

Reinforcement Learning-Based Correction of Indoor Path Loss Models for Wi-SUN Sub-GHz Communications

Hyunho Son¹ and Soyi Jung²

¹*Dept. Artificial Intelligence Convergence Network, Ajou University, Suwon, 16499, South Korea*

²*Dept. Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea*
{sohn8896, sjung}@ajou.ac.kr

Abstract—Reliable indoor wireless connectivity is essential for building management systems and smart city infrastructure. Wireless smart ubiquitous network, based on the IEEE 802.15.4g standard, provides low-power long-range communication, but performance degrades in corridors due to periodic ripples in the received signal strength indicator (RSSI) caused by multipath. Existing propagation models often fail to capture these fluctuations, leading to a large prediction error. This paper presents RSSI measurements at 917.1 MHz under frequency-shift keying (FSK) and orthogonal frequency-division multiplexing (OFDM) schemes and proposes reflection-aware path loss models whose coefficients are optimized via the soft actor-critic learning algorithm. The results show a significantly lower error in RSSI predictions, supporting more reliable indoor Internet of Things (IoT) planning.

Index Terms—Wi-SUN, path loss model, indoor wireless channel, ITU-R P.1238, reinforcement learning, RSSI prediction

I. INTRODUCTION

Stable wireless links are a prerequisite for indoor applications in building management systems (BMS), smart metering, energy monitoring, and smart city infrastructure [1], [2]. In such environments, wireless networks must guarantee high reliability, low latency, and extended coverage to support mission-critical services such as automated lighting control, HVAC management, real-time fault detection, and energy-efficient operations. Traditional solutions often face difficulties in achieving these requirements due to harsh propagation environments inside buildings. Corridors, mechanical rooms, underground spaces, and other indoor layouts frequently introduce extensive multipath fading, reflections from metallic surfaces, and diffraction around obstacles, leading to strong fluctuations in received signal strength (RSSI) and ultimately resulting in unstable wireless connectivity. [3], [4].

To address these challenges, the wireless smart ubiquitous network (Wi-SUN), built upon the IEEE 802.15.4g standard, has emerged as a promising solution. Wi-SUN provides a low-power, long-range wireless communication platform primarily operating in the sub-GHz industrial, scientific, and medical (ISM) band [5]. With a typical transmission power of around 20 mW, Wi-SUN enables coverage from several hundred meters up to multiple kilometers, depending on the deployment environment. This range extension, coupled with low energy consumption, makes Wi-SUN particularly attractive for large-

scale Internet of Things (IoT) deployments in smart grids, intelligent transportation systems, and building automation.

In February 2025, the Wi-SUN Alliance released the field area network (FAN) 1.1 specification, adding flexible modulation options—frequency-shift keying (FSK) and orthogonal frequency-division multiplexing (OFDM)—to meet diverse deployment needs. OFDM is resilient to frequency-selective fading in multipath environments, while FSK provides robust narrowband performance with simpler implementation. Additional features include carrier-sense multiple access with collision avoidance (CSMA/CA), frequency hopping (FH) for interference mitigation, and an IPv6 stack with the Routing Protocol for low-power and lossy networks (RPL) [6], [7]. Together, these capabilities enable scalable, interoperable deployments, positioning Wi-SUN as a core technology for smart cities and building management [8].

Despite these strengths, indoor propagation remains a major challenge. Sub-GHz signals penetrate walls better than higher-frequency bands but still suffer from RSSI fluctuations due to repeated reflections from walls, ceilings, and floors [9]. In corridor settings, standing wave patterns create deep fading poorly captured by conventional models. Widely adopted references such as ITU-R P.1238 and WINNER II fail to account for these effects, producing large prediction errors that hinder IoT planning. [10], [11]. To overcome these limitations, more accurate and adaptive propagation models are needed. Data-driven methods, especially reinforcement learning (RL), provide a promising approach by continuously optimizing parameters from empirical feedback.

This work addresses these issues with three contributions.

