SORA-Loc: Software-Defined Resource Allocation for Practical Multi-User 5G Uplink Localization

5G uplink localization has generally been be recognized as a promising technology, by exploiting user transmissions in cellular networks to support various mobile location-based services. Besides the spatial accuracy (SA), the temporal granularity (TG) of location updates is also necessary for estimating mobility trajectories, and the system has to make tradeoff between TG and SA due to diverse mobility and limited resource. Conventionally, this tradeoff is achieved by time-frequency resource allocation, which incurs prohibitive protocol overhead in multi-user scenarios. In this paper, we present *SORA-Loc*, a software-defined resource allocation approach that leverages software-only channel-state information (CSI) processing, optimizing spatio-temporal tradeoff by exploiting multi-user uplink transmissions. This approach reduce the resource allocation complexity of N users and K bandwidth parts from $O(N^K)$ to $O(NK^2)$, improving the quality of trajectory estimation with the given amount of total time-frequency resources. We design a comb-sliding multiplexing scheme to support a various range of TG and SA tradeoff for multi-user through uplink CSI combination. To further predict user mobility pattern and perform optimal spatio-temporal tradeoff, we propose a built-in mobility mapping method based on CSI dynamics that operates both online and offline. Extensive experiments and trace-driven simulation at scale demonstrate that given a pilot overhead of 14.3%, *SORA-Loc* facilitates trajectory estimation accuracy for 100 users with an improvement up to 4.84x compared with conventional 5G localization.

CCS Concepts: • **Networks → Location based services**; • **Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**.

Additional Key Words and Phrases: 5G localization, resource allocation

ACM Reference Format:
. 2018. SORA-Loc: Software-Defined Resource Allocation for Practical Multi-User 5G Uplink Localization. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 0, 0, Article x (August 2018), 21 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

5G localization enables numerous applications such as logistics management, intelligent transportation, and industrial automation [3, 32]. Compared with non-5G techniques like Global Navigation Satellite System (GNSS), WiFi and Inertial Measuring Unit (IMU) sensors, 5G localization reuses ubiquitous cellular infrastructure and can fully utilize the 5G new features at scale, such as wider bandwidth, multiple antennas, and ultra dense networking [12, 14, 23]. In the 3GPP 5G positioning standard [7], both uplink and downlink localization are supported. Compared with downlink cases, 5G uplink localization require base stations (BSs) receive arbitrary uplink transmission from multiple users and estimates their locations, which introduce less computation and radio capability requirements for the user equipments and significantly improve the timeliness of reporting location updates for upper-layer applications [16].

While promising progresses in 5G localization have been achieved, such as anchor selection [32] and signal scheduling [30], one observation is that these studies focus on downlink cases, improving the spatial accuracy on each individual location. For example, pioneer works like Tonetrack, Splicer and M3 [13, 48, 52] proposed CSI combination technologies for single user to improve spatial accuracy in Wi-Fi downlink positioning, which require the user to combine multiple frequency bandwidth in relatively stable scenarios. 5G can also benefit from CSI combination, which has been proven to effectively improve spatial accuracy in OFDM systems. However, compared with various mobile location-based services such as logistic tracking and factory automation [4, 57], we are more interested in locating multiple moving users by analyzing their uplink transmissions over a period of time in cellular networks, and formulating spatio-temporal mobility trajectories for upper-layer applications [1, 20]. Our key insight is that in 5G, not only the spatial accuracy of a single user needs to be considered, but

Author's address:

also the temporal granularity, thus requiring a flexible approach to management the spatio-temporal tradeoff for multiple users.

This paper explores how to generate high quality mobility trajectories for 5G uplink multi-user scenarios, especially with diverse mobility, which requires both high *temporal granularity* (TG) and *spatial accuracy* (SA). This system often has to make tradeoff between TG and SA due to dynamic mobility and resource-limited scenarios. For example, the mobility pattern of logistics trucks can vary over time, from waiting to load to driving under various traffic conditions [1, 4], and it may be desired to offer better TG when the truck is with high mobility and vise versa. This problem is more challenging when considering numerous users simultaneously connected to 5G, so frequent resource re-allocation on both bandwidth and time slots becomes necessary.

Conventional resource allocation approaches incur prohibitive costs in two aspects: (1) *Re-scheduling:* Allocating multi-dimensional resources for densely connected users with diverse service requirements is a complex optimization problem, which is hard to solve efficiently [31, 41]. (2) *Re-configuring:* Even if the system is able to schedule the resources dynamically, according to the radio resource control (RRC) protocols [16, 53], there is still an inevitable protocol overhead and time delay to configure down to the physical layer [39, 56]. In practice, the above two steps may iterate, leading to an even higher network load over time [31].

To solve this fundamental challenge, we pose the following question: *While resource allocation is necessary but costly, is there a way to provide flexible 5G uplink time-frequency resource allocation, so temporal granularity and spatial accuracy tradeoff in 5G uplink localization can be achieved without causing complex modifications and protocol overhead?*

This paper presents *SORA-Loc*, a **SO**ftware-Defined **R**esource **A**llocation approach for multi-user 5G uplink **Loc**alization (§4.1), which leverages software-only CSI processing to achieve a flexible and programmable spatio-temporal tradeoff according to user mobility, thus improving the quality of trajectory estimation with the given amount of total time-frequency resources. SORA-Loc exploits 5G multi-user uplink transmission, which combines coherent CSI measured across various time and frequency bandwidth, providing a fair range and granularity of the spatio-temporal tradeoff for each user. While the CSI combination technique was initially designed to enable low-cost devices with limited sampling rates to sense a larger downlink bandwidth [48, 52], we discover that in 5G uplink multi-user cases, it is possible to trade temporal granularity of CSI measurement for better spatial accuracy, and such a process can be fully software defined, thus triggering little protocol overhead.

5G standard offers a choice of frequency hopping to sense entire allocated uplink bandwidth, where multiple users can transmit sounding reference signals (SRS) in different bandwidth parts at different times [5], supporting channel dependent scheduling and link adaptation. We first validate the feasibility of CSI combination in 5G physical layer and demonstrate its performance gains in mobile scenarios (§4.2). Following such encouraging validation results, to fully leverage these opportunities to achieve desired spatio-temporal tradeoff in practical 5G localization, however, remains challenging on two technical aspects.

First, SORA-Loc needs to provide a flexible and 5G compatible spatio-temporal tradeoff functionality, which supports scalable multi-user multiplexing and is able to adapt to diverse tradeoff range and granularity requirements using software processing only. To address such a challenge, we design a *comb-sliding multiplexing* scheme (§4.3) to work with pilot transmissions without impacting actual data communications, which provides fine granularity for spatio-temporal tradeoff through CSI combination. In this way, a fully software-defined stack is built into 5G, where for an arbitrary user equipment, given any temporal requirements, SORA-Loc determines the achievable spatial accuracy.

Second, given that SORA-Loc operates on the base station side, achieving an optimal spatio-temporal tradeoff in practice necessitates the prediction of user mobility patterns by exploiting 5G pilot signals only. This prediction determines a most appropriate spatio-temporal tradeoff, which guides the proper combination of CSI. We design an efficient *mobility profiling and resource mapping* technique based on CSI dynamics that works for both online and offline mobility mapping applications (§4.4).

The proposed SORA-Loc is implemented using Xilinx UltraScale Radio Frequency System-on-Chip (RFSoC) hardware testbed. Full 5G standard is followed, with complete 5G signal generation and CSI estimation procedures performed on the testbed. Comprehensive experiments in both indoor and outdoor scenarios are conducted to evaluate SORA-Loc (§5). The results show that SORA-Loc achieves an average of 2.5 m trajectory estimation accuracy with only 1.4% pilot overhead, and achieves an improvement of up to 4.84x by extending existing resource allocation solutions as alternatives. Finally, large-scale, trace driven simulations are conducted to show system scalability and communication throughput with significant mobility dynamics and user access variations.

In summary, this paper makes the following contributions:

- We consider practical 5G uplink multi-user mobile localization scenario, where we define and aim to enhance a new evaluation metric considering both spatial accuracy and temporal granularity.
- We present SORA-Loc, to the best of our knowledge, the first programmable spatio-temporal tradeoff for 5G multi-user uplink localization, which is realized by a software-defined resource allocation approach, triggering little protocol overhead with reduced complexity from $O\left(N^K\right)$ to $O\left(NK^2\right)$.
- We design a flexible and 5G-compatible uplink multiplexing scheme to enable fair range and granularity of spatio-temporal tradeoff under user access variations, and along with a mobility mapping method to adapt significant mobility dynamics.
- We implement SORA-Loc and evaluate its performance with real-world experiments and trace-driven simulation at scale, supporting 100 users with up to 4.84x improvement in trajectory estimation accuracy over state of the art, given fixed pilot overhead of 14.3%.

2 PRIMER ON 5G LOCALIZATION

Localization procedures. According to the 3GPP standard [5], 5G localization can be summarized in three steps. 1) *Initialization*: The user equipment (UE) transmits location service requirements to the location management function (LMF), which will schedule and assign wireless resources for the UE and base station (BS) communication. 2) *CSI measurement*: The BS or UE estimates the CSI from the uplink (or downlink) physical layer transmission according to the pre-configured sounding reference signals (SRS). 3) *Location estimation*: The LMF estimates the UE location according to the measurements derived from the CSI, including the time-of-arrival (TOA), received-signal-strength (RSS), angle-of-arrival (AOA) and cell identity (CellID).

In detail, CSI measures the radio propagation channel as the sum of P attenuated and delayed paths, which in frequency domain can be expressed as

$$H_k = \sum_{p=1}^{P} \alpha_p \exp(-j2\pi \left(f_0 + k\Delta f\right) \tau_p), \tag{1}$$

where α_p and τ_p are the complex attenuation and propagation delay from p-th path, Δf is the subcarrier spacing, f_0 and f_k are the first and k-th carrier frequency, respectively.

For isometric subcarriers in OFDM symbol l, p-th path with TOA τ_p introduces a phase shift of $-2 \times \pi \times (k-1) \times \Delta f \times \tau_p$ at the k-th subcarrier relative to the first subcarrier of the received signal [25]. Hence, the Multiple Signal Classification (MUSIC) algorithm can be used to estimate the range between UE and BS [25, 52]. The LMF utilizes least-square method to estimate user locations and provide the information to upper-layer applications [42]. For simplicity, the location estimation function can be denoted as

$$\hat{X}_{jt_0}^{u_i} = \mathrm{loc}\left(\hat{H}_{\tilde{k},l}\right), \tag{2}$$

where $X_{jt_0}^{u_i}$ is the true location of user u_i during time period $((j-1)\,t_0, jt_0\,]$[1], where l is the corresponding transmitted symbol index and $\tilde{k}$ is the subcarrier set within $(f_k \pm B_0/2)$. We denote $\hat{H}_{\tilde{k},l}$ as the minimum time-frequency resources required for the location update.

