Joint Spatial/Temporal Measurement Prediction for L1/L2 Triggered Mobility

Izzet Emre Kucukkaya
Chair of Communication Networks
Technical University of Munich
Munich, Germany
Email: emre.kucukkaya@tum.de

Behnam Khodapanah
Nokia Standards
Munich, Germany
Email: behnam.khodapanah@nokia.com

Anna Prado
Chair of Communication Networks
Technical University of Munich
Munich, Germany
Email: anna.prado@tum.de

Fidan Mehmeti
Chair of Communication Networks
Technical University of Munich
Munich, Germany
Email: fidan.mehmeti@tum.de

Wolfgang Kellerer
Chair of Communication Networks
Technical University of Munich
Munich, Germany
Email: wolfgang.kellerer@tum.de

***Abstract*—The evolution of mobile networks toward 6th Generation of Cellular Network (6G) introduces new challenges for mobility management, especially in Ultra-Reliable Low-Latency Communication (URLLC) scenarios. Legacy handover mechanisms relying on higher-layer Radio Resource Control (RRC) signaling often suffer from high interruption times and suboptimal performance under high mobility. This paper investigates joint spatial-temporal forecasting methods for L1/L2 Triggered Mobility (LTM), exploiting temporal dependencies and spatial correlations of lower-layer measurements. We implement and evaluate statistical and Artificial Intelligence (AI)/Machine Learning (ML)-based forecasting models within a 6G system-level simulation framework. Results demonstrate that AI/ML models, particularly Long Short Term Memory (LSTM)-based architectures, significantly outperform the statistical Naive baseline in terms of prediction errors by up to 60%. Compared to the Release 18 LTM, AI/ML-based models exhibit a remarkable improvement in reducing signaling overhead by 80% while successfully predicting the similar number of handovers. The findings highlight the potential of predictive forecasting as a key enabler for next-generation mobility management.**

***Index Terms*—6G, Mobility Management, Handover, L1/L2 Triggered Mobility, LSTM.**

I. Introduction

The rapid growth of wireless systems has reached 5th Generation of Cellular Network (5G) deployments, with research now advancing toward the 6th Generation of Cellular Network (6G) [1]. Mobility management remains one of the most critical challenges in enabling Ultra-Reliable Low-Latency Communication (URLLC) and supporting mission-critical applications such as autonomous driving, industrial automation, and remote surgery [2], [3]. Conventional handover procedures, typically executed using Radio Resource Control (RRC) signaling, are prone to delays, service interruptions, and packet losses, especially under high-mobility scenarios [4], [5].

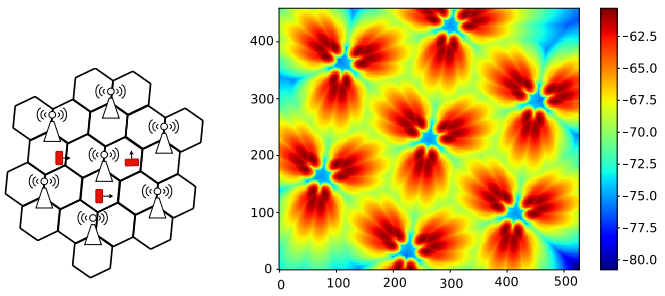
New mechanisms, such as Conditional Handover (CHO) and L1/L2 Triggered Mobility (LTM), have been standardized to reduce interruption time and signaling load [4], [5]. Both CHO and LTM use an A3 event to trigger Layer 3 (L3) measurement reporting for preparation of the candidate cells. A3 event is an event that is triggered when the measurement of the neighboring cell is better than the measurement of the serving cell by an offset. Based on L3 reports, the network decides to prepare the candidate cells if the requirements are met. After that, the network informs User Equipment (UE) about the pre-configured candidate cell configurations. In CHO, a UE compares the measurements directly without having further signaling from the network to make a handover decision. The handover is conducted without the need for another signaling with the serving cell. This procedure effectively reduces the handover interruption time by leveraging pre-configuration. However, since the exact timing of the handover is determined autonomously by the UE, the network lacks awareness of when it will occur, making coordination and resource management on the network side more challenging. In LTM, UE sends Layer 1 (L1) Reference Signal Received Power (RSRP) beam measurement reports to the network for those pre-configured candidate cells periodically. Until the network configures UE not to do that, UE continues these periodic measurement reports. Therefore, there exists a lot of messaging overhead in this situation. Moreover, these methods require making decisions early and accurately under uncertainty. Predictive mobility management, powered by Artificial Intelligence (AI) and Machine Learning (ML), yields a promising solution for decreasing the signaling overhead. In addition, based on the forecast of the future network condition, we believe we can prevent unnecessary handovers.

This paper focuses on spatial-temporal forecasting of mobility-related measurements for LTM, aiming to enhance prediction accuracy, reduce signaling overhead, and enable more proactive cell-switch decisions on a network illustrated in Fig. 1.

The contributions of this paper can be summarized as:

- We design and evaluate statistical and AI/ML models for

979-8-3195-0167-7/26/$31.00 ©2026 IEEE



(a) Cellular layout (b) Propagation map

Fig. 1: Simulation scenario with 21 cells across 7 Base Station (BS)s in Frequency Range 1 (FR1).

forecasting mobility-related measurements.

- We investigate the impact of temporal-only versus spatial-temporal approaches, incorporating UE velocity, position, and neighboring cell measurements.
- We analyze Key Performance Indicators (KPIs) such as signaling overhead, interruption time, and number of handovers, highlighting the trade-offs between model complexity and performance.

The remainder of this paper is organized as follows: In Section II, we provide some background information and describe related work highlighting what we do differently in this work. This is followed by Section III, in which we present the proposed algorithm in detail, including the dataset construction, forecasting models, online inference process, and the training parameters. In Section IV, we present results obtained using a system-level simulator of Nokia. Finally, Section v concludes this work.