- An empirical study of RSSI behavior for OFDM and FSK in a corridor setting.
- A RL calibration framework using the soft actor-critic (SAC) algorithm to tune the ITU-R P.1238 coefficients.
- Evidence of improved RSSI prediction accuracy over ITU-R P.1238 and WINNER II.

II. RELATED WORK

A. Indoor Propagation Models for Sub-GHz IoT

A wide range of indoor propagation models have been developed to characterize the behavior of the wireless channel and support the deployment of IoT systems. Among them,

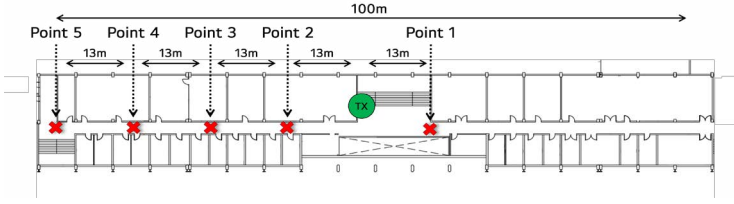


Fig. 1. Node deployment layout in the corridor environment.

the ITU-R P.1238 model has become a widely accepted reference due to its simple formulation and the availability of empirical coefficients tailored to various types of building such as residential, office and commercial environments [10]. Its frequency- and distance-dependent structure makes it easy to apply in preliminary planning studies.

In contrast, the WINNER II channel model provides a richer set of parameters, including fading distributions, delay spread statistics, and angular dispersion across multiple scenarios and frequencies [11]. This flexibility has made WINNER II highly influential in broadband wireless evaluations and channel emulation. However, its calibration largely targets frequency bands above 2 GHz, and little validation has been performed for sub-GHz IoT deployments, particularly in elongated or industrial environments.

Both ITU-R P.1238 and WINNER II share a common limitation: they do not explicitly model periodic fluctuations in received signal strength caused by repeated multipath reflections in corridor-like structures [12], [13]. As a result, predictions often diverge from empirical measurements when applied to low-frequency IoT links, underscoring the need for enhanced models that incorporate corridor-specific propagation effects.

B. RL for Propagation Model Calibration

RL has recently been introduced as a powerful tool for wireless communication research, offering the ability to optimize the behavior of the system under uncertain and time-varying conditions. Previous studies have applied RL to tasks such as adaptation of transmission power, spectrum allocation, and routing optimization [14]–[17]. These applications highlight the potential of RL to complement traditional model-based approaches by learning directly from interaction with the environment.

Advanced RL algorithms suited for continuous control problems, including SAC and the deep deterministic policy gradient (DDPG), have shown strong performance in optimizing system parameters in complex environments [18], [19]. SAC, in particular, leverages entropy-based regularization to balance exploration and exploitation, thereby providing improved convergence stability.

Despite this progress, most of the RL research in wireless networks has remained focused on network-level control and resource management [20]. Only limited efforts have explored the use of RL to refine propagation models themselves. Specifically, learning correction terms from measurement data to capture multipath corridor-specific ripples remains largely unexplored [21]. Bridging this gap opens new opportunities for data-driven propagation modeling, enabling models that adapt



Fig. 2. TX and RX setting for channel measurements.

TABLE I. Hardware specifications and PHY configurations used in measurements

(a) Hardware platform specifications		
Parameter	Specification	
Transceiver	EFR32FG25 (Silicon Labs)	
Main board	BRD4002A (Silicon Labs)	
Transmission power	16 dBm	
(b) Modulation scheme and channelization		
Modulation	Center frequency	Channel spacing
OFDM	917.1 MHz	200 kHz
FSK	917.1 MHz	200 kHz

to environment-specific channel characteristics rather than rely solely on generalized empirical formulas.

III. EXPERIMENTAL SETUP AND MEASUREMENT-BASED MODELING

A. Indoor Corridor Measurement Environment

The measurement campaign was conducted in the corridor of Woncheon Hall, Ajou University, with the objective of analyzing sub-GHz propagation in a controlled indoor setting. The corridor is 100 m long and 2 m wide, and includes a hall-like widening of approximately 10 m located around the 20 m mark, as illustrated in Fig. 1. All interior surfaces—walls, ceiling, and floor—are reinforced concrete, producing strong reflections and pronounced multipath fading.

