SPSTS-Pos: AI-Driven Indoor Positioning with Strongest-Path Space-Time Structural Features in O-RAN System

Anpei Li[1], Luhan Wang[1], Na Li[2], Minchi Ruan[1], Xiaonan Wang, Zhaoming Lu[1], Raymond Knopp[3], *Fellow, IEEE*
[1]Beijing Key Laboratory of Network System Architecture and Convergence,
School of Information and Communication Engineering,
Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Baicells Technologies Co. Ltd.
[3]Communication Systems Department, EURECOM, Biot, France
Email: [1]lianpei@bupt.edu.cn, [1]wluhan@bupt.edu.cn, [2]lina@baicells.com, [1]ruanminchi@bupt.edu.cn,
xwan4931@uni.sydney.edu.au, [1]lzy0372@bupt.edu.cn, [3]raymond.knopp@eurecom.fr

***Abstract*—With the adoption of larger bandwidths and massive antennas, mobile cellular networks exhibit substantial potential for supporting positioning functionalities by leveraging shared hardware resources. Based on this architecture, this paper presents a high-precision indoor positioning method for multi-antenna 5G systems, termed as SPSTS (Strongest-Path Space-Time Structure). Unlike traditional geometric or fingerprint-based positioning approaches, SPSTS extracts relative time-delay differences of the strongest multipath component (MPC) across antenna pairs using uplink SRS signals. Based on the extracted SPSTS data, we have designed a customized ResNet architecture with asymmetric convolutions to decouple spatiotemporal features and regress them into 2D coordinates. The method is implemented on the O-RAN-compliant platform using an OpenAirInterface BBU, a commercial O-RU and distributed antennas, operating 100 MHz bandwidth in an indoor office environment. Real-world experiments achieve state-of-the-art positioning accuracy of 0.21 m (CDF50%) in offline evaluation and 0.35 m (CDF90%) in real-time testing. A video demonstrating the real-time positioning is available at www.opensource5g.org/demo/SPSTS-Positioning.mp4.**

***Index Terms*—Integrated Sensing and Communication, 5G Positioning, Sounding Reference Signal, ResNet, AI-RAN**

I. Introduction

In modern urban environments, indoor positioning plays a critical role in diverse application scenarios such as commercial shopping malls, hospitals, and smart factories. The rapid advancement of integrated communication and sensing technologies has opened new avenues for indoor positioning with the use of 5G cellular networks, which is now receiving significant attention. Moreover, 3GPP has incorporated reference-signal-based positioning capabilities since Release 16 [1], [2].

Currently, 5G-based cellular positioning technologies demonstrate notable advantages across multiple dimensions. First, the strict air-interface time synchronization of 5G networks ensures a highly consistent time reference among base stations. Second, the wider bandwidth and higher sampling rates of 5G facilitate more precise distance measurements. Third, Massive MIMO technology effectively exploits multipath propagation and spatial diversity, enhancing spatial resolution [3]. According to the 3GPP Release 18 specifications, the positioning target for 5G NR in indoor scenarios aims to achieve absolute horizontal error of less than 10 m and an absolute vertical error of less than 3 m at a 95% confidence level [4].

Existing 5G-based positioning methods can be broadly classified into two categories according to their underlying principles: geometric positioning and fingerprint matching. Geometric positioning relies on the physical properties of signal propagation and primarily includes trilateration, triangulation, and joint estimation algorithms. For instance, [5] proposed an E-CID-based positioning method and validated its performance using the OAI 5G air-interface system; [6] developed a 5G positioning data optimization algorithm based on crossing azimuth to enable reliable data cleaning . However, in complex indoor environments, multipath propagation is highly variable and easily influenced by obstacles, making it difficult to establish accurate and robust geometric models, with positioning errors often on the order of several meters.

