Integrated Sensing and Communication Receiver Design for OTFS-Based MIMO System: A Unified Variational Inference Framework

Nan Wu, *Senior Member, IEEE*, Haoyang Li, Dongxuan He, *Member, IEEE*, Arumugam Nallanathan, *Fellow, IEEE*, and Tony Q. S. Quek, *Fellow, IEEE*

***Abstract*— This paper proposes a novel integrated sensing and communication (ISAC) receiver design framework for OTFS (orthogonal time frequency space)-based MIMO (multi-input-multi-output) systems from a unified perspective of variational inference. We first construct a factor graph representation for the OTFS-based MIMO system according to the factorization of the *a posteriori* probability (APP). This representation establishes a direct probabilistic link between sensing and communication, allowing both functionalities to benefit from their integration. On this basis, we develop a low computational complexity message passing algorithm by minimizing the variational free energy associated with the global APP. In particular, belief propagation, mean field, and expectation maximization algorithms for data detection, channel coefficient estimation, and kinematic parameter sensing are derived, respectively. To reduce the communication overhead for the implementation of ISAC algorithm, we propose a federated learning scheme for distributed kinematic parameter sensing. Specifically, by solving the sensing problem in different fashions, three federated learning modes are devised. Simulation results validate the superior performance of the proposed scheme.**

***Index Terms*— MIMO, OTFS, integrated sensing and communication, variational inference, federated learning.**

I. Introduction

With the advent of next generation networks, notably sixth generation (6G), high speed and ultra-reliable communications, along with stable service provision in high mobility scenarios, including, autonomous vehicles [1], industry automation [2], unmanned aerial vehicle (UAV) networks [3] and space-terrestrial integrated networks (STIN) [4], have received tremendous attention. These emerging applications necessitate not only high-quality wireless connectivity but also accurate and robust sensing capability. Against this background, integrated sensing and communications (ISAC) is poised to become one of the pivotal enablers of 6G for catering to the appealing services [5].

In the past, the sensing and communication functionalities were designed separately, leading to sub-optimal utilization of hardware, spectrum, and signaling resources [6]. This realization has sparked interest in the novel paradigm of ISAC systems. As early research efforts, the concept of radar-communication coexistence (RCC) emerged, advocating for the shared use of radar frequencies for communication purposes [7]. In a complementary manner, the wide millimeter wave (mmWave) and terahertz (THz) bands envisaged for 6G communication are capable of bearing sensing functionality [8], [9]. On this basis, a significant body of research is now focused on the development of ISAC systems that treat sensing and communication functionalities with equal importance, moving beyond the traditional approach of prioritizing one over the other [10], [11].

It is demonstrated that the ISAC systems achieve notable performance enhancements compared to individual sensing and communications systems. A pioneering work by [12] systematically characterized the tradeoff between sensing and communication performance, harnessing the perspective of maximum achievable rate of information theory and Cramér-Rao bound (CRB) of estimation theory. This paper revealed a fascinating insight that the wisely-designed ISAC systems outperforms their separated counterparts. For this purpose, the performance tradeoff for ISAC systems were discussed in the literature [13], [14]. In addressing the joint optimization problem, [13] considered weighted optimizations targeting a flexible tradeoff between sensing and communications performance, where the optimal solutions with closed-form for sensing were given. Additionally, [14] introduced a sensing information metric, termed estimation rate, facilitating the sensing and communication to be considered in an information theory context. In such scheme, the inner bound of the estimation and communication rates was derived.

In both sensing and communication filed, multi-input-multi-output (MIMO) system empowered by large-scale antenna array has been recognized for its substantial enhancement of communication capacity [15], [16] and sensing accuracy [17], [18]. Consequently, incorporating MIMO systems into ISAC systems presents a promising avenue to further elevate performance [19], [20]. Meanwhile, the orthogonal

Received 14 March 2024; revised 6 September 2024; accepted 7 November 2024. Date of publication 20 January 2025; date of current version 19 March 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFB2900600; and in part by the National Natural Science Foundation of China under Grant 62401047, Grant 62371045, and Grant 62301060. *(Corresponding author: Nan Wu.)*

Nan Wu, Haoyang Li, and Dongxuan He are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: wunan@bit.edu.cn; 3120225395@bit.edu.cn; dongxuan_he@bit.edu.cn).

Arumugam Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, U.K., and also with the Department of Electronic Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: a.nallanathan@qmul.ac.uk).

Tony Q. S. Quek is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore 487372 (e-mail: tonyquek@sutd.edu.sg).

Digital Object Identifier 10.1109/JSAC.2025.3531574

frequency division multiplexing (OFDM) modulation stands as the predominant modulation technique in MIMO-based ISAC system design [21], [22]. However, OFDM suffers performance degradation in the presence of Doppler shift induced by high mobility [23], [24]. In this context, a two-dimensional (2D) modulation scheme referred to as orthogonal time frequency space (OTFS) was proposed, where the information symbols are encoded in the delay-Doppler domain rather than the time-frequency domain [25], [26]. OTFS modulation converts the time-varying channel in time-frequency domain into a quasi-static channel in delay-Doppler domain, effectively countering the adverse effects of Doppler shift [23], [27]. It is demonstrated that OTFS signaling practically achieves full diversity under the high Doppler channel, which offers a promising solution to the problem of OFDM signaling [28], [29]. Moreover, OTFS signaling also exhibits the superior estimation performance than its counterpart, positioning it as a powerful candidate for MIMO-based ISAC system [29].

Significantly, the deployment of ISAC systems necessitates addressing two fundamental problems, parameter estimation and data detection [30], [31]. In [29], a maximum likelihood (ML) estimator was designed for parameter estimation, coupled with a message passing algorithm-based detector, incurring intense computation. Since only a few reflectors are involved in the signal propagation, the angle-delay-Doppler channel exhibits inherent sparsity, which can be harnessed to reduce the complexity of signal processing of ISAC significantly [32]. In particular, the parameter estimation problem can be formulated as a sparse representation problem, where sparse Bayesian learning can be used to estimate the angle, delay, and Doppler parameter [33], [34], [35]. An on-grid learning approach was proposed in [33], while [34], [35] resorted to the first order Taylor expansion to alleviate the mismatch in on-grid leaning. In addition, data detection also benefits from the sparsity. For this purpose, [36], [37], [38], [39] developed a family of low-complexity message passing algorithms for data detection, leveraging the perspective of variational inference.

Interestingly, the angle-delay-Doppler response of the effective channel of MIMO-OTFS signaling characterizes the angle, range, and radial velocity of the transceiver, thereby forming a vital link between the communication channel and the transceiver states. These parameters are referred to as kinematic parameters, as they usually contain kinematic quantities such as coordinate, velocity, and acceleration, etc. Hence, the mapping from kinematic parameters to channel parameters needs to be taken into consideration for the practical ISAC receiver design. This insight inspires an enormous amount of studies that characterize this kind of mapping, then optimize the ISAC performance based on the estimated kinematic parameters through beamforming [40], [41], [42], [43]. Within these literature, the unknown kinematic parameters was modeled directly based on the corresponding observation extracted from the received signals through ML estimator. In [40], an extended Kalman filter (EKF) was proposed to track and predict the parameter accurately, which suffers from intense computation. To reduce the complexity, the EKF was linearized, paving the way for adopting minimum mean square error (MMSE) estimator for parameter prediction [41]. Furthermore, a message passing algorithm is developed in [42] to predict the kinematic parameters based on a Bayesian framework with lower computational complexity. Otherwise, [43] adopted the deep learning (DL) approach to implicitly learn the features from historical estimated channels and directly predict the beamforming matrix.

Moreover, the interplay between parameter estimation and data detection is mutually beneficial, enhancing the performance of the entire ISAC system. Accordingly, a data-aided channel estimation approach was developed in [44], where channel estimates are refined by exploiting the detected data symbol. On the other hand, given the estimated angel, delay, and Doppler shift, the pair-wise error probability (PEP) can be further optimized by precoding and power allocation that matches the estimated channel [45]. Besides, the communication performance can be improved with the accurate channel state information (CSI) [24], [30]. Therefore, the development of advanced ISAC receiver for OTFS-based MIMO systems becomes crucial.

However, existing studies [29], [30], [31], [45] on ISAC receiver design for OTFS-based MIMO systems mainly focus on joint channel estimation and data detection (JCEDD), overlooking the direct relationship between practical kinematic parameter sensing and communication. on the other hand, inferring kinematic parameters from available channel parameters has been well investigated in literature [40], [41], [42], [43], which bears performance loss compared with direct inference from received signals [46]. This oversight has enlightened us to integrate these two functionalities. Specifically, we propose a unified ISAC receiver design framework for OTFS-based MIMO system, where the probabilistic information of sensing and communication are considered simultaneously. Our framework begins with the construction of a graphical model for ISAC systems. Then, efficient algorithm is derived to perform ISAC from the perspective of variational inference. Furthermore, we resort to the concept of federated learning to alleviate the communication overhead for sensing. Simulation results validate the superiority of the proposed framework. The main contributions of this paper can be summarized as follows.

- We derive the generalized probabilistic representation for the OTFS-based MIMO system, where the mapping from kinematic parameters to channel parameters is modeled as a delta function, establishing a direct probabilistic link between sensing and communication. Based on the factorization of the a *posteriori* probability (APP), we construct the corresponding multi-layer factor graph, on which various algorithms could be deployed efficiently to perform ISAC through message passing.
- As the calculation based on the original APP is not tractable, we resort to the efficient variational inference framework, in which the APP is approximated by a flexible trial distribution, termed variational distribution, by minimizing the variational free energy. In the framework, the optimization of the variational distribution is decomposed into three subproblems, namely, data detection, channel coefficient estimation, and kinematic

parameter sensing. By assuming different variational distributions, belief propagation, mean field, and expectation maximization algorithms for these subproblems are developed, with low computational complexity.

- To reduce the communication overhead, we introduce a communication-efficient federated learning scheme for implementing the ISAC algorithm. This method involves computing local sensing information at distributed processing units (DPUs) before transmitting it to a central processing unit (CPU) for aggregation and parameter sensing. Moreover, three different modes are devised by solving the sensing problem in different fashions. Additionally, we first introduce the concept of frequency difference of arrival (FDOA)-based localization into ISAC systems.

The remainder of this paper is organized as follows. The mathematical model of the considered OTFS-based MIMO system tailored for ISAC is introduced in Section II. Section III develops an advanced iterative ISAC receiver from a unified variational inference perspective. To reduce the communication overhead, a communication-efficient federated learning scheme is devised in Section IV. Simulation is conducted in Section V, while Section VI concludes this paper.

Notations: We use plain font letters to denote scalars, lowercase (uppercase) boldface letters to denote column vectors (matrices). The (m,n) element of a matrix $\mathbf{X}$ is denoted as $[\mathbf{X}]_{m,n}$. The superscripts $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, and $(\cdot)^{\mathrm{H}}$ represent the conjugate, transpose, and Hermitian transpose operators, respectively, while $\langle \mathbf{X} \rangle$ denotes the sum of all elements of $\mathbf{X}$, and $\dot{\mathbf{X}}(a)$ denotes the derivative of $\mathbf{X}(a)$ w.r.t. variable a; operators $\circ$, $./$ denote the element-wise multiplication and division, respectively; $\mathrm{vec}\,(\cdot)$ denotes the matrix vectorization, and $\mathrm{mat}\,(\cdot)$ is the inverse operation of it with appropriate size; $\mathbf{F}_N$, $\mathbf{I}_M$, and $\mathbf{0}_{M\times N}$ denote the discrete Fourier transform (DFT) matrix of size $N \times N$, the identity matrix of size $M \times M$, and the all zero matrix with size $M \times N$, respectively; $\mathcal{N}_{\mathbb{C}}\left(x, \mu, \sigma^2\right)$ denotes the complex Gaussian distribution of a random variable with mean μ and variance σ^2, and $\mathbb{E}_p\left[\cdot\right]$ denotes statistical expectation operator with respect to distribution p; $\delta\left(\cdot\right)$ denotes the Dirac delta function, and $\lceil\cdot\rceil$ denotes the ceil function.

II. System Model

Throughout this paper, we consider a MIMO system, where a number of BSs equipped with multiple antennas cooperate in the network to serve a vast area. The BSs are connected to the processing unit (i.e., CPU or DPU) via backhaul links, so that the information obtained at multiple BSs can be jointly processed to realize ISAC. The diagram of the proposed ISAC architecture is illustrated in Fig. 1.

Assuming that there are B BSs serving K users, and the location of the b-th BS and the k-th user equipment (UE) are $\tilde{\mathbf{u}}_b = [\tilde{u}_{x,b}, \tilde{u}_{y,b}]^{\mathrm{T}}$ and $\mathbf{u}_k = [u_{x,k}, u_{y,k}]^{\mathrm{T}}$, respectively. The BSs are stationary and the velocity of the k-th UE is $\mathbf{v}_k = [v_{x,k}, v_{y,k}]^{\mathrm{T}}$. Then we can calculate the parameters associated with the channel manifold. The relative distance, orientation, and radial velocity between the b-th BS and the k-th UE are

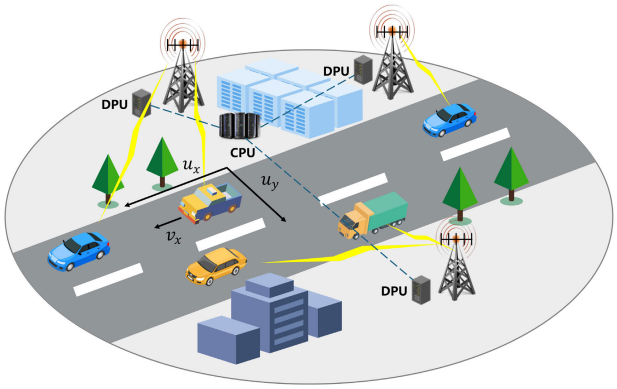


Fig. 1. The considered ISAC architecture with a number of BSs serving a vast area through cooperation.

given by

$$\rho_{b,k} = \rho(\Delta\mathbf{u}_{b,k}) \triangleq \|\Delta\mathbf{u}_{b,k}\|_2, \tag{1}$$

$$\theta_{b,k} = \theta(\Delta\mathbf{u}_{b,k}) \triangleq \arctan\frac{[1,0]\Delta\mathbf{u}_{b,k}}{[0,1]\Delta\mathbf{u}_{b,k}}, \tag{2}$$

$$v_{b,k} = v(\mathbf{v}_k, \Delta\mathbf{u}_{b,k}) \triangleq -\mathbf{v}_k^{\mathrm{T}}\frac{\Delta\mathbf{u}_{b,k}}{\|\Delta\mathbf{u}_{b,k}\|_2}, \tag{3}$$

where $\Delta\mathbf{u}_{b,k} = \mathbf{u}_k - \tilde{\mathbf{u}}_b$. For simplicity, we assume that the UE moves along the x-axis,[1] i.e., $v_{y,k} = 0$.