[1]This time period can be chosen according to the system requirements, which is typically longer than the duration of one OFDM symbol.

Evaluation metric. Spatial error is commonly used to evaluate the 5G localization accuracy. For example, the 3GPP standard adopts the cumulative distribution functions (CDF) of spatial errors with 50%, 67%, 80%, 90% percentiles [3]. Such a metric, however, only considers the location accuracy in the spatial domain, but not the temporal granularity of location updates, and thus may not comprehensively quantify the accuracy of a mobility trajectory derived from a time series of locations.

3 MOTIVATION

3.1 New spatio-temporal evaluation metric

For 5G users with diverse mobility patterns, when evaluating the reconstructed trajectories, the localization evaluation should consider both the spatial accuracy (SA) of location estimation and the temporal granularity (TG) for location updates. In Figure 1, we use an example to show the impact of TG and SA to the localization accuracy in mapping the trajectories of different mobile users. In this scenario, 100 users each transmits one pilot symbol to measure the CSI in turn, utilizing 100MHz bandwidth for every time slot of 10 ms[2]. Figure 1(a) to (c) plot the reconstructed trajectories of users moving along the same circular trajectory but with different speed of 3, 6, and 12 m/s, respectively. The results show that whenever the TG fails to catch up with the user mobility, lossy location information is obtained and thus leading to deteriorated trajectory accuracy.

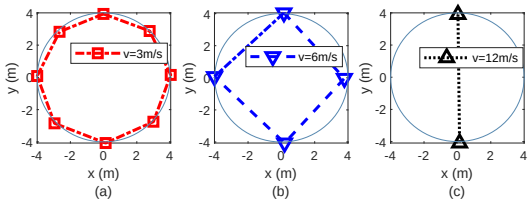

Fig. 1. Trajectory estimation with diverse user mobility using the same time-frequency resources.

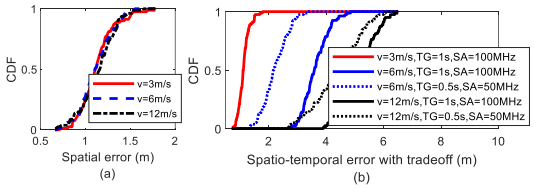

Fig. 2. (a) The traditional spatial error metric cannot quantify the impact from temporal granularity of location updates; (b) The STE quantifies the spatio-temporal performance, and making best tradeoff between TG and SA by resource allocation according to user mobility can improve the STE.

To comprehensively evaluate the performance of 5G localization in the presence of diverse mobility, we propose a new spatial-temporal error (STE) metric, which is formulated as

$$\mathrm{STE}\left(\hat{\xi}^{u_i}_{t_0,T}\right) = \frac{1}{T}\sqrt{\sum\nolimits_{j=1}^{T}\left(\hat{X}^{u_i}_{jt_0} - X^{u_i}_{jt_0}\right)^2}, \tag{3}$$

where $\hat{\xi}^{u_i}_{t_0,T}$ represents the mobility trajectory which contains a sequence of samples consisting of a time and a locations component, i.e., $\hat{\xi}^{u_i}_{t_0,T} = \{X^{u_i}_{t_0}, ..., X^{u_i}_{Tt_0}\}$ where $X^{u_i}_{jt_0}$ is the location of the user u_i sampled at time jt_0, $j \in \{1, \ldots, T\}$. The STE takes both TG and SA into consideration, corresponding to t_0 and $\left\|\hat{X}^{u_i}_{jt_0} - X^{u_i}_{jt_0}\right\|_2$, respectively.

We thereafter perform a Monte-Carlo simulation (100 rounds for each case) to evaluate the localization accuracy of the three different users. Figure 2(a) shows that with the traditional metric of only examining spatial errors, the above three users show indifferent localization performance, which significantly contrasts to what Figure 1(a) to (c) suggest. On the other hand, with the proposed STE metric, Figure 3(b) clearly shows the different qualities of the reconstructed trajectories with different user mobility levels (the user moving at 3m/s achieves the highest localization accuracy while the user moving at 12m/s achieves the lowest). We further experiment with a resource

[2]This setting is equivalent to TG=1s and SA~100MHz for each user's trajectory estimation.

allocation tradeoff for TG and SA, namely, with (i) SA from 100MHz bandwidth and TG of 1s location update interval, or (ii) SA from 50MHz bandwidth but TG of 0.5s location update interval for user 2 and 3. Figure 2(b) further suggests that allocating different time-frequency resources to trade TG for SA may improve overall STA performance in certain mobility conditions.

3.2 A fundamental challenge in multi-user 5G localization

To improve the STE, both higher frequency bandwidth (for better SA) and more time slots of location measurement (for better TG) are needed. However, for practical 5G scenarios with multiple users and limited total time-frequency resources, it is not feasible to allocate one user with more time and frequency resources simultaneously. Hence, it is necessary to make a tradeoff between TG and SA within the fair time-frequency share of one user to achieve the goal of minimizing the STE according to the user mobility pattern. For users with higher mobility, we may trade SA for better TG, e.g., with only 50MHz bandwidth but doubled time slots for location updates as shown in Figure 2(b), and vice versa. However, achieving such resource allocation for dense connected users under diverse mobility is challenging because it requires (1) re-scheduling and (2) re-configuration process, both of which may incur prohibitive cost and overhead.

Re-scheduling introduces a complex optimization problem. Users with diverse mobility exhibit heterogeneous spatial and temporal requirements. Figure 3 shows the workflow of traditional resource allocation approaches. Based on an existing allocation pattern (①), the system needs to explicitly re-schedule time and frequency resources to match the users mobility (②), which can be formulated as a complex searching problem. As one 5G base station concurrently serves multiple users in OFDMA time-frequency grids, for each candidate scheduling result (i.e., one possible OFDM resource grid for time-frequency sharing across the multiple users), the location management function (LMF) and access mobility function (AMF) in the core network need to evaluate the overall localization performance (e.g., STE) of all users. Assuming one clock cycle per evaluation, the total number of clock cycles, i.e., the computational cost brought by the re-scheduling is thus

$$\zeta_{\text{total}} = \mathrm{C}_{KN}^{K} + \mathrm{C}_{(K(N-1))}^{K} + \ldots + \mathrm{C}_{K(N-N+1)}^{K} = O\left(N^{K}\right) \tag{4}$$

where N and K are the number of users and sub-channels in frequency domain, respectively. The computational load increases significantly as N and K grow [31, 56].

Re-configuration incurs prohibitive connection latency. Even though the core network schedules the resource allocation problems, the system needs to configure the updated resource pattern from the 5G upper layer to the physical layer following the radio resource control (RRC) protocols (as shown in Figure 3 (③)). The 5G base station (BS) and user equipments (UEs) thereafter communicate on the assigned frequency bandwidth and time slots to obtain new CSI for location updates. Such a process incurs a significant overhead in 5G protocols [16], using time for configuration and frequent activation and deactivation of the radio front-end. Furthermore, due to diverse mobility, the system may need to repetitively perform the re-scheduling and re-configuration processes, resulting in prohibitive overhead.

To quantifying the latency caused by the above two processes, we denote the time used per clock circle (cc) as T_0^{d}, and the total delay can be modeled as

$$T_d = T_{init} + \zeta_{\text{total}} \cdot \left(T_0^{\mathrm{d}} + T_c\right) \tag{5}$$

where T_{init} is the initial access latency (around 160 ms [22, 44]), T_0^{d} is the scheduling latency per clock circle determined by the computing power [19]), and T_c is the configuration latency (around 2 ms [46]). Besides the computational cost, localization performance may be impaired when the mobility and user requirements change within T_d and the scheduling results based on previous observation become outdated.

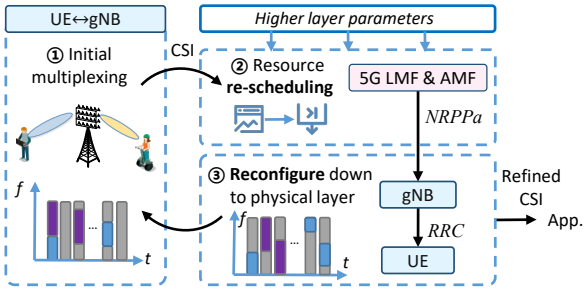

Fig. 3. Conventional resource allocation workflow requires repetitive re-scheduling and configuring through the 5G protocol, with significant computation overhead and latency.

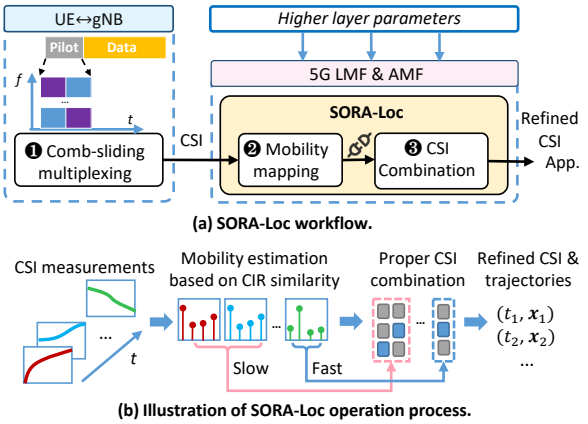

Fig. 4. SORA-Loc workflow: exploits uplink multi-user transmission, enabling flexible, fair, and programmable spatio-temporal tradeoffs according to user mobility.

4 SOFTWARE-DEFINED RESOURCE ALLOCATION

4.1 Design overview

As analyzed in Section 3, given the necessity and difficulties of resource allocation for practical 5G multi-user localization, we aim to design a software-based approach to achieve adaptable spatio-temporal tradeoff with upper-layer computations only, without going all the way through the 5G protocol stack to alter the frequency and time resources at the physical layer.

We leverage the knowledge of CSI combination, which provides the possibility to merge coherent CSI measurements taken at different times and frequency bands, creating an increased bandwidth for improved accuracy. While this technique has previously been proposed for low-cost devices to sense a broader bandwidth [48, 52], it also offers the capability to trade off temporal granularity for spatial accuracy. Compared to the conventional resource allocation workflow as illustrated in Figure 3, SORA-Loc can be seamlessly executed in software. By refining CSI in upper layer functions like the LMF, SORA-Loc allows flexible spatio-temporal tradeoff for multiple users and avoids the 5G protocol overhead for re-configuring the physical layer.