II. Background and Related Work

In this section, the background of mobility and AI/ML based implementations in cellular networks is discussed.

Mobility management [6], [7] ensures seamless connectivity as UE moves across cells. Legacy handovers rely on L3 measurements (RSRP) and event triggers such as A3, and many studies have proposed approaches to enhance the traditional handover procedure [8], [9]. Introduced in 5G Release 16, CHO improves performance by pre-configuring candidate cells, while Release 18 further reduces delay by allowing lower-layer triggered cell switches using Medium Access Control - Control Element (MAC-CE) signaling. Utilizing lower-layer signaling decreases handover interruption time and increases handover performance [10].

A. Reference Signals and Measurements

In 5G/6G, UEs measure Synchronization Signal Block (SSB) beams[1] and report RSRP, Reference Signal Received Quality (RSRQ), and Signal to Interference and Noise Ratio (SINR). Filtering is performed at both L1 (fast but noisy) and L3 (slower but smoothed). These measurements drive mobility events such as A1–A6, which trigger handovers or candidate preparation.

The A3 event is one of the main events that is used to control the mobility scenarios. It is defined as "A neighbor cell is offset-plus-better than the Special Cell (SpCell). Special Cell is Primary Cell (Serving Cell) in single-connectivity or in Master Cell Group (MSG)". In Secondary Cell Group (SCG), it is Primary Secondary Cell (PSC). This triggers handover preparation, which can be expressed as

$$\mathrm{RSRP}_{\mathrm{neighbor}} > \mathrm{RSRP}_{\mathrm{serving}} + \mathrm{O}_{\mathrm{A3}} + \mathrm{H}_{\mathrm{A3}}, \qquad (1)$$

where O_{A3} is the A3 Offset and H_{A3} is the A3 Hysteresis. The offset defines how much better the target cell's measurement is compared to the serving cell, while the hysteresis is used to filter out small fluctuations in the measurements.

B. Forecasting Time Series

The rapid growth of the use of attention-based models [11] increased the capability of AI-based modeling. Even though the performance of the transformer-based models in time series forecasting is fascinating, there is research going on whether we really need those complicated models. Reference [12] compared the performance of complex transformer-based models and the relatively simpler linear models.

Time series forecasting methods have been applied in wireless communications for traffic prediction, load balancing, and mobility management [13]–[16]. In these works, the authors either forecasted only one L3 Cell RSRP sample and predicted the handover using that, or they used the observed measurements to predict the handover directly. In this paper, we focused on the L1 beam measurement to incorporate spatial information. Statistical models such as Auto Regressive Integrated Moving Average (ARIMA) and exponential smoothing provide baselines for the AI/ML-based models. AI/ML models, particularly recurrent neural networks and Long Short Term Memory (LSTM) [17] architectures, demonstrate superior performance in capturing temporal dependencies. Extending these models with spatial correlations (neighboring cell measurements, UE position) provides more accurate forecasts for mobility events.

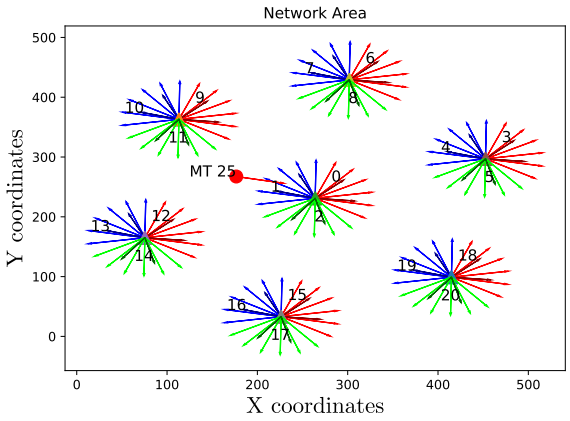
III. Proposed Algorithm

In this section, we describe the procedure for constructing the dataset used to train and evaluate the forecasting models. Furthermore, we present the forecasting algorithms and the online inference implementations considered in our study, along with their configurations and training parameters. Finally, we provide a discussion on deployment aspects.

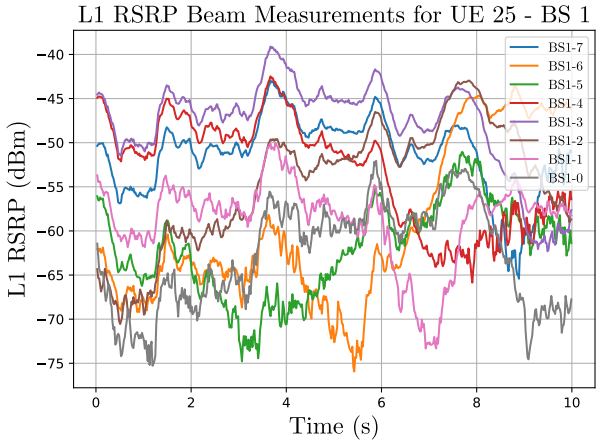
A. Dataset Construction

Datasets were generated by simulating UE mobility at varying speeds and trajectories. Observation Window (OW) is the size of the past input data samples for the AI model. Prediction Window (PW) is the size of the future output data samples forecasted by the AI model. In this work, both OW and PW are set to 16 time steps. The dataset was split into *training*, *validation*, and *testing subsets*. For training, 70%

[1]Judging from ongoing discussions among industrial players, SSBs will almost surely be part of 6G.