Fingerprinting methods, on the other hand, match real-time measurements against a pre-constructed database (fingerprints). In the offline phase, a radio map is built by collecting parameters such as Received Signal Strength Indicator (RSSI) or Channel State Information (CSI). In the online phase, mobile devices continuously measure and compare observations to estimate positions. For example, [7] proposed FLIT, a multi-dimensional data fusion system for single-BS mode, achieving an average positioning error of only 6.9 cm in an office environment; [8] proposed a similarity-learning CSI fingerprinting framework using convolutional neural networks that integrates spatial- and frequency-domain information from multiple base stations; [9] presented an EKF-based indoor fusion positioning algorithm combining 5G carrier-phase rang-

ing and CSI fingerprinting . Nevertheless, RSSI and CSI inherently lack explicit geometric location information. Sparse sampling in the offline phase can lead to poor accuracy for unlabeled points, and fingerprinting typically requires dense data collection. Furthermore, fingerprints are susceptible to variations caused by multipath effects and moving objects [10].

Therefore, achieving higher-precision positioning using 5G networks remains challenging due to: (1) the highly dynamic and complex indoor multipath environment, where small position changes can cause large multipath variations, making deterministic geometric modeling difficult; (2) the prevalence of non-line-of-sight (NLOS) transmissions due to indoor obstacles, causing distance-based measurements to fail; and (3) the low spatial discrimination and heavy data requirements of fingerprinting-based methods.

To address the above issues, we conducted extensive experimental measurements based on OpenAirInterface. Our observations indicate that, although in many scenarios the strongest transmission path measured by multiple base stations or multiple antennas does not correspond to the line-of-sight path, it exhibits a stable structural pattern across different antennas (spatial domain) and different OFDM symbols (temporal domain). Based on this observation, we propose the strongest-path space-time structure feature based in-door positioning method. We firstly abstract Sounding Reference Signal (SRS) from multiple base stations or antennas. Then the propogation delay of the strongest path for each link by computing time-domain progation delay. Subsequently, we construct the so-called strongest-path space-time structure feature dataset. Lastly, a customized Residual Network (ResNet) [11] is proposed to infer UE's position. The main contributions of are summarized as follows:

- We propose a novel Space-Time Structural Feature (SPSTS) set derived from the strongest propagation paths, which circumvents the need for Line-of-Sight (LoS) conditions. The proposed iterative reference selection scheme enhances robustness against systemic bias.
- We introduce a sliding window mechanism that enables a high-frequency positioning output even with sparse SRS transmissions, significantly improving the temporal consistency and applicability of the feature set in practical scenarios.
- We develop a lightweight ResNet architecture featuring an asymmetric convolutional layer. This design effectively decouples the spatial and temporal characteristics of the SPSTS features, allowing the network to efficiently learn the complex mapping to 2D coordinates.
- We build and validate our method on an end-to-end experimental platform integrating OpenAirInterface (OAI) with a commercial O-RAN Radio Unit (O-RU). Real-world experiments demonstrate state-of-the-art performance, achieving a positioning accuracy of 0.35 m at the 90th percentile CDF in real-time tests.

The remainder of this paper is organized as follows: Section II details the experimental setup and SPSTS Feature data acquisition process; Section III presents the SPSTS Feature data preprocessing scheme and neural network architecture; Section IV reports experimental results and analysis; and Section V concludes the paper.

II. System Architecture and Feature Data Set Construction

This research leverages the open source 5G stack OpenAir-Interface, a commercial O-RU and OpenAirInterface UE to construct an end-to-end O-RAN system.

A. System Architecture

The overall system architecture, depicted in Fig. 1, comprises a positioning server, an OAI BBU, a commercial O-RU, and the associated RF antennas. In typical indoor deployments, multiple RUs are often utilized to ensure adequate coverage. To reduce experimental complexity while preserving experimental relevance, we emulate a multi-RU scenario by deploying the four antennas of a single commercial O-RU in a distributed manner, which is reasonable, as multiple RUs are synchronized to the same BBU and keep synchronized.