Overall workflow. The proposed SORA-Loc system comprises three primary components and operates sequentially, as illustrated in Figure 4(a). (❶) To achieve flexible and scalable CSI combination in 5G, we propose a comb-sliding multiplexing pattern for multi-user pilot transmission. This pattern provides equitable range and granularity of time-frequency resources for CSI estimation, which are prepared for achieving the desired spatio-temporal tradeoff in subsequent processes. (❷) Leveraging the CSI reported by the physical layer, a mobility mapping module, designed for both online and offline scenarios, is deployed to forecast user mobility profiles and map the corresponding resources for subsequent combination. (❸) The CSI combination function refines CSI adaptively based on user mobility, achieving spatio-temporal tradeoff to enhance the STE performance. Based on the physical layer CSI extraction, SORA-Loc operates at the network-side to perform trajectory refinement and provides support to various location-based application requirements on TG and SA. Figure 4(b) illustrates the detailed procedure in CSI combination according to estimated mobility levels.

Complexity. Compared with traditional time-frequency resource allocation in physical layer, SORA-Loc transform the computation to CSI post-processing, which only introduce $O\left(NK^2\right)$ clock cycles caused by

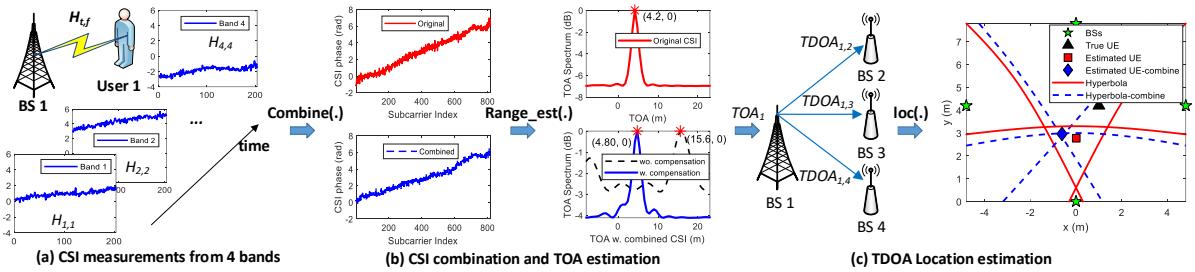

Fig. 5. Feasibility experiments of 5G range-based localization with CSI combination.

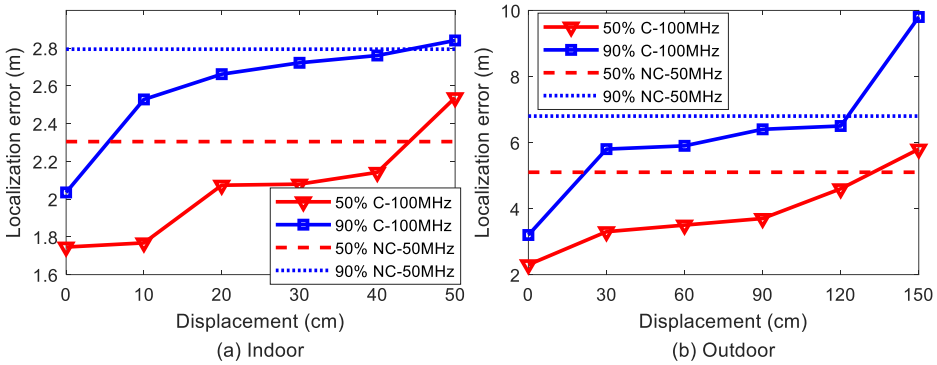

Fig. 6. For moving users, CSI combination can still increase accuracy within certain displacements.

$(K-1)$ $(K-1)$ CSI combination and $(K-1)$ $(K-1)$ mobility prediction operations, as compared with $O\left(N^K\right)$ in formula (4). See §A for detailed mathematical derivation.

The detailed design of SORA-Loc starts with the basis of CSI combination and a 5G proof-of-concept (§4.2). To make the design fully compatible and practical to 5G, we then introduce two design parts of SORA-Loc to support flexible and scalable CSI combination for 5G multi-user (§4.3) and adaptive mobility mapping in practical applications (§4.4).

4.2 CSI combination for 5G localization

Basis of CSI combination. The basic idea behind CSI combination is to combine the CSI from multiple frequency-agile transmissions, concatenating a larger bandwidth to yield improved accuracy [48]. However, merely stacking the reported CSI samples into a single vector is insufficient due to time and frequency offset. CSI combination aligns these CSI groups by using the first CSI group as reference to facilitate time-frequency offset compensation [52].

Note that CSI derived from non-overlapped or identical bandwidth parts (BWP) can also be weighted by the signal-to-noise-ratio (SNR) to mitigate errors over time. In this paper, we give a general CSI combination function

$$\text{combine}\left(\hat{H}_k, \hat{H}_{k'}\right) = \begin{cases} \text{concat}\left(\cdot\right), \text{if } |f_k - f_{k'}| \le B_0 \\ \text{weight}\left(\cdot\right), \text{otherwise}, \end{cases} \tag{6}$$

where $|f_k - f_{k'}| \le B_0$ stands for $\hat{H}_k$ and $\hat{H}_{k'}$ is from consecutive BWP with range B_0.

5G feasibility. CSI combination relies on consecutive frequency bands measured at different times. 5G operates as an OFDMA-based system, allowing multiple users to transmit their signals allocated within finely-grained

frequency and time blocks. This presents an opportunity to combine CSI according to user requirements for temporal granularity and spatial accuracy.

To validate the feasibility of CSI combination in 5G, we conduct an uplink time-difference-of-arrival (TDOA) range-based localization experiment in our hardware testbed. 5G standard supports range measurement with synchronized BSs as anchors. 3GPP standard anticipates ~3 ns synchronization error for meter level localization accuracy [3, 6] Our testbed incorporates an oven-controlled crystal oscillator, providing 1.5 ns synchronization accuracy, and the SA here is mainly determined by the bandwidth per estimation.

Combining the CSI obtained from four BWP, each of 25 MHz, in different time slots (Figure 5(a)), we are able to generate a 100 MHz CSI and estimate the ToA. 5G system allocates sub-carrier level comb structure to allow overlapped bandwidth for single user over consecutive time slots. We also implement received signal strength based ranging estimation to mitigate measurement error. After compensation, the combined CSI closely resembles the ToA that can be derived using the full-band CSI measurement (Figure 5(b)). We take BS 1 as the reference to measure the TDOA and employ the least square (LS) algorithm to estimate the user location. Figure 5(c) illustrates the localization error using the original full-band 100 MHz CSI, which is 1.7 m, while the combined CSI from the four 25 MHz BWP yields a similar performance of 1.9 m. CSI combination can immediately benefit other applications such as AOA estimation and human sensing with improved accuracy.

In mobile experiments, we observe that the CSI measured within certain displacements can still be combined to enhance accuracy. However, as user mobility increases, a gradual deterioration appears. The underlying rationale is that when the channel is relatively stable, CSI produces similar power delay profiles, allowing for effective combination. We conduct experiments with users moving within linear displacement ranging from $\Delta d = 10$ to 50 cm from the original point ($\Delta d = 0$) in indoor and from 30 to 150 cm in outdoor scenarios. To achieve an effective bandwidth of 100 MHz, we combine two consecutive 50 MHz CSI measured at the original point and another point located at Δd. As shown in Figure 6, the CDF of localization errors at the 50th and 90th percentiles for CSI combination (C-100) is lower than that for the 50 MHz bandwidth without CSI combination (NC-50) for displacements of up to 40 and 120 cm in indoor and outdoor scenarios. They correspond to a mobility level of 4 m/s indoors and 12 m/s outdoors, respectively, if we combine 5G SRS measurements over a duration of 100 ms.

4.3 Comb-sliding multiplexing for flexible and scalable CSI combination

After demonstrating the feasibility of CSI combination in the 5G physical layer, we subsequently investigate methods for guaranteeing various spatio-temporal tradeoff requirements through CSI combination (§4.3.1). Furthermore, in practical applications, the system also need to support dynamic user access (§4.3.2). While it is technically possible to combine CSI with any multiplexing pattern, it may result in substantial performance loss due to unfairness among users and inefficient utilization of CSI.

4.3.1 Flexibility to software-tunable combination

We address the first point by arranging time and frequency resources for each users' CSI combination with fair range and granularity. For example, the user transmits a signal in the first time slot, utilizing the entire bandwidth, and then remains inactive. Such a pattern would result in uneven TG and SA, potentially leading to a degradation in STE performance due to the lack of TG and SA tradeoff ability.

Hence, we introduce the comb-sliding pattern as the pilot multiplexing scheme for SORA-Loc. To illustrate, we select a scenario with $N = 4$ users, presuming that the initial access and synchronization processes have been successfully carried out. This ensures that the CSI from multiple users can be concurrently extracted by BSs [2, 6]. As depicted in Figure 7(a), the term **"comb"** signifies that *within* each SRS time slot, four distinct users will transmit their reference signals. This ensures that the minimum temporal granularity for each user. The term **"sliding"** denotes that *between* each consecutive SRS measurements, each user will slide and transmit on the next bandwidth segment. This ensures that after one period of the comb-sliding transmission, each user will gain an equal amount of frequency diversity, and their CSI can be combined using the concat $(\cdot)$ function. Then, the

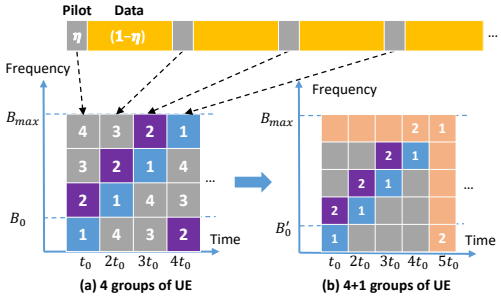

Fig. 7. The proposed comb-sliding pattern under: (a) fixed and (b) dynamic connections.

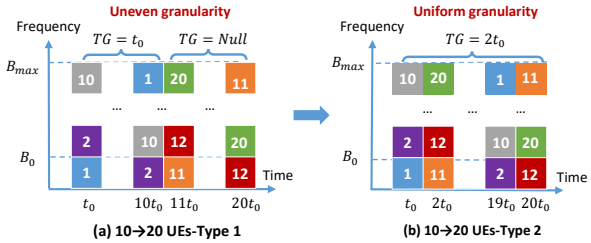

Fig. 8. Time-extended scalability support: (a) sequential and (b) alternative solution.

adjacency matrix of time(t)-frequency(k) multiplexing for user i can be logically expressed as

$$\mathbf{a}_{t,k}^{i} = \text{diag} \underbrace{(1,\ldots,1)}_{(a_{t_0,B_0},\ldots,a_{Nt_0,NB_0})}, \forall t \in ((m-1)\,Nt_0, mNt_0], \tag{7}$$

where $m \in \mathbb{Z}^+$, and the constraint on t implies that each user's multiplexing matrix during any complete combination period should be a diagonal matrix.