Remark 1: For scenarios involving multiple distributed UEs, techniques such as activity detection [47] and successive interference cancellation [48] can be employed. These techniques effectively reduce the problem to a single UE case. Consequently, to focus on the sensing and communication problem, we consider the single UE case, where B BSs equipped with M antennas uniform linear array (ULA) are employed to serve one single UE equipped with one antenna. For clarity, subscript k can be omitted. ■

A. OTFS Modulation

In this section, we describe the employed OTFS modulation. Suppose that a signal frame with duration N_1T occupies a bandwidth $N_2\Delta f$, where $T\Delta f = 1$ for orthogonality [32]. Hence, there are $N = N_1N_2$ resource elements in a transmission duration. First, the *inverse symplectic finite Fourier transform* (ISFFT) maps the information symbols $x_{k,l}$ in delay–Doppler domain to $\tilde{x}_{n_1,n_2}$ symbols in the time–frequency domain, i.e.,

$$\tilde{x}_{n_1,n_2} = \frac{1}{\sqrt{T}}\sum_{k=0}^{N_1-1}\sum_{l=0}^{N_2-1} x_{k,l}e^{j2\pi\left(\frac{n_1k}{N_1}-\frac{n_2l}{N_2}\right)}, \tag{4}$$

where $n_1 \in \{0,\ldots,N_1-1\}$ is the time index, $n_2 \in \{0,\ldots,N_2-1\}$ is the subcarrier index, $k \in \{0,\ldots,N_1-1\}$ is the Doppler index and $l \in \{0,\ldots,N_2-1\}$ is the delay index. After employing the transmitter pulse-shaping filter $h_{\mathrm{tx}}(t)$, the modulator generates the multicarrier time-domain baseband

[1]It is straightforward to extend the considered problem to a 2D case.

signal signal, i.e.,

$$s(t)=\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}\tilde{x}_{n_1,n_2}h_{\text{tx}}(t-n_1T)e^{j2\pi n_2\Delta f(t-n_1T)}, \quad (5)$$

which is also called *Heisenberg transform*.

B. OTFS-Based MIMO System

To enable the electromagnetic wave to propagate in wireless environment, the baseband signal (5) is modulated to high frequency band, i.e., $\tilde{s}(t)=s(t)e^{j2\pi f_0 t}$, where f_0 is the carrier frequency. For a linear time-variant channel, the received time domain signal at the m-th antenna of b-th BS is given by [9],

$$\tilde{r}_{b,m}(t)=\iint h_{b,m}(\nu,\tau)e^{j2\pi\nu(t-\tau)}\tilde{s}(t-\tau)d\tau d\nu+\tilde{w}_{b,m}(t), \quad (6)$$

where

$$\tilde{h}_{b,m}(\nu,\tau)=\sum_{p=0}^{P}\tilde{g}_{b,p}\delta(\tau-\tau_{b,p}-\frac{md\sin\theta_{b,p}}{c})\delta(\nu-\nu_{b,p}), \quad (7)$$

with $\tilde{g}_{b,p}\in\mathbb{C}$ denotes the complex channel gain and P is the number of multiple paths induced by reflectors. Variables $\nu_{b,p}=f_0v_{b,p}/c\in[-\nu_{\max},\nu_{\max}]$ and $\tau_{b,p}=\rho_{b,p}/c\in[0,\tau_{\max}]$ are the Doppler shift and delay of the p-th path at the b-th BS, with $v_{b,p}$ and $\rho_{b,p}$ denoting the corresponding redial velocity and relative distance, respectively, while c denoting the speed of light. The propagation path difference between two adjacent antennas of the p-th path at the b-th BS is $d\sin\theta_{b,p}$ with $\theta_{b,p}$ being the corresponding angle of arrival and d denoting the distance between two adjacent antennas. We assume that the maximum delay $\tau_{\max}$ and Doppler shift $\nu_{\max}$ are known in advanced. And $\tilde{w}_{b,m}(t)$ is the additive white Gaussian noise (AWGN). Without loss of generality, $p=0$ denotes the line of sight (LoS) path and the subscript 0 is omitted, while $p\neq 0$ denotes the non line of sight (NLoS) path. Specifically, we have $\theta_{b,0}=\theta_b$, $\rho_{b,0}=\rho_b$, and $v_{b,0}=v_b$ given by (1), (2), and (3), respectively.

After some algebra and demodulating $\tilde{r}_{b,m}(t)$ to baseband, we have

$$\begin{aligned}\bar{r}_{b,m}(t)&=\tilde{r}_{b,m}(t)e^{-j2\pi f_0t}\\&=\sum_{p=0}^{P}\tilde{g}_{b,p}r_b(t-\frac{md\sin\theta_{b,p}}{c})e^{-j2\pi f_0(\tau_{b,p}+\frac{md\sin\theta_{b,p}}{c})}+\bar{w}_{b,m}^{(t)}\\&\approx\sum_{p=0}^{P}\tilde{g}_{b,p}e^{-j2\pi f_0\tau_{b,p}}r_{b,p}(t)e^{-j2\pi\frac{md\sin\theta_{b,p}}{\lambda}}+\bar{w}_{b,m}(t), \quad (8)\end{aligned}$$

where

$$r_{b,p}(t)=\iint h_{b,p}(\nu,\tau)e^{j2\pi\nu(t-\tau)}s(t-\tau)d\tau d\nu, \quad (9)$$

with

$$h_{b,p}(\nu,\tau)=\delta(\tau-\tau_{b,p})\delta(\nu-\nu_{b,p}). \quad (10)$$

We assume that $\tilde{g}_{b,p}\sim\mathcal{N}_{\mathbb{C}}(\tilde{g}_{b,p};0,\gamma_{g_{b,p}}^{-1})$ and $\bar{w}_{b,m}(t)$ is the baseband AWGN. For simplicity, we set the $d=\lambda/2$. The approximation in (8) is reasonable as $\frac{md\sin\theta_{b,p}}{c}\ll T$, so that the effect on $r_{b,p}(t-\frac{md\sin\theta_{b,p}}{c})$ can be neglected [49]. Accordingly, we can write (8) in vector form, i.e.,

$$\mathbf{r}_b(t)=\sum_{p=0}^{P}\tilde{g}_{b,p}e^{-j2\pi f_0\tau_{b,p}}\mathbf{a}(\theta_{b,p})r_{b,p}(t)+\bar{\mathbf{w}}_b(t), \quad (11)$$

where $\mathbf{r}_b(t)=[\bar{r}_{b,1}(t),\ldots,\bar{r}_{b,M}(t)]^{\mathrm{T}}$ is the received array signal and $\mathbf{a}(\theta_{b,p})\triangleq\frac{1}{\sqrt{N}}[1,e^{-j\pi\sin\theta_{b,p}},\ldots,e^{-j\pi(M-1)\sin\theta_{b,p}}]^{\mathrm{T}}$ is the normalized array response with $\theta_{b,p}$ in $\mathbf{a}(\theta_{b,p})$ denoting that the vector is a function of it. We assume that $\bar{\mathbf{w}}_b(t)=[\bar{w}_{b,1}(t),\ldots,\bar{w}_{b,M}(t)]^{\mathrm{T}}$ is the independent identically distributed (i.i.d.) AWGN. In the massive MIMO scenario and the favourable channel condition, the array response $\mathbf{a}(\theta_{b,p})$ corresponding to different paths are asymptotically orthogonal [50], which reads[2]

$$\mathbf{a}^{\mathrm{H}}(\theta_{b,p})\mathbf{a}(\theta_{b,p})\to 0,\quad \forall p\neq p',\quad M\to\infty. \quad (12)$$

Besides, in practice, $\gamma_{g_b}^{-1}\gg\gamma_{g_{b,p}}^{-1},\forall p\neq 0$ due to the reflection loss.

Remark 2: It is noteworthy that the NLoS channel with higher path loss also characterizes the kinematic parameters of UE by introducing the virtual UEs, whose kinematic parameters are associated with that of UE and reflectors according to geometry. In this paper, we only consider the LoS channel, which can be distinguished by exploiting property (12) and imposing some physical constraints, e.g., the LoS channel can be select from multiple paths with minimal distance. ■

Accordingly, (11) can be rewritten as

$$\mathbf{r}_b(t)=\tilde{g}_be^{-j2\pi f_0\tau_b}\mathbf{a}(\theta_b)r_b(t)+\tilde{\mathbf{w}}_b(t), \quad (13)$$

where the noise vector $\mathbf{w}_b(t)$ incorporates the residual inter-beam interference

$$\tilde{\mathbf{w}}_b(t)=\mathbf{a}(\theta_b)\sum_{p=1}^{P}\tilde{g}_{b,p}e^{-j2\pi f_0\tau_{b,p}}\mathbf{a}^{\mathrm{H}}(\theta_b)\mathbf{a}(\theta_{b,p})r_{b,p}(t)+\bar{\mathbf{w}}_b^{(t)}. \quad (14)$$

As P is large, according to the central limit theorem (CLT), we assume that $\tilde{\mathbf{w}}_b(t)\sim\mathcal{N}_{\mathbb{C}}(\tilde{\mathbf{w}}_b(t);\mathbf{0}_{M\times 1},\sigma_w^2\mathbf{I}_M)$.

At the receiver, the optimal *Wigner transform* is employed at each antenna, which reads,

$$\tilde{\mathbf{y}}_{b,n_1,n_2}=\tilde{g}_be^{-j2\pi f_0\tau_b}\mathbf{a}(\theta_b)\tilde{z}_{b,n_1,n_2}+\tilde{\mathbf{w}}_{b,n_1,n_2}, \quad (15)$$

where

$$\tilde{z}_{b,n_1,n_2}=\int r_b(t)h_{\text{rx}}^*(t-n_1T)e^{-j2\pi n_2\Delta f(t-n_1T)}dt, \quad (16)$$

and

$$\tilde{\mathbf{w}}_{b,n_1,n_2}=\int\mathbf{w}_b(t)h_{\text{rx}}^*(t-n_1T)e^{-j2\pi n_2\Delta f(t-n_1T)}dt, \quad (17)$$

with $h_{\text{rx}}(t)$ denoting the receiver filter. Then the *symplectic finite Fourier transform* (SFFT) maps the received symbols

[2]For scenarios where two UEs are close to each other, the orthogonal property may not hold. In this circumstance, the inter-beam interference can be alleviated by using interference cancellation approach, e.g., [51].

from time-frequency domain to delay–Doppler domain, which reads,

$$\mathbf{y}_{b,k,l} = \frac{1}{\sqrt{T}} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \tilde{\mathbf{y}}_{b,n_1,n_2} e^{-j2\pi\left(\frac{kn_1}{N_1} - \frac{ln_2}{N_2}\right)}. \tag{18}$$

After some algebra, we have

$$\mathbf{y}_{b,k,l} = \tilde{g}_b e^{-j2\pi f_0 \tau_b} \mathbf{a}(\theta_b) z_{b,k,l} + \mathbf{w}_{b,k,l}, \tag{19}$$

where

$$z_{b,k,l} = \sum_{k'=0}^{N_1-1} \sum_{l'=0}^{N_2-1} x_{k',l'} \varphi_b(k-k', l-l'). \tag{20}$$

The effective noise $\mathbf{w}_{b,k,l} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{w}_{b,k,l}; \mathbf{0}_{M\times 1}, \sigma_w^2 \mathbf{I}_M), \forall b,k,l$ is i.i.d. following the derivation in [52]. The effective delay–Doppler domain channel response $\varphi_b(k-k', l-l')$ is given by

$$\varphi_b(k-k', l-l') = \tilde{\varphi}_b(\nu,\tau)\big|_{\nu=\frac{k-k'}{N_1 T}, \tau=\frac{l-l'}{N_2 \Delta f}}, \tag{21}$$

for $\tilde{\varphi}_b(\nu,\tau)$ being the 2D circular convolution of the channel response with $f(\nu,\tau)$, i.e.,

$$\tilde{\varphi}_b(\nu,\tau) = \iint h_b(\nu',\tau') f(\nu-\nu', \tau-\tau') e^{-j2\pi\nu'\tau'} d\tau' d\nu'. \tag{22}$$

In (22), $f(\nu,\tau)$ is the SFFT of a windowing function that is characterized by the ambiguity function. Interested readers may refer to [35] for details. For simplicity, we focus on the case of ideal pulse shaping that satisfies the *bi-orthogonal* condition for the transceiver, which is a commonly studied case in the OTFS literature [32], [35], [52]. In this way, $f(\nu,\tau)$ is the SFFT of a rectangular windowing function

$$f(\nu,\tau) = \frac{1}{N} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} 1 \cdot e^{-j2\pi(\nu n_1 T - \tau n_2 \Delta f)}, \tag{23}$$

By substituting (10), (22), and (23) into (21), we obtain

$$\varphi_b(k-k', l-l') = e^{-j2\pi\nu_b\tau_b} \varphi_{1,b}(k-k') \varphi_{2,b}(l-l'), \tag{24}$$

where

$$\begin{aligned}\varphi_{1,b}(k-k') &= \frac{1}{N_1} \sum_{n_1=0}^{N_1-1} e^{-j2\pi\frac{(k-k'-k_b)n_1}{N_1}} \\ &= \frac{e^{j(1-N_1)\pi\frac{k-k'-k_b}{N_1}}}{N_1} \frac{\sin(\pi(k-k'-k_b))}{\sin\left(\frac{\pi(k-k'-k_b)}{N_1}\right)}, \end{aligned} \tag{25}$$

$$\begin{aligned}\varphi_{2,b}(l-l') &= \frac{1}{N_2} \sum_{n_2=0}^{N_2-1} e^{j2\pi\frac{(l-l'-l_b)n_2}{N_2}} \\ &= \frac{e^{j(N_2-1)\pi\frac{l-l'-l_b}{N_2}}}{N_2} \frac{\sin\left(\pi\left(l-l'-l_b\right)\right)}{\sin\left(\frac{\pi(l-l'-l_b)}{N_2}\right)}, \end{aligned} \tag{26}$$

with $k_b = \nu_b N_1 T$ and $l_b = \tau_b N_2 \Delta f$. We can observe that the effective Doppler domain channel (25) and delay domain channel (26) exhibit periodicity, i.e., $\varphi_{1,b}(k-k') = \varphi_{1,b}((k-k')_{N_1})$ and $\varphi_{2,b}(l-l') = \varphi_{2,b}((l-l')_{N_2})$. Besides, when k_b and l_b are not integers, inter symbol interference occurs due to the non-orthogonality imposed by fractional Doppler shift and delay, which lead to the performance degradation. Hence, interference cancellation techniques are urgent.