(a) UE 25 initial location, antenna orientation, and movement direction at start of the simulation

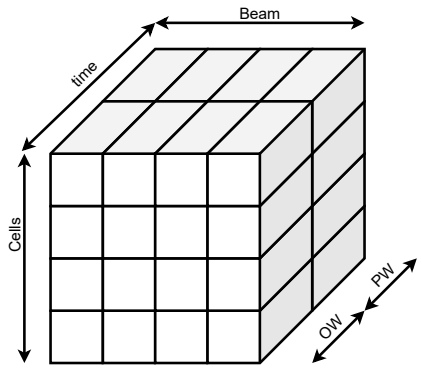


(b) L1 RSRP beam measurements of BS 1 at UE 25

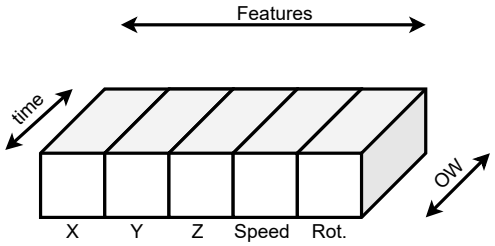
Fig. 2: A sample of a user trajectory and L1 RSRP beam measurements for UE 25.

of the UEs are used, whereas 10% of the UEs are used for validation. The remaining 20% of the UEs are used for testing. The train, validation, and test sets are constructed using the measurement of each UE group. Therefore, the data are not leaked through different datasets, even though the measurements are extracted within the same environment. The movement of the UE triggers fluctuations in the beam measurements, even in the case of a single cell. In Fig. 2a, the movement of a single UE (UE 25) is depicted, whereas in Fig.2b the L1 RSRP beam measurements of that user over time are shown. and measurement of a single UE are demonstrated. As can be seen from Fig. 2b, as the UE moves towards the BS, the beam that has the highest RSRP value changes. These fluctuations inherit the spatial information of that UE at that time.

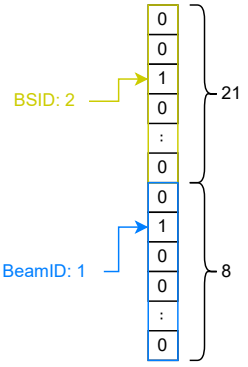
In these experiments, the simulation time is set to 10 seconds, and the SSB burst periodicity (the measurement



(a) RSRP Dataset



(b) Additional information of UE data



(c) One-hot embeddings of cell related Identifier (ID)s

Fig. 3: Dataset visualization.

periodicity) is chosen to be 20 ms.[2] This means that there are 499 data samples for each measured beam. In total, there are 147 UEs in the train set (70% of UEs) and for each UE there are 8 beams measured for the 21 BS. Hence, the number of OW-PW data pairs in training is $11,557,728$. The dataset structure is visualized in Fig. 3a. Moreover, additional information for each OW is also stored for the models that require this additional information. The UE-related information is stored as a time series; this can be observed

[2]These values were chosen as a trade-off between computation complexity and realistic behavior. Nevertheless, it is realistic to assume that the congested cells and slow moving UEs do not need extensively long simulations.

from Fig. 3b. The cell-related information is stored as a single vector, which is illustrated in Fig. 3c.

B. Forecasting Models

We experimented using statistical models, such as Naive forecaster, Moving Average, Exponential Smoothing, and ARIMA models, and the Naive forecaster outperformed the other statistical models. Therefore, the Naive forecaster is chosen as the baseline for our AI models.

In this paper, we investigate both Temporal-only and Spatial-Temporal Models, where LSTM is used as the backbone.

1) Temporal Forecasting: Temporal-only models utilize former observed measurements to forecast the future samples of the same observation window. The input of this model is considered as a single-time series. This single-time series does not inherit the relative position information of the UE to the BS or the angle information of the beams. Therefore, this modeling lacks the spatial information.

- **Statistical Models:** We utilized statistical models that only incorporate a single-time series and not exogenous variables, such as ARIMA, Moving Average, Exponential Smoothing, and Naive models. Based on our internal experiments, the Naive model outperforms all other statistical models, and it is chosen as the baseline for our AI/ML based models. The Naive model is a simple forecasting technique that assumes the value of the next time step is equal to the value of the current time step. Despite its simplicity, it often serves as a strong baseline, especially in scenarios where the data exhibits high autocorrelation or lacks complex patterns.
- **Temporal AI-ML Models:** LSTM is used as the backbone of our models. The architecture is illustrated in Fig. 4. For Temporal models, the input size of the LSTM (n) is set to 1. This value indicates the number of beams that are fed to the LSTM. There are 128 layers/units.[3] Sample normalization is used to normalize the measurement samples. Because the mean and standard deviation are also crucial for the forecasting, these values are concatenated with the output of the LSTM encoder. After that, four residual blocks are utilized to capture nonlinearities thanks to the non-linear activation function, Rectified Linear Unit (ReLU). The residual block is illustrated in Fig. 5. The output of the residual blocks is passed to the fully connected layer to reshape the output, and then, using the mean and standard deviation calculated before, the final predicted measurement samples are generated. Further, additional information is utilized to improve the model's performance. The model is illustrated in Fig. 6. Another LSTM block and a fully connected layer are utilized to encode the additional information. Then, the encoded inputs are concatenated.

[3]These hyperparameter values have been determined after comprehensive experimentation with different models and values.

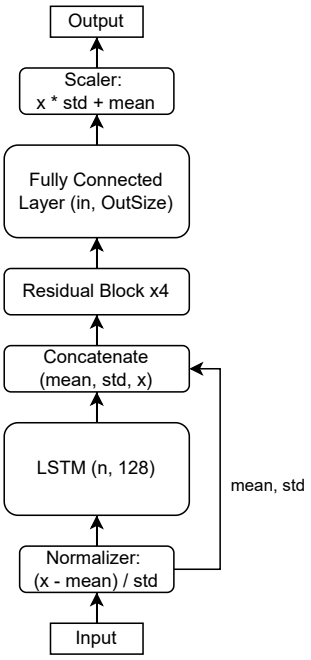


Fig. 4: AI/ML-based forecasting model.

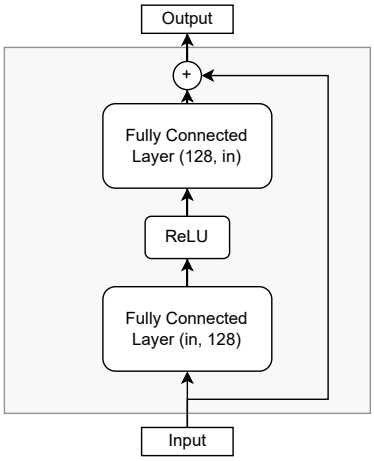


Fig. 5: Residual block.

