation updates process each observed entry, contributing $\mathcal{O}\big(N_{\mathrm{MS}}T(2K)\big)$. Hyperparameter updates and variational bound evaluation are of lower order. Combining these contributions yields the conservative per-iteration bound $\mathcal{O}\big(L^3(N_{\mathrm{MS}} + T + 2K) + L^2(N_{\mathrm{MS}} + T + 2K) + LN_{\mathrm{MS}}T(2K)\big)$.

6. Simulation results

This section evaluates the performance of the proposed SSBCP algorithm for parameter extraction and localization with the root mean square error (RMSE). The RMSE of the parameter estimate is defined as

$$\mathrm{RMSE}(\psi) = \sqrt{\frac{1}{L}\sum_{l=1}^{L}(\widehat{\psi}_l - \psi_l)^2}, \tag{67}$$

where $\psi = \{\theta, \phi, \tau\}$. Besides, the RMSE of the MS position estimate $\widehat{p}_{\mathrm{MS}}^{(i)}$ is defined as

$$\mathrm{RMSE}_{\mathrm{MS}} = \sqrt{\frac{1}{N_{\mathrm{MC}}}\sum_{i=1}^{N_{\mathrm{MC}}}\left\|\widehat{p}_{\mathrm{MS}}^{(i)} - p_{\mathrm{MS}}^{(i)}\right\|^2}, \tag{68}$$

where $\widehat{p}_{\mathrm{MS}}^{(i)}$ and $p_{\mathrm{MS}}^{(i)}$ denote the estimated and true MS positions in the i-th Monte Carlo trial respectively, and N_{MC} is the total number of trials. The RMSE of the SP position estimate is given by

$$\mathrm{RMSE}_{\mathrm{SP}} = \sqrt{\frac{1}{N_{\mathrm{MC}}}\sum_{i=1}^{N_{\mathrm{MC}}}\frac{1}{L}\sum_{l=1}^{L}\left\|\widehat{p}_{\mathrm{SP},l}^{(i)} - p_{\mathrm{SP},l}^{(i)}\right\|^2}. \tag{69}$$

The proposed SSBCP algorithm is compared with the ALS algorithm [20], BCPF algorithm [21], and BGCP algorithm [22]. The following parameters remain consistent across all compared algorithms and experiments: $N_{\mathrm{BS}} = 64$, $N_{\mathrm{MS}} = 8$, $T = 12$, and $N_{\mathrm{MC}} = 100$ Monte Carlo trials.

In the first experiment, we investigate the impact of the number of SP, *i.e.*, L, on the estimation performance. The parameter settings are as follows: $L \in \{2,3,4\}$ while fixing the number of subcarriers at $K = 12$ and the sparse noise ratio at SR = 0.01. Figure 3 gives the RMSE performance for different algorithms in estimating AoA/AoD and delay parameters across varying SNR levels for different L. As Figure 3 shows that the RMSE of all algorithms decreases with increasing SNR, demonstrating improved estimation accuracy under lower noise conditions. Under the more challenging condition of $L = 4$, the performance of the ALS algorithm deteriorates significantly. In contrast, Bayesian algorithms, including the proposed SSBCP algorithm, demonstrate robust stability. Overall, the proposed SSBCP algorithm remains stable for target numbers $L = \{2,3,4\}$, its estimation error is consistently lower than that of all other algorithms under comparison. The superior performance confirms the effectiveness and reliability of the proposed SSBCP algorithm for parameter extraction in challenging multi-path environments.