The transmitter (TX) and receiver (RX) nodes were mounted on tripods at a height of 1.4 m above the floor to ensure consistent antenna placement and to minimize body-shadow effects as illustrated in Fig. 2. The RX points were positioned as shown in Fig. 1. Point 1 maintained a line-of-sight (LoS) link to the TX, whereas points 2–5, deployed along the corridor at 13 m intervals, operated under non-line-of-sight (NLoS) conditions. This layout was designed to emphasize the reflection and diffraction effects commonly observed in corridor-type indoor channels.

B. Hardware Platform and PHY Configurations

The experiments employed EFR32FG25 transceivers (Silicon Labs) mounted on BRD4002A mainboards with a transmit

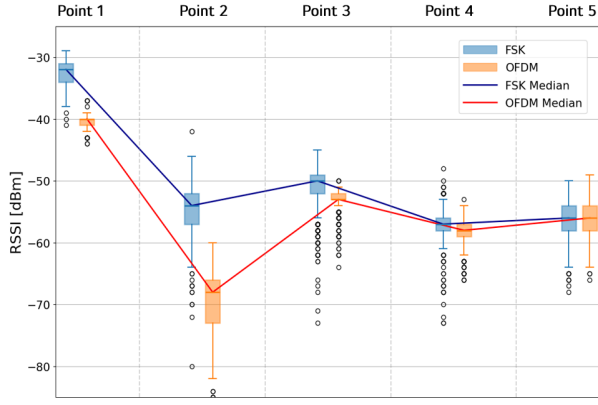


Fig. 3. Comparison of RSSI between FSK and OFDM across measurement points.

power of 16 dBm (Table I(a)). The devices comply with the IEEE 802.15.4g standard, ensuring compatibility with Wi-SUN FAN deployments.

For a fair comparison, two representative Wi-SUN PHY modes were selected: (i) OFDM Option 4 (MCS 4) and (ii) FSK, both operated at 917.1 MHz with a channel spacing of 200 kHz (Table I(b)).

At each RX point, 100 RSSI samples were recorded under identical conditions. The median was taken to suppress instantaneous fading fluctuations and to obtain a statistically stable representation of the propagation environment. This dataset provided the baseline for the PHY-level performance analysis and also served as input to the RL-based path loss calibration described in Section IV.

C. Comparative Analysis of PHY Mode Performance

Based on the measurement results, we analyze the performance gap between the two representative PHY modes in terms of RSSI. Fig. 3 presents the RSSI distributions observed at all receiver points. Contrary to the common expectation for multipath-rich corridors, our measurements show that the single-carrier (FSK) mode yields higher median RSSI than the multi-carrier (OFDM) mode at most receiver positions, with a particularly large gap at point 2.

In addition, the RSSI distributions of OFDM exhibit wider spreads and deeper fades compared to FSK. This counter-intuitive behavior can be explained by the interference-rich nature of the deployment environment. Strong in-band and adjacent channel interference, combined with residual carrier frequency and phase errors, generates intercarrier interference (ICI) that simultaneously degrades many OFDM subcarriers. In addition, the inherently high peak-to-average-power ratio (PAPR) of OFDM forces transmit power back-off, which reduces the effective link budget relative to the constant-envelope FSK scheme. As a result, the expected advantage of OFDM in mitigating frequency-selective fading is overshadowed by its vulnerability to interference and hardware impairments, whereas FSK remains comparatively robust under such conditions.

In summary, the measured ranking between PHYs in this deployment is interference-limited rather than purely multipath-

limited: FSK outperforms OFDM in most locations, whereas OFDM's inherent resilience to multipath fading does not translate into higher RSSI under strong interference. This motivates refined propagation and planning models that account for interference and hardware-impairment terms in addition to multipath.