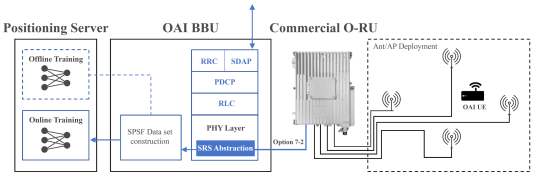

Fig. 1: System Architecture

The O-RU is connected to OAI BBU via the O-RAN Option 7-2 interface. We extract the uplink Sounding Reference Signals (SRS) from the four antenna streams within the OAI physical layer. These signals then undergo a processing pipeline, which includes channel estimation and strongest path extraction, to construct the final SPSTS Feature set. The Positioning Server is a GPU-equipped server responsible for both the offline training of the neural network and the real-time inference for positioning.

The OAI UE is employed as the positioning terminal. We equip it with an omnidirectional antenna to mitigate the impact of device posture. Key physical layer parameters are configured as follows: a subcarrier spacing (SCS) of 30 kHz, operation with a center frequency of 3.348 GHz, and a channel bandwidth of 100 MHz.

B. Sounding Reference Signal (SRS)

The 5G Sounding Reference Signal (SRS) is a reference signal transmitted by the User Equipment (UE) on the uplink in the 5G NR system. Its primary purpose is to enable the base station (gNB) to probe and measure the uplink channel state. The SRS is based on orthogonal Zadoff-Chu sequences and offers configurable transmission characteristics across the time, frequency, and spatial domains. Its allocation on the time-frequency resource grid is illustrated in Fig. 2, where the SRS is transmitted in a single OFDM symbol within a slot and is placed on interleaved subcarriers in the frequency domain.

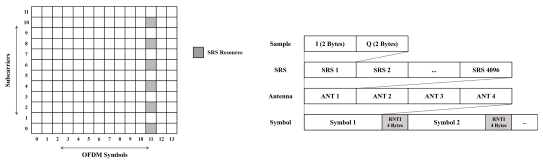

Fig. 2: Structure of SRS

The data structure of the SRS as extracted from the OAI protocol stack is depicted in Fig. 2. Data is collected and segmented in the time domain. Within each time slot, we extract the specific OFDM symbol designated for SRS transmission (located at symbol index 11). These extracted OFDM symbols, demarcated by the Radio Network Temporary Identifier (RNTI), are processed sequentially. Each OFDM symbol contains the received SRS data from all four antennas. For each antenna, this data comprises the complex SRS values from 4096 subcarriers, where each value consists of a 2-byte in-phase (I) and a 2-byte quadrature (Q) component.

C. SPSTS Feature Data Set Construction

The construction of the SPSTS Feature dataset involves four steps in Fig. 3: SRS signal extraction, channel estimation, time-domain CIR calculation, and strongest path difference calculation.

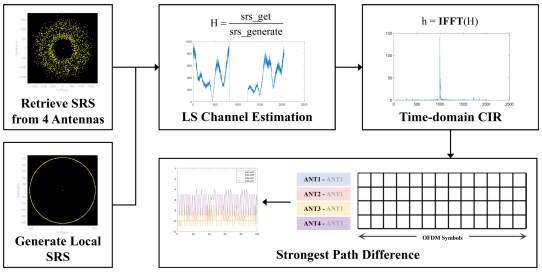

Fig. 3: Workflow of SPSTS feature calculation

First, the SRS reference signals are extracted from the designated resource elements in the uplink signal using the OAI protocol stack. Next, we perform Least Squares (LS) channel estimation by leveraging the known SRS sequence at the gNB and the received SRS signal, according to the formula $H = \frac{\text{srs}_{\text{get}}}{\text{srs}_{\text{generate}}}$. An IFFT is then applied to convert the frequency-domain channel response into the time-domain Channel Impulse Response (CIR). Finally, for each receive antenna, we identify the index of the peak corresponding to the strongest path in its CIR. The SPSTS data for a given OFDM symbol is then generated by calculating the difference between the peak index of each antenna and that of a reference antenna (e.g., antenna 1 in this paper).