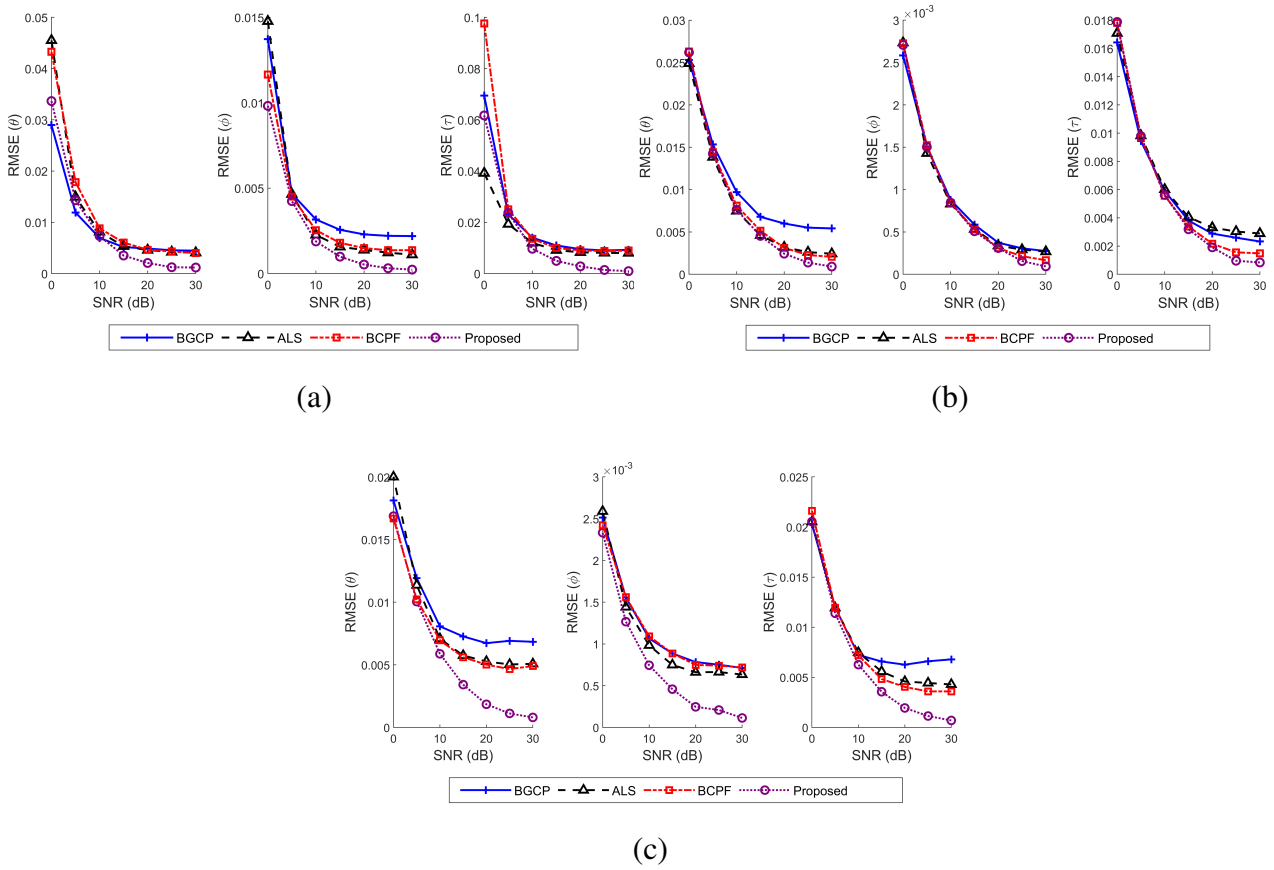

Figure 3. RMSE *versus* SNR for different K and SR; **(a)** $L = 2$, $K = 16$, SR = 0.01; **(b)** $L = 2$, $K = 8$, SR = 0.01; **(c)** $L = 2$, $K = 8$, SR = 0.05.

In the second experiment, we consider the effects of both the number of subcarriers K and the sparse noise ratio SR on algorithm performance under four distinct configurations: (a) $L = 2$, $K = 16$, SR = 0.01; (b) $L = 2$, $K = 8$, SR = 0.01; and (c) $L = 2$, $K = 8$, SR = 0.05. As demonstrated in Figure 4, the proposed SSBCP algorithm consistently achieves convergence while accurately recovering the target AoA/AoD and delay parameters for L targets, regardless of whether the subcarrier number is halved or doubled. This

robust performance demonstrates the proposed SSBCP algorithm's excellent adaptability to dimensional expansion in tensor processing and confirms its stability against subcarrier quantity variations. Notably, BGCP maintains a relatively high RMSE even at high SNR, suggesting weaker stability in low-noise environments. ALS demonstrates a more consistent downward trend across mid-to-high SNR ranges, showing stronger robustness under moderate noise. In contrast, the proposed SSBCP algorithm exhibits the most significant RMSE reduction across the entire range, particularly in low-SNR scenarios, where it effectively suppresses noise and maintains higher precision. This demonstrates the proposed SSBCP algorithm's superior capability in handling sparse noise and signal decomposition, ensuring high accuracy even in challenging low-SNR environments.