D. ITU-R Path Loss Model Calibration via Coefficient Optimization

The baseline model used in this study is the ITU-R P.1238 indoor path loss model, which is expressed as [10]:

$$PL_{\text{ITU-R}}(d, f, n_f) = 20\log_{10}(f) + N\log_{10}(d) + L_f(n_f) + C, \quad (1)$$

where PL denotes the path loss in dB, f is the frequency in MHz, d is the transmitter-receiver distance in meters, N is the distance power loss coefficient, n_f is the number of floors between nodes and $L_f(n_f)$ represents floor penetration loss. The term C is a constant offset that compensates for empirical fitting. In the original ITU-R model for indoor office environments, it is typically set to -28 dB.

Although this model captures generalized indoor propagation behaviors, it lacks adaptation to the corridor-specific fading dynamics previously discussed in Section II-A. These effects introduce systematic deviations between predicted and measured RSSI values, which degrade the accuracy of conventional planning models.

To address this, we do not introduce a new path loss formulation but instead retain the ITU-R model structure and aim to improve its accuracy through empirical calibration. Specifically, we optimize two key coefficients— N and C —based on RSSI measurements collected in the corridor environment of the Woncheon Hall, Ajou University. This is achieved using the RL-based framework introduced in Section IV, where the agent learns to minimize the RSSI prediction error.

This data-driven coefficient optimization enables the classical ITU-R model to more precisely reflect propagation characteristics in sub-GHz corridor scenarios, without altering its analytic form. As a result, the calibrated model can offer improved performance in practical deployment planning for indoor Wi-SUN systems.

E. Measurement Dataset for RL Calibration

A dedicated dataset D was constructed from the corridor measurements described in Section III-A. Each record consists of transmitter and receiver coordinates along with the median RSSI measured under specific PHY and visibility conditions. The dataset was explicitly designed as input for the RL-based calibration of the ITU-R model. Formally, each entry can be expressed as a tuple:

$$(P_{\text{tx}}, P_{\text{rx}}, m, f, n_f, \tilde{r}), \quad (2)$$

where P_{tx} and P_{rx} denote the two-dimensional transmitter and receiver coordinates in meters. The mode variable m represents

the PHY-layer transmission scheme and the LoS condition, defined as:

$$m \in \{\text{LoS/FSK}, \text{LoS/OFDM}, \text{NLoS/FSK}, \text{NLoS/OFDM}\}, \quad (3)$$

$\tilde{r}$ denotes the median RSSI aggregated at each receiver location; f and n_f follow the definitions in Section III-D.

To ensure statistical diversity, approximately 300 receiver positions were randomly sampled along the corridor. For each position, independent tuples were generated across all four PHY/visibility conditions, thereby forming a dataset that comprehensively reflects environmental-specific propagation variability. This construction ensures sufficient state-space coverage for SAC training.

The preprocessing steps consisted of the following:

- 1) Outlier suppression at each point using an IQR-based method,
- 2) Median aggregation to obtain $\tilde{r}$,
- 3) Normalization of coordinates to SI units (meters),
- 4) Train/validation split by spatial locations to prevent data leakage.

$$D = \bigcup \{(P_{\text{tx}}, P_{\text{rx}}, f, n_f, \tilde{r})_i\}_m. \quad (4)$$

IV. RL-BASED PATH LOSS MODEL CORRECTION USING SAC

A. MDP Formulation for Path Loss Calibration

The calibration problem is formulated as a Markov decision process (MDP), where the agent aims to optimize the coefficients of the ITU-R P.1238 path loss model.

- **State:** $s_t = [f, d, n_f]$, where f is the carrier frequency in MHz, d is the Euclidean distance between the transmitter and receiver, and n_f is the floor separation.
- **Action:** $a_t = [N, C]$, where for each scenario m in Eq. (3), the policy $\pi^{(m)}$ learns a global pair $(N^{(m)}, C^{(m)})$ that remains fixed within that scenario.
- **Reward:** The reward is defined as the negative mean squared error (MSE) between predicted and measured RSSI values:

$$R_t = -\text{MSE}(PL_{\text{proposed}}(d, f, n_f; N, C), \tilde{r}). \quad (5)$$

This formulation enables the RL agent to directly minimize the RSSI prediction error while preserving the analytical structure of the ITU-R model.