Unfolding the delay-Doppler domain, the above transmitting and receiving process can be summarized as vector form, following

$$\mathbf{Y}_b = g_b \mathbf{a}(\theta_b) \operatorname{vec}(\mathbf{Z}_b)^{\mathrm{T}} + \mathbf{W}_b, \tag{27}$$

where

$$\mathbf{Z}_b = \mathbf{\Phi}_1(v_b) \mathbf{X} \mathbf{\Phi}_2^{\mathrm{T}}(\rho_b), \tag{28}$$

and $\mathbf{Y}_b \in \mathbb{C}^{M\times N}$ is the concatenation of all received symbols at the b-th BS, i.e., $[\mathbf{Y}_b]_{:,n} = \mathbf{y}_{b,k,l}$, while $\mathbf{W}_b \in \mathbb{C}^{M\times N}$ is the AWGN in space-delay-Doppler domain with $[\mathbf{W}_b]_{:,n} = \mathbf{w}_{b,k,l}$. For simplicity, the index transformation for n and (k,l), namely,

$$n = k + lN_1, \tag{29}$$

is *omitted* in the subsequent sections. In (28), $\mathbf{X} \in \mathbb{C}^{N_1\times N_2}$ and $\mathbf{Z}_b \in \mathbb{C}^{N_1\times N_2}$ are the transmitted symbols at UE and the received effective OTFS symbols at the b-th BS with their elements given by $[\mathbf{X}]_{k,l} = x_{k,l}$ and $[\mathbf{Z}_b]_{k,l} = z_{b,k,l}$, respectively. The constant $e^{-j2\pi\nu_b\tau_b}$ in (24) is incorporated in the complex channel gain for notational simplicity. The effective complex channel gain $g_b \triangleq \tilde{g}_b e^{-j2\pi f_0\tau_b} e^{-j2\pi\nu_b\tau_b}$ has the same probability density function as $\tilde{g}_b$ due to the circular symmetry of complex Gaussian function, i.e., $g_b \sim \mathcal{N}_{\mathbb{C}}(g_b; 0, \gamma_{g_b}^{-1})$. Accordingly, $\mathbf{\Phi}_1(v_b) \in \mathbb{C}^{N_1\times N_1}$ and $\mathbf{\Phi}_2(\rho_b) \in \mathbb{C}^{N_2\times N_2}$ are the effective channel response in Doppler domain and delay with their elements given by $[\mathbf{\Phi}_1(v_b)]_{k,k'} = \varphi_{1,b}(k-k')$ and $[\mathbf{\Phi}_2(\rho_b)]_{l,l'} = \varphi_{2,b}(l-l')$, respectively. Note that the variables v_b and ρ_b in $\mathbf{\Phi}_1(v_b)$ and $\mathbf{\Phi}_2(\rho_b)$ indicate that the matrix is a function of the them.

Interestingly, $\mathbf{\Phi}_1(v_b)$ and $\mathbf{\Phi}_2(\rho_b)$ are the circulant matrices, which can be rewritten following eigenvalue decomposition of the form

$$\mathbf{\Phi}_1(v_b) = \mathbf{F}_{N_1} \mathbf{D}_1(v_b) \mathbf{F}_{N_1}^{\mathrm{H}}, \tag{30}$$

$$\mathbf{\Phi}_2(\rho_b) = \mathbf{F}_{N_2}^{\mathrm{H}} \mathbf{D}_2(\rho_b) \mathbf{F}_{N_2}, \tag{31}$$

where $\mathbf{D}_1(v_b)$ and $\mathbf{D}_2(\rho_b)$ are diagonal matrices with their diagonal elements given by $[\mathbf{D}_1(v_b)]_{k,k} = e^{j2\pi k k_b/N_1}$ and $[\mathbf{D}_2(\rho_b)]_{l,l} = e^{-j2\pi l l_b/N_2}$. This phenomenon occurs in the matrix, whose elements is a given function of the corresponding elements of the circulant matrix, e.g.,

$$\dot{\mathbf{\Phi}}_1(v_b) = \mathbf{F}_{N_1} \dot{\mathbf{D}}_1(v_b) \mathbf{F}_{N_1}^{\mathrm{H}}, \tag{32}$$

$$\dot{\mathbf{\Phi}}_2(\rho_b) = \mathbf{F}_{N_2}^{\mathrm{H}} \dot{\mathbf{D}}_2(\rho_b) \mathbf{F}_{N_2}, \tag{33}$$

and

$$\mathbf{\Xi}_1(v_b) \triangleq \mathbf{\Phi}_1(v_b) \circ \mathbf{\Phi}_1^*(v_b) = \mathbf{F}_{N_1} \mathbf{E}_1(v_b) \mathbf{F}_{N_1}^{\mathrm{H}}, \tag{34}$$

$$\mathbf{\Xi}_2(\rho_b) \triangleq \mathbf{\Phi}_2(\rho_b) \circ \mathbf{\Phi}_2^*(\rho_b) = \mathbf{F}_{N_2}^{\mathrm{H}} \mathbf{E}_2(\rho_b) \mathbf{F}_{N_2}, \tag{35}$$

with $\mathbf{E}_1(v_b)$ and $\mathbf{E}_2(\rho_b)$ are diagonal matrices with their diagonal elements determined by the circular convolution of the sequences $\{e^{j2\pi k k_b/N_1}\}$ and $\{e^{-j2\pi l l_b/N_2}\}$, respectively. This inspires us to develop low-complexity algorithms by exploiting *fast Fourier transform* (FFT).

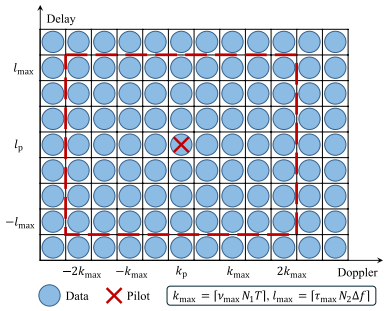


Fig. 2. OTFS frame design without guard symbols.

C. Frame Design for ISAC

For the sake of channel coefficient estimation and kinematic parameter sensing, pilots are embedded in OTFS frame. To achieve high spectral efficiency, we adopt an approach that superimposes pilots on the data symbols with no guard symbols [44], which reads

$$x_{k,l} = \begin{cases} x_{\mathrm{d},k,l} + x_{\mathrm{p}}, & k = k_{\mathrm{p}}, l = l_{\mathrm{p}}, \\ x_{\mathrm{d},k,l}, & \text{otherwise}, \end{cases} \tag{36}$$

where $x_{\mathrm{d},k,l}$ takes a value from constellation $\mathcal{S}$, and we consider only one pilot for simplicity. The spectral efficiency of the superimposed approach is $\eta = \log|\mathcal{S}|$ bps/Hz. The frame design for ISAC is illustrated in Fig. 2.

III. Variational Inference Based ISAC

In this section, we develop computationally efficient iterative ISAC receivers from a unified variational inference perspective for the OTFS-based MIMO system, where the global joint probability distribution is exploited. Therefore, we refer to this scheme as centralized ISAC (C-ISAC).

A. Graphical Model for ISAC

A generalized probabilistic representation[3] for the OTFS-based MIMO system (27) is given by

$$\begin{aligned} &p(\mathbf{X}, \mathbf{g} \mid \mathbf{Y}; \mathbf{p}) \\ &= p(\mathbf{X}) \prod_b p(g_b) \int_{\mathbf{h}_b} p(\mathbf{Y}_b \mid \mathbf{X}; g_b, \mathbf{h}_b) \delta(\mathbf{h}_b - f_b(\mathbf{p})), \end{aligned} \tag{37}$$

where $\mathbf{Y} = [\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_B^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{g} = [g_1, \ldots, g_b]^{\mathrm{T}}$, $\mathbf{h}_b = [\rho_b, \theta_b, v_b]^{\mathrm{T}}$, $\mathbf{p} = [\mathbf{u}^{\mathrm{T}}, v_x]^{\mathrm{T}}$ denotes the collection of all received symbols, channel coefficients, channel parameters and kinematic parameters, respectively. The mapping function $f_b(\mathbf{p})$ is given by (1)-(3). The likelihood function is

$$\begin{aligned} &p(\mathbf{Y}_b \mid \mathbf{X}; g_b, \mathbf{h}_b) \\ &= \int_{\mathbf{Z}} p(\mathbf{Y}_b \mid \mathbf{Z}_b; g_b, \theta_b) \times \delta(\mathbf{Z}_b - \mathbf{\Phi}_1(v_b)\mathbf{X}\mathbf{\Phi}_2(\rho_b); \rho_b, v_b) \\ &= \prod_n \int_{z_{b,n}} \mathcal{N}_{\mathbb{C}}([\mathbf{Y}_b]_{:,n}; g_b \mathbf{a}(\theta_b) z_{b,n}, \sigma_w^2 \mathbf{I}_M) \\ &\quad \times \delta(z_{b,n} - [\mathbf{\Phi}_1(v_b)]_{k,:} \mathbf{X} [\mathbf{\Phi}_2(\rho_b)]_{l,:}^{\mathrm{T}}), \end{aligned} \tag{38}$$

[3]Without any prior of $\mathbf{p}$, $p(\mathcal{X}, \mathbf{g} \mid \mathbf{Y}; \mathbf{p})$ is equivalent to $p(\mathcal{X}, \mathbf{g}, \mathbf{p} \mid \mathbf{Y})$ by assuming the uniform prior. This representation can directly be generalized to conditions with prior.

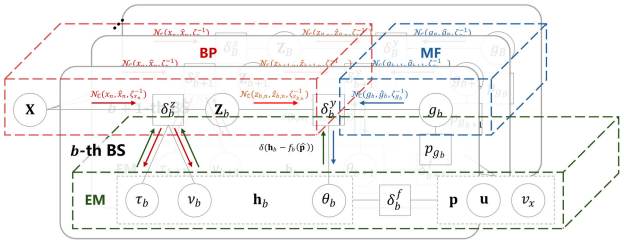


Fig. 3. Multi-layer factor graph model for OTFS-based MIMO system.

where $p(g_b) = \mathcal{N}_{\mathbb{C}}(g_b; 0, \gamma_{g_b}^{-1})$. The symbol prior of the superimposed frame (36) is

$$p(\mathbf{X}) = \prod_{k,l} p(x_{k,l}) = p_0(x_{k_{\mathrm{p}},l_{\mathrm{p}}} - x_{\mathrm{p}}) \prod_{(k,l)\neq(k_{\mathrm{p}},l_{\mathrm{p}})} p_0(x_{k,l}), \tag{39}$$

where $p_0(x) = \frac{1}{|\mathcal{S}|}\sum_{s_i\in\mathcal{S}} \delta(x - s_i)$ and the symbol $x_{k_{\mathrm{p}},l_{\mathrm{p}}}$ can serve as a pseudo data symbol which takes value on $\mathcal{S}_{\mathrm{p}} \triangleq \{s_i + x_{\mathrm{p}} \mid s_i \in \mathcal{S}\}$. The concept of pseudo data symbol can be used in a variety of OTFS frame designs. The function $\delta(\mathbf{h}_b - f_b(\mathbf{p}))$ maps the space of the kinematic parameter to that of the communication channel, which reads

$$\begin{aligned} \delta(\mathbf{h}_b - f_b(\mathbf{p})) = \delta(\theta_b - \theta(\Delta\mathbf{u}_b))\delta(\rho_b - \rho(\Delta\mathbf{u}_b)) \\ \times \delta(v_b - v(\mathbf{v}, \Delta\mathbf{u}_b)) \end{aligned} \tag{40}$$

The elements of $\delta(\mathbf{h}_b - f_b(\mathbf{p}))$ have an impact on the process for parameter sensing. Heuristically, $\delta(\theta_b - \theta(\Delta\mathbf{u}_b))$, $\delta(\rho_b - r(\Delta\mathbf{u}_b))$ and $\delta(v_b - v(\mathbf{v}, \Delta\mathbf{u}_b))$ can be interpreted as direction of arrival (DOA), time of arrival (TOA) and FDOA-based localization, respectively. The concept of FDOA-based localization was proposed in [53], which will be elaborated in Section IV.

According to the factorization in (37), we construct the multi-layer factor graph to characterize the statistical relationships between variables as depicted in Fig. 3.

B. A Free Energy Approach

It is well known that the maximum *a posteriori* (MAP) algorithm is optimal for sensing and communication. Nevertheless, the maximization problem of (37) is not tractable due to its inherent integral w.r.t. the variables that are discrete and highly coupled. For this reason, we resort to the efficient variational inference-based method to approximate $p(\mathbf{X}, \mathbf{g} \mid \mathbf{Y}; \mathbf{p})$ to a flexible trial distribution $q(\mathbf{X}, \mathbf{g}, \mathbf{p})$, which is chosen from a family of tractable distributions $\mathcal{Q}$ to minimize the variational free energy, i.e.,

$$\begin{aligned} q(\mathbf{X}, \mathbf{g}, \mathbf{p}) &= \arg\min_{q\in\mathcal{Q}} \mathrm{KL}\left(q(\mathbf{X}, \mathbf{g}, \mathbf{p}) \,||\, p(\mathbf{X}, \mathbf{g} \mid \mathbf{Y}; \mathbf{p})\right) \\ &= \arg\min_{q\in\mathcal{Q}} \sum_{\mathbf{X},\mathbf{g},\mathbf{p}} q(\mathbf{X}, \mathbf{g}, \mathbf{p}) \ln \frac{q(\mathbf{X}, \mathbf{g}, \mathbf{p})}{p(\mathbf{X}, \mathbf{g} \mid \mathbf{Y}; \mathbf{p})}, \end{aligned} \tag{41}$$

where

$$q(\mathbf{X}, \mathbf{g}, \mathbf{p}) = q(\mathbf{X}) q(\mathbf{g}) q(\mathbf{p}), \tag{42}$$

with

$$q(\mathbf{g}) = \prod_b q(g_b) \triangleq \prod_b \mathcal{N}_{\mathbb{C}}(g_b; \hat{g}_b, \zeta_{g_b}^{-1}), \tag{43}$$

$$q(\mathbf{p}) = \delta(\mathbf{p} - \hat{\mathbf{p}}) \triangleq \delta(\mathbf{u} - \hat{\mathbf{u}})\delta(v_x - \hat{v}_x). \tag{44}$$

Then the optimal solution of the MAP problem can be directly obtained based on the approximate distribution by exploiting its properties.