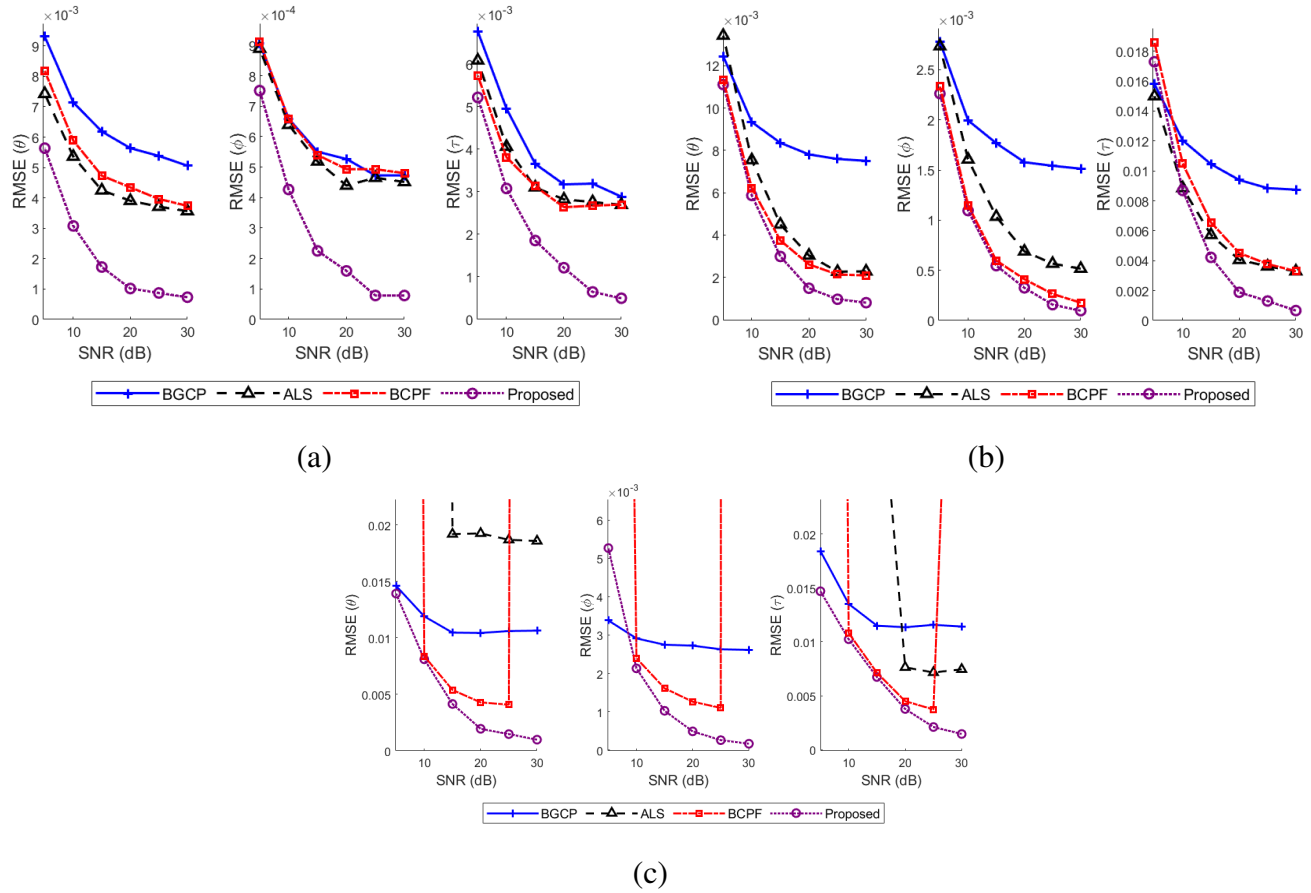

Figure 4. RMSE *versus* SNR for different L, where $K = 12$ and SR = 0.01; **(a)** $L = 2$; **(b)** $L = 3$; **(c)** $L = 4$.