Following this pattern, all users have the opportunity to measure across the entire bandwidth after $4t_0$ with a step of $1t_0$. As a result, both the range and granularity for flexible spatio-temporal tradeoff through CSI combination can be ensured. Note that the proposed comb-sliding multiplexing is embedded in the SRS pilot transmission only, which does not impact data communication, thus of little overhead.

4.3.2 *Scalability for dynamic user connection*

Moving beyond the fixed comb-sliding pattern for fixed user connections, we then explore how to extend our design to fit both small and large-scale dynamic scenarios.

For a small-scale increase in the number of connected users, as shown in Figure 7(b), it is advisable to expand the original matrix. Specifically, the logical multiplexing matrix should be extended from 4×4 to 5×5. Note that each user's transmission period is not a straightforward option due to hardware limitations. Consequently, we divide the frequency bandwidth into five sub-channels and extend the time slots within one combination period from $4t_0$ to $5t_0$. However, when faced with a large-scale connection increase, as depicted in Figure (8), configuring a finer-grained time-frequency pattern is not a viable solution. This is because the spatial accuracy with each measurement time is inherently constrained by the available bandwidth and would decrease as the number of users increases, and this will also elevating operational complexity due to frequent frequency hopping.

In such situations, a necessary compromise must be made in the time domain to increase the TG while maintaining the required minimum frequency bandwidth B_0. We refer to this compromise as the time-extended user scalability support. As shown in Figure 8(a), sequential transmissions between two groups will lead to uneven TG and unstable performance. Therefore, the alternative transmission is recommended, as shown in Figure 8(b), where users in two groups experience consistent TG over time. Here, the adjacency matrix of time-frequency multiplexing can be extended to

$$\tilde{\mathbf{a}}_{t,k}^{i} = \mathbf{a}_{t,k}^{i} \otimes [1\ 0] = \left[\mathbf{a}_{t_0}^{i}, \mathbf{0}, \mathbf{a}_{2t_0}^{i}, \mathbf{0}, \ldots, \mathbf{a}_{Nt_0}^{i}, \mathbf{0}\right], \tag{8}$$

where $\otimes$ is the Kronecker product to demonstrate the time-extended operation by inserting a vector $\mathbf{0}$ for other users between each column of the original $\mathbf{a}_{t,k}^{i}$ in formula (7).

To generalize the time-extended case in scalable comb-sliding multiplexing, we define that users n and $(n+10)$ belong to a single user group in a logical manner. Therefore, the comb-sliding matrix is able to set a *mathematical boundary* for the TG and SA tradeoff ranges. In the context of N groups of users with a $T \times B$ OFDM grid, the maximum and minimum values of TG and SA are as follows

$$\begin{cases} TG_{\min} = \dfrac{T}{N} \leftrightarrow TG_{\max} = T, \\ SA_{\max} = \dfrac{N}{B} \leftrightarrow SA_{\min} = \dfrac{1}{B}, \end{cases} \tag{9}$$

where we regard the available bandwidth as the achievable SA for simplicity, as it determines the resolution. To streamline the analysis, we introduce the term *combination window length* as w. This represents the quantity of combined CSI in units of the minimum TG, i.e., $w = lt_0, \forall l \in \mathbb{Z}^+$.

As a result, we can realize software-defined resource allocation by establishing an API, where the input includes the parameter w and the original CSI, and the output is the refined CSI. To provide further clarity, the time slots used for CSI combination projects the quantity of bandwidth and CSI available for combination and localization process.

Impact on communication. SORA-Loc arranges pilot transmission only with the given amount of total time-frequency resources for 5G localization, which reflects negligible impact on communication. As Figure 7 shows, we denote the percentage of pilot signal resources as η, and the equivalent communication throughput can be expressed by

$$R = (1 - \eta) \cdot B \cdot \log_2 (1 + SNR) . \tag{10}$$

In this paper, we set the minimum TG of 4 users as 20 ms to measure 100 MHz bandwidth, i.e., 4 OFDM symbol every 2 radio frames and η equals to 1/70 [5]. This overhead reflects only a 1.43% reduction of the communication throughput.

4.4 Mobility mapping for proper spatio-temporal tradeoff

Now that the CSI combination function in SORA-Loc has been built, we must consider the most appropriate way to combine CSI, since CSI can only be effectively combined when the channel conditions are relatively stable [17]. In practical applications, users may exhibit diverse mobility patterns, which significantly impact the coherence of CSI [48]. For instance, given an empirical coherency range, user speed determines the coherence time, and CSI within that period can be combined to improve the STE. Inappropriate CSI combination does not contribute much and may even lead to performance degradation. Therefore, it becomes imperative to estimate the user's mobility and map to the proper combination window length. This approach effectively bridges the gap of applying SORA-Loc in practice.

One simple way to achieve mobility mapping is by reporting the user's IMU readings to the BS. However, this method is not always feasible due to error accumulation and extended battery drain caused by the on-board sensor, as well as large communication overhead. In this study, we design a general framework of mobility estimation for 5G users, which only requires the estimated CSI, cooperating with the design idea of SORA-Loc. For online and offline operations, the system has varying information availability, computational resources constraints and latency requirements, necessitating tailored solutions for each.

4.4.1 Online mobility mapping

For online operation, based on the current and previous CSI, the system needs during runtime to quickly decide whether to combine them to increase the accuracy or process the current CSI individually to derive an updated location. Firstly, we chose an indicator capable of representing changes in CSI coherence from user mobility. Then, we design the resource mapping between the indicator dynamics and the appropriate CSI combination.

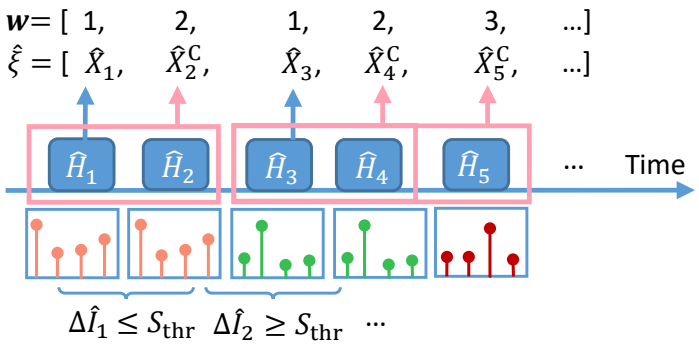

Fig. 9. Illustration of online mobility mapping.

Indicator selection. The traditional Doppler frequency shift f_d is not suitable for the localization over a period of time [35]. This is because a trajectory may encompass multiple stationary points with very similar f_d, which is beyond its capabilities and cannot precisely reflect the mobility. Using the similarity between CSI $\hat{H}_{jt_0}$ and $\hat{H}_{(j-1)t_0}$ is also infeasible[3], since they may contain time and frequency offsets. Even if we can compensate these offsets for $\hat{H}_{jt_0}$ based on the reference of $\hat{H}_{(j-1)t_0}$, certain errors may persist if they are not highly correlated.

It is noted that the power delay profiles (PDP) derived from the CSI directly reflects the similarity of radio propagation channels, which is suitable to be the indicator. Hence, the corresponding indicator can be calculated as

$$\Delta\hat{I}_{jt_0}^{\mathrm{PDP}} = \rho^{-1}(\hat{H}_{jt_0}, \hat{H}_{(j-1)t_0}) = \left\| \mathrm{P}(\hat{H}_{jt_0}) - \mathrm{P}(\hat{H}_{(j-1)t_0}) \right\|_2, \tag{11}$$

where $\rho(\hat{H}_{jt_0}, \hat{H}_{(j-1)t_0})$ is the PDP similarity as defined in [48], and a smaller value of the indicator indicates that $\hat{H}_{jt_0}$ and $\hat{H}_{(j-1)t_0}$ are more similar, suggesting that the user is likely to have undergone a small displacement. The indicator can also be integrated with other measurements like the IMU readings if available, which can be expressed as

$$\Delta\hat{I}_{jt_0}^{\mathrm{IMU}} = \mathrm{f}\left(\hat{X}_{(j-1)t_0}, \varepsilon_{(j-1)t_0}, e_{(j-1)t_0}\right), \tag{12}$$

where $\mathrm{f}(\cdot)$ is the state transition function, $\varepsilon_{(j-1)t_0}$ is the sensor readings related to the user motion between time $(j-1)t_0$ and jt_0, and $e_{(j-1)t_0}$ is the process noise. The similarity threshold S_{thr} can be determined through empirical findings concerning range, time, or ratios derived from literature or experimental tests. Depending on the relationship between the indicator and the threshold, the system can estimate the user's mobility profile.

Resource mapping. Subsequently, we map the mobility estimation to the resources that require combination, as represented by the combination window length w. Figure 9 presents an illustrative example, where the combination window vector is $\mathbf{w} = [1, 2, 1, 2, 3 \ldots]$, and a higher value refers to a lower mobility level. As a result, the combination window length is used as input for the CSI combination function, in conjunction with the original CSI samples.

We consolidate and outline the mobility estimation and resource mapping process in Appendix §B *Algorithm 1.*

4.4.2 *Offline mobility mapping*

For offline operation, where the 5G SRS measurements are stored for offline post processing. The system thus has full access to CSI traces. In contrast to the online mobility mapping scenario where mobility has to be estimated solely based on the current CSI measurements, in offline scenarios, the system gains full time series of CSI measurements for optimized mobility estimation.

[3]Here, we simplify the symbol of CSI measured in $((j-1)\,t_0, jt_0\,]$ by substituting subcarrier-OFDM symbol index $\tilde{k}, l$ as jt_0.

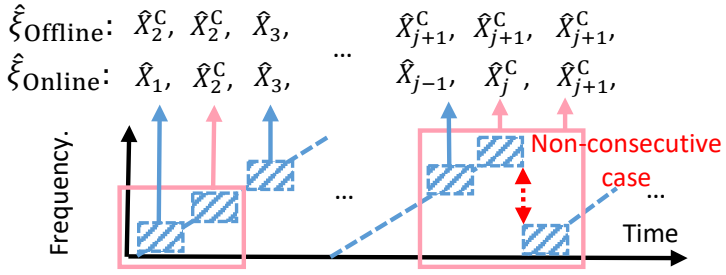

Fig. 10. Illustration of offline mobility mapping.