To seek the solution of problem (41), we resort to coordinate descent method which is to find the optimal $q(\mathbf{X})$, $q(\mathbf{g})$, and $q(\mathbf{p})$, iteratively. Substituting (42) into (41), we obtain

$$\begin{aligned}
&\mathrm{KL}\left(q(\mathbf{X},\mathbf{g},\mathbf{p}) \,||\, p(\mathbf{X},\mathbf{g} \mid \mathbf{Y};\mathbf{p})\right) \\
&= \int_{\mathbf{X},\mathbf{g},\mathbf{p}} q(\mathbf{X})q(\mathbf{g})q(\mathbf{p}) \ln \frac{q(\mathbf{X})q(\mathbf{g})q(\mathbf{p})}{p(\mathbf{X},\mathbf{g} \mid \mathbf{Y};\mathbf{p})} \\
&= \int_{\mathbf{X}} q(\mathbf{X}) \ln q(\mathbf{X}) + \int_{\mathbf{g}} q(\mathbf{g}) \ln q(\mathbf{g}) + \int_{\mathbf{p}} q(\mathbf{p}) \ln q(\mathbf{p}) \\
&\quad - \int_{\mathbf{X},\mathbf{g},\mathbf{p}} q(\mathbf{X})q(\mathbf{g})q(\mathbf{p}) \ln p(\mathbf{X},\mathbf{g} \mid \mathbf{Y};\mathbf{p}).
\end{aligned} \tag{45}$$

Hence, problem (41) can be solved by solving the following subproblems iteratively,

$$q(\mathbf{X}) = \arg\min_{q \in \mathcal{Q}_{\mathbf{X}}} \mathrm{KL}\left(q(\mathbf{X}) \,||\, \tilde{p}(\mathbf{X} \mid \mathbf{Y})\right), \tag{46}$$

$$q(\mathbf{g}) = \arg\min_{q \in \mathcal{Q}_{\mathbf{g}}} \mathrm{KL}\left(q(\mathbf{g}) \,||\, \tilde{p}(\mathbf{g} \mid \mathbf{Y})\right), \tag{47}$$

$$q(\mathbf{p}) = \arg\min_{q \in \mathcal{Q}_{\mathbf{p}}} \mathrm{KL}\left(q(\mathbf{p}) \,||\, \tilde{p}(\mathbf{p} \mid \mathbf{Y})\right), \tag{48}$$

where

$$\tilde{p}(\mathbf{X} \mid \mathbf{Y}) \propto \exp(\mathbb{E}_{\mathbf{g}\sim q(\mathbf{g}),\mathbf{p}\sim q(\mathbf{p})}\{\ln p(\mathbf{X},\mathbf{g} \mid \mathbf{Y};\mathbf{p})\}), \tag{49}$$

$$\tilde{p}(\mathbf{g} \mid \mathbf{Y}) \propto \exp(\mathbb{E}_{\mathbf{X}\sim q(\mathbf{X}),\mathbf{p}\sim q(\mathbf{p})}\{\ln p(\mathbf{X},\mathbf{g} \mid \mathbf{Y};\mathbf{p})\}), \tag{50}$$

$$\tilde{p}(\mathbf{p} \mid \mathbf{Y}) \propto \exp(\mathbb{E}_{\mathbf{X}\sim q(\mathbf{X}),\mathbf{g}\sim q(\mathbf{g})}\{\ln p(\mathbf{X},\mathbf{g} \mid \mathbf{Y};\mathbf{p})\}). \tag{51}$$

Here, we give the following proposal for the convergence performance of the proposed approach.

Proposition 1 (Convergence Property): In each iteration, if the solutions to subproblems (46), (47), and (48) exist, the variational free energy does not increase.

Proof: See Section A in the Appendix.

According to Proposition 1, the proposed free energy approach is always guaranteed to converge if the solutions to these subproblems exist.

We refer to (46), (47), and (48) as data detection, channel coefficient estimation, and kinematic parameter sensing, respectively. And the belief $q(\mathbf{X})$, $q(\mathbf{g})$, and $q(\mathbf{p})$ serve as the messages passing on the constructed factor graph. In the following subsections, we derive the efficient message passing algorithm for these subproblems.

C. Belief Propagation for Data Detection

Given $q(\mathbf{g})$ and $q(\mathbf{p})$, by substituting (37) into (49) and neglecting the terms that are unrelated to $\mathbf{X}$, we have

$$\begin{aligned}
&\frac{\tilde{p}(\mathbf{X} \mid \mathbf{Y})}{p(\mathbf{X})} \\
&\propto \exp(\mathbb{E}_{\mathbf{g}\sim q(\mathbf{g}),\mathbf{p}\sim q(\mathbf{p})}\{\sum_b \ln p(\mathbf{Y}_b \mid \mathbf{X}; g_b, f_b(\mathbf{p}))\}) \\
&\propto \prod_b \exp(\mathbb{E}_{g_b\sim\mathcal{N}_{\mathbb{C}}(g_b;\hat{g}_b,\zeta_{g_b}^{-1})}\{\ln p(\mathbf{Y}_b \mid \mathbf{X}; g_b, f_b(\hat{\mathbf{p}}))\}).
\end{aligned} \tag{52}$$

Considering the term of (52) w.r.t. b, we obtain

$$\begin{aligned}
&\exp(\mathbb{E}_{g_b\sim\mathcal{N}_{\mathbb{C}}(g_b;\hat{g}_b,\zeta_{g_b}^{-1})}\{\ln p(\mathbf{Y}_b \mid \mathbf{X}; g_b, f_b(\hat{\mathbf{p}}))\}) \\
&= \exp(\sum_n \mathbb{E}_{g_b}\{\ln \int_{z_{b,n}} \mathcal{N}_{\mathbb{C}}([\mathbf{Y}_b]_{:,n}; g_b\mathbf{a}(\hat{\theta}_b)z_{b,n}, \sigma_w^2\mathbf{I}_M) \\
&\quad \times \delta(z_{b,n} - [\mathbf{\Phi}_1(\hat{v}_b)]_{k,:}\mathbf{X}[\mathbf{\Phi}_2(\hat{\rho}_b)]_{l,:}^{\mathrm{T}})\}) \\
&= \prod_n \int_{z_{b,n}} \exp(\mathbb{E}_{g_b}\{\ln \mathcal{N}_{\mathbb{C}}([\mathbf{Y}_b]_{:,n}; g_b\mathbf{a}(\hat{\theta}_b)z_{b,n}, \sigma_w^2\mathbf{I}_M)\}) \\
&\quad \times \delta(z_{b,n} - [\mathbf{\Phi}_1(\hat{v}_b)]_{k,:}\mathbf{X}[\mathbf{\Phi}_2(\hat{\rho}_b)]_{l,:}^{\mathrm{T}}) \\
&\propto \prod_n \mathcal{N}_{\mathbb{C}}(\bar{z}_{b,n}; [\mathbf{\Phi}_1(\hat{v}_b)]_{k,:}\mathbf{X}[\mathbf{\Phi}_2(\hat{\rho}_b)]_{l,:}^{\mathrm{T}}, \sigma_{z_b}^2) \\
&\triangleq \tilde{p}(\bar{\mathbf{Z}}_b \mid \mathbf{X}),
\end{aligned} \tag{53}$$

where

$$\bar{z}_{b,n} = \frac{\hat{g}_b^*[\mathbf{r}(\hat{\theta}_b, \mathbf{Y}_b)]_n}{\zeta_{g_b}^{-1} + |\hat{g}_b|^2}, \tag{54}$$

$$\sigma_{z_b}^2 = \frac{\sigma_w^2}{\zeta_{g_b}^{-1} + |\hat{g}_b|^2}, \tag{55}$$

with $\mathbf{r}^{\mathrm{T}}(\theta, \mathbf{Y}) = \mathbf{a}^{\mathrm{H}}(\theta)\mathbf{Y}$ denoting the beam alignment operator.

Remark 3: The digital beam alignment process $\mathbf{r}(\hat{\theta}_b, \mathbf{Y}_b)$ can also be achieved by performing analog beam alignment on (13) first, i.e., $\mathbf{a}^{\mathrm{H}}(\hat{\theta}_b)\mathbf{r}_b(t)$, followed by *Wigner transform*. This schedule does not incur additional hardware overhead for signal processing. ■

It is noteworthy that $\bar{z}_{b,n}$ can be seen as an observation of SISO OTFS signal corrupted by an AWGN, i.e., $\bar{z}_{b,n} = z_{b,n} + \bar{w}_{z_{b,n}}$, where $z_{b,n} = [\mathbf{\Phi}_1(\hat{v}_b)]_{k,:}\mathbf{X}[\mathbf{\Phi}_2(\hat{\rho}_b)]_{l,:}^{\mathrm{T}}$ and $\bar{w}_{z_{b,n}} \sim \mathcal{N}_{\mathbb{C}}(\bar{w}_{z_{b,n}}; 0, \sigma_{z_b}^2)$. This can be rewritten as,

$$\bar{\mathbf{Z}}_b = \mathbf{\Phi}_1(\hat{v}_b)\mathbf{X}\mathbf{\Phi}_2^{\mathrm{T}}(\hat{\rho}_b) + \bar{\mathbf{W}}_{z_b}, \tag{56}$$

where $[\bar{\mathbf{Z}}_b]_{k,l} = \bar{z}_{b,n}$ and $[\bar{\mathbf{W}}_{z_b}]_{k,l} = \bar{w}_{z_{b,n}}$. Hence, we obtain that

$$q(\mathbf{X}) = \arg\min_{q \in \mathcal{Q}_{\mathbf{X}}} \mathrm{KL}(q(\mathbf{X}) \,||\, \prod_b \tilde{p}(\bar{\mathbf{Z}}_b \mid \mathbf{X})p(\mathbf{X})). \tag{57}$$

Minimizing the variational free energy w.r.t. trial probability functions $q(\mathbf{X})$ is an exact procedure with exponential complexity. Hence, we resort to Bethe method, which is an efficient way for this problem by minimizing the region-based free energy. Constructing the corresponding Lagrangian function of (57) subject to a set of normalization constraints, the variational free energy is minimized to obtain the BP rule [54], i.e.,

$$\mu_{a\to i}(x_i) \propto \int_{\mathbf{x}_a \setminus x_i} f_a(\mathbf{x}_a) \prod_{j\in\mathcal{M}(a)\setminus i} n_{j\to a}(x_j), \tag{58}$$

$$\mu_{i\to a}(x_i) = \prod_{c\in\mathcal{M}(i)\setminus a} m_{c\to i}(x_i), \tag{59}$$

$$q(x_i) = \prod_{c\in\mathcal{M}(i)} m_{c\to i}(x_i), \tag{60}$$

Algorithm 1 The GAMP Algorithm for Data Detection

1: **Initialization**: Set $[\hat{\mathbf{X}}]_{k,l}=0$, $[\mathbf{\Sigma}_x]_{k,l}=1$, and $[\hat{\mathbf{S}}]_{k,l}=0$. Then, set $[\hat{\mathbf{X}}]_{k_\mathrm{p},l_\mathrm{p}}=x_\mathrm{p}$.
2: **for** $i_\mathrm{d}=1$ to $I_\mathrm{d,end}$ **do**
3: *//Output linear step*
4: **for** $b=1$ to B **do**
5: $\hat{\mathbf{P}}_b=\mathbf{\Phi}_1(\hat{v}_b)\hat{\mathbf{X}}\mathbf{\Phi}_2^{\mathrm{T}}(\hat{\rho}_b)-\mathbf{\Sigma}_{p_b}\circ\hat{\mathbf{S}}_b$,
6: $\mathbf{\Sigma}_{p_b}=\mathbf{\Xi}_1(\hat{v}_b)\mathbf{\Sigma}_x\mathbf{\Xi}_2^{\mathrm{T}}(\hat{\rho}_b)$.
7: **end for**
8: *//Output nonlinear step*
9: **for** $b=1$ to B **do**
10: $\hat{\mathbf{Z}}_b=\mathbf{\Sigma}_{z_b}\circ(\bar{\mathbf{Z}}_b/\sigma_{z_b}^2+\hat{\mathbf{P}}_b./\mathbf{\Sigma}_{p_b})$,
11: $\mathbf{\Sigma}_{z_b}=1./(1/\sigma_{z_b}^2+1./\mathbf{\Sigma}_{p_b})$,
12: $\sigma_{z_b}^2=(\|\bar{\mathbf{Z}}_b-\hat{\mathbf{Z}}_b\|_F^2+\langle\mathbf{\Sigma}_{z_b}\rangle)/N$,
13: $\hat{\mathbf{S}}_b=\mathbf{\Gamma}_{s_b}\circ(\bar{\mathbf{Z}}_b-\hat{\mathbf{P}}_b)$,
14: $\mathbf{\Gamma}_{s_b}=1./(\sigma_{z_b}^2+\mathbf{\Sigma}_{p_b})$.
15: **end for**
16: *//Input linear step*
17: $\hat{\mathbf{R}}=\hat{\mathbf{X}}+\mathbf{\Sigma}_r\circ\sum_b(\mathbf{\Phi}_1^{\mathrm{H}}(\hat{v}_b)\hat{\mathbf{S}}_b\mathbf{\Phi}_2^{*}(\hat{\rho}_b))$,
18: $\mathbf{\Sigma}_r=1./\sum_b\mathbf{\Gamma}_{r_b}$,
19: $\mathbf{\Gamma}_{r_b}=\mathbf{\Xi}_1^{\mathrm{T}}(\hat{v}_b)\mathbf{\Gamma}_{s_b}\mathbf{\Xi}_2(\hat{\rho}_b)$.
20: *//Input nonlinear step*
21: **for** $\forall k,l$ **do**
22: $[\hat{\mathbf{X}}]_{k,l}=\mathbb{E}_{x_{k,l}\sim b(x_{k,l})}\{x_{k,l}\}$,
23: $[\mathbf{\Sigma}_x]_{k,l}=\mathbb{V}_{x_{k,l}\sim b(x_{k,l})}\{x_{k,l}\}$,
24: where $b(x_{k,l})\propto p(x_{k,l})\mathcal{N}_{\mathbb{C}}(x_{k,l};[\hat{\mathbf{R}}]_{k,l},[\mathbf{\Sigma}_r]_{k.l})$.
25: **end for**
26: **end for**
27: **return** $\hat{x}_{k,l}=[\hat{\mathbf{X}}]_{k,l}$, $\zeta_{x_{k,l}}^{-1}=[\mathbf{\Sigma}_x]_{k,l}$, $\hat{z}_{b,k,l}=[\hat{\mathbf{Z}}_b]_{k,l}$, and $\zeta_{z_{b,k,l}}^{-1}=[\mathbf{\Sigma}_{z_b}]_{k,l}$.

where $f_a(\mathbf{x}_a)$ denotes the factors of $\tilde{p}(\bar{\mathbf{Z}}_b\mid\mathbf{X})p(\mathbf{X})$ with the set of its variables denoting by $\mathcal{M}(a)$, and $\mathcal{M}(i)$ denotes the set of factors that include variable x_i. To further reduce the computational complexity, by adopting expectation propagation (EP) and first-order approximation, the efficient generalized approximate message passing (GAMP) algorithm for problem (57) can be derived [55], as outlined in **Algorithm 1**. This algorithm has demonstrated its excellent tradeoff between BER performance and complexity for data detection [56]. Besides, the devised GAMP algorithm can be implemented with lower complexity by FFT.