A meticulous data acquisition procedure was implemented to ensure the independence and diversity of the measurements. The process was structured into three distinct rounds of collection across 10 representative measuring points. For each point, the UE would first establish a connection with the gNB, after which uplink signals spanning 2000 consecutive OFDM symbols were captured to compute and store the SPSTS features. To ensure spatial independence, the UE was manually disconnected and physically moved to the next location before re-initiating the collection process. A full sweep of all 10 points constituted a single collection round. To introduce temporal variation, a 30-minute waiting period was observed between the end of the first round and the beginning of the second. To further ensure hardware state independence, the gNB was completely powered down for 30 minutes after the second round before being rebooted for the third and final round. This comprehensive procedure yielded a total of 30 independent datasets (10 points 3 rounds).

III. Neural Network Design

A. Model Input Data Design

To enhance the influence of the temporal continuity of the SPSTS data during model training, we employ a sliding window mechanism to serialize the signal features as shown in Fig. 4. This approach constructs a time-correlated channel feature sequence, which improves the spatial distinguishability of the channel state by capturing the dynamic relationships between consecutive symbols.

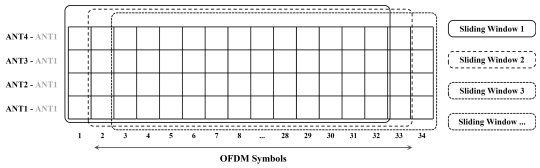

Fig. 4: Sliding window of SPSTS

The processing of each OFDM symbol produces a [1, 4] feature vector containing the relative SPSTS delays across the four receive antennas. To enhance spacial feature robustness, each antenna is sequentially designated as the reference, yielding an antenna tensor of shape [1, 16]. To capture temporal dynamics, we apply a sliding window of 32 symbols, advancing one symbol at a time. This procedure stacks 32 consecutive feature vectors, forming the final input tensor of shape [32, 16]. This tensor structure preserves the spatial information from the four antenna channels while encapsulating the channel's evolution over the 32-symbol window.

B. ResNet Model Design

We propose a lightweight convolutional neural network as shown in Fig. 5, which is a customized version of the classic ResNet-18, specifically adapted for our positioning task.

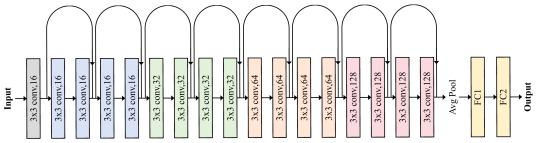

Fig. 5: ResNet model structure

The model, which consists of 17 convolutional layers and 2 fully connected layers, takes the [32, 16] SPSTS feature tensor as input. The input signal first passes through a 3x3 convolutional layer with an asymmetric stride for initial feature extraction and spatial downsampling. This is followed by four stacked residual blocks, each containing two units. The

channel dimensions progressively increase ($16 \rightarrow 32 \rightarrow 64 \rightarrow 128$) as the spatial resolution is halved, thereby enhancing the semantic representation. Finally, an adaptive global average pooling layer compresses the high-dimensional feature tensor into a 128-dimensional vector, which is then mapped to the final coordinate estimate by two fully connected layers. The entire network contains only about 0.3 million parameters, balancing computational efficiency with discriminative power.

The network is trained in a supervised manner using a loss function based on the Euclidean distance error. The optimization objective is to minimize the spatial deviation between the predicted coordinates $\hat{\mathbf{P}} = (\hat{x}_n, \hat{y}_n)$ and the ground-truth coordinates $\mathbf{P} = (x_n, y_n)$. The Mean Squared Error (MSE) is used as the loss function for a training set of N samples:

$$L = \frac{1}{N}\sum_{n=1}^{N}[(\hat{x}_n - x_n)^2 + (\hat{y}_n - y_n)^2] \tag{1}$$

This model is optimized for small SPSTS datasets and has demonstrated high accuracy in our experiments, even with a limited number of training samples.