B. SAC-Based Framework Description

The SAC algorithm optimizes a stochastic policy in an off-policy manner by updating an actor and two critic networks [18]. Its learning objective augments the expected return with an entropy term, with a temperature α (automatically tuned) to balance exploration and exploitation. In our offline setting, empirical RSSI measurements define a fixed dataset of transitions $\{(s_t, a_t, r_t, s_{t+1})\}$ collected without further environment interaction, which is stored in a replay buffer.

Fig. 4 describes the proposed calibration pipeline. A state s_t is sampled from the dataset, where s_t encodes the propagation condition (e.g. carrier frequency f , distance d , floor separation n_f , and the PHY/visibility mode m). The policy network

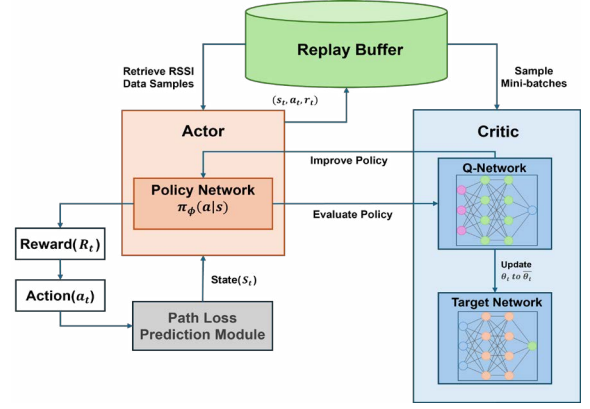


Fig. 4. Proposed RL framework for ITU-R model correction using SAC.

TABLE II. RL parameters for SAC-based calibration

Parameter	Value
Discount factor (γ)	0.99
Learning rate (η)	3×10^{-4}
Batch size	512
Replay buffer size	10^6
Total training steps	5×10^5

$\pi_\phi(a|s)$ (actor) outputs a pair of coefficients $a_t = (N, C)$, which were applied to the ITU-R model to produce a predicted RSSI. This prediction is compared with the measured RSSI to calculate the reward r_t . The transition (s_t, a_t, r_t, s_{t+1}) is added to the buffer. Twin critic networks Q_{θ_1} and Q_{θ_2} (with target networks) are updated to estimate action values under the current policy, and the actor is then updated to maximize the entropy-regularized objective. For numerical stability, the continuous action (N, C) is squashed and rescaled to predefined valid ranges. Through these iterations, the agent learns coefficients that minimize the prediction error while preserving the analytical structure of the ITU-R model. The key hyperparameters used for training are summarized in Table II. To reduce sampling bias, mini-batches are stratified across distance bins and LoS/NLoS groups, and rewards are normalized with a running mean/variance to stabilize critic targets. We also apply a mild ℓ_2 regularization on network weights and a small penalty on large coefficient updates $\Delta(N, C)$ to discourage oscillatory policies during offline training.

V. PERFORMANCE EVALUATION

A. Convergence and Coefficient Optimization

For baseline comparison, the ITU-R P.1238 model was configured using the recommended parameters for office environments, with a distance exponent $N = 30$ and an offset term $C = -28$ dB. The WINNER II office model was also evaluated using the standard coefficients for indoor environments [11]. The general form of the WINNER II path loss model is expressed as:

$$PL(d, f) = A \cdot \log_{10}(d) + B + C \cdot \log_{10}\left(\frac{f}{5}\right), \quad (6)$$

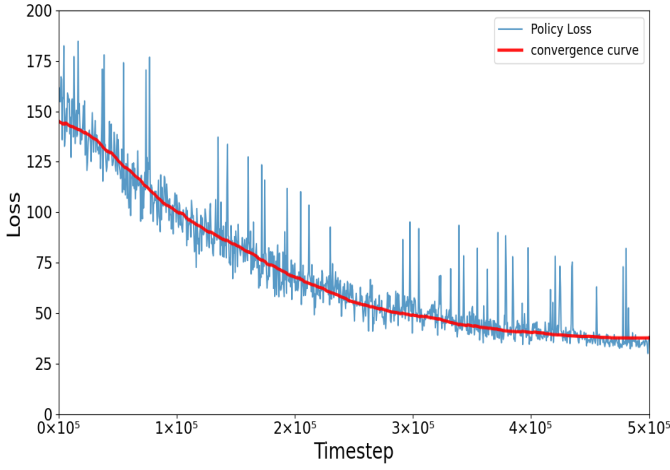


Fig. 5. Policy loss convergence during training.

