

Demo: Vision-Language Model-Guided Online Calibration of an Electromagnetic Digital Twin

Zerui Kang¹, Yishen Lim¹, Zhouyou Gu¹, Seungnyun Kim¹, Seung-Woo Ko², Tony Q.S. Quek¹, Jihong Park¹

¹ Singapore University of Technology and Design

Singapore, Singapore

²Inha University

Incheon, Republic of Korea

zerui_kang@mymail.sutd.edu.sg; swko@inha.ac.kr

{yishen_lim,zhouyou_gu,seungnyun_kim,tonyquek,jihong_park}@sutd.edu.sg

Abstract

An electromagnetic (EM) digital twin gives mobile robots wireless situational awareness but depends on material conductivities that change with the environment. Online calibration faces initialization sensitivity and measurement travel costs. We demonstrate a vision-language model (VLM)-guided framework using a Unitree G1 robot and NVIDIA Sionna, with two VLM calls: material classification maps visible materials through ITU-R P.2040 to conductivity priors for Sionna's gradient descent on accumulated received signal strength (RSS) measurements; waypoint planning selects the next measurement location online using residual RSS calibration error and image coverage. In a real indoor scenario, the framework achieves a normalized mean absolute conductivity error of 1.74×10^{-4} within 20 m of travel; random initialization never converges, while random waypoints require over twice the travel.

CCS Concepts

• **Networks** → **Network measurement**; *Network simulations*; • **Computer systems organization** → *Robotics*.

Keywords

Electromagnetic digital twin, vision-language model, online calibration, differentiable ray tracing

ACM Reference Format:

Zerui Kang¹, Yishen Lim¹, Zhouyou Gu¹, Seungnyun Kim¹, Seung-Woo Ko², Tony Q.S. Quek¹, Jihong Park¹. 2026. Demo: Vision-Language Model-Guided Online Calibration of an Electromagnetic Digital Twin. In *The Twenty-seventh International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '26)*, November 23–26, 2026, Tokyo, Japan. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3842721.3850098>

1 Introduction

Electromagnetic (EM) digital twins model the radio propagation characteristics of a physical environment to provide mobile robots with situational awareness of the wireless environment. Their fidelity relies on estimates of environmental EM properties, which must be recalibrated from radio measurements as the environment

changes. While existing methods estimate EM properties via differentiable ray tracing, e.g., NVIDIA Sionna [2], they typically operate *offline* on pre-collected measurements. For example, Hoydis et al. [3] learn material and antenna parameters from recorded indoor data, while Ai et al. [1] densify sparse measurements through a Bayesian channel map. To reduce the human effort of repeated measurement campaigns, mobile robots with onboard sensing offer a promising approach for *online calibration*, where measurements are acquired progressively and travel-efficient selection becomes important. Our prior work [5] uses a VLM to initialize material parameters and select measurement positions from a scene view, but these positions are fixed in advance and cannot adapt to evolving RSS calibration errors. Measurement positions should instead be selected from preceding observations to target poorly calibrated regions.

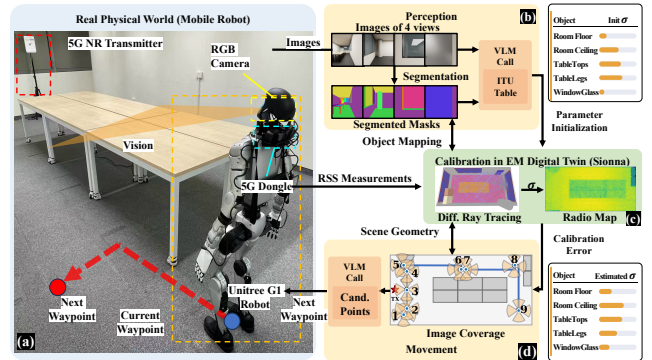


Figure 1: The proposed online calibration framework.

An aim of this paper is to propose a VLM-based online calibration framework for an EM digital twin. The key idea is to couple two VLM calls in a perception-calibration-movement loop. Unlike [5], where the VLM initializes all parameters once from a complete view and fixes the measurement positions before calibration starts, both calls are made online in every round. First, since one viewpoint reveals only a few objects, the material parameters cannot be initialized once for the whole scene. A *material-classification* call therefore initializes the objects newly observed at each waypoint from the camera images and their segmentation masks, which lets the VLM assign a material to each object rather than to the whole image. Priors close to the true values place the parameters where the RSS gradient points toward the true values, so every RSS measurement moves the parameters closer to the true values and the calibration converges within a few rounds. Second, since one measurement calibrates only the materials along its propagation paths, measuring



This work is licensed under a Creative Commons Attribution 4.0 International License. *MobiHoc '26, Tokyo, Japan*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2975-1/2026/11

<https://doi.org/10.1145/3842721.3850098>

where the nearby materials are already calibrated wastes travel. A *waypoint-planning* call therefore selects the next waypoint from the RSS calibration error and the image coverage, so that each measurement reaches the materials still uncertain and each camera view covers the objects not yet observed, and the calibration converges with the least travel. Together, these two calls close the loop: visual observations initialize newly observed materials, RSS-based calibration errors guide the next waypoint, and newly collected images and RSS measurements feed into the next round.

