

OTFS-Enabled Coarse-to-Fine Joint Synchronization and Localization under High Mobility

Qidong Chen, Rui Zhang, Suzhi Bi and Zhi Quan

College of Electronics and Information Engineering, Shenzhen University

Email: 2510044042@mails.szu.edu.cn, {zhrrygh, bsz, zquan}@szu.edu.cn

Abstract—Achieving precise joint synchronization and localization (JSL) in 6G integrated sensing and communication (ISAC) networks is fundamentally challenged by severe space-time coupling and ranging ambiguity under extreme mobility. While physical-layer measurements such as orthogonal time frequency space (OTFS) fractional delays offer high resolution, they suffer from periodic modulo ambiguity. Conversely, medium access control (MAC) timestamps are unambiguous but provide coarse temporal resolution. To bridge this gap, this paper proposes an OTFS-assisted cross-layer JSL architecture utilizing an 8-dimensional extended Kalman filter. Our approach models the continuous-time coupled evolution of multi-order kinematics and clock dynamics to track spatial and temporal states. We introduce a dual-update mechanism that uses MAC-layer observations for coarse spatial bounding, thereby resolving the periodic ambiguity of OTFS measurements during the fine update stage. Simulation results demonstrate that the proposed framework achieves decimeter-level localization accuracy for target speeds up to 150 m/s and exhibits strong robustness against MAC-layer timing noise, while maintaining nanosecond-level synchronization accuracy.

Index Terms—Joint synchronization and localization, orthogonal time frequency space, extended Kalman filtering

I. INTRODUCTION

The evolution of 6G networks has witnessed a paradigm shift toward Distributed Integrated Sensing and Communication (D-ISAC) architectures [1]. By geographically dispersing sensing nodes and base stations, D-ISAC overcomes the spatial coverage and Line-of-Sight (LoS) blockage limitations inherent in traditional monostatic radar configurations. However, this decentralized topology disrupts the natural synchronization advantage of shared local oscillators, inevitably introducing continuous time offsets and carrier frequency offsets among asynchronous nodes. In extreme mobility scenarios, such as unmanned aerial vehicle swarms and high-speed trains, this clock asynchrony tightly intertwines with intense physical kinematics, inducing severe space-time coupling. Consequently, achieving ultra-high-precision joint synchronization and localization (JSL) remains a formidable challenge for highly dynamic networks.

To tackle the JSL task, early methods heavily rely on static snapshot assumptions, utilizing time-difference-of-arrival lin-

earization or weighted least squares solutions [2]–[4]. To enable dynamic state tracking, extended Kalman filter (EKF) frameworks have been proposed [5], [6], which primarily depend on medium access control (MAC)-layer protocols and packet-level timestamps. However, as analyzed in [7], while MAC-layer observations provide globally unambiguous delays, their precision is fundamentally bottlenecked by the quantization limits of analog-to-digital converters and hardware clock jitter. This inherent coarse temporal resolution restricts the accuracy of MAC-centric estimation to the sub-meter level, which falls critically short of the centimeter-level requirements envisioned for 6G applications.

To transcend these limits, recent research explores multi-source and cross-layer fusion frameworks, such as integrating angle-of-arrival features [8] and fusing spatial-temporal measurements via nonlinear filtering for joint 3D positioning and synchronization [9], combining radio with LiDAR-IMU sensors [10], and tracking carrier phase and Doppler shifts [11]. Concurrently, orthogonal time frequency space (OTFS) modulation has emerged as a promising ISAC candidate due to its resilience against extreme Doppler spreads [12], inspiring novel heterogeneous architectures that fuse OTFS communication data with physical sensing features [13]. By transforming doubly-selective fading channels into a sparse delay-Doppler (DD) representation, OTFS extracts sub-symbol waveform features [14] and ultra-high-resolution fractional delays [15]. However, integrating these fine-grained physical-layer (PHY) measurements into continuous dynamic tracking introduces a fatal flaw, i.e., periodic ranging ambiguity. Because the extracted fractional delays are modulo-truncated by the symbol duration, the measurement space exhibits severe sawtooth discontinuities. Under extreme dynamic conditions, even marginal prediction errors can cause the estimated state to fall on the opposite side of a modulo boundary. This instantly invalidates the local linearization, leading to an inconsistent local innovation and potentially causing subsequent tracking failure.

Consequently, existing JSL solutions fall into a siloed layer dilemma: MAC-layer methods are globally stable but coarse, whereas OTFS PHY-layer methods are extremely precise but highly vulnerable to periodic ambiguity. Motivated by the statistical complementarity between these two layers, we aim to fuse MAC and PHY information and thus surpass the accuracy limits of single-layer estimation. By utilizing MAC-

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62471311 and No. 62271325, and in part by the Guangdong Basic and Applied Basic Research Foundation under Project 2024B1515020089, and in part by Shenzhen Science and Technology Program under Project JCYJ20250604181912016. (Corresponding author: Rui Zhang.)

layer data to construct a high-confidence spatial bounding box, the global ambiguity is statistically suppressed, creating a safe subspace to fully exploit the resolution of OTFS observations. Building upon this philosophy, the core contributions of this paper are given as follows:

- We fuse unambiguous MAC-layer timestamps with ultra-high-resolution OTFS fractional delays. This integration effectively resolves the periodic ranging ambiguity inherent in single-layer physical measurements.
- We establish an 8-dimensional extended Kalman filter (8D-EKF) based on continuous-time stochastic modeling. This framework rigorously captures the coupled evolution of multi-order kinematics and local oscillator dynamics, providing a robust engine against severe space-time coupling.
- To robustly incorporate the highly nonlinear OTFS measurements, we design a two-stage filtering algorithm. By combining coarse ambiguity capture with wrapped-residual refinement, the proposed method reduces the risk of ambiguity-induced tracking failures and achieves decimeter-level localization accuracy for target speeds up to 150 m/s.