- **Spatial-Temporal Forecasting:** Temporal-only models rely on historical RSRP measurements and UE-cell related information, while spatial-temporal models exploit not only UE position, velocity, but also neighboring beam and neighboring cell measurements. We designed both Single Cell Spatial-Temporal (SC ST) and MC ST models.
  - *Single Cell Spatial Temporal Models (SC ST):* The same architecture as the Temporal models is used. However, the number of beams (n) is increased to 8, which is the number of beams each cell has. Furthermore, the model that utilizes additional information

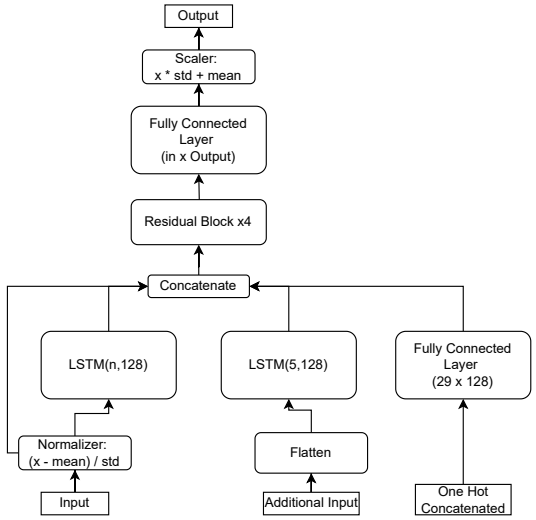


Fig. 6: AI/ML-based forecasting model that uses additional information.

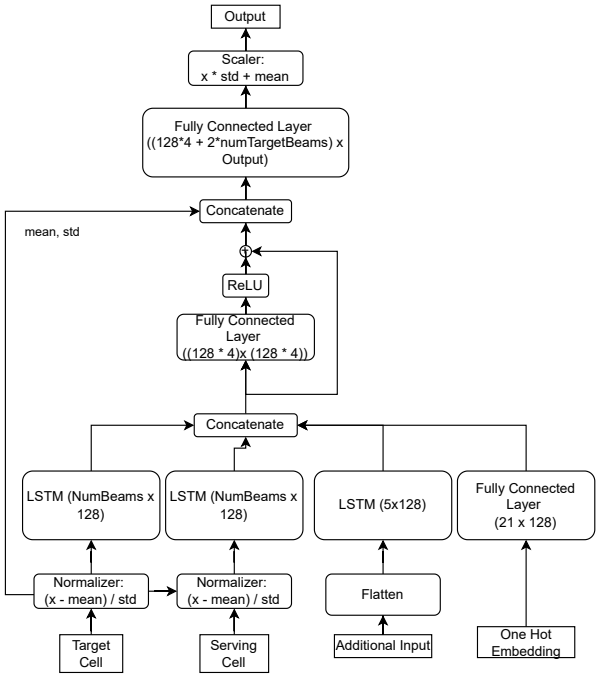


Fig. 7: AI/ML based Multiple Cell Spatial-Temporal (MC ST) forecasting model that uses additional information.

is implemented. Moreover, we implemented a variant of this model that forecasts all future 8 beams at once. For this beam, one hot embedding is not utilized. The model output consists of multiple beams in this scenario. In this scenario, the normalization is done for each beam separately, and all mean and standard deviations are concatenated.

– *Multiple Cell Spatial Temporal Models (MC ST):* The model architecture is similar to the SC ST models. The difference stems from having serving cell beams in addition to the target cell beams as the input. Another LSTM block is utilized with the same size as the LSTM block for the target cell beams. This model is trained on samples where a UE has a serving cell. This model architecture, which utilizes additional information as well as the serving cell beam measurements, is illustrated in Fig. 7. There are also variations of this model where additional information is utilized, and all 8 beams of the target cell are forecasted at once (multiple beams).

C. Online Inference

The forecasting models were integrated into the simulation for real-time inference. LTM decisions were made based on predicted measurements, enabling proactive mobility management.

The AI/ML based LTM execution signaling chart is illustrated in Fig. 8. UE stores the measurements and forecasts the future samples. Then, the UE creates a L1 measurement report, which consists of forecasted samples. Based on those future samples, the network decides when the cell switch occurs. Moreover, the decision window is implemented. The network decides to switch the cell only if the next N consecutive samples of the target cell are better than the serving cell on those time steps, where N is the decision window size. This implementation is similar to the Time-To-Trigger (TTT).

D. Training Parameters

The models are implemented using PyTorch, and the PyTorch implementation of Adaptive Moment Estimation (ADAM) optimizer [18], is used in training. Batch sizes are set to 2048 for Temporal models and 256 for Spatial Temporal models. Mean Squared Error (MSE) loss function is used, where the averaging is done through all prediction windows, beams, cells, and UE combinations. This is written as

$$L_{\text{MSE}} = \frac{1}{P \cdot B \cdot C \cdot U} \sum_{p=1}^{P} \sum_{b=1}^{B} \sum_{c=1}^{C} \sum_{u=1}^{U} \left(Y_{p,b,c,u} - \hat{Y}_{p,b,c,u} \right)^2, \tag{2}$$

where P is the PW size, B is the number of beams for a single cell, C is the number of cells, U is the number of UEs, $Y_{p,b,c,u}$ is the actual RSRP value, and $\hat{Y}_{p,b,c,u}$ is the predicted RSRP value. The models are trained through 25 epochs. We use early stopping, where the training stops if the model does not improve after 5 epochs, based on its performance on the validation data.