IV. Implementation and Results

A. Performance Metrics

To comprehensively evaluate the performance of our proposed method, we employ a set of standard metrics based on the Euclidean distance error, which is the distance between the predicted coordinates (x_n, y_n) and the ground-truth coordinates for $(\hat{x}_n, \hat{y}_n)$each test sample i.

$$e_n = \sqrt{(x_n - \hat{x}_n)^2 + (y_n - \hat{y}_n)^2} \tag{2}$$

For accuracy metrics, we use the Mean Absolute Error (MAE) to represent the average positioning error and the Root Mean Square Error (RMSE), which is more sensitive to large deviations. They are defined as:

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N} e_n \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} e_n^2} \tag{4}$$

For stability metrics, we assess the consistency of our method using the Standard Deviation of the Error (STDE) and the worst-case performance using the Maximum Error (MAXE). They are defined as:

$$\text{STDE} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(e_n - \text{MAE})^2} \tag{5}$$

$$\text{MAXE} = max(e_n), n = 1, 2, 3...N \tag{6}$$

For statistical analysis, we utilize the Cumulative Distribution Function (CDF) to provide a holistic view of the error distribution. For a test set of N samples with corresponding Euclidean errors$e_1, e_2, ..., e_N$, the CDF, $F(e)$,is defined as the fraction of samples with an error equal to a value e:

$$F(e) = \frac{i}{N}, e_i \leq e \leq e_{i+1} \tag{7}$$

From the CDF, we specifically report the 50th percentile (CDF50) as the median error and the 90th percentile (CDF90) as a benchmark for high-reliability performance.They are defined as:

$$\text{CDF50\%} = F^{-1}(0.5) = e_{\lfloor 0.5*N \rfloor} \tag{8}$$

$$\text{CDF90\%} = F^{-1}(0.9) = e_{\lfloor 0.9*N \rfloor} \tag{9}$$

B. SPSTS Feature Stability Verification

The robustness of the SPSTS feature was validated through a multi-stage experiment designed to assess its stability over time and across link reconnections. At each of the 10 fixed locations, we performed a three-stage data collection: an initial baseline capture of 500 SRS symbols, a second capture after a 10-minute delay to assess temporal stability, and a third capture following a UE disconnection and reconnection to evaluate resilience to link state changes. For each stage, we computed the relative time delay $\Delta\tau_k = \tau_k - \tau_1, k = 1, 2, 3, 4$, where τ_k is the strongest path delay for antenna k.

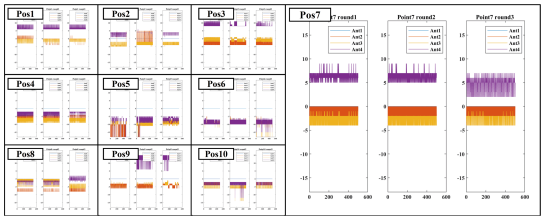

Fig. 6: SPSTS feature stability verifciation

Fig. 6illustrates the results of this experiment. As exemplified by the data from Point 7, the SPSTS features exhibit remarkable consistency. The subplots for Point 7 (initial, time-delayed, and post-reconnection) show that the relative time delays maintain their characteristic patterns. Specifically, after a 10-minute delay, the feature values remain stable: $\Delta\tau_2$ fluctuates within [-2, 0], $\Delta\tau_3$ within [-4, -2], and $\Delta\tau_4$ within [5, 7], with only occasional outliers. Even after a full link reconnection, the features show minimal deviation; for instance, $\Delta\tau_4$ settles into a new stable range of [2, 6], but its fundamental structure is preserved. This high degree of stability was consistently observed across all 10 test locations, confirming that the SPSTS forms a distinct and repeatable spatial fingerprint for each position.