II. SYSTEM MODEL

We consider a cooperative high-mobility network consisting of N reference nodes, denoted as R_i for $i \in \{1, 2, \dots, N\}$, and M high-speed target nodes. In line with practical cooperative positioning setups, we assume that the reference nodes are anchored at known coordinates and maintain perfect synchronization with the global physical time t , while the target nodes have unknown spatial and temporal states. Furthermore, instead of having an ideal hardware clock, each target node is equipped with an imperfect local oscillator such that its observed local time T inevitably skews, creating a continuous clock deviation relative to t .

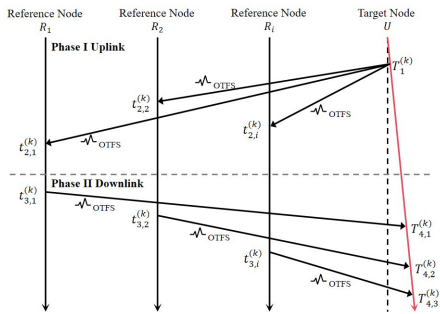


Fig. 1. Timing diagram of the proposed cross-layer two-way interaction protocol during the k -th TDMA round, illustrating the OTFS signal exchange and timestamp definitions across the uplink and downlink phases.

In this paper, we propose a cross-layer two-way interaction protocol for the considered high-mobility network, as shown in Fig. 1. In the considered network, we assume that multiple target nodes access the channel via TDMA. Since the transmitted signals from different targets are temporally orthogonal and do not interfere with each other, without loss of generality,

we focus our subsequent analysis on a specific target node U . To elaborate, in the k -th interaction round, we divide the complete interaction between the target node and the reference nodes into two distinct phases:

- **Phase I Uplink:** Target node U broadcasts the OTFS sounding signal at its local time $T_1^{(k)}$. The i -th reference node R_i receives it at the true system time $t_{2,i}^{(k)}$.
- **Phase II Downlink:** Each reference node sends its response in a predefined order with a communication interval t_d (e.g., $t_{3,i}^{(k)} = t_{3,i-1}^{(k)} + t_d$). Node R_i transmits at system time $t_{3,i}^{(k)}$, embedding timestamps $t_{2,i}^{(k)}$ and $t_{3,i}^{(k)}$. The target node U receives this signal at local time $T_{4,i}^{(k)}$.

Due to the high mobility of the target node during the interaction interval, its significant spatial displacement renders the uplink and downlink geometric distances ($D_{k,i}^u$ and $D_{k,i}^d$) strictly asymmetric.

A. Joint Clock and Kinematic State Model

a) *Clock model:* Let the target node's clock offset at time t be $\theta(t)$ and its clock skew be $\delta(t)$. The local clock evolution is modeled as

$$T(t) = t + \theta(t). \quad (1)$$

In the continuous-time domain, the clock offset and skew are coupled; their evolution is modeled by

$$\dot{\theta}(t) = \delta(t) + w_{\theta}(t), \quad (2)$$

$$\dot{\delta}(t) = w_{\delta}(t), \quad (3)$$

where $\dot{\theta}(t)$ is the first derivative of the clock offset $\theta(t)$ with respect to time t , characterizing the instantaneous evolution rate of the local time phase. $\dot{\delta}(t)$ is the first derivative of the clock skew $\delta(t)$, representing the instantaneous rate of change of the oscillator's frequency skew. The initial conditions are given by $\theta(0) = \theta_0$ and $\delta(0) = \delta_0$. $w_{\theta}(t)$ and $w_{\delta}(t)$ represent phase noise and frequency random walk white noise, assumed to be independent zero-mean Gaussian white noises with power spectral densities denoted as q_{bias} and q_{skew} .

According to this clock model, the target node's local timestamps satisfy $T_1^{(k)} = T(t_1^{(k)})$ and $T_{4,i}^{(k)} = T(t_{4,i}^{(k)})$. To ensure rigorous state estimation, the full-dimensional system state is aligned to true time $t_{4,i}^{(k)}$ when the target node receives the signal. The 8-dimensional joint state vector at $t_{4,i}^{(k)}$ is defined as

$$\mathbf{X}_{k,i} = [\mathbf{p}_{k,i}^T, \mathbf{v}_{k,i}^T, \mathbf{a}_{k,i}^T, \theta_{k,i}, \delta_{k,i}]^T, \quad (4)$$

where $\mathbf{p}_{k,i} = [p_{x,k,i}, p_{y,k,i}]^T$ is the 2D position (m), $\mathbf{v}_{k,i} = [v_{x,k,i}, v_{y,k,i}]^T$ is the 2D velocity (m/s), $\mathbf{a}_{k,i} = [a_{x,k,i}, a_{y,k,i}]^T$ is the 2D acceleration (m/s²), $\theta_{k,i} = \theta(t_{4,i}^{(k)})$ is the clock offset (s), and $\delta_{k,i} = \delta(t_{4,i}^{(k)})$ is the clock skew.