E. Deployment Considerations

In principle, the L1 measurement predictions can be performed at the UE side or at the network side. The UE-sided models have access to delay-free L1 measurements. Moreover, the UE's model can have access to reference signal measurements that may not be configured to be reported. On the other hand, running the AI/ML models at the UE side incurs excess memory, computation resources, and energy consumption, which is much more constrained at the UE side compared to the network side implementation. For the

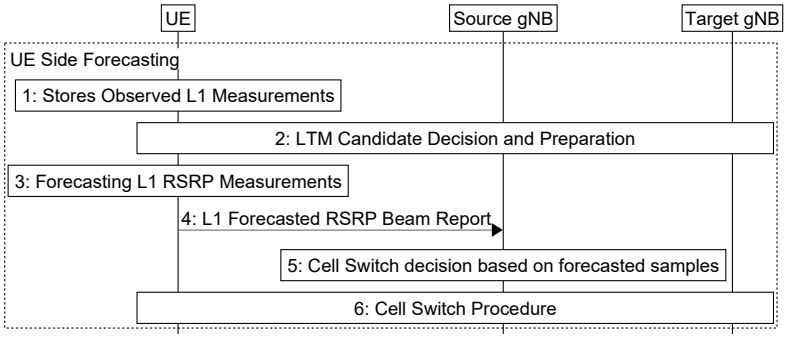
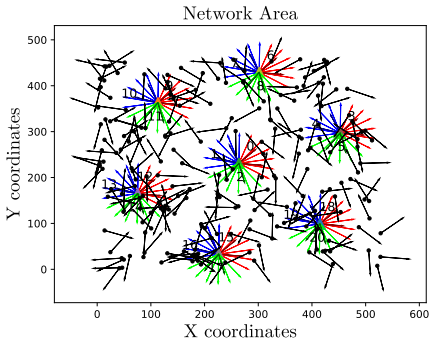
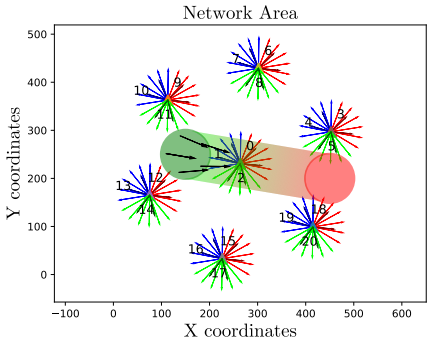


Fig. 8: Signaling chart of the AI/ML-based LTM approach.



(a) Randomized Path



(b) Street Path

Fig. 9: Considered mobility scenarios.

network-sided models, the UE has to perform measurement and reporting and the inference takes place at the network side. This limits the measurements that the model receives, as it is not feasible for the UE to report every measurement. Additionally, the measurement reports have to be transmitted over the air, which inevitably will introduce delays.

For the network-controlled mobility, the predicted future measurements should be available at the source node, regardless of the UE- or network-based models. However, for conditional mobility, where the UE decides when to perform the handover, only the UE-sided model is useful. In those scenarios, the network should be able to verify the performance of the AI/ML model extensively to avoid call drops and mobility failures. In later generations of the cellular networks, the specifications of the Life Cycle Management (LCM) of the AI/ML models will facilitate the data collection, training, compliance, validation, and performance monitoring. Moreover, the topic of using AI/ML for measurement prediction is currently being discussed in initial 3GPP meetings for 6G in terms of for inter-frequency predictions, L3 cell-level prediction, CSI-RS predictions, etc.

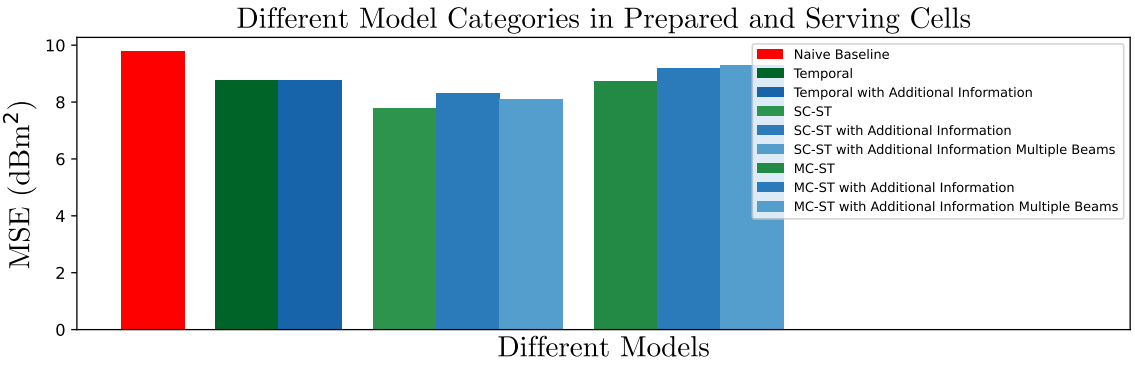
IV. Performance Evaluation

The evaluation of forecasting models was conducted under both the Randomized Path and Street Path mobility scenarios. We compare statistical baselines against AI/ML-based models, with a particular focus on LSTM-based spatial-temporal architectures. Moreover, other AI/ML-based models with linear backbones and transformer backbones were also tested. In this work, due to space limitations, we present only the best-performing models, which are the models with LSTM backbones.