TABLE III. Optimized parameters (N, C) for the proposed model in four scenarios

Scenario	N	C [dB]
LoS, FSK	17.890	-34.020
LoS, OFDM	20.560	-31.720
NLoS, FSK	28.020	-23.110
NLoS, OFDM	30.170	-13.320

where d is the transmitter-receiver separation in meters and f is the center frequency in GHz. The coefficients A , B , and C differ depending on the propagation condition:

- **LoS:** $A = 18.7$, $B = 46.8$, $C = 20.0$
- **NLoS:** $A = 20.0$, $B = 46.4$, $C = 20.0$

Although optimized for office environments, these coefficients were directly applied to the sub-GHz (917.1 MHz) scenario for comparison.

The proposed RL-based calibration framework was validated using the measurement dataset obtained from controlled corridor experiments with both FSK and OFDM modes. Fig. 5 illustrates the convergence of policy loss during SAC training. Although minor fluctuations were observed in the early stages due to stochastic sampling, the policy loss steadily decreased and converged after approximately 5×10^5 time steps. This convergence behavior confirms that the SAC agent successfully learned an optimal policy to minimize the RSSI prediction error through data-driven coefficient optimization. The smooth convergence trend also indicates the stability of the learning process and the robustness of the chosen hyperparameters.

The optimized path loss coefficients derived from the SAC-trained policy are summarized in Table III. These coefficients— N (distance power loss exponent) and C (constant offset)—were independently learned for each communication scenario defined in Eq. (3). In particular, NLoS cases yielded higher values N , which aligns with theoretical expectations due to increased attenuation and severe multipath dispersion in obstructed environments.

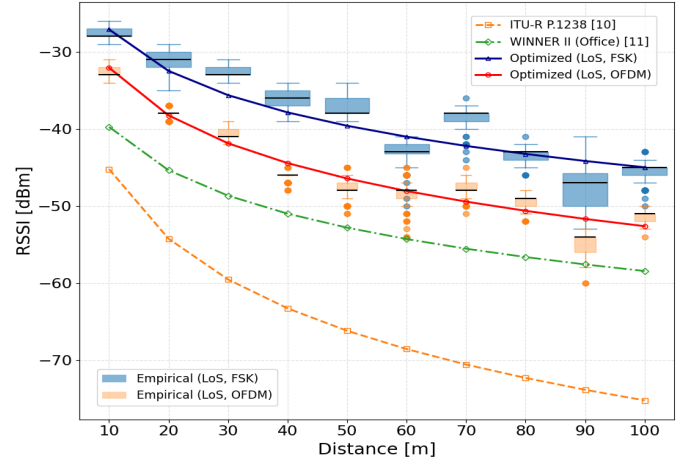


Fig. 6. RSSI comparison of measurement, ITU-R, WINNER II, and optimized model (LoS).

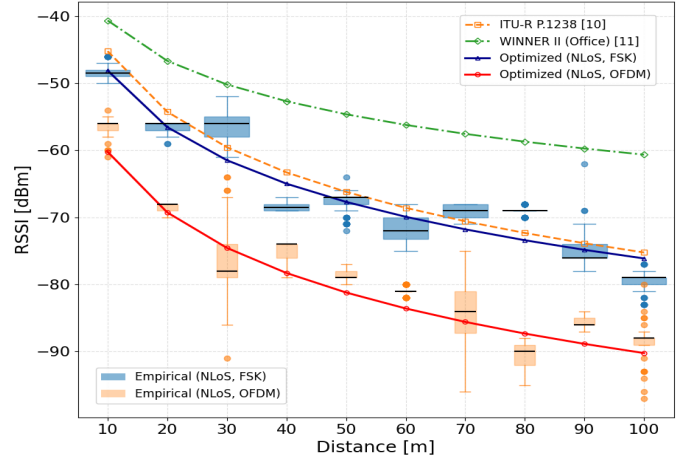


Fig. 7. RSSI comparison of measurement, ITU-R, WINNER II, and optimized model (NLoS).