2 System Design

We demonstrate the framework in the indoor scenario of Fig. 1(a) with a fixed 5G NR base station and a Unitree G1 humanoid robot carrying a 5G dongle and an RGB camera. A desktop with an NVIDIA RTX Pro 6000 GPU hosts the EM digital twin in Sionna and GPT-5.4 as the VLM. The dongle records RSS measurements from the fixed transmitter, while the Sionna configuration uses the same carrier frequency and antenna locations as the physical setup. The following three stages run once per round at each waypoint. For simplicity, we calibrate only the material conductivities.

Perception: The scene images in the four directions around the robot and the segmentation masks are provided to the **material-classification** call, which labels the material of each segmented object, and each label is mapped through the ITU-R P.2040 database to a conductivity value [4], as shown in Fig. 1(b). The VLM outputs are conductivity priors for the visible objects and an image-coverage record of the objects observed so far, while objects outside all camera views retain their estimates from the previous round.

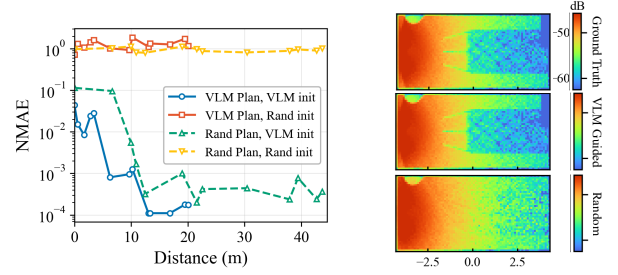
Calibration: Sionna in Fig. 1(c) refines the conductivities by gradient descent on the error between measured and simulated RSS, and reports a convergence indicator, namely the steadiness of the loss and the parameters over the last 50 iterations. Since the VLM initializes the parameters close to the true values, gradient descent can quickly converge before the next measurement arrives, leaving small RSS-based calibration errors to the next round.

Movement: The RSS at a waypoint depends predominantly on the materials along the propagation paths from the base station to the robot, so a measurement can calibrate only those materials. The *waypoint-planning* call therefore takes the room layout, the RSS-based calibration error, and the image coverage, and picks the next waypoint whose propagation paths interact with the materials of still-uncertain conductivity and whose camera view covers objects not yet observed. Fig. 1(d) shows the waypoints selected online.

3 Observations

We ablate each VLM call with a random baseline that draws initial conductivities at random or picks waypoints uniformly in the room. Fig. 2(a) plots the normalized mean absolute error (NMAE) of the estimated conductivity against traveled distance. For evaluation, the ground-truth conductivity of each material is predefined and used by Sionna to generate the radio map, while material parameters remain unknown to the digital twin throughout the calibration.

Initialization decides convergence. Both randomly initialized runs stay at an NMAE near unity (100%) regardless of the planner, since the random values lie far from the true values, where gradient descent cannot update them in the correct direction. With the VLM



(a) NMAE vs. traveled distance.

(b) Calibrated radio maps.

Figure 2: Conductivity error and radio maps for the planning and initialization combinations.

priors, both planners converge, confirming that a material label places each conductivity where the RSS gradient points toward the nearby true value in the parameter space.

Planning decides the travel needed to converge. Given the priors, the waypoint-planning call reaches an NMAE of 1.74×10^{-4} within 20.06 m, whereas the random planner needs 43.69 m to reach 3.66×10^{-4} because it wastes travel where nearby materials are already calibrated. Fig. 2(b) shows the calibrated radio maps, where the proposed scheme attains a mean absolute error of 7×10^{-3} dB against the ground truth, versus 2.027 dB for the random baseline.

4 Demonstration

The demonstration runs the full pipeline in a photorealistic NVIDIA Isaac Sim replica of our meeting room, reproducing its layout, object placement, robot viewpoints, and measurement locations; real-world sensing noise and radio-hardware effects require separate validation. The EM digital twin and VLM run remotely in our laboratory. A live dashboard displays each round's material labels, robot movement to the selected waypoint, and calibrated radio map alongside the ground truth. Attendees can change object materials to trigger reclassification, recalibration, and map updates, or select measurement positions and compare performance with VLM planning. The setup requires network access and AC power for one laptop and one monitor. Video: <https://youtu.be/hwMRUpDA1PU>.

Acknowledgments

This work was supported in part by the NVIDIA Academic Grant Program, in part by A*STAR under its IAF-ICP (H25-MCP3438), and in part by the National Research Foundation, Singapore, and the Infocomm Media Development Authority under the Communications and Connectivity Bridging Funding Initiative.

References

- [1] Junjie Ai et al. 2026. Propagation-consistent wireless environment digital twin construction under sparse measurements. *arXiv preprint arXiv:2605.22361* (2026).
- [2] Jakob Hoydis et al. 2023. Sionna RT: Differentiable ray tracing for radio propagation modeling. In *Proc. IEEE Globecom Workshops*. IEEE, Kuala Lumpur, Malaysia, 317–321. doi:10.1109/GCWkshps58843.2023.10465179
- [3] Jakob Hoydis et al. 2024. Learning radio environments by differentiable ray tracing. *IEEE Trans. Mach. Learn. Commun. Netw.* 2 (2024), 1527–1539. doi:10.1109/TMLCN.2024.3474639
- [4] ITU-R. 2023. *Effects of Building Materials and Structures on Radiowave Propagation above about 100 MHz*. Recommendation ITU-R P.2040-3. ITU.
- [5] Zerui Kang et al. 2026. Vision-Language-Model-Guided Differentiable Ray Tracing for Fast and Accurate Multi-Material RF Parameter Estimation. In *Proc. AAAI Workshop Mach. Learn. Wireless Commun. Netw. (ML4Wireless)*. AAAI, Singapore.