It is evident that there exists an estimation error in $\hat{\mathbf{p}}$. Given our assumption $q(\mathbf{p})=\delta(\mathbf{p}-\hat{\mathbf{p}})$, this error is disregarded, resulting in mismatch in the matrices $\mathbf{a}(\hat{\theta}_b)$, $\mathbf{\Phi}(\hat{v}_b)$, and $\mathbf{\Phi}(\hat{\rho}_b)$. Consequently, the variance of the effective noise is underestimated. By introducing the effective noise $\sigma_{z_b}^2$ as a variable and harnessing mean field approximation, the variance updating formula is obtained [37], which is shown in Line 12. Afterwards, we obtain the belief of symbol $\mathbf{X}$, which reads

$$q(\mathbf{X})\approx\prod_{k,l}\mathcal{N}_{\mathbb{C}}(x_{k,l};\hat{x}_{k,l},\zeta_{x_{k,l}}^{-1})\triangleq\prod_{k,l}q(x_{k,l}). \tag{61}$$

D. Mean Field for Channel Coefficient Estimation

Given $q(\mathbf{X})$ and $q(\mathbf{p})$, by substituting (37) into (50) and neglecting the terms that are unrelated to $\mathbf{g}$, we have

$$\begin{aligned}&\frac{\tilde{p}(\mathbf{g}\mid\mathbf{Y})}{q(\mathbf{g})}\\&\propto\exp(\mathbb{E}_{\mathbf{X}\sim q(\mathbf{X}),\mathbf{p}\sim q(\mathbf{p})}\{\sum_b\ln p(\mathbf{Y}_b\mid\mathbf{X};g_b,f_b(\mathbf{p}))\})\\&\propto\prod_b\exp(\mathbb{E}_{\mathbf{X}\sim q(\mathbf{X})}\{\ln p(\mathbf{Y}_b\mid\mathbf{X};g_b,f_b(\hat{\mathbf{p}}))\})\end{aligned} \tag{62}$$

Assuming the mean field approximation, i.e., $q(\mathbf{g})\triangleq\prod_b q(g_b)$ are fully factorized. After some algebra, (47) can be rewritten as

$$q(g_b)=\arg\min_{q\in\mathcal{Q}_g}\mathrm{KL}\left(q(g_b)\,||\,\tilde{p}(g_b\mid\mathbf{Y})\right), \tag{63}$$

where

$$\frac{\tilde{p}(g_b\mid\mathbf{Y})}{p(g_b)}\propto\exp(\mathbb{E}_{\mathbf{X}\sim q(\mathbf{X})}\{\ln p(\mathbf{Y}_b\mid\mathbf{X};g_b,f_b(\hat{\mathbf{p}}))\}). \tag{64}$$

Considering the right term of (64), we have

$$\begin{aligned}&\exp(\mathbb{E}_{\mathbf{X}\sim q(\mathbf{X})}\{\ln p(\mathbf{Y}_b\mid\mathbf{X};g_b,f_b(\hat{\mathbf{p}}))\})\\&=\exp(\sum_n\mathbb{E}_{\mathbf{X}\sim q(\mathbf{X})}\{\ln\int_{z_{b,n}}\mathcal{N}_{\mathbb{C}}([\mathbf{Y}_b]_{:,n};g_b\mathbf{a}(\hat{\theta}_b)z_{b,n},\sigma_w^2\mathbf{I}_M)\\&\quad\times\delta(z_{b,n}-[\mathbf{\Phi}_1(\hat{v}_b)]_{k,:}\mathbf{X}[\mathbf{\Phi}_2(\hat{\rho}_b)]_{l,:}^{\mathrm{T}})\})\\&=\exp(\sum_n\int_{\mathbf{X}}q(\mathbf{X})\int_{z_{b,n}}\ln\mathcal{N}_{\mathbb{C}}([\mathbf{Y}_b]_{:,n};g_b\mathbf{a}(\hat{\theta}_b)z_{b,n},\sigma_w^2\mathbf{I}_M)\\&\quad\times\delta(z_{b,n}-[\mathbf{\Phi}_1(\hat{v}_b)]_{k,:}\mathbf{X}[\mathbf{\Phi}_2(\hat{\rho}_b)]_{l,:}^{\mathrm{T}}))\\&=\exp(\sum_n\mathbb{E}_{z_{b,n}\sim q(z_{b,n})}\{\ln\mathcal{N}_{\mathbb{C}}([\mathbf{Y}_b]_{:,n};g_b\mathbf{a}(\hat{\theta}_b)z_{b,n},\sigma_w^2\mathbf{I}_M)\})\\&\propto\mathcal{N}_{\mathbb{C}}(\bar{g}_b;g_b,\sigma_{g_b}^2),\end{aligned} \tag{65}$$

where

$$\bar{g}_b=\frac{\sum_n\hat{z}_{b,n}^*[\mathbf{r}(\hat{\theta}_b,\mathbf{Y}_b)]_n}{\sum_n\zeta_{z_{b,n}}^{-1}+|\hat{z}_{b,n}|^2}, \tag{66}$$

$$\sigma_{g_b}^2=\frac{\sigma_w^2}{\sum_n\zeta_{z_{b,n}}^{-1}+|\hat{z}_{b,n}|^2}, \tag{67}$$

with $q(z_{b,n})=\int_{\mathbf{X}}q(\mathbf{X})\delta(z_{b,n}-[\mathbf{\Phi}_1(\hat{v}_b)]_{k,:}\mathbf{X}[\mathbf{\Phi}_2(\hat{\rho}_b)]_{l,:}^{\mathrm{T}})\triangleq\mathcal{N}_{\mathbb{C}}(z_{b,n};\hat{z}_{b,n},\zeta_{z_{b,n}}^{-1})$ calculated at the data detection. Then, the solution of problem (63) is

$$q(g_b)\propto p(g_b)\mathcal{N}_{\mathbb{C}}(\bar{g}_b;g_b,\sigma_{g_b}^2)\propto\mathcal{N}_{\mathbb{C}}(g_b;\hat{g}_b,\zeta_{g_b}^{-1}), \tag{68}$$

where

$$\hat{g}_b=\frac{\bar{g}_b}{\zeta_{g_b}\sigma_{g_b}^2}, \tag{69}$$

$$\zeta_{g_b}=\gamma_{g_b}+1/\sigma_{g_b}^2. \tag{70}$$

E. Expectation Maximization for Kinematic Parameter Sensing

Given $q(\mathbf{X})$ and $q(\mathbf{g})$, by substituting (37) into (51) and neglecting the terms that are unrelated to $\mathbf{g}$, we have

$$\tilde{p}(\mathbf{p} \mid \mathbf{Y}) \propto \exp(L(\mathbf{p}, \hat{\mathbf{p}})), \tag{71}$$

where

$$L(\mathbf{p}, \hat{\mathbf{p}}) = \mathbb{E}_{\mathbf{X}\sim q(\mathbf{X}), \mathbf{g}\sim q(\mathbf{g})}\{\ln p(\mathbf{X}, \mathbf{g} \mid \mathbf{Y}; \mathbf{p}), \tag{72}$$

with the variable $\hat{\mathbf{p}}$ indicating that the variational distributions $q(\mathbf{X})$ and $q(\mathbf{g})$ are obtained based on the latest estimated parameters $\hat{\mathbf{p}}$, which agrees with the notation of E-step in expectation maximization algorithm.

Here, we give the following proposal for calculating the delta belief that minimizes the variational free energy.

Proposition 2 (Delta Belief): Assume that $q(x) = \delta(x - \hat{x})$, then the value $\hat{x}$ that minimizes $\mathrm{KL}(q(x) \mid p(x))$ is as

$$\hat{x} = \arg\max_{x} \ln p(x). \tag{73}$$

Proof: See Section VI-B in the Appendix.

Hence, minimizing the variational free energy is equivalent to

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} -L(\mathbf{p}, \hat{\mathbf{p}}), \tag{74}$$

which agrees with the notation of M-step in expectation maximization algorithm.

Considering (72), we have

$$L(\mathbf{p}, \hat{\mathbf{p}}) \triangleq \sum_{b} L_b(\mathbf{p}, \hat{\mathbf{p}}), \tag{75}$$

where

$$\begin{aligned} &L_b(\mathbf{p}, \hat{\mathbf{p}}) \\ &\triangleq \mathbb{E}_{\mathbf{X}\sim q(\mathbf{X}), g_b\sim \mathcal{N}_{\mathbb{C}}(g_b;\hat{g}_b,\zeta_{g_b}^{-1})}\{\ln p(\mathbf{Y}_b \mid \mathbf{X}; g_b, f_b(\mathbf{p}))\} \\ &= -\frac{1}{\sigma_w^2}\sum_{n}\left\|[\mathbf{Y}_b]_{:,n} - \hat{g}_b\mathbf{a}(\theta_b)[\mathbf{\Phi}_1(v_b)]_{k,:}\hat{\mathbf{X}}[\mathbf{\Phi}_2(\rho_b)]_{l,:}^{\mathrm{T}}\right\|_2^2 \\ &\quad - \frac{1}{\sigma_w^2}\sum_{n}\zeta_{g_b}^{-1}\left|[\mathbf{\Phi}_1(v_b)]_{k,:}\hat{\mathbf{X}}[\mathbf{\Phi}_2(\rho_b)]_{l,:}^{\mathrm{T}}\right|^2, \\ &\quad - \frac{1}{\sigma_w^2}\sum_{n}(\zeta_{g_b}^{-1} + |\hat{g}_b|^2)[\mathbf{\Xi}_1(v_b)]_{k,:}\mathbf{\Sigma}_{\mathbf{X}}[\mathbf{\Xi}_2(\rho_b)]_{l,:}^{\mathrm{T}}, \end{aligned} \tag{76}$$

where $\theta_b = \theta(\Delta\mathbf{u}_b)$, $\rho_b = \rho(\Delta\mathbf{u}_b)$ and $v_b = v(v, \Delta\mathbf{u}_b)$ are functions of the kinematic parameters. We assume that the scenario satisfies the localizability criterion [57]. Then, we can learn parameters $\mathbf{p}$ by exploiting gradient descent algorithms for problem (74). According to (76), we give the objective functions of the channel parameters, which reads,

$$J(\theta_b) = \left\|\mathbf{Y}_b - \hat{g}_b\mathbf{a}(\theta_b)\,\mathrm{vec}(\check{\mathbf{Z}}_b)^{\mathrm{T}}\right\|_F^2, \tag{77}$$

$$\begin{aligned} J(\rho_b) = &\left\|\check{\mathbf{R}}_b - \hat{g}_b\check{\mathbf{Z}}_{1,b}\mathbf{\Phi}_2^{\mathrm{T}}(\rho_b)\right\|_F^2 \\ &+ \zeta_{g_b}^{-1}\left\|\check{\mathbf{Z}}_{1,b}\mathbf{\Phi}_2^{\mathrm{T}}(\rho_b)\right\|_F^2 \\ &+ (\zeta_{g_b}^{-1} + |\hat{g}_b|^2)\langle\mathbf{\Sigma}_{z_{1,b}}\mathbf{\Xi}_2^{\mathrm{T}}(\rho_b)\rangle, \end{aligned} \tag{78}$$

$$\begin{aligned} J(v_b) = &\left\|\check{\mathbf{R}}_b - \hat{g}_b\mathbf{\Phi}_1(v_b)\check{\mathbf{Z}}_{2,b}\right\|_F^2 \\ &+ \zeta_{g_b}^{-1}\left\|\mathbf{\Phi}_1(v_b)\check{\mathbf{Z}}_{2,b}\right\|_F^2 \\ &+ (\zeta_{g_b}^{-1} + |\hat{g}_b|^2)\langle\mathbf{\Xi}_1(v_b)\mathbf{\Sigma}_{z_{2,b}}\rangle, \end{aligned} \tag{79}$$

where $\check{\mathbf{Z}}_b = \mathbf{\Phi}_1(\hat{v}_b)\hat{\mathbf{X}}\mathbf{\Phi}_2^{\mathrm{T}}(\hat{\rho}_b)$, $\check{\mathbf{R}}_b = \mathrm{mat}(\mathbf{r}(\hat{\theta}_b, \mathbf{Y}_b))$, $\check{\mathbf{Z}}_{1,b} = \mathbf{\Phi}_1(\hat{v}_b)\hat{\mathbf{X}}$, $\mathbf{\Sigma}_{z_{1,b}} = \mathbf{\Xi}_1(\hat{v}_b)\mathbf{\Sigma}_x$, $\check{\mathbf{Z}}_{2,b} = \hat{\mathbf{X}}\mathbf{\Phi}_2^{\mathrm{T}}(\hat{\rho}_b)$ and $\mathbf{\Sigma}_{z_{2,b}} = \mathbf{\Sigma}_x\mathbf{\Xi}_2^{\mathrm{T}}(\hat{\rho}_b)$. Notably, leveraging variational inference, the derived objective function for kinematic parameter sensing directly infers kinematic parameters from received signals, accounting for the variance of data symbols and channel coefficients, i.e., the second and third terms in (78) and (79). We refer to these terms as *variance-based regularizer*. As detailed in Section V, this thoughtful consideration enables the proposed method outperform classic ML-based and nonlinear least square (NLS)-based approaches.