In order to ensure the precise operation of the state vector and the filter, it is necessary to define the physical interaction interval $\Delta t_{k,i} = t_{4,i}^{(k)} - t_1^{(k)}$ and the physical filter update interval $\Delta t_{k,i}^{\text{upd}}$. Due to the cyclic TDMA scheduling among N reference nodes, the update interval bridges either two

adjacent slots within the same round or the boundary between consecutive rounds, which is rigorously formulated as

$$\Delta t_{k,i}^{\text{upd}} = \begin{cases} t_{4,1}^{(k)} - t_{4,N}^{(k-1)}, & i = 1, k > 1 \\ t_{4,i}^{(k)} - t_{4,i-1}^{(k)}, & i > 1, k > 1. \end{cases} \quad (5)$$

It is mathematically crucial to emphasize that these true physical intervals ($\Delta t_{k,i}$ and $\Delta t_{k,i}^{\text{upd}}$) are inherently inaccessible to the unsynchronized target node. In our proposed 8D-EKF, all state transitions and geometric back-tracing operations are directly driven by the corresponding intervals measured by the local oscillator, defined as the local interaction interval $\Delta T_{k,i} = T_{4,i}^{(k)} - T_1^{(k)}$ and the local filter update interval $\Delta T_{k,i}^{\text{upd}}$ (which identically mirrors the piecewise boundary logic of $\Delta t_{k,i}^{\text{upd}}$ using local timestamps T). Since the target clock skew satisfies $|\delta_{k,i}| \ll 1$, the locally measured intervals $\Delta T_{k,i}$ and $\Delta T_{k,i}^{\text{upd}}$ are used in the state propagation and kinematic back-tracing.

b) Kinematic Model: To characterize the highly dynamic spatial evolution of the target node U , a two-dimensional constant-acceleration kinematic model is adopted. The kinematic state is compactly represented by a 6-dimensional vector

$$\mathbf{Y}_{k,i} = [\mathbf{p}_{k,i}^T, \mathbf{v}_{k,i}^T, \mathbf{a}_{k,i}^T]^T. \quad (6)$$

Based on Newtonian kinematics, the one-step state prediction equation can be expressed as a linear matrix multiplication

$$\mathbf{Y}_{k,i} = \mathbf{F}_{\text{kin}}(\Delta T_{k,i}^{\text{upd}}) \mathbf{Y}_{\text{prev}} + \mathbf{q}_{\text{prev}}, \quad (7)$$

where $\mathbf{Y}_{\text{prev}}$ denotes the state at the previous update epoch, and the state transition matrix $\mathbf{F}_{\text{kin}}(\Delta T_{k,i}^{\text{upd}})$ is given by

$$\mathbf{F}_{\text{kin}}(\Delta T_{k,i}^{\text{upd}}) = \begin{bmatrix} \mathbf{I}_2 & \Delta T_{k,i}^{\text{upd}} \mathbf{I}_2 & \frac{1}{2} (\Delta T_{k,i}^{\text{upd}})^2 \mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{I}_2 & \Delta T_{k,i}^{\text{upd}} \mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{I}_2 \end{bmatrix}. \quad (8)$$

The term $\mathbf{q}_{\text{prev}}$ denotes the discrete-time process noise vector, which is derived by modeling the derivative of acceleration as a continuous white noise process (jerk noise) with power spectral density q_{jerk} .

Governed by the cyclic TDMA schedule, the joint state prediction must distinguish between inter-round and intra-round updates. The piecewise recursive equation is given by

$$\begin{cases} \mathbf{X}_{k,1} = \mathbf{F}(\Delta T_{k,1}^{\text{upd}}) \mathbf{X}_{k-1,N} + \mathbf{q}_{k-1,N}, & i = 1, k > 1 \\ \mathbf{X}_{k,i} = \mathbf{F}(\Delta T_{k,i}^{\text{upd}}) \mathbf{X}_{k,i-1} + \mathbf{q}_{k,i-1}, & i > 1, k > 1, \end{cases} \quad (9)$$

where $\mathbf{X}_{k,i}$ is defined in (4), the state transition matrix is $\mathbf{F}(\Delta T_{k,i}^{\text{upd}}) = \text{diag}(\mathbf{F}_{\text{kin}}(\Delta T_{k,i}^{\text{upd}}), \mathbf{F}_{\text{clock}}(\Delta T_{k,i}^{\text{upd}}))$, with $\mathbf{F}_{\text{kin}}(\Delta T_{k,i}^{\text{upd}})$ as in (8) and $\mathbf{F}_{\text{clock}}(\Delta T_{k,i}^{\text{upd}}) = [1, \Delta T_{k,i}^{\text{upd}}, 0, 1]$ is the clock evolution block. The term $\mathbf{q}$ denotes the discrete-time process noise vector driven by the kinematic noise q_{jerk} and the clock noises q_{bias} and q_{skew} .

B. OTFS Physical-Layer Channel Model

Let $x[m, n]$ be the transmitted symbols in the DD domain. Through standard OTFS transforms, $x[m, n]$ is converted to

a time-domain signal, propagated through the time-varying channel, and mapped back to the received DD symbols $y[m, n]$. By performing a 2D peak search to identify the dominant LoS path, the estimated grid indices $(\hat{m}, \hat{n})$ are mapped to the physical delay $\hat{\tau}$ and Doppler shift $\hat{\nu}$ [16] as

$$\hat{\tau} = \frac{\hat{m}}{M_{\tau} \Delta f}, \quad \hat{\nu} = \frac{\hat{n}}{N_{\nu} T_{\text{sym}}}, \quad (10)$$

where M_{τ} and N_{ν} denote the total number of delay and Doppler grids, respectively.