In the third experiment, to further evaluate the localization performance of the proposed SSBCP algorithm under NLoS conditions, we design a 2D localization experiment based on SPs. In the experiment, the SPs are uniformly sampled within the region $[0, 160] \times [-50, 20]$ meters. The MS orientation σ is set to zero, and the clock bias B is set to 20 meters. All experiments are conducted at a fixed SNR, *i.e.*, 25 dB. Two configurations are tested: one uses $L = 3$, and the other uses $L = 5$. As shown in Figure 5, with 3 NLoS SPs, the proposed SSBCP algorithm combined with geometric consistency achieves reliable and accurate positioning. Even when the number of SPs is increased to 5, the proposed SSBCP algorithm maintains a high level of localization accuracy, as demonstrated in Figure 6. These results demonstrate that the proposed SSBCP algorithm delivers robust parameter estimation and localization for both MS and SPs under NLoS conditions in both tested configurations.

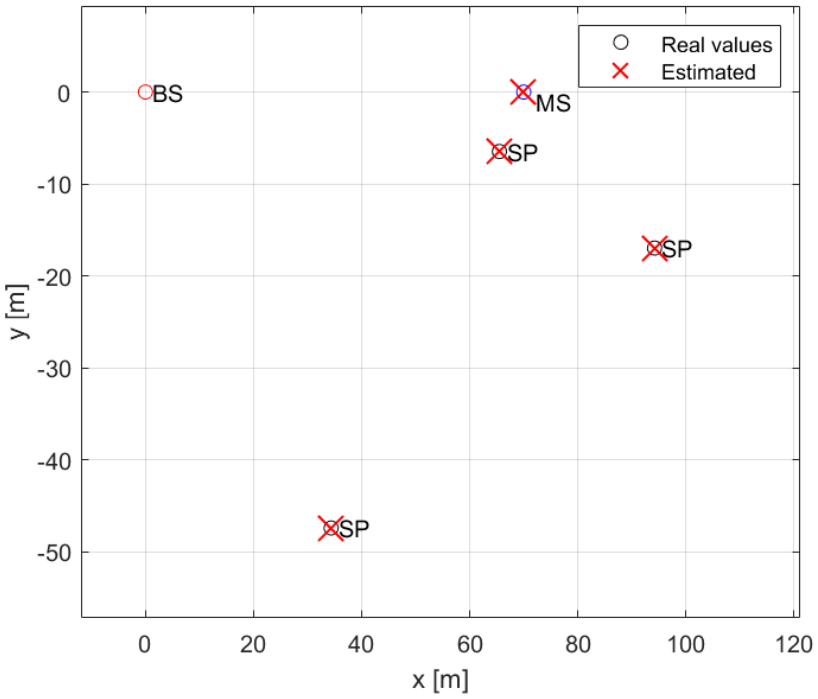

Figure 5. Localization RMSE with 3 NLoS SPs.

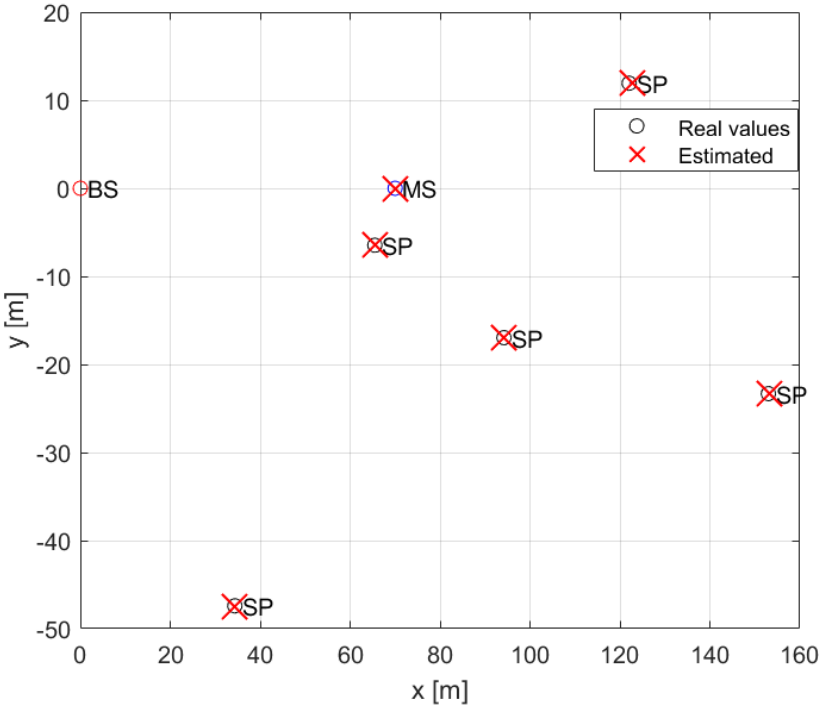

Figure 6. Localization RMSE with 5 NLoS SPs.