We first calculate the derivatives of (76) w.r.t. the channel parameters θ_b, ρ_b and v_b. The derivative of (76) w.r.t. the DOA parameter θ_b is given by

$$\frac{\partial J(\theta_b)}{\partial \theta_b} = -2\Re\left\{\hat{g}_b\,\mathrm{vec}(\check{\mathbf{Z}}_b)^{\mathrm{T}}\mathbf{Y}_b^{\mathrm{H}}\dot{\mathbf{a}}(\theta_b)\right\}. \tag{80}$$

The derivative of (76) w.r.t. the TOA parameter ρ_b is

$$\begin{aligned} \frac{\partial J(\rho_b)}{\partial \rho_b} = &-2\Re\left\{\langle(\hat{g}_b^*\check{\mathbf{Z}}_{1,b}^{\mathrm{H}}(\check{\mathbf{R}}_b - \hat{g}_b\check{\mathbf{Z}}_{1,b}\mathbf{\Phi}_2^{\mathrm{T}}(\rho_b))) \circ \dot{\mathbf{\Phi}}_2^{\mathrm{H}}(\rho_b)\rangle\right\} \\ &+ 2\zeta_{g_b}^{-1}\Re\left\{\langle(\check{\mathbf{Z}}_{1,b}^{\mathrm{H}}\check{\mathbf{Z}}_{1,b}\mathbf{\Phi}_2^{\mathrm{T}}(\rho_b)) \circ \dot{\mathbf{\Phi}}_2^{\mathrm{H}}(\rho_b)\rangle\right\} \\ &+ (\zeta_{g_b}^{-1} + |\hat{g}_b|^2)\langle\mathbf{\Sigma}_{z_{1,b}}\dot{\mathbf{\Xi}}_2^{\mathrm{T}}(\rho_b)\rangle. \end{aligned} \tag{81}$$

The derivative of (76) w.r.t. the FDOA parameter v_b is

$$\begin{aligned} \frac{\partial J(v_b)}{\partial v_b} = &-2\Re\left\{\langle(\hat{g}_b^*(\check{\mathbf{R}}_b - \hat{g}_b\mathbf{\Phi}_1(v_b)\check{\mathbf{Z}}_{2,b})\check{\mathbf{Z}}_{2,b}^{\mathrm{H}}) \circ \dot{\mathbf{\Phi}}_1^*(v_b)\rangle\right\} \\ &+ 2\zeta_{g_b}^{-1}\Re\left\{\langle(\mathbf{\Phi}_1(v_b)\check{\mathbf{Z}}_{2,b}\check{\mathbf{Z}}_{2,b}^{\mathrm{H}}) \circ \dot{\mathbf{\Phi}}_1^*(v_b)\rangle\right\} \\ &+ (\zeta_{g_b}^{-1} + |\hat{g}_b|^2)\langle\dot{\mathbf{\Xi}}_1(v_b)\mathbf{\Sigma}_{z_{2,b}}\rangle. \end{aligned} \tag{82}$$

Additionally, we calculate the derivatives of the channel parameters w.r.t. the kinematic parameters. The derivative of θ_b w.r.t. $\mathbf{u}$ is

$$\frac{\partial \theta_b}{\partial \mathbf{u}} = \frac{\partial \theta(\Delta\mathbf{u}_b)}{\partial \mathbf{u}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\frac{\Delta\mathbf{u}_b}{\|\Delta\mathbf{u}_b\|_2^2}. \tag{83}$$

The derivative of ρ_b w.r.t. $\mathbf{u}$ is

$$\frac{\partial \rho_b}{\partial \mathbf{u}} = \frac{\partial r(\Delta\mathbf{u}_b)}{\partial \mathbf{u}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\frac{\Delta\mathbf{u}_b}{\|\Delta\mathbf{u}_b\|_2}. \tag{84}$$

The derivative of ρ_b w.r.t. $\mathbf{p} = [\mathbf{u}^{\mathrm{T}}, v_x]^{\mathrm{T}}$ is

$$\frac{\partial v_b}{\partial \mathbf{u}} = \frac{\partial v(v_x, \Delta\mathbf{u}_b)}{\partial \mathbf{u}} = [0, -v_x]\frac{\Delta\mathbf{u}_b}{\|\Delta\mathbf{u}_b\|_2}\frac{\partial \theta_b}{\partial \mathbf{u}}, \tag{85}$$

and

$$\frac{\partial v_b}{\partial v_x} = \frac{\partial v(v_x, \Delta\mathbf{u}_b)}{\partial \mathbf{u}} = [1, 0]\frac{\Delta\mathbf{u}_b}{\|\Delta\mathbf{u}_b\|_2}. \tag{86}$$

Then, we can use a variety of gradient descent algorithms to find the global solution of (74).

The objective functions (77), (78) and (79) have a intuitive insight that could be interpreted as DOA, TOA and FDOA-based localization, respectively. When localizability[4] is satisfied [57], the corresponding localization mode can be instantiated by considering one of the objective functions independently.

[4]The localizability of FDOA-based localization have not been studied yet. In the next section, we gives an example of how to use radial velocity information for localization.

Algorithm 2 The Proposed C-ISAC Algorithm

Initialization: Set $[\hat{\mathbf{X}}]_{k,l} = 0$ and $[\mathbf{\Sigma}_x]_{k,l} = 1$. Then, set $[\hat{\mathbf{X}}]_{k_\mathrm{p},l_\mathrm{p}} = x_\mathrm{p}$. Set $\hat{\mathbf{p}}$ to value that maximizes (75) with coarse on-grid search. Set $\hat{\mathbf{Z}}_b = \mathbf{\Phi}_1(\hat{v}_b)\hat{\mathbf{X}}\mathbf{\Phi}_2^{\mathrm{T}}(\hat{\rho}_b)$ and $\mathbf{\Sigma}_{z_b} = \mathbf{\Xi}_1(\hat{v}_b)\mathbf{\Sigma}_x\mathbf{\Xi}_2^{\mathrm{T}}(\hat{\rho}_b)$.
for $i = 1$ to I_{end} **do**
  //sensing cycle
  for $i_\mathrm{s} = 1$ to $I_{\text{s,end}}$ **do**
    calculate $\hat{\mathbf{p}}$ for problem (74) using gradient descent with the gradients specified by (80)–(86).
  end for
  //communication cycle
  for $i_\mathrm{c} = 1$ to $I_{\text{c,end}}$ **do**
    calculate $\hat{g}$ and ζ_{g_b} using (69) and (70),
    calculate $\hat{x}_{k,l}$, $\zeta_{x_{k,l}}$, $\hat{z}_{b,n}$ and $\zeta_{z_{b,n}}^{-1}$ using **Algorithm 1**.
  end for
end for
return $\hat{x}_{k,l}$, $\zeta_{x_{k,l}}^{-1}$, $\hat{g}$, $\hat{\mathbf{p}}$, and $\{\hat{\rho}_b, \hat{\theta}_b, \hat{v}_b\}$.

F. Message Passing

So far, we derive the messages passing on the constructed multi-layer factor graph, which is shown in Fig. 3. As the variables are highly coupled, the proposed algorithm need to update the messages iteratively. After sufficient number of iterations, the messages converge to a fixed point. Then, the transmitted data and the kinematic parameters of the UE can be extracted. Therefore, the message passing schedule has a crucial effect on the performance [58]. Heuristically, we propose an ISAC message passing schedule consisted of two iterative cycles, namely, communication cycle and sensing cycle. If the messages of one cycle converge, the messages of the other cycle start updating. In the communication cycle, data detection and channel coefficient estimation are performed iteratively. In the sensing cycle, the channel parameter sensing is performed.

Significantly, at the initial stage, as no information about the data symbols is known, we can only estimate the parameters by considering the pilot symbol. However, after sufficient number of iterations, a confident belief on the data symbols will bring the performance gain of the estimation by regarding them as *pseudo* pilot symbols. Thus, we propose a data-aided estimation method, which heuristically select confident received symbol $[\mathbf{Y}_b]_{:,n}$ for estimation. The selection criterion is as follows,

$$\mathcal{M}_i = \begin{cases} \{(k,l) \mid |k-k_\mathrm{p}| \le k_{\max}, 0 \le l - l_\mathrm{p} \le l_{\max}\}, & i \le I, \\ \{(k,l) \mid |\forall k,l\}, & i > I, \end{cases} \tag{87}$$

where i is the iteration index, and I is set experimentally. Then, the summation in (76)-(82), including norms and operator $\langle\cdot\rangle$, is performing on $n \in \mathcal{M}_t$ along the delay-Doppler index.

The summarized algorithm, referred to as centralized ISAC (C-ISAC), is shown in **Algorithm 2**.

IV. A Communication-Efficient Federated Learning Scheme

A. Communication-Efficient Federated Learning

In the proposed C-ISAC scheme, the received symbol $\mathbf{Y}$ is transmitted via backhaul links to the CPU, resulting in significant communication overhead. To address this issue, we resort to the advanced federated learning approach, where ISAC algorithm is processed in a distributed manner. Specifically, local sensing and communication results are obtained at DPUs equipped on each BS. These results are then aggregated at the CPU to produce a global ISAC result that can be subsequently downloaded by each DPU. The federated learning scheme also helps alleviate the burden of CPU. However, since the communication overhead for aggregating communication results remains similar to that of the C-ISAC scheme, we focus on a more communication-efficient federated learning scheme, where only the sensing results are transmitted to the CPU.

In the proposed federated learning scheme, at each DPU, data detection and the channel coefficient estimation are performed similarly to C-ISAC scheme, with the exception of setting $\mathbf{Y} = \mathbf{Y}_b$. Subsequently, (74) is optimized in a distributed manner with the kinematic parameter $\mathbf{p}$ representing a global model learned from aggregated local parameters. Based on the processing of different local parameters, three federated modes are proposed, namely classic FedGD-ISAC [59], FedAvg-ISAC [60] and novel FedLoc-ISAC.

B. Classic Federated Learning

FedGD-ISAC is a relatively direct federated learning method, which is essentially a distributed version of gradient descent algorithm. In FedGD-ISAC, each DPU uses its local data to calculate the local gradient of $L_b(\mathbf{p}, \hat{\mathbf{p}})$ w.r.t. $\mathbf{p}$. Then, the local gradients are transmitted to the CPU to update the local estimate according to the global gradient that averages the local gradients, i.e., $\frac{\partial L(\mathbf{p},\hat{\mathbf{p}})}{\partial \mathbf{p}} = \sum_b \frac{\partial L_b(\mathbf{p},\hat{\mathbf{p}})}{\partial \mathbf{p}}$. In spite of the fact that the gradient of FedGD-ISAC is in the same form as that of C-ISAC scheme, the two diagram have different performance due to the different objective function when adopting some method, e.g., adaptive step-size method.

FedAvg-ISAC is an improvement on FedGD-ISAC aimed at reducing communication rounds to improve efficiency. In FedAvg-ISAC, each DPU updates the local estimate. Afterwards, the estimate are aggregated at the CPU to give a global estimate. This process reduces the necessary communication rounds as each DPU can perform sensing cycles locally. For FedAvg-ISAC, we give the following proposal for kinematic parameter sensing on single BS.

Proposition 3 (Parameter Equivalence): Assume that $J(\mathbf{h})$ is the objective function and a function

$$\begin{aligned} f : \mathbb{P} &\to \mathbb{H} \\ \mathbf{p} &\mapsto \mathbf{h} \end{aligned} \tag{88}$$

is well-posed pointwise. Then, $\hat{\mathbf{p}} = f^{-1}(\hat{\mathbf{h}})$ is the solution that minimizes $J(f(\mathbf{p}))$, where $\hat{\mathbf{h}}$ is the solution that minimizing $J(\mathbf{h})$.

Proof: This can be proved by contradiction. As f is well-posed, $\forall \bar{\mathbf{p}} \neq \hat{\mathbf{p}}$, we have $\bar{\mathbf{h}} \neq \hat{\mathbf{h}}$, which does not minimize $J(\mathbf{h})$. ■

Obviously, the coordinate transformation $f_b(\mathbf{p})$ satisfies the parameter equivalence at single BS. Therefore, we first use gradient descent to find the solution $\hat{\mathbf{h}}_b = [\hat{\rho}_b, \hat{\theta}_b, \hat{v}_b]^{\mathrm{T}}$ that minimizes $L_b(f_b(\mathbf{h}_b), f_b(\hat{\mathbf{p}}))$, namely, (77), (78), and (79). Then the local estimate at the b-th BS is $\hat{\mathbf{p}}_b = [\hat{\mathbf{u}}_b^{\mathrm{T}}, \hat{v}_{x,b}]^{\mathrm{T}}$ with

$$\hat{\mathbf{u}}_b = \hat{\rho}_b \begin{bmatrix} \sin(\hat{\theta}_b) \\ \cos(\hat{\theta}_b) \end{bmatrix} + \tilde{\mathbf{u}}_b, \tag{89}$$

$$\hat{v}_{x,b} = \hat{v}_b / \sin(\hat{\theta}_b). \tag{90}$$

Then, the global estimate of the kinematic parameters is the average of all local estimates, i.e., $\hat{\mathbf{p}} = \frac{1}{B}\sum_b \hat{\mathbf{p}}_b$.

C. Novel FedLoc-ISAC Mode

We propose a novel federate learning mode, referred to as FedLoc-ISAC, where the DOA, TOA and FDOA information could be used for learning parameters $\mathbf{p}$. The objective function (75) can be minimized through a two-step process. First, we obtain the solution $\hat{\mathbf{h}}_b$ that minimizes (77)–(79). Then, the sub-optimal solution can be calculated by classic localization approach, e.g., NLS-based method [46]. However, the proposed approach outperforms its counterpart in [46] by incorporating the uncertainty in data detection and channel coefficient estimation. In the proposed FedLoc-ISAC mode, each DPU learning the channel parameter $\mathbf{h}_b$ at single BS. Then the estimated channel parameter $\hat{\mathbf{h}}_b$ rather than the local estimated kinematic parameters $\hat{\mathbf{p}}_b$ are aggregated at CPU to perform localization.