In highly dynamic systems, mitigating severe inter-carrier interference requires a large subcarrier spacing Δf . However, since the symbol duration is $T_{\text{sym}} = 1/\Delta f$, this spectral configuration strictly limits the maximum unambiguous propagation distance. Any physical distance exceeding this boundary, defined as the ambiguity range $\lambda_{\text{sym}} = c \cdot T_{\text{sym}}$, inevitably folds back into the principal delay grid. This fundamental time-frequency trade-off induces periodic ranging ambiguity, mathematically necessitating the disambiguation tracking mechanism in our subsequent filtering framework.

C. Problem Formulation

In the proposed JSL framework, the core task is to obtain precise location and timing estimates, which can be formulated as a recursive state estimation problem for the 8-dimensional joint state vector $\mathbf{X}_{k,i}$. The primary objective is to compute the optimal a posteriori state estimate $\hat{\mathbf{X}}_{k,i}$ that minimizes the minimum mean square error (MMSE), which is given by

$$\hat{\mathbf{X}}_{k,i} = \arg \min_{\hat{\mathbf{X}}} \mathbb{E}[(\mathbf{X}_{k,i} - \hat{\mathbf{X}})^T (\mathbf{X}_{k,i} - \hat{\mathbf{X}}) | T_1^{(k)}, t_{2,i}^{(k)}, t_{3,i}^{(k)}, T_{4,i}^{(k)}, \hat{\tau}, \hat{\nu}]. \quad (11)$$

To rigorously evaluate the algorithm's performance in highly dynamic environments, we define the root mean square error (RMSE) evaluation metrics $\text{RMSE}_{\chi_{k,i}}$. To eliminate the influence of initial transient states and ensure academic rigor, these metrics are calculated exclusively over the steady-state period after filter convergence. Let $\mathcal{K}_{\text{steady}}$ denote the set of evaluation epochs after convergence, and K_{steady} be its total number of samples. The generalized RMSE for a specific state variable $\chi_{k,i}$ is defined as

$$\text{RMSE}_{\chi_{k,i}} = \sqrt{\frac{1}{K_{\text{steady}}} \sum_{k \in \mathcal{K}_{\text{steady}}} \|\hat{\chi}_{k,i} - \chi_{k,i}\|^2}, \quad (12)$$

where the target variable $\chi_{k,i} \in \{\mathbf{p}_{k,i}, \theta_{k,i}, \delta_{k,i}\}$ represents the 2D position, clock offset, and clock skew, respectively.

The fundamental obstacle to MMSE estimation lies in the severe non-linearity of the OTFS delay observations $\hat{\tau}$. Because the fractional delay observations entail a modulo operation $\text{mod}(\cdot, T_{\text{sym}})$, the measurement space exhibits severe sawtooth discontinuities. Under extreme dynamic conditions, even marginal state prediction errors can cause the estimated state to fall on the opposite side of a modulo boundary relative

to the true state. This instantly invalidates the local linearization, leading to Jacobian matrix singularity and catastrophic filter collapse. Therefore, the core focus of this study is to design a robust cross-layer mechanism that harnesses high-precision OTFS measurements while mathematically guaranteeing tracking stability across these ambiguous boundaries.

III. CROSS-LAYER OBSERVATION

This section establishes a cross-layer observation framework, serving as a mathematical interface between physical-layer signals and the 8-dimensional joint state space. Leveraging the complementarity between the stable MAC measurements and the highly precise but ambiguous OTFS observations, we formulate a joint observation model as

$$\mathbf{Z}_{k,i} = \mathbf{h}(\mathbf{X}_{k,i}) + \mathbf{w}_{k,i}, \quad (13)$$

where $\mathbf{h}(\cdot)$ denotes the non-linear function mapping the joint state into the cross-layer measurement space, and $\mathbf{w}_{k,i}$ is the composite Gaussian measurement noise accounting for both MAC jitter and OTFS estimation errors.

1) *MAC-Layer Observations*: In the MAC layer, the standardized round-trip time (RTT) and one-way time of flight (TOF) are measured to extract geometric ranging and anchor the absolute clock offset $\theta_{k,i}$, respectively. Due to extreme mobility during the interaction interval $\Delta T_{k,i}$, the propagation distances are strictly asymmetric. Defining the target's state at reception time $t_{A,i}^{(k)}$ as $\mathbf{p}_{k,i}$, the true downlink distance is $D_{k,i}^d = \|\mathbf{p}_{k,i} - \mathbf{p}_{R,i}\|$. Conversely, the physical position $\mathbf{p}_{k,i}^u$ at the uplink transmission time must be inversely derived from the target's kinematics as $\mathbf{p}_{k,i}^u = \mathbf{p}_{k,i} - \mathbf{v}_{k,i}\Delta T_{k,i} + \frac{1}{2}\mathbf{a}_{k,i}\Delta T_{k,i}^2$, yielding the uplink distance $D_{k,i}^u = \|\mathbf{p}_{k,i}^u - \mathbf{p}_{R,i}\|$.