These findings underscore the dual nature of the SPSTS feature's robustness. First, it is resilient to short-term environmental perturbations, as the dominant multipath geometry remains largely unchanged, ensuring sub-sample consistency in the relative delays. Second, it is invariant to link state changes, as the physical path-length differences between antennas are not altered by the reconnection process. This inherent stability

makes the SPSTS a reliable and powerful feature source for robust, learning-based positioning systems.

C. Performance Comparison

To demonstrate the efficacy of our proposed SPSTS feature, we conducted a comparative analysis against two alternative feature sets derived from the same raw SRS data. The first, termed CIR-256, uses the raw Channel Impulse Response (CIR) data, which contains more comprehensive channel information. The second, CIR-BMP, transforms the CIR data into a 2D image format, designed to leverage the visual processing strengths of ResNet architectures.

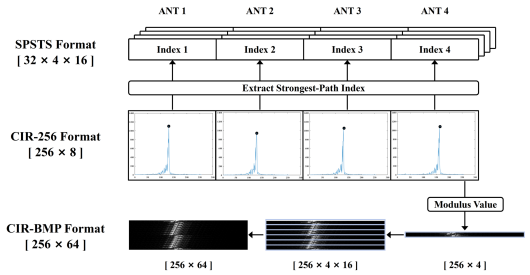

Fig. 7: Different feature-format structures

As Fig. 7 shows, the CIR-256 dataset is generated by applying a 2048-point IFFT to the frequency-domain channel estimate, yielding a high-resolution CIR. We extract the central 256 points, which contain the vast majority of the signal energy. Each complex sample is then split into its real and imaginary components, resulting in a [256 × 8] feature vector per OFDM symbol for the four antennas. The CIR-BMP dataset is created from this CIR-256 data; the modulus of each complex CIR sample is calculated, normalized, and then mapped to pixel brightness. A sliding window of 64 symbols is used to construct a [256 × 64] pixel image.

Using the data collected as described in Section II, we trained our ResNet model on all three datasets under identical hyperparameter configurations. The first two collection rounds served as the training set, with the third round used for validation. The performance metrics on the validation set are presented in Table I, and the corresponding CDF curves are shown in Fig. 8.

TABLE I: Performance Comparison of Feature Formats

Metric	SPSTS	CIR-256	CIR-BMP
MAE (m)	0.153	0.623	0.353
RMSE (m)	0.370	0.846	0.439
STDE (m)	0.336	0.572	0.262
MAXE (m)	2.743	3.900	2.358
CDF50 (m)	0.054	0.422	0.310
CDF90 (m)	0.228	1.416	0.706

From an accuracy perspective, the superiority of the SPSTS feature format is immediately apparent. The MAE and RMSE of SPSTS are lower than those of CIR-256 by 75.4% and 56.3%, respectively. From a stability perspective, the SPSTS feature also demonstrates more consistent performance. Its STDE is substantially lower than that of CIR-256. From a statistical perspective, the analysis further solidifies these findings. SPSTS achieves a breakthrough CDF90 of 0.228 m, outperforming CIR-256 by 84.2% and CIR-BMP by 67.7%.

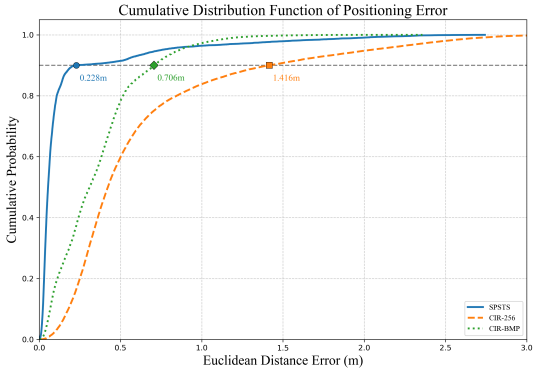

Fig. 8: CDF Comparison of different features

In summary, the SPSTS feature, through its spatio-temporal decoupling, enables highly efficient extraction of a discriminative positioning fingerprint in multipath environments, providing an optimal feature representation for sub-meter positioning.