B. Model Comparison and Result Analysis

In order to evaluate the effectiveness of the optimized model, its performance was compared with two widely adopted indoor path loss models, ITU-R P.1238 and WINNER II (Office). Fig. 6 and Fig. 7 depict the RSSI prediction results for the LoS and NLoS scenarios, respectively. The empirical RSSI measurements are represented in the form of boxplots and are accompanied by prediction curves obtained from the three models under consideration.

In the LoS case, as shown in Fig. 6, the proposed model is closely aligned with the empirical data for both the FSK and OFDM modulations. In contrast, the ITU-R model consistently overestimates path loss, leading to overly pessimistic predictions, particularly in OFDM transmissions, where frequency selectivity is more pronounced. In the NLoS case, illustrated in Fig. 7, the RL-optimized model provides an accurate representation of steep signal degradation as well as non-linear loss patterns that arise from multipath propagation and shadowing. However, both ITU-R P.1238 and WINNER II fail to capture the severity of the attenuation, producing higher RSSI values

than those observed in the measurements. These results suggest that the proposed RL-calibrated model not only surpasses ITU-R P.1238 in terms of accuracy but also generalizes better than the WINNER II formulation, particularly in corridor-type NLoS scenarios where complex multipath effects dominate signal propagation.

To further validate these observations, we computed the MSE between the empirical median RSSI and the model predictions, where the error was first calculated at each measurement distance and then averaged over all distances. In the LoS group, ITU-R P.1238 exhibited a large average error of 556.93, while WINNER II reduced this to 119.03. However, the proposed model achieved an MSE of only 3.26, corresponding to a reduction of 99.42% and 97.26% compared to ITU-R and WINNER II, respectively. In the NLoS group, ITU-R showed an error of 98.99, and WINNER II diverged further with 395.80. The proposed model reduced this to just 8.43, which translates to an improvement of 91.48% and 97.87% over ITU-R and WINNER II, respectively.

These results confirm that the RL-calibrated framework substantially reduces prediction errors under both LoS and NLoS conditions, highlighting its robustness against severe multipath fading. Furthermore, the findings demonstrate that classical models, originally optimized for higher-frequency scenarios, fail to generalize to sub-GHz corridor deployments. In contrast, the proposed approach adapts effectively through measurement-driven learning, thereby achieving both accuracy and generalization across diverse indoor environments.

VI. CONCLUSION

This study proposed a RL-based method for calibrating the ITU-R P.1238 path loss model in sub-GHz indoor corridor environments. Using optimization of the model parameters through the SAC algorithm, the proposed approach improves the accuracy of RSSI prediction while preserving the analytic structure. Empirical measurements in OFDM and FSK modes under both LoS and NLoS conditions demonstrate that the learned parameters effectively captured environment-specific fading characteristics. Compared with baseline models such as ITU-R and WINNER II, the proposed RL-calibrated model consistently achieves lower RSSI prediction errors in all test scenarios, validating its applicability to realistic corridor environments. In particular, calibration achieves substantial error reductions without altering the functional form of ITU-R or introducing site-specific features beyond (N, C) , thus preserving interpretability and simplicity of deployment.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government [Ministry of Science and ICT (Information and Communications Technology) (MSIT)] under Grant RS-2024-00358662.