By mapping the temporal measurements into equivalent pseudo-ranges via the speed of light c , the MAC-layer observation vector is formulated as

$$z_{k,i,1} \triangleq c \cdot \text{RTT} = D_{k,i}^u + D_{k,i}^d + c\delta_{k,i}\Delta T_{k,i} + w_{\text{RTT}}, \quad (14)$$

$$z_{k,i,2} \triangleq c \cdot \text{TOF} = D_{k,i}^d + c\theta_{k,i} + w_{\text{TOF}}, \quad (15)$$

where the equivalent distance measurement noises are defined as $w_{\text{RTT}} \sim \mathcal{N}(0, 2\sigma_{\text{MAC},\rho}^2)$ and $w_{\text{TOF}} \sim \mathcal{N}(0, \sigma_{\text{MAC},\rho}^2)$.

2) *OTFS Physical-Layer Observations*: Building upon the geometric distances, the OTFS physical layer extracts ultra-high-resolution delays and Doppler shifts via a 2D peak search. These fractional delays are then mapped into equivalent pseudo-ranges to maintain dimensional consistency while preserving their strict coupling with the kinematic and clock states.

For **Phase I Uplink**, the target transmits at velocity $\mathbf{v}_{k,i}^u = \mathbf{v}_{k,i} - \mathbf{a}_{k,i}\Delta T_{k,i}$. The Doppler shift is induced by the radial velocity and clock skew $\delta_{k,i}$. Furthermore, the pseudo-range observation is fundamentally modulo-truncated by the ambiguity range λ_{sym} . The observations are formulated as

$$z_{k,i,3} \triangleq \nu_{k,i}^u = -\frac{f_c}{c}\mathbf{v}_{k,i}^u \cdot \mathbf{u}_{k,i}^u + f_c\delta_{k,i} + w_\nu, \quad (16)$$

$$z_{k,i,4} \triangleq cT_{k,i}^u = \text{mod}(D_{k,i}^u - c\theta_{k,i} + c\delta_{k,i}\Delta T_{k,i}, \lambda_{\text{sym}}) + w_\rho, \quad (17)$$

where $\mathbf{u}_{k,i}^u = (\mathbf{p}_{k,i}^u - \mathbf{p}_{R,i})/D_{k,i}^u$ is the unit direction vector.

Similarly, for **Phase II Downlink**, with the target receiving at velocity $\mathbf{v}_{k,i}$, the observations are

$$z_{k,i,5} \triangleq \nu_{k,i}^d = -\frac{f_c}{c}\mathbf{v}_{k,i} \cdot \mathbf{u}_{k,i}^d - f_c\delta_{k,i} + w_\nu, \quad (18)$$

$$z_{k,i,6} \triangleq cT_{k,i}^d = \text{mod}(D_{k,i}^d + c\theta_{k,i}, \lambda_{\text{sym}}) + w_\rho, \quad (19)$$

where $\mathbf{u}_{k,i}^d = (\mathbf{p}_{k,i} - \mathbf{p}_{R,i})/D_{k,i}^d$, and the measurement noises are $w_\nu \sim \mathcal{N}(0, \sigma_{\text{OTFS},f}^2)$ and $w_\rho \sim \mathcal{N}(0, \sigma_{\text{OTFS},\rho}^2)$.

The severe non-linearity introduced by this modulo truncation $\text{mod}(\cdot, \lambda_{\text{sym}})$ mathematically necessitates our proposed coarse-to-fine dual-update EKF architecture to ensure robust tracking across ambiguous boundaries.

IV. PROPOSED OTFS-ASSISTED 8D-EKF JSL FRAMEWORK

This section elaborates on the proposed 8D-EKF architecture. This framework not only captures the physical coupling between kinematics and clock evolution through continuous-time integration but also incorporates a coarse-to-fine dual-update mechanism to robustly resolve the modulo ambiguity inherent in OTFS measurements.

A. Process Noise Covariance Matrix $\mathbf{Q}$ and Measurement Noise Covariance Matrix $\mathbf{R}$

Derived through exact integration of the continuous-time stochastic models over the local update interval $\Delta T_{k,i}^{\text{upd}}$, the discrete process noise covariance matrix is defined as $\mathbf{Q}_{k,i} = \text{diag}(\mathbf{Q}_{k,i}^{\text{kin}}, \mathbf{Q}_{k,i}^{\text{clock}})$, where

$$\mathbf{Q}_{k,i}^{\text{kin}} = q_{\text{jerk}} \begin{bmatrix} \frac{(\Delta T_{k,i}^{\text{upd}})^5}{20} \mathbf{I}_2 & \frac{(\Delta T_{k,i}^{\text{upd}})^4}{8} \mathbf{I}_2 & \frac{(\Delta T_{k,i}^{\text{upd}})^3}{6} \mathbf{I}_2 \\ \frac{(\Delta T_{k,i}^{\text{upd}})^4}{8} \mathbf{I}_2 & \frac{(\Delta T_{k,i}^{\text{upd}})^3}{3} \mathbf{I}_2 & \frac{(\Delta T_{k,i}^{\text{upd}})^2}{2} \mathbf{I}_2 \\ \frac{(\Delta T_{k,i}^{\text{upd}})^3}{6} \mathbf{I}_2 & \frac{(\Delta T_{k,i}^{\text{upd}})^2}{2} \mathbf{I}_2 & \Delta T_{k,i}^{\text{upd}} \mathbf{I}_2 \end{bmatrix}, \quad (20)$$

$$\mathbf{Q}_{k,i}^{\text{clock}} = \begin{bmatrix} q_{\text{bias}}\Delta T_{k,i}^{\text{upd}} + q_{\text{skew}}\frac{(\Delta T_{k,i}^{\text{upd}})^3}{3} & q_{\text{skew}}\frac{(\Delta T_{k,i}^{\text{upd}})^2}{2} \\ q_{\text{skew}}\frac{(\Delta T_{k,i}^{\text{upd}})^2}{2} & q_{\text{skew}}\Delta T_{k,i}^{\text{upd}} \end{bmatrix}. \quad (21)$$