Taking the example of T original CSI samples with TG=t_0, by traversing combination window lengths ranges from t_0 to Tt_0 with plenty of computational resources, we can systematically assess and determine a proper spatio-temporal tradeoff. The above process can be expressed as

$$\mathbf{w}^* = \arg\min_{w_j \in \mathbf{w}} \text{STE}\left(\hat{\xi}^{u_i}_{\mathbf{w},T}\right), \tag{13}$$

where $\hat{\xi}_{\mathbf{w},T}$ is the reconstructed mobility trajectory with a combination window vector $\mathbf{w}$ and $w_j \in \{t_0, 2t_0, \ldots, Tt_0\}$.

Constraining the offline optimization problem. In scenarios characterized by dense user connections and long-time trajectory tracking, the number of possible combinations for $\mathbf{w}$ can grow exponentially. This results in a substantial computational burden that may not be alleviated even with the assistance of offline resources. Therefore, certain constraints for w_j must be taken into consideration. Besides the objective of reducing the computational overhead, another goal of constraining the problem is to fully exploit the consecutive CSI to increase the resolution.

Recall the mathematical bound formulated in formula (9). Valid values of w_j (i.e., the TG after CSI combination) are between T/N and T, while can be additionally constrained. We suggest, the available value of combined CSI in this range be divisible by N (or an integer multiple of N), because the combination of consecutive (overlapping) bandwidth can improve the resolution. Other types of combination like weight $(\cdot)$ may only benefit by reducing the error variance, as analyzed in formula (6). Taking $N = 4$ as an example, by setting $w_j = t_0$, $2t_0$ or $4t_0$, the system can use the function of concat $(\cdot)$, while $w_j = 3t_0$ will leave the last part CSI alone and can only perform weight $(\cdot)$ but cannot be concatenated with the next bandwidth part to increase the resolution. Based on this, we can add the constraint C1 : $Nt_0/w_j \in \mathbb{Z}^+$ to the optimization problem in (13).

5 EVALUATION

5.1 Experimental setup

Following the 5G standard, we implement SORA-Loc with complete 5G signal configuration and processing procedures in Xilinx UltraScale Radio Frequency System-on-Chip (RFSoC) hardware testbed [50]. As shown in Figure11, multiple users' SRS pilot signals are configured and generate the aggregated waveform along with the uplink shared channel, as defined by the 3GPP NR standard TS 38.211 [5]. Each radio frame comprises one OFDM symbol with SRS, on a subcarrier spacing of 30 KHz. Then, the received signals are demodulated to extract the corresponding OFDM symbols for CSI estimation. For the software part, we modify the codes from Matlab 5G Toolbox to continuously extract 5G CSI [36]. For the hardware part, we employ four radio chains of the RFSoC board as receivers (base stations) and additional four radio chains as the transmitters (user equipments). Base stations are well synchronized via a LMK04828 oven-controlled crystal oscillator (OCXO) with less than 1.5 ns synchronization error [21]. These devices are equipped with omni-directional antennas, low-noise-amplifier and low-loss LMR-195 coaxial cables. The central frequency is set to 2.7 GHz within 5G frequency range one (FR1,

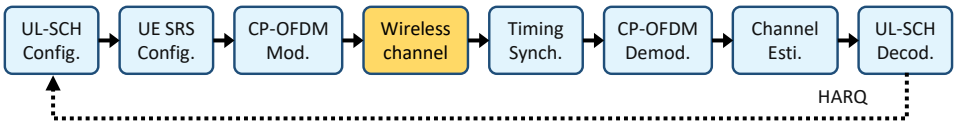

Fig. 11. Complete 5G signal generation and CSI estimation procedures performed on the testbed.

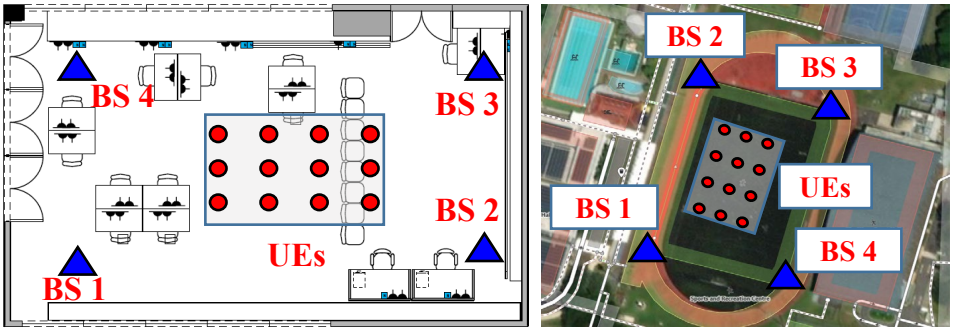

Fig. 12. Indoor and outdoor experiment scenarios.

sub-6GHz). All the data recorded at the BSs are transferred back via a PCIe 4.0×8 interface connecting the RFSoC to the x86-64 based server.

Figure 12(a) and (b) show our testbed deployments in indoor office and outdoor playground environments, spanning an area of 75 m^2 and 3600 m^2, respectively. Our objective is to evaluate SORA-Loc across diverse scenarios to illustrate its performance improvement. In indoor areas, the presence of nearby walls and furniture results in abundant multipath propagation, while outdoor areas are less affected.

5.2 Benefit of the comb-sliding multiplexing

We evaluate the STE performance gain induced by the proposed comb-sliding multiplexing scheme. Trace-driven simulation is conducted using the experiment data from Figure 6. We apply the localization error distribution, which has a zero mean and variances of 8, 4, and 2 m for cases that combine 25, 50 and 100 MHz bandwidth, respectively.

We consider a mobility trajectory including four time slots as $[X_1, X_1 + \Delta d, X_2, X_2 + \Delta d]$. As shown in Figure 13(a), for the comb-sliding case, the user measures 25 MHz of consecutive overlapped bandwidth in each slot. During the first and last two slots, the user moves within the effective combination range (i.e., $\Delta d \leq S_{\text{thr}}$). We vary the displacement between the second and third slots to assess the impact of increased mobility, denoted as $\Delta D = \|X_1 - X_2\|$, to 2, 4, and 8 m. The required TG is set at $2t_0$ and three benchmarks are conducted: "TG-limit" as in [3] (100 MHz in a single slot),"TG-limit with filter" as in [3, 42] (using a linear interpolation filter to mitigate the degraded TG) and "SA-limit" as in [47] (25 MHz in each slot).

Figure 13(b) shows that the proposed comb-sliding method achieves the lowest STE at $\Delta D = 4m$, marking a 40% and 70% improvement over the TG and SA-limit benchmarks, respectively. This improvement is attributed to SORA-Loc's efficient aggregation of two 50 MHz CSI in two dominant locations, X_1 and X_2. The TG-limit benchmark achieves the highest SA at X_1 but struggles to track subsequent mobility changes. In contrast, the SA-limit benchmark updates the user's location every slot but suffers from a lack of frequency diversity, which limits its ability to detect the direct path. Nevertheless, by applying the weight$(\cdot)$ operation to the SA-limit method, the system can reduce errors over time, outperforming scenarios without such combination. Figure 13(c)

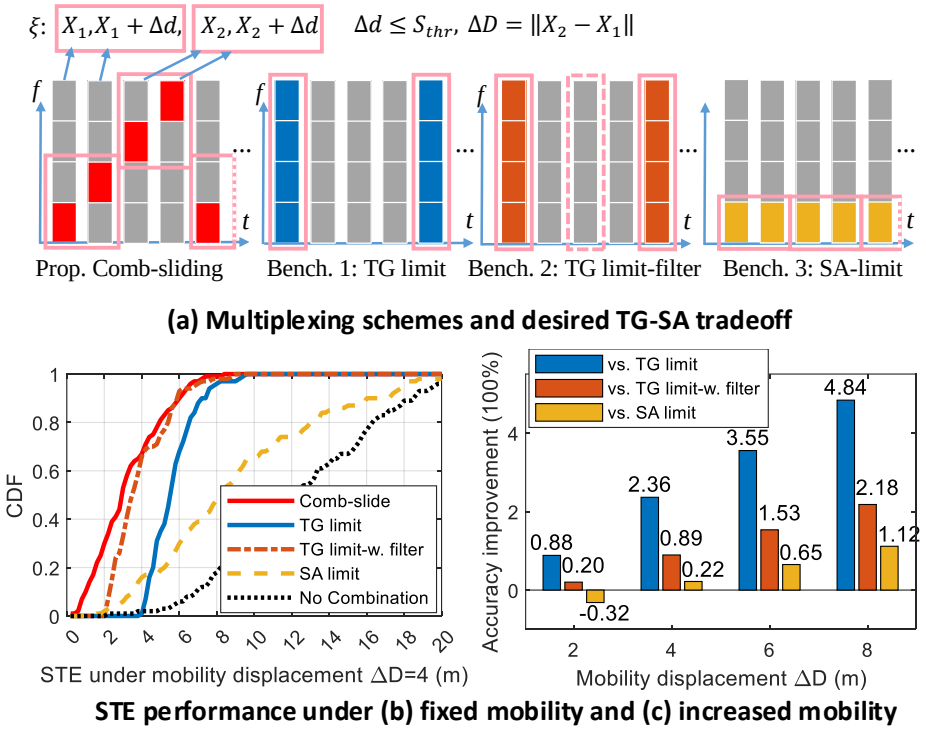

Fig. 13. STE performance influenced by different multiplexing schemes, where the TG-limit, SA-limit and filter-based benchmarks cannot make flexible tradeoff under diverse mobility.

shows that even with a linear filter, the performance of benchmark methods still degrades significantly up to 4.84x with the increase of mobility.

5.3 Performance of the mobility mapping

Leveraging comb-sliding multiplexing, SORA-Loc enables flexible CSI combination to meet diverse spatio-temporal tradeoff demands. We then assess the mobility mapping technique for its ability to adaptively select the appropriate combination based on user mobility.