REFERENCES

- [1] D. Minoli, K. Sohraby, and B. Occhiogrosso, "IoT considerations, requirements, and architectures for smart buildings—energy optimization and next-generation building management systems," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 269–283, Feb. 2017.
- [2] M. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, Dec. 2015.
- [3] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on smart grid communication infrastructures: Motivations, requirements and challenges," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 5–20, Mar. 2013.
- [4] V. C. Güngör, D. Sahin, T. Koçak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "Smart grid technologies: Communication technologies and standards," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 529–539, Nov. 2011.
- [5] *IEEE Std 802.15.4g-2012: Low-rate wireless personal area networks (LR-WPANs)—Amendment 3: Physical layer (PHY) specifications for smart utility networks*, IEEE Std., Apr. 2012, amendment to IEEE Std 802.15.4-2011. [Online]. Available: https://standards.ieee.org/standard/802_15_4g-2012.html
- [6] R. Hirakawa *et al.*, "Specification and performance analysis of Wi-SUN FAN," in *Proc. 2023 International Conference on Emerging Technologies*, Nov. 2023, pp. 1–6.
- [7] S. Hara and R. Prasad, "Overview of multicarrier CDMA," *IEEE Communications Magazine*, vol. 35, no. 12, pp. 126–133, Dec. 1997. [Online]. Available: <https://ieeexplore.ieee.org/document/642841>
- [8] N. Tangsunantham and C. Pirak, "Experimental performance analysis of Wi-SUN channel modelling applied to smart grid applications," *Energies*, vol. 15, no. 7, p. 2417, Mar. 2022. [Online]. Available: <https://doi.org/10.3390/en15072417>
- [9] D. Puccinelli and M. Haenggi, "Multipath fading in wireless sensor networks: Measurements and interpretation," in *Proc. of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. Florence, Italy: ACM, May 2006, pp. 314–321.
- [10] ITU-R, "Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz," International Telecommunication Union (ITU), Geneva, Switzerland, Recommendation ITU-R P.1238-10, July 2019.
- [11] WINNER II Consortium, "WINNER II Channel Models," IST-4-027756 WINNER II D1.1.2 V1.2, Tech. Rep., Sep. 2007, <https://www.ist-winner.org/>.
- [12] C.-C. Pu and W.-Y. Chung, "Mitigation of multipath fading effects to improve indoor RSSI performance," *IEEE Sensors Journal*, vol. 8, no. 11, pp. 1884–1886, Nov. 2008.
- [13] J. Zhang, "Indoor channels," in *LTE-Advanced and Next Generation Wireless Networks*. Wiley, 2012, summarizes WINNER II indoor models and their applicability (office-like environments, 2–6 GHz). [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/9781118410998.ch3>
- [14] N. Nguyen Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133–3174, Oct. 2019.
- [15] H. Zhang, S. Guo, C. Qiu, X. Chen, and L. T. Yang, "Deep reinforcement learning for wireless networks: A review," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1344–1379, Jun. 2021.
- [16] W. J. Yun, J. P. Kim, S. Jung, J.-H. Kim, and J. Kim, "Quantum multiagent actor-critic neural networks for internet-connected multirobot coordination in smart factory management," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 9942–9952, Jun. 2023.
- [17] C. Park, W. J. Yun, J. P. Kim, T. K. Rodrigues, S. Park, and S. Jung, "Quantum multiagent actor-critic networks for cooperative mobile access in multi-UAV systems," *IEEE Internet of Things Journal*, vol. 10, no. 22, pp. 20 033–20 048, Nov. 2023.
- [18] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. of the 35th International Conference on Machine Learning (ICML)*, Jul. 2018, pp. 1861–1870.
- [19] E. H. Sumiea, S. J. Abdulkadir, H. S. Alhussian, S. M. Al-Selwi, A. Alqushaibi, M. G. Ragab, and S. M. Fati, "Deep deterministic policy gradient algorithm: A systematic review," *Heliyon*, vol. 10, no. 9, p. e30697, May 2024.
- [20] Y. Liang, G. Wu, H. Xu, P. Xu, and S. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2282–2292, Oct. 2019.
- [21] M. Vasudevan, M. Tran, M. Samimi, and T. S. Rappaport, "Machine learning for radio propagation modeling," *IEEE Communications Magazine*, vol. 62, no. 5, pp. 74–80, May 2024.