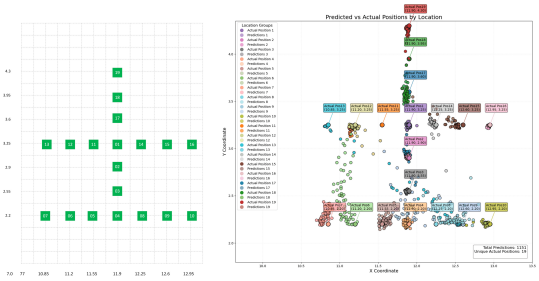

Fig. 9: Supplementary Experiment

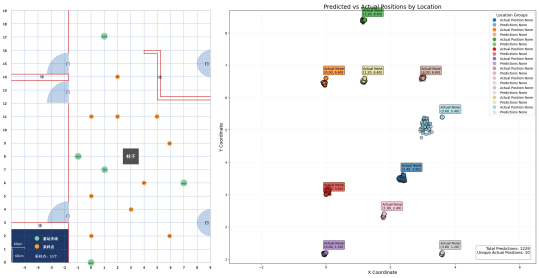

Fig. 10: Positioning result on validation set

To further assess the models generalization capability, we conducted an additional experiment in a different room featuring a finer and denser sampling grid, as illustrated in Fig. 9. Following the identical acquisition protocol, three rounds of SPSTS data were collected; the corresponding validation results are presented in Fig. 9.

D. Real-Time Demonstration

To validate the system's performance in a real-world operational environment, we deployed the complete positioning system in a 9.0 m × 11.4 m 5 m indoor office space (Fig. 10). The 2D layout of the site is illustrated in Fig. 10. The area is divided into a grid of 285 cells, each measuring

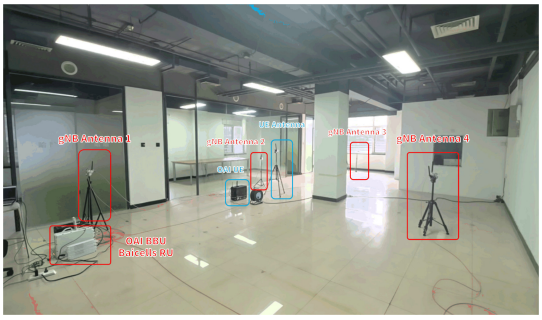

Fig. 11: Experimental scenario

60 cm 60 cm. The environment presents significant multipath propagation and Non-Line-of-Sight (NLOS) challenges. Four dual-polarized antennas (marked as ant1 to ant4) are deployed near the geometric center of the four walls at a height of 1 m to establish an omnidirectional coverage pattern. Strong reflectors, such as surrounding metal door frames and densely packed metal desk legs, create a rich multipath environment. During data collection, the UE was placed at 10 distinct sampling points (marked as orange circles) to capture a diverse range of channel conditions.

The real-time positioning workflow follows the architecture shown in Fig. 1: while maintaining a persistent connection, the system extracts SRS signals from uplink packets in real-time, generates the SPSTS features, and feeds them into the pre-trained ResNet model for position prediction. The final coordinates are then rendered on a visualization interface.

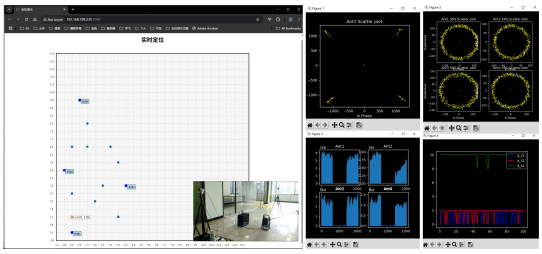

Fig. 12: Real-time Visualized Positioning System

Our visualization system consists of two main components: a real-time trajectory plot and a key parameter display. The trajectory plot (Fig. 12) clearly shows the movement of the UE. The parameter display (Fig. 12) provides insight into four critical metrics: the Original SRS Constellation Graph, whose ring-like shape confirms the correct reception of SRS signals; the SRS Spectrum, which allows for a qualitative observation of channel variations; the Compensated PUSCH Constellation Graph, which which shows the status of data plane transmission. and the SPSTS Feature plot, which is used to monitor the feature stability and the impact of environmental changes.