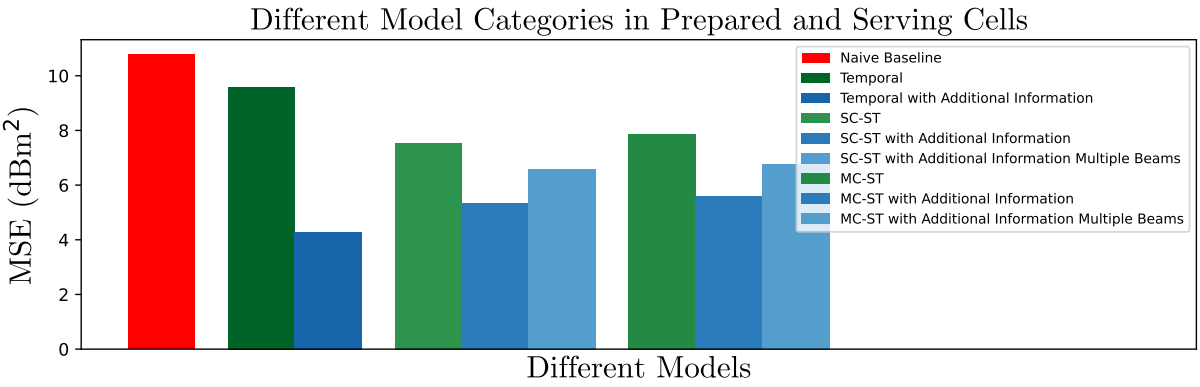
A. System-Level Simulation

The simulations are performed in a 6G System-Level Simulator, which is developed by Nokia. The details of the simulator can be found in [19]. The simulator incorporates reference signal measurement modeling, user trajectory generation, and lower-layer signaling architecture. On top of it, a dataset-generating framework is implemented to gather realistic measurement datasets across a multi-cell layout in different mobility scenarios. This simulator also includes realistic signaling models for L3 messages, while L1 reports and MAC-CE cell switch commands are transmitted ideally without any delay. The simulation is configured to operate in FR1 region with a central frequency of 4 GHz, carrier bandwidth of 10 MHz, and Subcarrier Spacing (SCS) of 15 kHz. UE employs an isotropic antenna with a maximum transmission power of 23 dBm, while BSs use 2×2 MIMO $2\times4\times4$ antenna panel configuration with 8 beams and 44 dBm/bandwidth.

1) Network Configuration: In a 3rd Generation Partnership Project (3GPP)-compliant urban macro scenario with a 200-meter inter-site distance in the FR1 band, we evaluated the network performance. The setup includes 21 cells across seven BSs, each with three sectors, and every cell is equipped with 8 beams, each one of them targeting different azimuth/elevation angles. The network setup that is used is demonstrated in Fig. 1a. In Fig. 1b, the propagation map is visualized based



(a) Randomized Path



(b) Street Path

Fig. 10: Forecasting performance of models across different mobility scenarios. The loss is used as mean squared error to penalize the larger differences in training of AI models. The error is calculated as the (linear) difference of forecasted and actual levels.

on the beam propagation measurements in the simulation environment. For slow fading, spatially correlated log-normal fading is used [20]. For channel model, 3GPP 38.901 is used [21]. Furthermore, the box scenario is used instead of the wrap-around scenario [22]. The simulation is run for 10 seconds. The SSB burst periodicity, i.e., the measurement periodicity is set to 20 ms. In each simulation, there are 21 UEs. The fading randomization is kept similar, and this simulation configuration was run 10 times with different mobility randomization. Therefore, in total, there are 210 UEs moving in the same environment in total, which experience the same fading effects.

2) Mobility Scenarios: Each UE moves with a constant speed with a predetermined orientation. UE antenna orientations are also predetermined randomly at the start of the simulation. UE speed is randomly chosen at the start of the simulation from one of these three values: 30 km/h, 60 km/h, and 90 km/h.

- **Randomized Path:** In this scenario, UEs are placed at completely random locations, and each UE follows a randomly generated path. The movement is linear with a constant speed along the assigned path. This type of mobility represents highly diverse and uncoordinated movement patterns. An illustration of Randomized Path Mobility is shown in Fig. 9a.
- **Street Path:** To better reflect real-world conditions where UEs tend to cluster around specific areas and follow common routes, the Street Path mobility model is introduced. Each UE starts from a randomly selected point within a designated start circle and moves toward a randomly chosen target point within a target circle. Although UEs share the same general route (e.g., a street), each UE has a unique path and starting position. As shown in Fig. 9b, UEs originate from the green circle (radius 50 m) and travel toward the red circle (radius 50 m). The center of the target circle is located at the position (450, 200).

B. Forecasting Performance

Fig. 10 illustrates the MSE across different model categories. Statistical baselines such as ARIMA and Exponential Smoothing [23] consistently underperformed, with the Naive predictor serving as the most competitive among them in our internal experiments. Naive model outperforms the ARIMA(4,1,5) model by more than 50% in terms of MSE loss. We utilized the Exponential Smoothing model with parameter α of 0.6, 0.8, and 0.9. The best performing value was 0.9, and it performed similarly to the Naive model (as expected). Therefore, as stated before, the Naive Model is used as

the baseline in the further experiments, where we evaluate mobility-related KPIs. Darts' implementations of these models are used [23]. Darts is a Python library for forecasting and anomaly detection. It contains models from classics such as ARIMA to deep neural networks. In contrast, AI/ML based models reduced MSE by at least 30% compared to the Naive baseline. Including additional input features (UE position, speed, and the antenna orientation) further decreased the loss up to 50%, highlighting the importance of contextual information such as the UE position and velocity, as can be observed from Fig. 10b. Spatial-temporal models (SC ST and MC ST) achieved a lower prediction error than the Temporal models. In Fig. 10a, these models outperform Temporal models by 8%, the models with additional information 2%, and the Naive baseline by 17%. As can be seen from Fig. 10a, the effect of the additional information is highly effective in the scenarios where the randomness is reduced. For the Street Path, the prediction error is 50% smaller than for Randomized Path mobility (for comparing the Temporal with Additional Information performance, refer to Fig. 10), which results in fewer ping-pong handovers. The handover rate and the number of reported beams are similar across both scenarios because of the network setup. Moreover, even though forecasting all target beams at once decreases the inference time, the loss is higher by approximately 7% than the single beam predictors.

Moreover, we tested different OW and PW sizes in internal experiments. We observed that the increase in PW size increases the averaged error as expected. However, increasing the OW size does not decrease the averaged error. On the contrary, it is slightly increased (by up to 5%). The reason is the same as why the Naive model outperformed the other statistical baselines – *the latest samples inherit more information*. As the size is increased, the unnecessary samples are used, which have the potential to confuse and mislead the models. This explains why we set both OW and PW to the last 8 measurements.