Here we give an example for FDOA-based localization, which has not been studied before to the best of our knowledge. For simplicity, we assume that $B = 3$ BSs and all the BSs are located along the x-axis, i.e., $\tilde{u}_{y,b} = 0$. Given the FDOA parameter $\{v_{b,k}\}$, we have

$$\beta_k = \frac{v_{b,1}^2}{v_{b,k}^2} = \frac{(u_x - \tilde{u}_{x,1})^2((u_x - \tilde{u}_{x,k})^2 + u_y^2)}{(u_x - \tilde{u}_{x,k})^2((u_x - \tilde{u}_{x,1})^2 + u_y^2)}, \quad k = 2, 3, \tag{91}$$

After some algebra, it reads

$$\frac{u_y^2}{(u_x - \tilde{u}_{x,1})^2} = \frac{(1 - \beta_k)(u_x - \tilde{u}_{x,k})^2}{\beta_k(u_x - \tilde{u}_{x,k})^2 - (u_x - \tilde{u}_{x,1})^2}, \quad k = 2, 3, \tag{92}$$

which is equivalent to a single variable cubic equation, w.r.t. u_x. After that, the kinematic parameters $\mathbf{p}$ can be obtained. Compared with DOA and TOA-based localization, the FDOA -based localization is able to learn the velocity parameter.

D. Computational Complexity

In the following, we analyze the computational complexity[5] of our proposed algorithms. First, we consider the C-ISAC scheme. According to (30)-(35), the multiplication of circulant matrices can be implemented by low-complexity FFT operation.[6] For data detection, the complexity of calculating $\bar{z}_{b,n}$ in (54) and $\sigma_{z_b}^2$ in (55) is in the order of $\mathcal{O}(BMN)$, and the complexity of GAMP is in the order of $\mathcal{O}(4BN(\log_2 N_1 + \log_2 N_2) + 4BN + N + 3N|\mathcal{S}|)$. For channel coefficient estimation, the complexity of calculating $\bar{g}_b$ in (66) and $\sigma_{g_b}^2$ in (67) is in the order of $\mathcal{O}(BMN + BN)$, and the complexity of calculating $\hat{g}_b$ in (69) and ζ_{g_b} in (70) is in the order of $\mathcal{O}(B)$. For kinematic parameter sensing, the complexity of calculating the objective function is in the order of $\mathcal{O}(BN(M + 2\log_2 N_1 + 2\log_2 N_2))$, and the complexity of calculating the corresponding gradient w.r.t. the kinematic parameters is in the order of $\mathcal{O}(BN(M + 2N_1 + 2N_2 + 2\log_2 N_1 + 2\log_2 N_2 + 5))$, respectively. It is noteworthy that the computational complexity of the communication-efficient federated learning scheme can be analyzed similarly, except for setting $B = 1$ at each DPU.

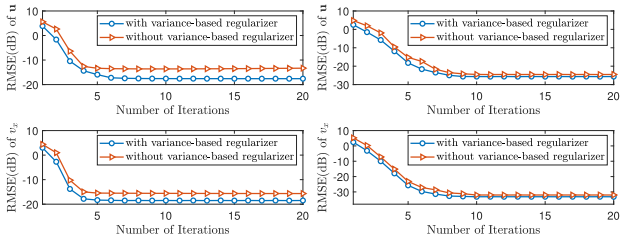
V. Simulation Results

In this section, we evaluate the ISAC performance of the proposed variational inference-based approaches. Unless specified otherwise, the system parameters used in the simulations are given as follows. The number of time slots and subcarriers are $N_1 = 32$ and $N_2 = 256$, respectively. The bandwidth is 10 MHz with the carrier frequency $f_0 = 10$ GHz. We consider three BSs equipped with $M = 16$ antennas. For the conciseness of notations, we use the standard unit system, with units omitted. The positions of $B = 3$ BSs are $[0, 0]$, $[60, 0]$, $[120, 0]$. The UE are randomly and uniformly distributed within an area $u_x \in [0, 120], u_y \in [10, 30]$, while its velocity are randomly and uniformly distributed within $v_x \in [30, 60]$. The channel coefficients g_b are generated by $g_b \sim \mathcal{N}_{\mathbb{C}}(g_b; 0, \gamma_{g_b}^{-1})$ with $\gamma_{g_b}^{-1} = 1$. The power of pilot symbol is $P_{\mathrm{p}} = \frac{|x_{\mathrm{p}}|^2}{N}$ and the power of data symbols is $P_{\mathrm{d}} = \mathbb{E}\{|x_{\mathrm{d},k,l}|^2\}$, while quadrature phase shift keying (QPSK) modulation is adopted. The signal-to-noise ration (SNR) is defined as $\mathrm{SNR} = \frac{P_{\mathrm{p}} + P_{\mathrm{d}}}{\sigma_w^2}$. For comparison, we extend the JCEDD algorithm [41] to the ISAC scenario by incorporating NLS-based localization and evaluate its performance. This approach is referred to as JCEDD-NLS. Besides, perfect communication case (where $\mathbf{X}$ and $\mathbf{g}$ are known) and perfect sensing case (where $\mathbf{p}$ is known) are considered as benchmarks.

Fig. 4 illustrates the effect of the variance-based regularizer in (78) and (79) on sensing. Without the regularizer, the sensing performance degrades due to overfitting on unreliable detected data and estimated channel coefficients, especially at low SNR. This issue can be mitigated by accounting for the uncertainty in data detection and channel coefficient estimation. Consequently, the proposed algorithm with the regularizer achieves improved sensing performance. However, this improvement diminishes as SNR increases, due to the reduced impact of uncertainty at higher SNR. Additionally, the regularizer accelerates convergence by providing an adaptive

[5]Here, we mainly focus on the maximum number of required matrix multiplications.

[6]For example, the complexity for $\mathbf{A} \in \mathbb{C}^{M \times N}$ multiplexing $\mathbf{x} \in \mathbb{C}^N$ is $\mathcal{O}(N^2)$. If $\mathbf{A}$ is circulant matrix, the multiplication can be implemented by FFT with its complexity reducing to an order of $\mathcal{O}(N \log_2 N)$.



(a) SNR = 15dB (b) SNR = 30dB

Fig. 4. The effect of the variance-based regularizer for sensing.

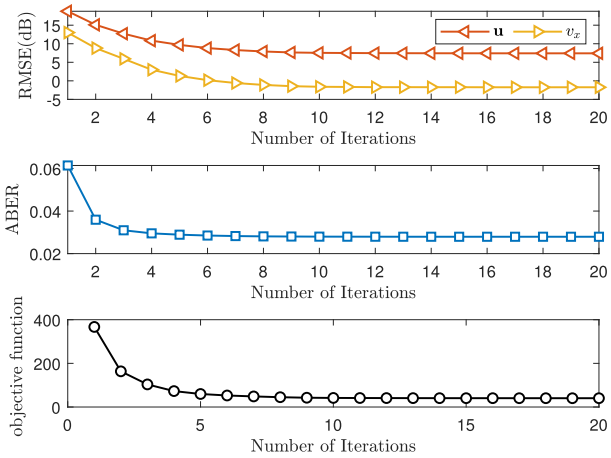


Fig. 5. The convergence performance of the proposed FDOA-based FedLoc-ISAC mode for estimation.

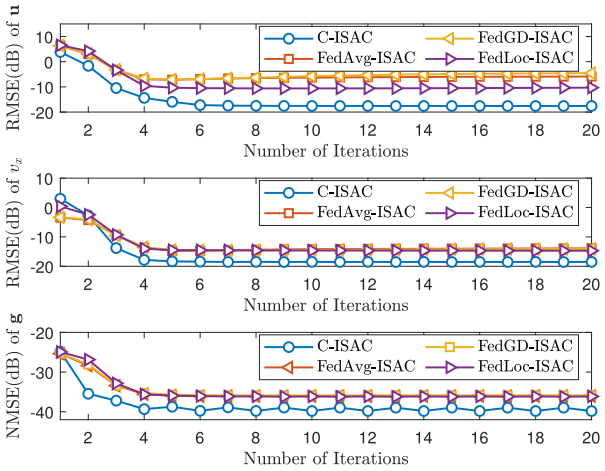


Fig. 6. The convergence performance of the proposed ISAC schemes for estimation.

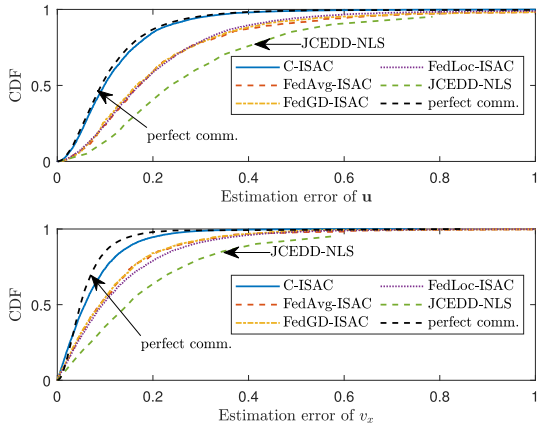


Fig. 7. The CDF of the sensing error for the proposed ISAC schemes.

gradient, which incorporates extra gradient information based on the uncertainty.

Then, we evaluate the effect of velocity information on ISAC performance in Fig. 5. Note that the objective function of velocity is (79). We can observe that as the number of iterations increases, the FDOA-based FedLoc-ISAC mode achieves better root mean square error (RMSE), while the value of the objective function decreases. Even without any DOA and TOA information, FDOA information still can give a available results. This phenomenon validates the assumption that the FDOA information be utilized for ISAC.

Fig. 6 illustrates the convergence performance of the proposed schemes for position, velocity, and channel coefficient estimation. It is observed that the C-ISAC scheme achieves the best performance, as it exploits the received symbols of all BSs. Therefore, a global optimal solution for sensing problem (74) can be obtained. On the other hand, the communication-efficient federated learning scheme does not obtain the full diversity gain within all BSs, leading to a communication performance loss, so that the sensing performance degrades due to the lack of the detected data symbols. Nevertheless, the proposed novel FedLoc-ISAC method achieves the best performance among three federate learning modes. It is because that the FedLoc-ISAC mode harnesses the sensing information from all BSs, rather than the local estimates in FedAvg-ISAC mode and the local gradients in FedGD-ISAC mode. As the number of iterations increases, the position estimation performance of the two classic federated learning modes degrades due to the myopic focus on local objective function.

In Fig. 7, the cumulative distribution function (CDF) of the kinematic parameter sensing error is given. We can observe that, the proposed ISAC scheme achieves better performance. This improvement is owing to the variational inference, which allows the scheme to directly infer kinematic parameters from received signals and incorporate the variance of data symbols and channel coefficients in kinematic parameter sensing. Since C-ISAC optimizes the global objective function and benefits from enhanced communication performance due to its diversity, the performance gap compared to the perfect communication case is minimal.

The sensing and the average bit error rate (ABER) performance for the proposed ISAC scheme is depicted in Fig. 8 and Fig. 9, respectively. As discussed earlier, leveraging variational inference, the proposed ISAC schemes outperform JCEDD-NLS. Moreover, C-ISAC achieves the optimal ISAC performance by utilizing the full communication diversity and solving the global optimization problem. Despite of the same communication diversity, the FedLoc-ISAC mode outperforms the FedAvg-ISAC mode and the FedGD-ISAC mode for kinematic parameter sensing owing to the full use of sensing information. Accordingly, the FedLoc-ISAC mode has a slight

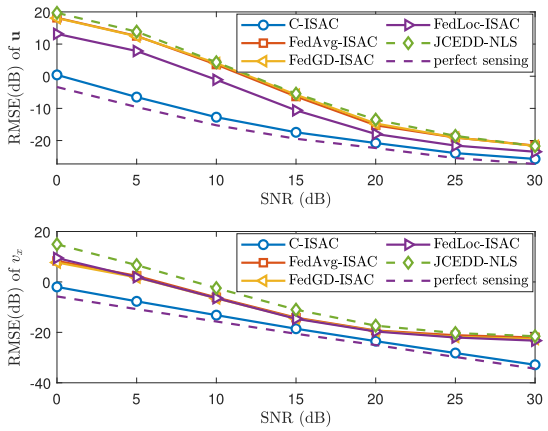


Fig. 8. The sensing performance versus SNR for the proposed ISAC scheme.

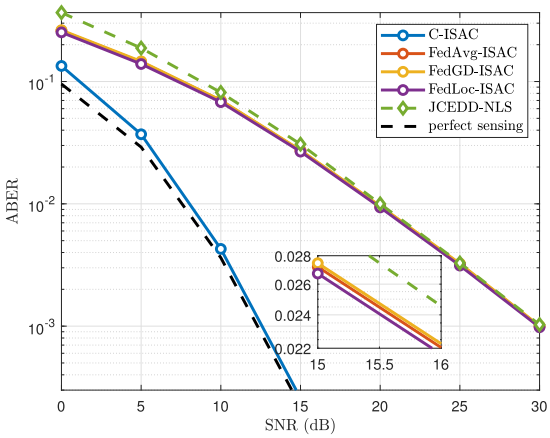


Fig. 9. The ABER performance versus SNR for the proposed ISAC scheme.

ABER performance improvement compared to its federated learning-based counterparts. Furthermore, as data and channel coefficients are unknown, C-ISAC experiences a slight reduction in sensing performance compared to the perfect communication case, which provides accurate sensing result at high SNR. Consequently, this slight sensing performance loss also leads to a minor decrease in communication performance.

VI. Conclusion

In this paper, harnessing the concept of advanced variational inference, we proposed a novel integrated sensing and communication receiver design framework for OTFS-based MIMO systems. By exploiting the factorization of the APP, we constructed a multi-layer factor graph, in which the mapping from kinematic parameters to channel parameters is incorporated. To circumvent the intractable original APP, we approximated it by a flexible form. Accordingly, the ISAC problem was decomposed to data detection, channel coefficient estimation, and kinematic parameter sensing, which has been solved by derived belief propagation, mean filed, and expectation maximization algorithms. Moreover, to economize the communication overhead for transmitting the received signals, we devised a communication-efficient federated learning scheme comprising three modes. Simulation results validated that the proposed framework can achieve an excellent ISAC performance.