7. Conclusion

This paper has introduced an SSBCP algorithm for channel parameter extraction and localization in mmWave massive MIMO-OFDM systems. The proposed SSBCP algorithm has explicitly modeled receiver-side hardware impairments as a combination of sparse impulse noise and Gaussian noise,

departing from traditional i.i.d assumptions. By constructing a real-valued tensor representation and incorporating structured sparsity through latent binary variables, the proposed SSBCP algorithm has been designed to effectively mitigate the impact of sparse noise and enables robust parameter estimation. The algorithm has leveraged hierarchical priors and variational Bayesian inference to achieve uncertainty quantification and noise suppression. Theoretical analysis has confirmed the uniqueness of the decomposition, while simulation results have demonstrated strong performance across varying SNR levels and sparse noise conditions. Nevertheless, its direct applicability to other communication systems requires further validation, as the underlying assumptions may not hold in those contexts. Additionally, the algorithm's effectiveness depends on the sparse noise ratio staying within a reasonable range, an excessively high ratio can severely corrupt the observations and potentially compromise the convergence stability of variational inference. Overall, the proposed SSBCP algorithm has provided a unified and scalable solution for joint channel estimation and positioning under realistic hardware constraints, offering enhanced reliability for next-generation wireless systems. In future work, we will focus on extending the proposed algorithm to RIS-assisted communication systems [42,43], including advanced variants such as active simultaneously transmitting and reflecting surfaces (ASTARS/STARS) [44,45]. Additionally, we will develop higher-order Bayesian tensor decomposition algorithms to adapt to more complex propagation environments.

Data availability statement

No supplementary or additional data were generated in this study.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used generative AI tools only to improve language and readability. Specifically, the authors used DeepSeek and DeepL for language polishing and readability enhancement in limited sections of the manuscript, and Grammarly for grammar checking throughout the manuscript. These tools were used solely for language editing and not for content generation. The authors take full responsibility for the content of the manuscript.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62471444 and U2141233.

Authors' contribution

Yizhuo Xia: conceptualization, methodology, investigation, software, validation, writing—original and draft preparation. Weijia Yu: writing—original, draft preparation, methodology and writing—reviewing. Yuyang Xu: writing—reviewing and editing. Xin Luo: writing—reviewing and editing. Jianhe Du: conceptualization, supervision and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Balti E, Akoum S, Alfalujah I, Evans BL. Hybrid beamforming design for full-duplex millimeter wave massive MIMO systems. *IEEE Trans. Veh. Technol.* 2024, 73(11):17041–17058.

[2] An Z, Zhang T, Shen M, De Carvalho E, Ma B, *et al.* Series-constellation feature based blind modulation recognition for beyond 5G MIMO-OFDM systems with channel fading. *IEEE Trans. Cogn. Commun. Netw.* 2022, 8(2):793–811.

[3] Li X, Gao X, Liu Y, Huang G, Zeng M, *et al.* Overlay CR-NOMA assisted intelligent transportation system networks with imperfect SIC and CEEs. *Chin. J. Electron.* 2023, 32(6):1258–1270.

[4] Liu J, Zhang W, Jiang Y. Channel estimation considerate precoder design for multi-user massive MIMO-OFDM systems: the concept and fast algorithms. *IEEE Trans. Commun.* 2025, 73(6):3820–3832.

[5] Xiao J, Wang J, Wang Z, Wang J, Xie W, *et al.* Multi-task learning for near/far field channel estimation in STAR-RIS networks. *IEEE Trans. Commun.* 2024, 72(10):6344–6359.

[6] Xu Y, Tian Q, Chen Q, Wu Q, Huang C, *et al.* Robust secure beamforming design for Multi-RIS-Aided MISO systems with hardware impairments and channel uncertainties. *IEEE Trans. Commun.* 2024, 73(3):1517–1530.

[7] de Jesus Torres A, Sanguinetti L, Björnson E. Electromagnetic interference in RIS-aided communications. *IEEE Wireless Commun. Lett.* 2021, 11(4):668–672.