Determination of the threshold for CSI combination. Building on the mobile experiments depicted in Figure 6, we delve into the benefits of CSI combination in Section (§4.2) and examine the changes in indicators that inform the mobility decision-making process. We employ the similarity of normalized PDP, alongside the Channel Impulse Response (CIR), as defined in formula (11), to serve as this indicator. Figure 14(a) and (b) reveal that the similarity of both indoor and outdoor CIR decreases as displacements increase, albeit at varying degrees depending on the wireless environment. The effective combination range in the outdoor scenarios (S_{thr} = 1.2m) is larger than that in the indoor case (0.4 m), because the outdoor playground is more spacious than the indoor office environment, leading to reduced multipath fading. In case of unavailable IMU, the CIR similarity decrease ratio S_{thr} = 30% and 66% can be set as thresholds, respectively. Thus, the similarity threshold can be applied in the decision step (6) in Algorithm 1. In the absence of empirical CSI coherence values, the above process can be applied to estimate the threshold for CSI combination.

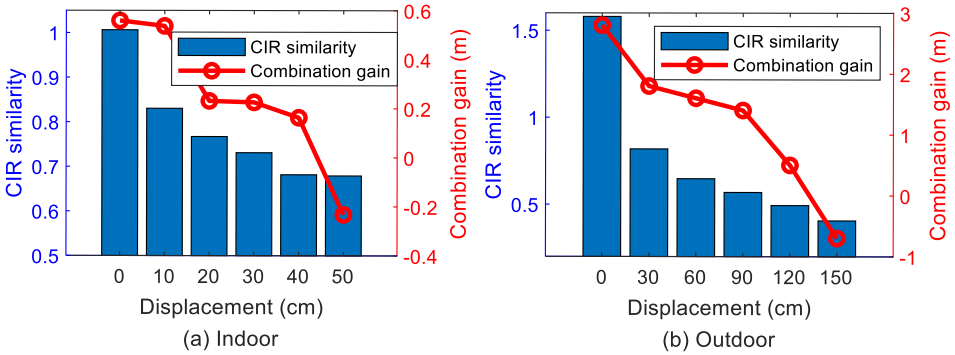

Fig. 14. Changes of the CIR similarity and corresponding CSI combination gains in indoor and outdoor scenarios along different displacement scales.

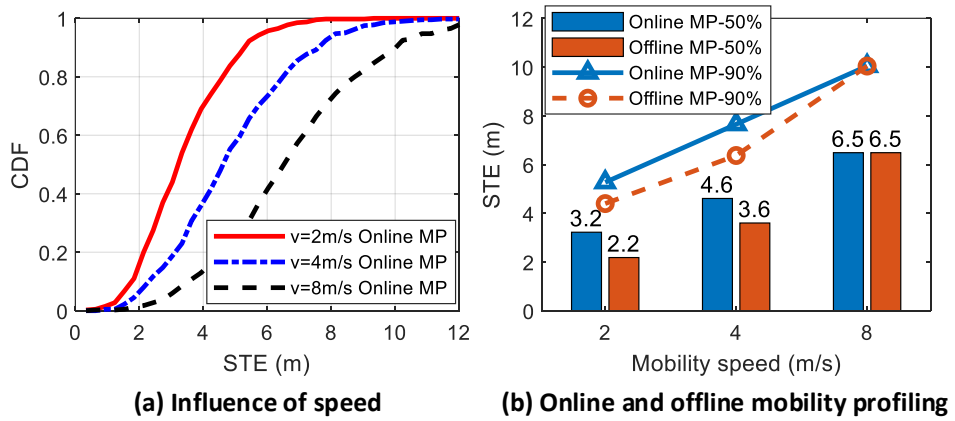

Fig. 15. STE vs. user speed using online and offline mobility mapping.

Effect of online and offline mobility mapping. The user's transmission period is configured at 20 ms to scan a 100 MHz bandwidth using four sequential 25 MHz subchannels, corresponding to movement speeds of 2, 4, and 8 m/s. This translates to a displacement of 0.4, 0.8, and 1.6 m, respectively, during each transmission period. Drawing from the effective combination ranges identified in indoor experiments, as depicted in Figure 14, the optimal combination window lengths are established as 4, 2, and 1, leading to effective bandwidths of 100, 50, and 25 MHz, respectively. We do not consider the case where $w = 3$, because this window length does not meet the constraint $Nt_0/w \in \mathbb{Z}^+$, leading to less efficient combinations due to non-consecutive bandwidths. Figure 15(a) shows the CDF of STE at various speeds using online mobility mapping. With a fixed transmission period, the STE performance deteriorates as mobility increases due to the reduction in combined CSI.

As analyzed in Section §4.4, online operation generates real-time locations based on the current reported CSI, combined with previous CSI when necessary. For the offline stage, the system leverages historical observations from the entire trajectory, enabling the use of optimal combinations to generate trajectories with improved STE. Figure 15(b) shows that the offline-based combination yields better STE in scenarios of lower mobility. In higher mobility cases, both the online and offline methods choose individual combinations, leading to negligible difference in STE performance.

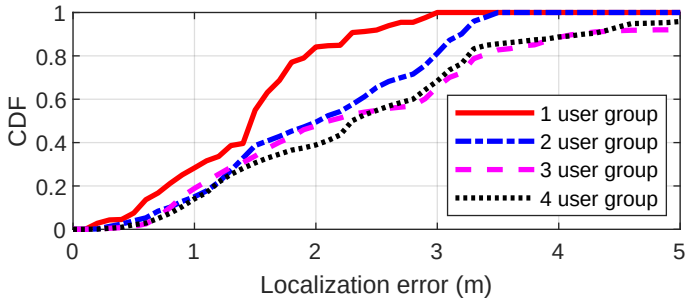

Fig. 16. Performance of the multi-user localization.

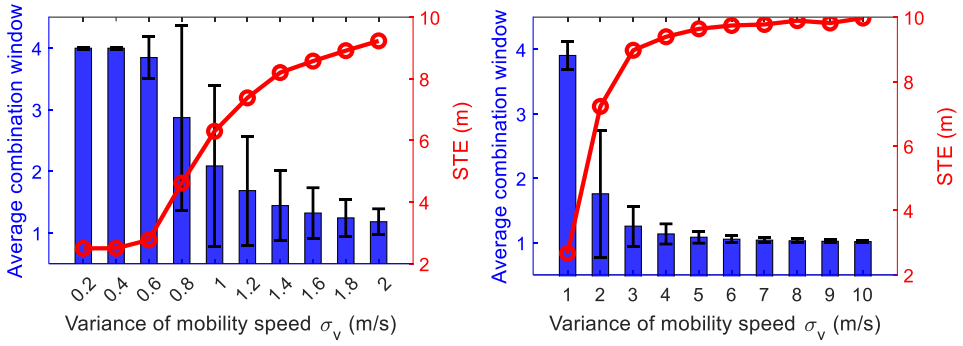

Fig. 17. Average combination window length and the corresponding median STE vs. user mobility. Error bars show the variance of the window length among the entire trajectory.

5.4 Performance of the multi-user localization

Experiment. We expand the single-user experiment to include multiple users in indoor scenarios, where each user employs the comb-sliding multiplexing scheme to transmit their SRS, with the center frequencies changing over time. Each device is assigned a unique frame tag and reference signal sequence, known to the network side [5], enabling the base station to extract the CSI for every user. As the number of connected users increases, the system allocates more subchannels to enable flexible and scalable CSI combinations. Consequently, the division of subchannels expands from 1 to 4, indicating the system's capacity to support groups of users ranging from 1× to 4×, updating their location with 100 MHz of frequency diversity. Figure 16 illustrates that with the increase of user groups, there is a slight decline in localization accuracy. This decline can be attributed to hardware imperfections such as frequency leakage and accumulated errors during the CSI combination process.

Trace-driven simulation at scale We reuse key parameters obtained from real-world experiments to build a simulator capable of evaluating SORA-Loc at scale. This simulator integrates data on effective combination ranges and error variances from various CSI combinations to facilitate large-scale simulations. It generates between 4 to 100 simultaneous users, each displaying different mobility patterns. The simulator creates trajectories based on random walks for indoor environments and vehicle movements for outdoor scenarios, covering 500 points. User speeds are modeled following a normal distribution $\mathcal{N}(0, \sigma_v^2)$, and orientation angles adhere to $\mathcal{N}(0, \sigma_a^2)$, with σ_a set to 90 and 45 degrees, respectively. We vary the value of σ_v to simulate different mobility scenarios. Utilizing the comb-sliding multiplexing scheme, we assign four consecutive subchannels per slot t_0, scaling the

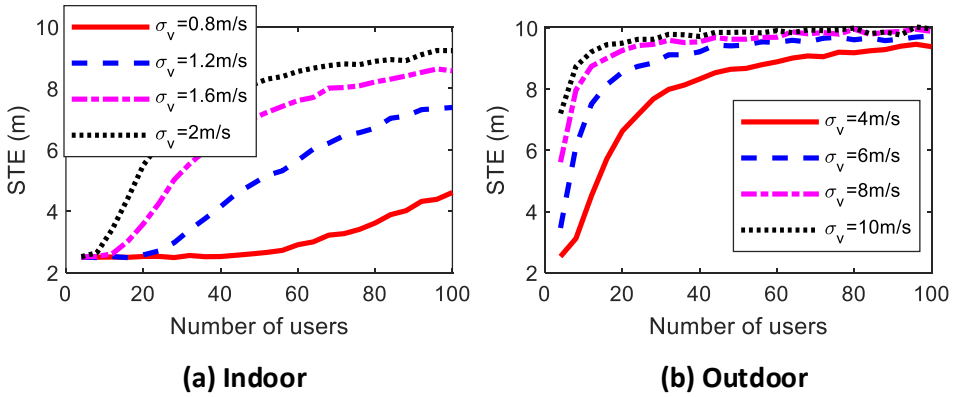

Fig. 18. STE vs. number of accessed users. Limited by fixed available resources, increased users in the time-extended multiplexing suffer from longer TG.

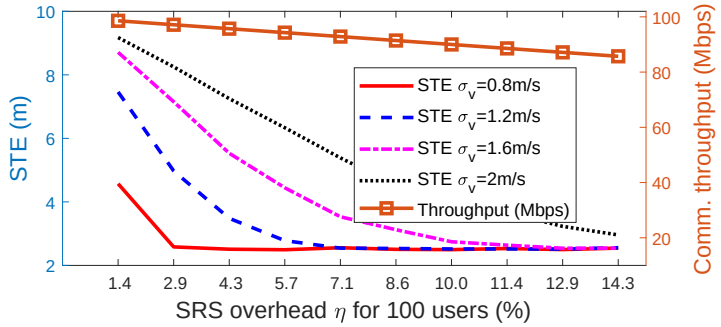

Fig. 19. System equivalent communication throughput and STE performance vs. increased pilot overhead when SNR=0 dB in indoor 100 users case.

number of users in each group to 25 for the time-extended assignment, as depicted in Figure 8(b). The original transmission period for a single user to scan the entire bandwidth is set at 20 ms (i.e., $t_0 = 5$ ms). We then apply mobility mapping method to correlate user mobility with the combination window length and assess the STE.