The measurement noise covariance matrix is given by

$$\mathbf{R} = \text{diag}(2\sigma_{\text{MAC},\rho}^2, \sigma_{\text{MAC},\rho}^2, \sigma_{\text{OTFS},f}^2, \sigma_{\text{OTFS},\rho}^2, \sigma_{\text{OTFS},f}^2, \sigma_{\text{OTFS},\rho}^2). \quad (22)$$

B. Dual Update Mechanism and Disambiguation

To mitigate the numerical instability caused by modulo wrap-around in the OTFS pseudo-range observations, we design a two-stage coarse-to-fine update mechanism instead of directly assimilating all measurements in a single EKF update.

a) *Stage-1 (Coarse Update)*: The predicted state $\mathbf{X}_{k|k-1,i}$ and covariance $\mathbf{P}_{k|k-1,i}$ are first conditioned on the globally unambiguous MAC-layer observations and Doppler measurements, denoted as $\mathbf{z}_1 = [z_{k,i,1}, z_{k,i,2}, z_{k,i,3}, z_{k,i,5}]^T$, using a standard EKF update. This yields the intermediate state estimate $\hat{\mathbf{X}}_{k,i}^{(1)}$ and covariance $\mathbf{P}_{k,i}^{(1)}$. Although relatively coarse, this update reduces the state uncertainty and increases the likelihood that the subsequent OTFS ambiguity decision remains within the correct half-period capture region.

Cross-Layer Ambiguity Capture Condition: The Stage-1 update is free of OTFS ranging ambiguity and therefore provides a coarse prediction for resolving the periodically folded OTFS pseudo-ranges. For $s \in \{u, d\}$, define

$$\varphi_{k,i}^s \triangleq \begin{cases} D_{k,i}^u - c\theta_{k,i} + c\delta_{k,i}\Delta T_{k,i}, & s = u, \\ D_{k,i}^d + c\theta_{k,i}, & s = d. \end{cases} \quad (23)$$

The OTFS pseudo-range observation satisfies

$$z_{k,i,j_s} = \text{mod}(\varphi_{k,i}^s, \lambda_{\text{sym}}) + w_{\rho,k,i}^s, \quad j_u = 4, j_d = 6. \quad (24)$$

Using the Stage-1 prediction $\hat{\varphi}_{k,i}^{s,(1)}$, the ambiguity index is estimated by the nearest-integer rule

$$\hat{\kappa}_{k,i}^s = \text{round} \left(\frac{\hat{\varphi}_{k,i}^{s,(1)} - z_{k,i,j_s}}{\lambda_{\text{sym}}} \right). \quad (25)$$

Define the corresponding ambiguity-decision error as

$$\varepsilon_{k,i}^s \triangleq \hat{\varphi}_{k,i}^{s,(1)} - \varphi_{k,i}^s - w_{\rho,k,i}^s. \quad (26)$$

Then, apart from the boundary tie case, correct ambiguity resolution is achieved if and only if

$$\hat{\kappa}_{k,i}^s = \kappa_{k,i}^{s,*} \iff |\varepsilon_{k,i}^s| < \frac{\lambda_{\text{sym}}}{2}. \quad (27)$$

Assuming independent Gaussian Stage-1 prediction and OTFS delay-estimation errors, the ambiguity-resolution probability for each link is

$$\mathcal{P}_{\text{unwrap}} = 2\Phi \left(\frac{\lambda_{\text{sym}}}{2\sqrt{\sigma_{\text{MAC},\rho}^2 + \sigma_{\text{OTFS},\rho}^2}} \right) - 1, \quad (28)$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

Let $\mathcal{A}_{k,i}$ denote correct joint uplink and downlink ambiguity resolution. The joint capture condition is then

$$\mathcal{A}_{k,i} \iff \max_{s \in \{u,d\}} |\varepsilon_{k,i}^s| < \frac{\lambda_{\text{sym}}}{2}. \quad (29)$$

b) Stage-2 (Fine Update): Based on the intermediate estimate $\hat{\mathbf{X}}_{k,i}^{(1)}$, the filter subsequently assimilates the high-resolution but ambiguous OTFS pseudo-range observations $\mathbf{z}_2 = [z_{k,i,4}, z_{k,i,6}]^T$. To maintain a locally consistent innovation within the selected ambiguity branch, the Stage-2 residual is wrapped into the principal interval $[-\lambda_{\text{sym}}/2, \lambda_{\text{sym}}/2)$ as

$$\mathbf{y}_2 = \text{mod} \left(\mathbf{z}_2 - \mathbf{h}_{k,\{4,6\}} \left(\hat{\mathbf{X}}_{k,i}^{(1)} + \frac{\lambda_{\text{sym}}}{2}, \lambda_{\text{sym}} \right) - \frac{\lambda_{\text{sym}}}{2} \right). \quad (30)$$

Here, $\text{mod}(\cdot, \lambda_{\text{sym}})$ acts element-wise. The final state and covariance are obtained by the standard EKF correction using $\mathbf{y}_2$ and the Stage-2 Kalman gain $\mathbf{K}_2$.