[8] Shi E, Zhang J, Ng DWK, Ai B. Uplink performance of RIS-aided cell-free massive MIMO system with electromagnetic interference. *IEEE J. Sel. Areas Commun.* 2023, 41(8):2431–2445.

[9] Long W, Moretti M, Bacci G, Sanguinetti L. Near-field MMSE channel estimation for THz RIS-aided communications with electromagnetic interference. *IEEE Wireless Commun. Lett.* 2025, 14(12):4152–4156.

[10] Björnson E, Matthaiou M, Debbah M. Massive MIMO with non-ideal arbitrary arrays: hardware scaling laws and circuit-aware design. *IEEE Trans. Wireless Commun.* 2015, 14(8):4353–4368.

[11] Fu H, Si W, Kim I. Deep learning-based joint pilot design and channel estimation for OFDM systems. *IEEE Trans. Commun.* 2023, 71(8):4577–4590.

[12] Tang J, Gao X, You L, Shi D, Yang J, *et al.* Massive MIMO-OFDM channel acquisition with time-frequency phase-shifted pilots. *IEEE Trans. Commun.* 2024, 73(6):4520–4535.

[13] Zhu L, Xiong Y, Li Z, Guan Y, Chu Z, *et al.* A novel gridless uplink/downlink channel estimation method for millimeter wave MIMO-OFDM systems. *IEEE Trans. Wireless Commun.* 2025, 24(5):3780–3793.

[14] Bian X, Xu W, Wang Y, Yuen C, Cai Z. A sparse-based channel estimation scheme for XL-RIS-Assisted wireless communications with low pilot overheads. *IEEE Trans. Commun.* 2025, 74:367–380.

[15] Li X, Shan C, Zhao H, Yuan W, Zhang R. A modified structured SAMP channel estimation method for FDD MIMO-OTFS systems. *IEEE Wireless Commun. Lett.* 2024, 13(11):3005–3009.

[16] Du J, Chen Y, Zhang P, Mumtaz S, Li X, *et al.* An effective simultaneous channel estimation and sensing algorithm for mmWave MIMO-OFDM systems. *IEEE Trans. Wireless Commun.* 2024, 23(11):17054–17069.

[17] Huang H, Zhang J, Jiang J. Low complexity hybrid-Field channel estimation based on simultaneous weighted OMP algorithm in extreme large-scale MIMO systems. *IEEE Access* 2024, 12:46551–46561.

[18] Tao Y, Gao Z, Li Z, Wan Z, Li T, *et al.* Pseudo-random TDM-MIMO FMCW-based millimeter-wave sensing and communication integration for UAV swarm. *IEEE Internet Things J.* 2025, 13(1):549–564.

[19] Sidiropoulos ND, De Lathauwer L, Fu X, Huang K, Papalexakis EE, *et al.* Tensor decomposition for signal processing and machine learning. *IEEE Trans. Signal Process.* 2017, 65(13):3551–3582.

[20] Zhou Z, Fang J, Yang L, Li H, Chen Z, *et al.* Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems. *IEEE J. Sel. Areas Commun.* 2017, 35(7):1524–1538.

[21] Zhao Q, Zhang L, Cichocki A. Bayesian CP factorization of incomplete tensors with automatic Rank determination. *IEEE Trans. Pattern Anal. Mach. Intell.* 2015, 37(9):1751–1763.

[22] Chen X, He Z, Sun L. A bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Transp. Res. Part C: Emerg. Technol.* 2019, 98:73–84.

[23] Lee J, Gil G, Lee YH. Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications. *IEEE Trans. Commun.* 2016, 64(6):2370–2386.

[24] Albataineh Z, Al Bataineh M, Hayajneh KF, Al Athamneh R. Advanced 5G channel estimation in mmWave MIMO systems: leveraging compressive sensing for enhanced performance. *IEEE Access* 2025, 13:72104–72115.

[25] Choi JW, Shim B, Chang SH. Downlink pilot reduction for massive MIMO systems via compressed sensing. *IEEE Wireless Commun. Lett.* 2015, 19(11):1889–1892.

[26] Zhang S, Ning X, Zheng X, Shi Q, Chang T, *et al.* A physics-based and data-driven approach for localized statistical channel modeling. *IEEE Trans. Wireless Commun.* 2023, 23(6):5409–5424.