Figure 17 illustrates that the combination window length decreases with increased user mobility (indicated by the speed variance σ_v), as the system needs to update locations using the minimum time slot to keep pace with higher mobility. This leads to a reduction in resource availability for individual groups under the time-extended assignment, as shown in Figure 18. The STE performs adequately at lower levels of mobility but tends to worsen as mobility increases. It is important to note that these results are based on very limited and fixed SRS resource configurations. In the scenario where $t_0 = 5$ ms, the corresponding reference signal overhead is merely $\eta = 1/70$, in accordance with the 5G frame structure [5]. We then adjust the SRS pilot overhead η from 1 to 10 times to evaluate the system's localization capacity. Figure 19 indicates that allocating additional 12.86% pilot resources allows the system to support more than 100 users and accommodate higher mobility while maintaining the required STE around 2.9 m. The system equivalent communication throughput only drops by 14.1% and still achieves 85.71 Mbps at a typical SNR=0 dB, introducing an acceptable communication overhead.

6 RELATED WORKS

5G-dependent Localization. 5G localization is an emerging technology that has been standardized in 3GPP release 15 [32]. Release 16 further listed AOA [16], TDOA [30] and round-trip-time-based localization [24]. For reduced capability and limited observation cases, cell-identity can be used to get a rough position estimation [9]. To exploit the opportunities in 5G, such as larger bandwidth and dense small cells, ISLA [23] facilitates low-cost IoT devices to achieve higher range resolution by introducing a novel filter.

5G-independent Localization. Besides the cellular networks, numerous systems like IMU [15, 45], Wi-Fi [25, 49], RFID [28, 34], Bluetooth [18, 26], LoRa [8, 29] and GNSS [33, 38] have been exploited to enable efficient localization services. As the 3GPP standard suggests, a system can integrate both 5G-dependent and independent techniques to enhance robustness [3], exploiting scheme diversity to achieve continuous high accuracy localization [9]. Our design can be regarded as an enhanced 5G-dependent system and fusing with others to support various applications [37].

Resource allocation. For communication, traditional resource allocation schemes typically require dynamic power and sub-channel changes for improved throughput [27, 40, 55]. For localization, pilot transmission scheduling [20, 30] and data usage [38] have been studied to enhance the localization accuracy and latency [32]. However, these approaches generally rely on the re-configuration integrated into radio protocols, thus posing non-negligible overhead. SORA-Loc can be integrated in any existing resource allocation setup, which we go a step further by refining the CSI in the upper layer by software processing only.

CSI combination. Tone-track [52] and Splicer [48] combine CSI from multiple frequency bands to increase accuracy of the range-based localization [43]. CSI combination can also benefit the AOA [25, 51] and fingerprint-based localization [10, 11] with finer-grained information. M3 [13] extends the CSI combination to non-adjacent channels based on likelihood accumulation. These works focus on enabling low-cost device to sense a larger bandwidth and enhance the spatial accuracy. SORA-Loc is the first work to achieve a software-defined, spatio-temporal tradeoff based on CSI combination. Carrier aggregation combines distinct frequency resources, primarily with a focus on enhancing communication throughput and coverage [54].

7 CONCLUSION

This paper presents SORA-Loc, a software-defined resource allocation approach designed to achieve flexible spatio-temporal tradeoff for enhanced mobility trajectory estimation, triggering little protocol overhead. Our experiments on the hardware testbed and trace-driven simulation at scale demonstrate that, given fixed pilot resources in cellular networks, SORA-Loc outperforms several traditional localization benchmarks with reduced spatio-temporal errors. We believe this is a new efficient way to enable mobile location-based services for densely connected users, moving a critical step towards practical 5G localization.

REFERENCES

[1] 3GPP TR 22.872. 2018. Study on positioning use cases. (V16.1.0, Sept. 2018).
[2] 3GPP TR 38.801. 2017. Study on new radio access technology: Radio access architecture and interfaces. (V14.0.0, Mar. 2017).
[3] 3GPP TR 38.855. 2019. Study on NR positioning support. (v16.0.0, Mar. 2019).
[4] 3GPP TS 22.261. 2019. Service requirements for the 5G system. (V16.7.0, Mar. 2019).
[5] 3GPP TS 38.211. 2022. Physical channels and modulation. (V17.4.0, Dec. 2022).
[6] 3GPP TS 38.300. 2023. NR and NG-RAN Overall Description; Stage 2. (V17.5.0, Jun. 2023).
[7] 3GPP TS 38.455. 2021. NG-RAN; NR Positioning Protocol A (NRPPa). (Std., Rel. 16, July. 2021).
[8] Atul Bansal, Akshay Gadre, Vaibhav Singh, Anthony Rowe, Bob Iannucci, and Swarun Kumar. 2021. OwLL: Accurate LoRa Localization Using the TV Whitespaces. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)* (Nashville, TN, USA) *(IPSN '21)*. Association for Computing Machinery, New York, NY, USA, 148–162.
[9] Yang Cai and Yuan Shen. 2019. An integrated localization and control framework for multi-agent formation. *IEEE Transactions on Signal Processing* 67, 7 (2019), 1941–1956.

[10] Chen Chen, Yan Chen, Yi Han, Hung-Quoc Lai, and KJ Ray Liu. 2016. Achieving centimeter-accuracy indoor localization on WiFi platforms: A frequency hopping approach. *IEEE Internet of Things Journal* 4, 1 (2016), 111–121.
[11] Chen Chen, Yan Chen, Hung-Quoc Lai, Yi Han, and KJ Ray Liu. 2016. High accuracy indoor localization: A WiFi-based approach. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 6245–6249.
[12] Yunfei Chen, Jie Zhang, Wei Feng, and Mohamed-Slim Alouini. 2021. Radio sensing using 5G signals: Concepts, state of the art, and challenges. *IEEE Internet of Things Journal* 9, 2 (2021), 1037–1052.
[13] Zhe Chen, Guorong Zhu, Sulei Wang, Yuedong Xu, Jie Xiong, Jin Zhao, Jun Luo, and Xin Wang. 2019. M3: Multipath assisted Wi-Fi localization with a single access point. *IEEE Transactions on Mobile Computing* 20, 2 (2019), 588–602.
[14] José A. del Peral-Rosado, Ronald Raulefs, José A. López-Salcedo, and Gonzalo Seco-Granados. 2018. Survey of Cellular Mobile Radio Localization Methods: From 1G to 5G. *IEEE Communications Surveys & Tutorials* 20, 2 (2018), 1124–1148.
[15] Ashutosh Dhekne, Ayon Chakraborty, Karthikeyan Sundaresan, and Sampath Rangarajan. 2019. TrackIO: Tracking First Responders Inside-Out. In *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*. USENIX Association, Boston, MA, 751–764.
[16] Satyam Dwivedi, Ritesh Shreevastav, Florent Munier, Johannes Nygren, Iana Siomina, Yazid Lyazidi, Deep Shrestha, Gustav Lindmark, Per Ernström, Erik Stare, et al. 2021. Positioning in 5G networks. *IEEE Communications Magazine* 59, 11 (2021), 38–44.
[17] Andrea Goldsmith. 2005. *Wireless communications*. Cambridge university press.
[18] Ioannis Gouzouasis, Stylianos Papaharalabos, Mohamad Abou Nasa, and Peter Karlsson. 2023. A Novel Antenna System for Direction Finding Applications Using BLE 5.1 Technology. In *2023 17th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 1–5.
[19] Ali Hosseinghorban and Akash Kumar. 2022. A Partial-Reconfiguration-Enabled HW/SW Co-Design Benchmark for LTE Applications. *Electronics* 11, 7, 978 (2022).
[20] H. Huang, G. Gartner, J. M. Krisp, M. Raubal, and Nvd Weghe. 2018. Location based services: ongoing evolution and research agenda. *Journal of location based services* (2018).
[21] Texas Instruments. 2024. LMK04208 Low-Noise Clock Jitter Cleaner with Dual Loop PLLs datasheet. Retrieved Jan 20, 2024 from https://www.ti.com/product/LMK04208#features
[22] Ruqing Jin, Xiaofeng Zhong, and Shidong Zhou. 2016. The Access Procedure Design for Low Latency in 5G Cellular Network. In *2016 IEEE Globecom Workshops (GC Wkshps)*. 1–6.
[23] Suraj Jog, Junfeng Guan, Sohrab Madani, Ruochen Lu, Songbin Gong, Deepak Vasisht, and Haitham Hassanieh. 2022. Enabling IoT Self-Localization Using Ambient 5G Signals. In *19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22)*. 1011–1026.
[24] Ryan Keating, Mikko Säily, Jari Hulkkonen, and Juha Karjalainen. 2019. Overview of positioning in 5G new radio. In *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 320–324.
[25] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. 2015. SpotFi: Decimeter Level Localization Using WiFi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication* (London, United Kingdom) *(SIGCOMM '15)*. Association for Computing Machinery, New York, NY, USA, 269–282.
[26] Patrick Lazik, Niranjini Rajagopal, Oliver Shih, Bruno Sinopoli, and Anthony Rowe. 2015. ALPS: A Bluetooth and Ultrasound Platform for Mapping and Localization. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems* (Seoul, South Korea) *(SenSys '15)*. Association for Computing Machinery, New York, NY, USA, 73–84.
[27] Erran Li Li, Pal Martin, and Richard Yang Yang. 2008. Proportional Fairness in Multi-Rate Wireless LANs. In *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*. 1004–1012. https://doi.org/10.1109/INFOCOM.2008.154
[28] Bo Liang, Purui Wang, Renjie Zhao, Heyu Guo, Pengyu Zhang, Junchen Guo, Shunmin Zhu, Hongqiang Harry Liu, Xinyu Zhang, and Chenren Xu. 2023. RF-Chord: Towards Deployable RFID Localization System for Logistic Networks. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. 1783–1799.
[29] Jun Liu, Jiayao Gao, Sanjay Jha, and Wen Hu. 2021. Seirios: Leveraging Multiple Channels for LoRaWAN Indoor and Outdoor Localization. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking* (New Orleans, Louisiana) *(MobiCom '21)*. Association for Computing Machinery, New York, NY, USA, 656–669.
[30] Qirui Liu, Rongke Liu, Yang Zhang, Yanli Yuan, Zijie Wang, Haolan Yang, Lin Ye, Mohsen Guizani, and John S Thompson. 2023. Management of Positioning Functions in Cellular Networks for Time-Sensitive Transportation Applications. *IEEE Transactions on Intelligent Transportation Systems* (2023).
[31] Xiangnan Liu, Haijun Zhang, Keping Long, Arumugam Nallanathan, and Victor CM Leung. 2021. Energy efficient user association, resource allocation and caching deployment in fog radio access networks. *IEEE Transactions on Vehicular Technology* 71, 2 (2021), 1846–1856.
[32] Ying Liu, Xiufang Shi, Shibo He, and Zhiguo Shi. 2017. Prospective positioning architecture and technologies in 5G networks. *IEEE Network* 31, 6 (2017), 115–121.
[33] Debiao Lu and Eckehard Schnieder. 2015. Performance Evaluation of GNSS for Train Localization. *IEEE Transactions on Intelligent Transportation Systems* 16, 2 (2015), 1054–1059.