V. Conclusion

This paper proposed and validated a high-precision indoor positioning method for multi-antenna 5G systems based on a novel Strongest-Path Space-Time Structural (SPSTS) feature. The method constructs a spatial fingerprint by extracting the relative time delay differences of the strongest multipath component from uplink SRS signals and employs a lightweight ResNet with asymmetric convolutions to decouple the spatio-temporal features. On an O-RAN-based OAI testbed operating at 3.5 GHz with 100 MHz bandwidth, experimental results demonstrate that our method achieves breakthrough accuracy in a typical NLOS indoor environment. The offline evaluation yielded a MAE of 0.21 m and a CDF90 of 0.33 m, while the real-time test maintained a CDF90 performance of 0.35 m, an improvement of over 59.7% compared to traditional CIR-based methods. The SPSTS feature is proven to be highly robust against environmental changes and independent of link state, offering a new paradigm for sub-meter positioning in 5G-Advanced and 6G Integrated Sensing and Communication (ISAC) systems. Furthermore, we observed that using commercial smartphones introduces performance fluctuations, as device posture significantly affects antenna characteristics. Future work will focus on enhancing the system's robustness and consistency against this factor.

Acknowledgment

This work was supported by the National Key Research and Development Program of China under Grant 2024YFE0200300, the Beijing Natural Science Foundation under Grant L222003 and BUPT-SICE Excellent Student Creative Foundation under Grant yccx-2004-007.

References

[1] 3GPP, "Study on NR positioning support," 3rd Generation Partnership Project (3GPP), Tech. Rep. 38.855, 2019, version 16.0.0. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38series/38.855/38855-g00.zip.

[2] ——, "Physical channels and modulation," 3rd Generation Partnership Project (3GPP), Tech. Rep. 38.211, 2022, version 16.8.0. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38series/38.211/38211-g80.zip.

[3] L. Italiano, B. C. Tedeschini, M. Brambilla, H. Huang, M. Nicoli, and H. Wymeersch, "A tutorial on 5g positioning," *arXiv preprint arXiv:2311.10551*, 2023.

[4] 3GPP Working Group, "Service requirements for cyber-physical control applications in vertical domains," 3GPP Working Group, Sophia Antipolis, France, 2021, rep. TS 22.104 V18.3.0.

[5] D. Li, X. Chu, L. Wang, Z. Lu, S. Zhou, and X. Wen, "Performance evaluation of e-cid based positioning on oai 5g-nr testbed," in *2022 IEEE/CIC international conference on communications in china (ICCC)*. IEEE, 2022, pp. 832–837.

[6] B. Song and B. Wang, "A crossing azimuth based optimal base station selection algorithm for 5g positioning," in *2023 35th Chinese Control and Decision Conference (CCDC)*. IEEE, 2023, pp. 3724–3729.

[7] J. Du, L. Wang, Z. Lu, H. Niu, and X. Wen, "Flit: Multidimensional data fusion localization system based on vision transformer in single base station mode," in *2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2024, pp. 1–7.

[8] Q. Li, X. Liao, M. Liu, and S. Valaee, "Indoor localization based on csi fingerprint by siamese convolution neural network," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 12 168–12 173, 2021.

[9] X. Li, L. Chen, X. Zhou, Y. Ruan, and Z. Liu, "Fusion of 5g carrier phase ranging and 5g csi fingerprinting for indoor positioning," in *2022 14th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2022, pp. 1–5.

[10] M. T. Hoang, "Wifi fingerprinting based indoor localization with autonomous survey and machine learning," Ph.D. dissertation, 2020.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.