[27] Dey A, Pattanayak P. Efficient pilot contamination reduction in spatially correlated cell-free massive MIMO systems with hardware impairments. *Wireless Netw.* 2025, 31(3):2817–2826.

[28] Giroto L, Long X, Karle C, Utku Erdem U, Jeong T, *et al.* System concept and demonstration of bistatic MIMO-OFDM-based ISAC. *IEEE Trans. Radar Syst.* 2025, 4:232–246.

[29] Brihuega A, Anttila L, Valkama M. Beam-level frequency-domain digital predistortion for OFDM massive MIMO transmitters. *IEEE Trans. Microw. Theory Tech.* 2022, 71(4):1412–1427.

[30] Ceniklioglu B, Tubail DA, Canbilen AE, Develi I, Ikki SS. Error analysis of the joint localization and synchronization of RIS-assisted mm-Wave MISO-OFDM under the effect of hardware impairments. *IEEE Open J. Commun. Soc.* 2022, 3:2151–2161.

[31] Mokhtari Z, Sabbaghian M, Dinis R. A survey on massive MIMO systems in presence of channel and hardware impairments. *Sensors* 2019, 19(1):164.

[32] Taghizadeh O, Radhakrishnan V, Cirik AC, Mathar R, Lampe L. Hardware impairments aware transceiver design for bidirectional full-duplex MIMO OFDM systems. *IEEE Trans. Veh. Technol.* 2018, 67(8):7450–7464.

[33] Yu W, Du J, Chen Y, Jin L, Li X, *et al.* Vehicle localization based on bayesian tensor decomposition in intelligent transportation systems. *IEEE Trans. Intell. Transp. Syst.* 2025, 26(10):16204–16221.

[34] Cheng L, Chen Z, Wu YC. Bayesian tensor CPD: performance and real-world applications. In *Bayesian Tensor Decomposition for Signal Processing and Machine Learning: Modeling, Tuning-Free Algorithms, and Applications*, 1st ed. Cham: Springer International Publishing, 2023. pp. 59–75.

[35] Belaoura W, Ghanem K, Nedil M, Bousbia-Salah H. On the impact of hardware impairments on mm-Wave MIMO underground channel estimation. In *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)*, Mostaganem, Algeria, June 3, 2022.

[36] Haardt M, Nossek J. Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational burden. *IEEE Trans. Signal Process.* 2002, 43(5):1232–1242.

[37] Du J, Han M, Chen Y, Jin L, Gao F. Tensor-based joint channel estimation and symbol detection for time-Varying mmWave massive MIMO systems. *IEEE Trans. Signal Process.* 2021, 69:6251–6266.

[38] Zhao Q, Zhou G, Zhang L, Cichocki A, Amari S. Bayesian robust tensor factorization for incomplete multiway data. *IEEE Trans. Neural Netw. Learn. Syst.* 2015, 27(4):736–748.

[39] Xu W, Liu A, Zhou B, Zhao M. Successive linear approximation VBI for joint sparse signal recovery and dynamic grid parameters estimation. *IEEE Trans. Wireless Commun.* 2025, 24(11):9645–9659.

[40] Nassirpour S, Nguyen T, Nguyen DHN. Variational bayesian inference for time-varying massive MIMO channels: estimation and detection. *IEEE Trans. Wireless Commun.* 2025, 24(11):9112–9126.

[41] Kruskal JB. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* 1977, 18(2):95–138.

[42] Wang F, Jiang W, Li R, Li X, Nallanathan A. Optical reconfigurable intelligent surfaces assisted visible light positioning and communications: performance analysis and design optimization. *Adv. Inf. Commun.* 2026, 1(1):1–35.

[43] Li X, Ti Z, Zhang J, Xu X, Peng H, *et al.* Performance analysis of covert communication in IRS-assisted NOMA networks. *Sci. China Inf. Sci.* 2024, 54(6):1502–1515.

[44] Yue X, Song X, Zhao J, Huang C, Jin H, *et al.* ASTARS aided NOMA covert communication networks. *IEEE Trans. Veh. Technol.* 2025, 74(7):10661–10673.

[45] Yue X, Xie J, Ouyang C, Liu Y, Shen X, *et al.* Active simultaneously transmitting and reflecting surface assisted NOMA networks. *IEEE Trans. Wireless Commun.* 2024, 23(8):9912–9926.