V. SIMULATION RESULTS

We compare the proposed two-stage 8D-EKF with LS-I [3], a Stage-1-only EKF, and a single-stage wrapped EKF. The simulations evaluate the estimation accuracy, filter consistency, and ambiguity-resolution behavior of the proposed framework. The main simulation parameters are summarized in Table I.

TABLE I
SIMULATION PARAMETERS

Param.	Description	Value
f_c	Carrier frequency	0.6 GHz
$\sigma_{\text{MAC},\rho}$	STD of MAC pseudo-range meas.	0.15 m
$\sigma_{\text{OTFS},\rho}$	STD of OTFS pseudo-range	5.1×10^{-2} m
$\sigma_{\text{OTFS},f}$	STD of OTFS Doppler	3.3 Hz
v	Target speed	150 m/s
t_d	Communication interval	0.2 s
d_0	Initial range	4.5 km
T_{sym}	OTFS symbol duration	10^{-6} s
q_{jerk}	PSD, kinematic jerk noise	6×10^{-3}
q_{bias}	PSD, clock phase noise	5×10^{-10}
q_{skew}	PSD, frequency skew noise	1×10^{-10}
θ_0	Initialization clock offset	$[-0.5, 0.5] \times 10^{-3}$ s
δ_0	Initialization clock skew	$[-1, 1] \times 10^{-9}$
N	Number of reference nodes	3

TABLE II
INNOVATION CONSISTENCY OF THE TWO-STAGE 8D-EKF

Update	Dimension	Expected	Mean ANIS	95% CI
Stage-1	4	4	3.9305	[3.8266, 4.1772]
Stage-2	2	2	2.0841	[1.8779, 2.1258]

A. Tracking Performance and Filter Consistency

Fig. 2 compares the four schemes for target speeds from 10 to 150 m/s. The LS-I estimator degrades rapidly with increasing speed because its snapshot formulation cannot account for the intra-round target motion. In contrast, the Stage-1-only EKF remains stable over the entire speed range, achieving a position RMSE of approximately 0.19–0.42 m.

The single-stage wrapped EKF suffers from incorrect ambiguity-branch selection under the considered initialization, resulting in severe estimation errors. By first constraining the state through the ambiguity-free Stage-1 update, the proposed two-stage EKF achieves a 100% joint ambiguity-resolution success rate for all tested speeds. Its position RMSE remains below 0.16 m and reaches approximately 0.15 m at 150 m/s, substantially outperforming the Stage-1-only EKF. Similar improvements are observed in clock-offset and clock-skew estimation. These results verify that the proposed coarse-to-fine update effectively combines robust ambiguity capture with high-resolution OTFS refinement.

To further assess the statistical consistency of the proposed estimator, Table II reports the ANIS statistics of the two EKF updates. The Stage-1 and Stage-2 values remain close to their nominal degrees of freedom, indicating approximate innovation consistency under the considered nonlinear tracking conditions.

B. Ambiguity Capture and Tracking Robustness

Fig. 3 shows that increasing MAC-layer timing noise degrades ambiguity resolution and increases the risk of tracking divergence. For $\lambda_{\text{sym}} = 300$ m, the degradation becomes pronounced beyond approximately 200 ns; at $\sigma_{\text{MAC},t} = 500$ ns, the ambiguity-resolution probability decreases to about 70%, while the divergence probability reaches about 23%. Increasing λ_{sym} substantially improves robustness: at 500 ns,

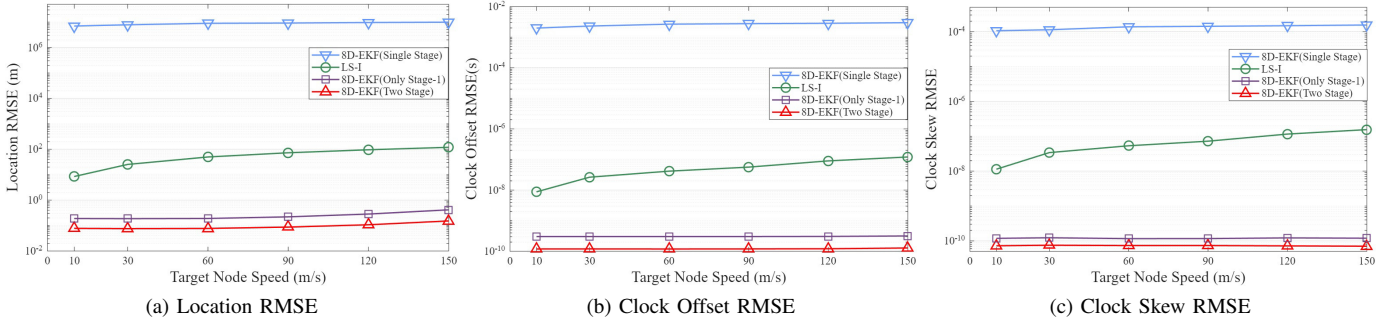


Fig. 2. Comparative performance evaluation of different algorithms under varying motion speeds.