C. Impact on LTM Key Performance Indicators

The integration of forecasting into the LTM framework was evaluated through multiple KPIs: the average number of handovers per UE per second (ratio of number of handovers over total time of active UEs), average number of L1 measurement reports per UE per second (ratio of number of L1 measurement reports over total time of active UEs), and average number of Ping Pong (PP) handovers. PP handovers refer to the handovers that occur back to the previous serving cell within 1 second. For example, after a handover from cell A to cell B, if another handover takes place within 1 second from cell B back to cell A, this second handover is classified as a PP handover.

From Fig. 10, we can observe that Temporal model with Additional Information (T-Add), which is a model in which we incorporate spatial information as an additional vector, outperforms the other models especially for the Street Path scenario. Therefore, in this section we compare only R18 LTM, Naive baseline, Temporal, and Temporal with Additional Information due to space limitations. As shown in Fig. 11a, even though the Release 18 LTM implementation triggered a similar number of handovers to the other forecasting models, the number of L1 reports decreased by more than 80% compared to the Temporal with additional information model, resulting in higher signaling overhead when using R18 LTM. In contrast, AI/ML-based forecasting models substantially reduced the number of L1 reports by predicting a similar number of handovers. Specifically, Temporal with additional information achieved more than 80% reduction in L1 reports compared to the R18 LTM, while maintaining a similar number of handovers.

Moreover, the averaged forecasting error has a direct impact on the mobility-related KPIs. We observe the same trend in the model performances, both in terms of forecasting error and mobility-related KPIs. Furthermore, we investigated KPIs across different mobility scenarios. As shown in Fig. 11a, the Temporal with Additional Information models result in an increased number of PP handovers. This is because of the poor forecasting performance in the Randomized Path mobility scenario. In contrast, the Temporal model reduces the number of PPs by 25% and the number of measurement reports by 80% compared to R18 LTM, although it performs similarly to the Naive model in this scenario. For the Street Path mobility scenario, shown in Fig. 11b, the AI/ML-based models achieve 62% fewer PP handovers compared to R18 LTM. These results demonstrate that predictive models enable more proactive and stable handover decisions.

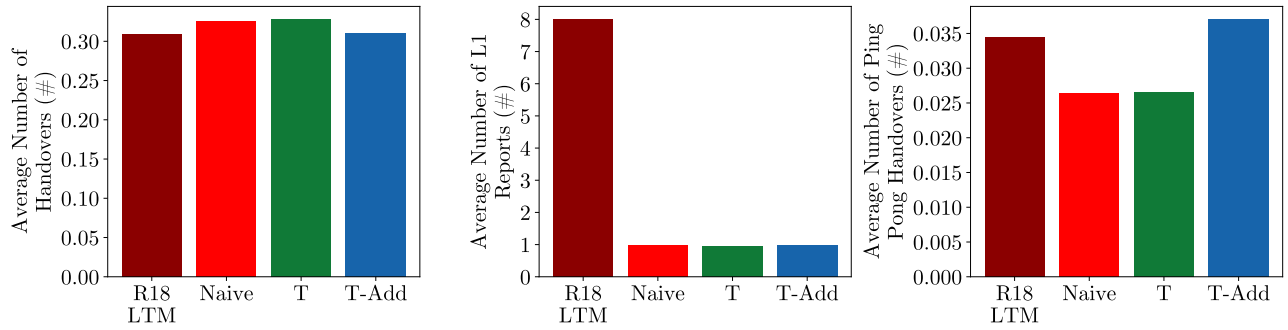
V. Conclusion

This paper presented a comprehensive study of joint spatial-temporal forecasting methods for LTM in cellular networks. By leveraging both statistical and AI/ML-based approaches within a system-level simulation, we demonstrated that LSTM-based architectures remarkably outperform baselines. The proposed models significantly improve forecasting accuracy by up to 60%, reduce unnecessary handovers by 10%, and decrease signaling overhead by 80% while maintaining seamless service continuity. Our findings underline the potential of predictive mobility management as a critical enabler for URLLC and mission-critical applications. Our future work will extend this framework toward multi-user optimization, online learning for dynamic environments, and integration with higher-layer mobility mechanisms to further enhance adaptability and scalability in next-generation networks. Another direction we plan to pursue is to implement the approaches proposed in this work in a campus network.