Appendices

A. Proof of Convergence Performance

For t-th iteration, considering subproblem (46), given $q^t(\mathbf{X})$, $q^t(\mathbf{g})$, and $q^t(\mathbf{p})$, we have

$$q^{t+1}(\mathbf{X}) = \underset{q\in\mathcal{Q}_{\mathbf{X}}}{\arg\min}\, \mathrm{KL}\left(q(\mathbf{X}) \,||\, \tilde{p}^t(\mathbf{X}\mid\mathbf{Y})\right) \tag{93}$$

where

$$\tilde{p}^t(\mathbf{X}\mid\mathbf{Y}) \propto \exp(\mathbb{E}_{\mathbf{g}\sim q^t(\mathbf{g}),\mathbf{p}\sim q^t(\mathbf{p})}\{\ln p(\mathbf{X},\mathbf{g}\mid\mathbf{Y};\mathbf{p})\}) \tag{94}$$

Then we can observe that

$$\mathrm{KL}\left(q^{t+1}(\mathbf{X}) \,||\, \tilde{p}^t(\mathbf{X}\mid\mathbf{Y})\right) \le \mathrm{KL}\left(q^t(\mathbf{X}) \,||\, \tilde{p}^t(\mathbf{X}\mid\mathbf{Y})\right) \tag{95}$$

This property also holds for subproblems (47) and (48). ■

B. Proof of Delta Belief

Using the property of delta function, the variational free energy of $p(x)$ on distribution family $q(x) = \delta(x-\hat{x})$ is given by

$$\begin{aligned}\mathrm{KL}(q(x)\,||\,p(x)) &= \int_x \delta(x-\hat{x})\ln\delta(x-\hat{x}) \\ &\quad - \int_x \delta(x-\hat{x})\ln p(x) \\ &= \int_x \delta(x-\hat{x})\ln\delta(x-\hat{x}) - \ln p(\hat{x}),\end{aligned} \tag{96}$$

where the first term is a constant. Therefore, minimizing (96) is equivalent to (73). ■

References

[1] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 829–846, Apr. 2018.

[2] A. Mukherjee, P. Goswami, M. A. Khan, L. Manman, L. Yang, and P. Pillai, "Energy-efficient resource allocation strategy in massive IoT for industrial 6G applications," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5194–5201, Apr. 2021.

[3] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proc. IEEE*, vol. 107, no. 12, pp. 2327–2375, Dec. 2019.

[4] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714–2741, 4th Quart., 2018.

[5] J. A. Zhang et al., "Enabling joint communication and radar sensing in mobile networks—A survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 306–345, 1st Quart., 2022.

[6] Z. Wei et al., "Integrated sensing and communication signals towards 5G-A and 6G: A survey," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11068–11092, Jan. 2023.

[7] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang, "Radar and communication coexistence: An overview: A review of recent methods," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 85–99, Sep. 2019.

[8] W. Chen, L. Li, Z. Chen, T. Quek, and S. Li, "Enhancing THz/mmWave network beam alignment with integrated sensing and communication," *IEEE Commun. Lett.*, vol. 26, no. 7, pp. 1698–1702, Jul. 2022.

[9] H. Sarieddeen, N. Saeed, T. Y. Al-Naffouri, and M.-S. Alouini, "Next generation terahertz communications: A rendezvous of sensing, imaging, and localization," *IEEE Commun. Mag.*, vol. 58, no. 5, pp. 69–75, May 2020.

[10] F. Liu, C. Masouros, A. Li, T. Ratnarajah, and J. Zhou, "MIMO radar and cellular coexistence: A power-efficient approach enabled by interference exploitation," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3681–3695, Jul. 2018.

[11] F. Liu et al., "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.

[12] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, "On the fundamental tradeoff of integrated sensing and communications under Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5723–5751, Jun. 2023.

[13] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4264–4279, Aug. 2018.

[14] A. R. Chiriyath, B. Paul, and D. W. Bliss, "Radar-communications convergence: Coexistence, cooperation, and co-design," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 1, pp. 1–12, Mar. 2017.

[15] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[16] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[17] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 106–114, Sep. 2007.

[18] J. Li, P. Stoica, L. Xu, and W. Roberts, "On parameter identifiability of MIMO radar," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 968–971, Dec. 2007.

[19] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, Apr. 2018.

[20] F. Liu, Y.-F. Liu, A. Li, C. Masouros, and Y. C. Eldar, "Cramér–Rao bound optimization for joint radar-communication beamforming," *IEEE Trans. Signal Process.*, vol. 70, pp. 240–253, 2022.

[21] D. Ma, N. Shlezinger, T. Huang, Y. Liu, and Y. C. Eldar, "Joint radar-communication strategies for autonomous vehicles: Combining two key automotive technologies," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 85–97, Jul. 2020.

[22] K. Wu, J. A. Zhang, X. Huang, and Y. J. Guo, "Integrating low-complexity and flexible sensing into communication systems," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1873–1889, Jun. 2022.

[23] F. Hlawatsch and G. Matz, *Wireless Communications Over Rapidly Timevarying Channels*. New York, NY, USA: Academic, 2011.

[24] Y. G. Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for wireless communications: Signal detection with enhanced channel estimation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1471–1477, Sep. 2002.

[25] R. Hadani et al., "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[26] R. Hadani and A. Monk, "OTFS: A new generation of modulation addressing the challenges of 5G," 2018, *arXiv:1802.02623*.

[27] Z. Wei et al., "Orthogonal time-frequency space modulation: A promising next-generation waveform," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 136–144, Aug. 2021.

[28] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Effective diversity of OTFS modulation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 249–253, Feb. 2020.

[29] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "On the effectiveness of OTFS for joint radar parameter estimation and communication," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5951–5965, Sep. 2020.

[30] W. Yuan, Z. Wei, S. Li, J. Yuan, and D. W. K. Ng, "Integrated sensing and communication-assisted orthogonal time frequency space transmission for vehicular networks," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 6, pp. 1515–1528, Nov. 2021.

[31] Y. Wu, C. Han, and Z. Chen, "DFT-spread orthogonal time frequency space system with superimposed pilots for terahertz integrated sensing and communication," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7361–7376, Mar. 2023.

[32] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501–6515, Oct. 2018.

[33] F.-G. Yan, M. Jin, S. Liu, and X.-L. Qiao, "Real-valued MUSIC for efficient direction estimation with arbitrary array geometries," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1548–1560, Mar. 2014.

[34] T. Ma, Y. Xiao, and X. Lei, "Channel reconstruction-aided MUSIC algorithms for joint AoA&AoD estimation in MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 12, no. 2, pp. 322–326, Feb. 2023.

[35] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, "Off-grid channel estimation with sparse Bayesian learning for OTFS systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7407–7426, Sep. 2022.

[36] W. Yuan, Z. Wei, J. Yuan, and D. W. K. Ng, "A simple variational Bayes detector for orthogonal time frequency space (OTFS) modulation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7976–7980, Jul. 2020.

[37] Z. Yuan, F. Liu, W. Yuan, Q. Guo, Z. Wang, and J. Yuan, "Iterative detection for orthogonal time frequency space modulation with unitary approximate message passing," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 714–725, Feb. 2022.

[38] Y. Ge, Q. Deng, P. C. Ching, and Z. Ding, "Receiver design for OTFS with a fractionally spaced sampling approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4072–4086, Jul. 2021.

[39] S. Li, W. Yuan, Z. Wei, and J. Yuan, "Cross domain iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2227–2242, Apr. 2022.

[40] F. Liu, W. Yuan, C. Masouros, and J. Yuan, "Radar-assisted predictive beamforming for vehicular links: Communication served by sensing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7704–7719, Nov. 2020.

[41] Z. Du et al., "Integrated sensing and communications for V2I networks: Dynamic predictive beamforming for extended vehicle targets," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3612–3627, Nov. 2023.

[42] W. Yuan, F. Liu, C. Masouros, J. Yuan, D. W. K. Ng, and N. González-Prelcic, "Bayesian predictive beamforming for vehicular networks: A low-overhead joint radar-communication approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1442–1456, Mar. 2021.

[43] C. Liu et al., "Learning-based predictive beamforming for integrated sensing and communication in vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2317–2334, Aug. 2022.

[44] W. Yuan, S. Li, Z. Wei, J. Yuan, and D. W. K. Ng, "Data-aided channel estimation for OTFS systems with a superimposed pilot and data transmission scheme," *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1954–1958, Sep. 2021.

[45] S. Li et al., "A novel ISAC transmission framework based on spatially-spread orthogonal time frequency space modulation," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1854–1872, Jun. 2022.

[46] M. F. Keskin, S. Gezici, and O. Arikan, "Direct and two-step positioning in visible light systems," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 239–254, Jan. 2018.

[47] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[48] X. Meng et al., "Advanced NOMA receivers from a unified variational inference perspective," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 934–948, Apr. 2021.

[49] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[50] H. Q. Ngo, *Massive MIMO: Fundamentals and System Designs*, vol. 1642. Linköping, Sweden: University Electronic Press, 2015.

[51] J. Bechter, M. Rameez, and C. Waldschmidt, "Analytical and experimental investigations on mitigation of interference in a DBF MIMO radar," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1727–1734, May 2017.

[52] Z. Wei, W. Yuan, S. Li, J. Yuan, and D. W. K. Ng, "Transmitter and receiver window designs for orthogonal time-frequency space modulation," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2207–2223, Apr. 2021.

[53] K. C. Ho and Y. T. Chan, "Geolocation of a known altitude object from TDOA and FDOA measurements," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 3, pp. 770–783, Jul. 1997.

[54] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

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

[56] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, Oct. 2014.

[57] Z. Yang and Y. Liu, "Understanding node localizability of wireless ad hoc and sensor networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 8, pp. 1249–1260, Aug. 2012.

[58] Q. Su and Y.-C. Wu, "On convergence conditions of Gaussian belief propagation," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1144–1155, Mar. 2015.

[59] J. Chen, X. Pan, R. Monga, S. Bengio, and R. Jozefowicz, "Revisiting distributed synchronous SGD," 2016, *arXiv:1604.00981*.

[60] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

Nan Wu (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Beijing Institute of Technology (BIT), Beijing, China, in 2003, 2005, and 2011, respectively.

He was a Visiting Ph.D. Student with the Department of Electrical Engineering, The Pennsylvania State University, USA, from 2008 to 2009. He is currently a Professor with the School of Information and Electronics, BIT. His research interests include signal processing in wireless communication networks.

Dr. Wu was a recipient of the National Excellent Doctoral Dissertation Award by MOE of China in 2013. He serves as an Editorial Board Member for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE WIRELESS COMMUNICATIONS LETTERS.

Haoyang Li received the B.S. degree from Beijing Institute of Technology (BIT), Beijing, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Information and Electronics. His research interests include statistical inference on graphical models and its application to wireless communications.

Dongxuan He (Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in information and communication systems from Beijing Institute of Technology (BIT) in 2013 and 2019, respectively.

From 2017 to 2018, he was a Visiting Student with Singapore University of Technology and Design (SUTD). From 2019 to 2022, he was a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University. He is currently an Assistant Professor with the School of Information and Electronics, BIT. His current interests include integrated sensing and communication (ISAC), terahertz communication, and AI empowered wireless communications.

Dr. He is currently serving as a Guest Editor for IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, *Electronics*, and *Space: Science & Technology*. He was an Exemplary Reviewer of the IEEE WIRELESS COMMUNICATIONS LETTERS.

Arumugam Nallanathan (Fellow, IEEE) received the Ph.D. degree from The University of Hong Kong, Hong Kong, in 2000.

He has been a Professor in wireless communications and the Head of the Communication Systems Research (CSR) Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, since September 2017. He was with the Department of Informatics, King's College London, from December 2007 to August 2017, where he was a Professor in wireless communications from April 2013 to August 2017 and a Visiting Professor from September 2017 to August 2020. He was an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, from August 2000 to December 2007. He has published nearly 700 technical papers in scientific journals and international conferences. His research interests include artificial intelligence for wireless systems, beyond 5G wireless networks, and the Internet of Things (IoT).

Dr. Nallanathan was a co-recipient of the Best Paper Awards presented at the IEEE International Conference on Communications 2016 (ICC'2016), IEEE Global Communications Conference 2017 (GLOBECOM'2017), and IEEE Vehicular Technology Conference 2018 (VTC'2018). He was also a co-recipient of IEEE Communications Society Leonard G. Abraham Prize in 2022. He is an IEEE Distinguished Lecturer. He has been selected as a Web of Science Highly Cited Researcher in 2016, 2022, and 2023. He was a Senior Editor of IEEE WIRELESS COMMUNICATIONS LETTERS; and an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE SIGNAL PROCESSING LETTERS. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society and the technical program chair and a member of technical program committees in numerous IEEE conferences. He received the IEEE Communications Society SPCE Outstanding Service Award in 2012 and IEEE Communications Society RCC Outstanding Service Award in 2014.

Tony Q. S. Quek (Fellow, IEEE) received the B.E. and M.E. degrees in electrical and electronics engineering from Tokyo Institute of Technology in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering and computer science from Massachusetts Institute of Technology in 2008.

Currently, he is the Cheng Tsang Man Chair Professor with Singapore University of Technology and Design (SUTD) and a ST Engineering Distinguished Professor. He is also the Director of the Future Communications Research and Development Programme, the Head of ISTD Pillar, and the AI on RAN Working Group Chair in AI-RAN Alliance. His current research topics include wireless communications and networking, network intelligence, non-terrestrial networks, open radio access networks, and 6G.

Dr. Quek has been actively involved in organizing and chairing sessions and has served as a member of the technical program committee and the symposium chairs in a number of international conferences. He is an IEEE Fellow, a WWRF Fellow, and a fellow of the Academy of Engineering Singapore. He is currently serving as an Area Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was honored with the 2008 Philip Yeo Prize for Outstanding Achievement in Research, the 2012 IEEE William R. Bennett Prize, the 2015 SUTD Outstanding Education Awards—Excellence in Research, the 2016 IEEE Signal Processing Society Young Author Best Paper Award, the 2017 CTTC Early Achievement Award, the 2017 IEEE ComSoc AP Outstanding Paper Award, the 2020 IEEE Communications Society Young Author Best Paper Award, the 2020 IEEE Stephen O. Rice Prize, the 2020 Nokia Visiting Professor, the 2022 IEEE Signal Processing Society Best Paper Award, and the 2024 IIT Bombay International Award For Excellence in Research in Engineering and Technology.