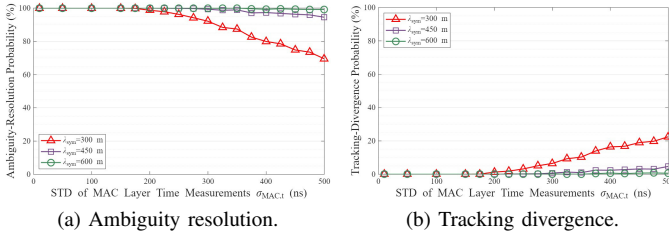


Fig. 3. Ambiguity-resolution and tracking robustness versus MAC-layer noise for different OTFS ambiguity ranges λ_{sym} .

the ambiguity-resolution probabilities remain approximately 95% and 99% for $\lambda_{sym} = 450$ and 600 m, respectively, with divergence probabilities below 5% and 1%. Moreover, ambiguity failures do not necessarily cause immediate divergence, although frequent incorrect branch selections increase the risk of loss of track. These results are consistent with (29), since a larger λ_{sym} enlarges the admissible Stage-1 decision-error interval before an incorrect ambiguity branch is selected.

VI. CONCLUSION

This paper proposed an OTFS-assisted coarse-to-fine framework for joint synchronization and localization in high-mobility wireless networks. An 8D-EKF tracks the target's kinematic and clock states using ambiguity-free MAC/Doppler observations for coarse conditioning and OTFS pseudo-ranges for fine refinement. Simulations demonstrate decimeter-level localization accuracy at speeds up to 150 m/s and confirm innovation consistency in both update stages. Ambiguity-resolution experiments show that increasing λ_{sym} improves tracking robustness, while observed first-failure errors closely match the theoretical half-period boundary $\lambda_{sym}/2$. The current study is limited to simulation-based evaluation under the considered motion, clock, and measurement models, and its robustness still depends on the accuracy of the Stage-1 prediction. Future work will investigate more complex maneuvering dynamics, imperfect reference-node timing, and experimental validation in realistic wireless environments.

REFERENCES

[1] K. Han, K. Meng, X.-Y. Wang, and C. Masouros, "Network-level ISAC design: State-of-the-art, challenges, and opportunities," *IEEE Journal of*

Selected Topics in Electromagnetics, Antennas and Propagation, vol. 1, no. 1, pp. 65–83, 2025.

- [2] R. M. Vaghefi and R. M. Buehrer, "Cooperative joint synchronization and localization in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3615–3627, 2015.
- [3] J. Zheng and Y.-C. Wu, "Joint time synchronization and localization of an unknown node in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1309–1320, 2010.
- [4] S. Zhu and Z. Ding, "Joint synchronization and localization using TOAs: A linearization based WLS solution," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1016–1025, 2010.
- [5] Z. Liu, J. Zhan, T. Ju, Z. Cao, B. Yang, and H. Liang, "TDOA based extended kalman filter fusion approach for UAV localization in GPS-denied environment," in *2024 4th International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC)*, 2024, pp. 664–669.
- [6] J. Cano, S. Chidami, and J. Le Ny, "A kalman filter-based algorithm for simultaneous time synchronization and localization in UWB networks," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 1431–1437.
- [7] G. Giorgi and C. Narduzzi, "Performance analysis of kalman-filter-based clock synchronization in IEEE 1588 networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 8, pp. 2902–2909, 2011.
- [8] M. Koivisto, M. Costa, J. Werner, K. Heiska, J. Talvitie, K. Leppänen, V. Koivunen, and M. Valkama, "Joint device positioning and clock synchronization in 5G ultra-dense networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2866–2881, 2017.
- [9] M. Koivisto, M. Costa, A. Hakkarainen, K. Leppänen, and M. Valkama, "Joint 3D positioning and network synchronization in 5G ultra-dense networks using UKF and EKF," 2016.
- [10] F. Li, L. Wang, W. Qing, Y. Li, Y. Liu, and Y. Niu, "Adaptive-weighted EKF and LiDAR-IMU tight-coupling enabled GPS-denied UAV navigation," in *2025 International Conference on Intelligent Robotics and Automatic Control (IRAC)*, 2025, pp. 150–154.
- [11] Z. Yang, Y. Shen, X. Shi, and Y. Wang, "Joint time synchronization and ranging method aided by doppler frequency shift," in *2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2022, pp. 1616–1621.
- [12] A. Nordio, C. F. Chiasserini, and E. Viterbo, "Joint communication and sensing in OTFS-based UAV networks," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 6, pp. 9445–9460, 2025.
- [13] D. Hu, J. Nakazato, K. Maruta, R. Dinis, O. Abbassi Aghda, and M. Tsukada, "OTFS based RIS-assisted vehicle positioning and tracking in V2X scenario," *IEEE Transactions on Communications*, vol. 73, no. 11, pp. 10 676–10 691, 2025.
- [14] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "On the effectiveness of OTFS for joint radar parameter estimation and communication," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 5951–5965, 2020.
- [15] J. Wang, C. Li, N. Zhang, and J. Yang, "Fractional parameter estimation and Cramér-Rao bound for OTFS-based ISAC systems," *IEEE Transactions on Vehicular Technology*, vol. 75, no. 5, pp. 8747–8752, 2026.
- [16] K. R. Murali and A. Chockalingam, "On OTFS modulation for high-doppler fading channels," in *2018 Information Theory and Applications Workshop (ITA)*, 2018, pp. 1–10.