[34] Yunfei Ma, Nicholas Selby, Manish Singh, and Fadel Adib. 2017. Fine-Grained RFID Localization via Ultra-Wideband Emulation *(SIGCOMM Posters and Demos '17)*. Association for Computing Machinery, New York, NY, USA, 116–118.
[35] Sean F Mason, Christian R Berger, Shengli Zhou, and Peter Willett. 2008. Detection, synchronization, and Doppler scale estimation with multicarrier waveforms in underwater acoustic communication. *IEEE Journal on selected areas in communications* 26, 9 (2008), 1638–1649.
[36] MathWorks. 2023. Matlab 5G Toolbox. Retrieved March 2, 2023 from https://ww2.mathworks.cn/products/5g.html
[37] Flavio Morselli, Stefania Bartoletti, Moe Z Win, and Andrea Conti. 2021. Localization in 5G ecosystem with Wi-Fi. In *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 441–445.
[38] A.B.M. Musa, James Biagioni, and Jakob Eriksson. 2016. Trading Off Accuracy, Timeliness, and Uplink Usage in Online GPS Tracking. *IEEE Transactions on Mobile Computing* 15, 8 (2016), 2124–2136. https://doi.org/10.1109/TMC.2015.2457432
[39] Arvind Narayanan, Xumiao Zhang, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, Xiao Zhu, Xiaoxuan Zhang, Denis Rybkin, Zhengxuan Yang, Zhuoqing Morley Mao, Feng Qian, and Zhi-Li Zhang. 2021. A Variegated Look at 5G in the Wild: Performance, Power, and QoE Implications. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference* (Virtual Event, USA) *(SIGCOMM '21)*. Association for Computing Machinery, New York, NY, USA, 610–625.
[40] Cunhua Pan, Maged Elkashlan, Jiangzhou Wang, Jinhong Yuan, and Lajos Hanzo. 2018. User-centric C-RAN architecture for ultra-dense 5G networks: Challenges and methodologies. *IEEE Communications Magazine* 56, 6 (2018), 14–20.
[41] Jie Ren, Bradford D. Boyle, Gwanmo Ku, Steven Weber, and John MacLaren Walsh. 2016. Overhead Performance Tradeoffs—A Resource Allocation Perspective. *IEEE Transactions on Information Theory* 62, 6 (2016), 3243–3269.
[42] Louis L Scharf and Cédric Demeure. 1991. *Statistical signal processing: Detection, estimation, and time series analysis*. Vol. 63. Addison-Wesley Reading, MA.
[43] Souvik Sen, Jeongkeun Lee, Kyu-Han Kim, and Paul Congdon. 2013. Avoiding Multipath to Revive Inbuilding WiFi Localization *(MobiSys '13)*. Association for Computing Machinery, New York, NY, USA, 249–262.
[44] M. Sesia, M. Toufik, and M. Baker. 2009. LTE - The UMTS Long Term Evolution: From Theory to Practice, 2nd Edition. *Wiley Publishing* (2009).
[45] Guobin Shen, Zhuo Chen, Peichao Zhang, Thomas Moscibroda, and Yongguang Zhang. 2013. Walkie-Markie: Indoor Pathway Mapping Made Easy. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*. USENIX Association, Lombard, IL, 85–98.
[46] L. Valcarenghi, K. Kondepu, and P. Castoldi. 2016. Analytical and experimental evaluation of CPRI over Ethernet dynamic rate reconfiguration. In *2016 IEEE International Conference on Communications (ICC)*. 1–6.
[47] Zijie Wang, Rongke Liu, Qirui Liu, John S Thompson, and Michel Kadoch. 2019. Energy-efficient data collection and device positioning in UAV-assisted IoT. *IEEE Internet of Things Journal* 7, 2 (2019), 1122–1139.
[48] Yaxiong Xie, Zhenjiang Li, and Mo Li. 2015. Precise Power Delay Profiling with Commodity WiFi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (Paris, France) *(MobiCom '15)*. Association for Computing Machinery, New York, NY, USA, 53–64.
[49] Yaxiong Xie, Jie Xiong, Mo Li, and Kyle Jamieson. 2019. MD-Track: Leveraging Multi-Dimensionality for Passive Indoor Wi-Fi Tracking. In *The 25th Annual International Conference on Mobile Computing and Networking* (Los Cabos, Mexico) *(MobiCom '19)*. Association for Computing Machinery, New York, NY, USA, Article 8.
[50] AMD Xilinx. 2023. Zynq UltraScale RFSoC. Retrieved Feb 22, 2023 from https://www.xilinx.com/products/silicon-devices/soc/rfsoc.html
[51] Jie Xiong and Kyle Jamieson. 2013. ArrayTrack: A Fine-Grained Indoor Location System. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*. USENIX Association, Lombard, IL, 71–84.
[52] Jie Xiong, Karthikeyan Sundaresan, and Kyle Jamieson. 2015. ToneTrack: Leveraging Frequency-Agile Radios for Time-Based Indoor Wireless Localization *(MobiCom '15)*. Association for Computing Machinery, New York, NY, USA, 537–549.
[53] Dongzhu Xu, Anfu Zhou, Xinyu Zhang, Guixian Wang, Xi Liu, Congkai An, Yiming Shi, Liang Liu, and Huadong Ma. 2020. Understanding Operational 5G: A First Measurement Study on Its Coverage, Performance and Energy Consumption *(SIGCOMM '20)*. Association for Computing Machinery, New York, NY, USA, 479–494.
[54] Guangxiang Yuan, Xiang Zhang, Wenbo Wang, and Yang Yang. 2010. Carrier aggregation for LTE-advanced mobile communication systems. *IEEE communications Magazine* 48, 2 (2010), 88–93.
[55] Haijun Zhang, Yu Qiu, Xiaoli Chu, Keping Long, and Victor CM Leung. 2017. Fog radio access networks: Mobility management, interference mitigation, and resource optimization. *IEEE Wireless Communications* 24, 6 (2017), 120–127.
[56] Haijun Zhang, Yuxin Zhang, Xiangnan Liu, Chao Ren, Haojin Li, and Chen Sun. 2023. Time Allocation Approaches for a Perceptive Mobile Network Using Integration of Sensing and Communication. *IEEE Transactions on Wireless Communications* (2023).
[57] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. 2008. Understanding Mobility Based on GPS Data. In *Proceedings of the 10th International Conference on Ubiquitous Computing* (Seoul, Korea) *(UbiComp '08)*. Association for Computing Machinery, New York, NY, USA, 312–321.

Algorithm 1: Online mobility mapping

Input : CSI samples $[\hat{H}_{t_0}, \hat{H}_{2t_0}, ..., \hat{H}_{Tt_0}]$, start index of the combination $c_s \leftarrow 1$, similarity threshold S_{thr}
Output: Mobility profile represented by the combination window $\mathbf{w} = [w_1, \ldots, w_T]$, estimated trajectory $\hat{\xi}$

for $j \leftarrow 1$ ***to*** T **do**
  if $j == 1$ **then**
    Report $\hat{X}_{jt_0} = \text{loc}\left(\hat{H}_{jt_0}\right)$ to the LMF, add $\hat{X}_{jt_0}$ to $\hat{\xi}$;
  else
    Calculate the indicator changes $\Delta\hat{I}_{j-1}$ according to $\hat{H}_{jt_0}$ and $[\hat{H}_{(c_s)t_0}, ..., \hat{H}_{(j-1)t_0}]$;
    if $\Delta\hat{I}_{(j-1)t_0} \leqslant S_{thr}$ **then**
      `// Combine CSI to increase accuracy`
      $\hat{X}^{\text{C}}_{jt_0} = \text{loc}(\text{combine}(\hat{H}_{(c_s)t_0}, ..., \hat{H}_{jt_0}))$;
      Report $\hat{X}_{jt_0} = \hat{X}^{\text{C}}_{jt_0}$ and add $\hat{X}_{jt_0}$ to $\hat{\xi}$;
    else
      `// Process CSI individually`
      Report $\hat{X}_{jt_0} = \text{loc}(\hat{H}_{jt_0})$ and add $\hat{X}_{jt_0}$ to $\hat{\xi}$;
      `// Update the combination start point`
      $c_s \leftarrow j$;
  Combination window length $w_j \leftarrow (j - c_s + 1)$;

A COMPLEXITY ANALYSIS OF SORA-LOC

Taking the example of the time-extended comb-sliding multiplexing as proposed in Section 4.3, when dealing with N user group and K bandwidth parts (usually K is limited and $N \gg K$), SORA-Loc needs to perform $(K-1)\,(K-1)$ CSI combination and $(K-1)\,(K-1)$ mobility prediction operations for each user group. Therefore, the total computational load equals to $O\left(N \times 2 \times (K-1)\,(K-1)\right) = O\left(NK^2\right)$, leading to a significant reduction as compared with $O\left(N^K\right)$ brought by the traditional time-frequency resource allocation method.

B ONLINE MOBILITY MAPPING ALGORITHM

The detail of online mobility mapping process analyzed in §4.4 and also Figure 10 are summarized as Algorithm 1. The start index for CSI combination is denoted as c_s, where c_s to $(j-1)$ stands for the index of previous CSI that can be combined with the current CSI sample $\hat{H}_{jt_0}$ if $\Delta\hat{I}_{(j-1)t_0} \leq S_{\text{thr}}$. Otherwise, $\hat{H}_{jt_0}$ will be employed for location estimation individually and retained as backup data for potential combination with subsequent CSI samples. At each time instance jt_0 during the online stage, the system can provide an update location $\hat{X}_{jt_0}$ either from individual CSI samples $\hat{H}_{jt_0}$, or from the combination of previous CSI as $\hat{X}^{\text{C}}_{jt_0}$.

Received x May 2025; revised xx March 2009; accepted xx June 2009