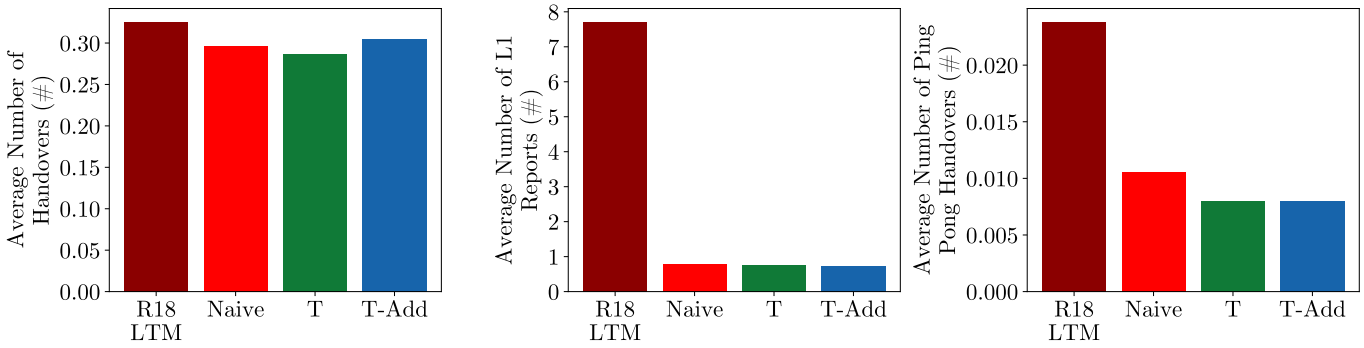
References

[1] M. Hoffmann, G. Kunzmann, T. Dudda, R. Irmer, A. Jukan, G. Macher, A. Ahmad, F. R. Beenen, A. Bröring, F. Fellhauer, G. P. Fettweis, F. H. P. Fitzek, N. Franchi, F. Gast, B. Haberland, S. Hoppe, S. Joodaki, N. P. Kuruvatti, C. Li, M. Lopez, F. Mehmeti, T. Meyerhoff, L. Miretti, G. T. Nguyen, M. Parvini, R. Pries, R. F. Schaefer, P. Schneider, D. A. Schupke, S. Strassner, H. Stubbe, and A. M. Voicu, "A secure and resilient 6G architecture vision of the german flagship project 6G-ANNA," *IEEE Access*, vol. 11, pp. 102 643–102 660, 2023.

[2] W. Jiang and B. Han, *Evolution to Fifth-Generation (5G) Mobile Cellular Communications*. Cham: Springer Nature Switzerland, 2024.



(a) Randomized Path mobility scenario



(b) Street Path mobility scenario

Fig. 11: Mobility-related KPIs for different AI/ML-based models and the Release 18 LTM (denoted as R18 LTM) across multiple mobility scenarios. 'T' denotes the Temporal model, whereas 'T-Add' denotes the Temporal with Additional Information model.

[3] ——, *Evolution Toward Sixth-Generation (6G) Mobile Cellular Systems*. Cham: Springer Nature Switzerland, 2024.

[4] 3GPP, "TS 38.331: NR; Radio Resource Control (RRC); Protocol specification," 3GPP, Technical Specification 38.331, Dec. 2024, release 18. [Online]. Available: https://www.3gpp.org/DynaReport/38331.htm

[5] ——, "TS 38.300: NR and NG-RAN Overall description; Stage-2," 3GPP, Technical Specification 38.300, Jan. 2025, release 19. [Online]. Available: https://www.3gpp.org/DynaReport/38300.htm

[6] A. Prado, W. Kellerer, and F. Mehmeti, "α-fair mobility management in 5G networks," *IEEE Transactions on Network and Service Management*, vol. 22, no. 5, pp. 5118–5136, 2025.

[7] A. Prado, F. Mehmeti, and W. Kellerer, "Reducing mobility-related signaling with network sum throughput maximization in 5G," *IEEE Transactions on Network and Service Management*, vol. 22, no. 6, pp. 6048–6065, 2025.

[8] A. Prado, F. Stöckeler, F. Mehmeti, and W. Kellerer, "Enabling proportionally fair mobility management in 5G networks," in *2023 IEEE 20th Consumer Communications Networking Conference (CCNC)*, 2023.

[9] A. Prado, Z. Ding, F. Mehmeti, and W. Kellerer, "Mobility management for computation-intensive tasks in cellular networks with SD-RAN," in *2024 20th International Conference on Network and Service Management (CNSM)*, 2024.

[10] B. Khodapanah, S. Goyal, M. Gursu, J. Stanczak, A. Kakkavas, R. Temelli, A. Badalioglu, P. Spapis, and C. Majumdar, "On L1/L2-triggered mobility in 3GPP release 18 and beyond," *IEEE Access*, vol. 12, 2024.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023. [Online]. Available: https://arxiv.org/abs/1706.03762

[12] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proc. of AAAI'23/IAAI'23/EAAI'23*, 2023.

[13] D. Kulkarni, M. Venkatesan, and A. V. Kulkarni, "Deep learning traffic prediction and resource management for 5G RAN slicing," *Journal of The Institution of Engineers (India): Series B*, vol. 106, no. 2, 2025.

[14] H.-S. Park, H. Kim, C. Lee, and H. Lee, "Mobility management paradigm shift: From reactive to proactive handover using AI/ML," *IEEE Network*, vol. 38, no. 2, Mar. 2024.

[15] J. P. S. H. Lima, A. A. M. de Medeiros, E. P. de Aguiar, E. F. Silva, V. A. de Sousa, M. L. Nunes, and A. L. Reis, "Deep learning-based handover prediction for 5G and beyond networks," in *ICC 2023 - IEEE International Conference on Communications*.

[16] A. Masri, T. Veijalainen, H. Martikainen, S. Mwanje, J. Ali-Tolppa, and M. Kajó, "Machine-learning-based predictive handover," in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, 1997.

[18] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," *CoRR*, vol. abs/1912.01703, 2019. [Online]. Available: http://arxiv.org/abs/1912.01703

[19] K. I. Pedersen, R. Maldonado, G. Pocovi, E. Juan, M. Lauridsen, I. Z. Kovács, M. Brix, and J. Wigard, "A tutorial on radio system-level simulations with emphasis on 3GPP 5G-advanced and beyond," *IEEE Communications Surveys Tutorials*, vol. 26, no. 4, pp. 2290–2325, 2024.

[20] D. Catrein and R. Mathar, "Gaussian random fields as a model for spatially correlated log-normal fading," in *Australasian Telecommunication Networks and Applications Conference*, 2008.

[21] 3GPP, "TR 38.901 Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Technical Report 38.901, Jul. 2025, release 19. [Online]. Available: https://www.3gpp.org/DynaReport/38215.htm

[22] M. Iridon and D. Matula, "Symmetric cellular network embeddings on a torus," in *Proc. of 7th International Conference on Computer Communications and Networks (Cat. No.98EX226)*, 1998.

[23] J. Herzen, F. Lässig, S. G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T. V. Pottelbergh, M. Pasieka, A. Skrodzki, N. Huguenin, M. Dumonal, J. Kościsz, D. Bader, F. Gusset, M. Benheddi, C. Williamson, M. Kosinski, M. Petrik, and G. Grosch, "Darts: User-friendly modern machine learning for time series," 2022.