Differentiable Wireless Simulation with Geometric Transformers

Thomas Hehn [1] **Markus Peschl** [1] **Tribhuvanesh Orekondy** [1] **Arash Behboodi** [1] **Johann Brehmer** [1]

Abstract

Modelling the propagation of electromagnetic signals is critical for designing modern communication systems. While there are precise simulators based on ray tracing, they do not lend themselves to solving inverse problems or the integration in an automated design loop. We propose to address these challenges through differentiable neural surrogates that exploit the geometric aspects of the problem. We introduce the Wireless Geometric Algebra Transformer (Wi-GATr), a generic, equivariant backbone architecture for simulating wireless propagation in a 3D environment. Further, we introduce two datasets of wireless signal propagation in indoor scenes. On these datasets, we show the data-efficiency of our model on signal prediction and applicability to inverse problems based on differentiable predictive modelling.

1. Introduction

Innovations in modern wireless communication systems build upon electromagnetic wave propagation. Therefore, modelling and understanding wave propagation in space is a core research area in wireless communication. Wireless signal propagation follows Maxwell's equations of electromagnetism and is traditionally modelled by ray-tracing simulators. However, these traditional wireless simulators take substantial time to evaluate for each scene, cannot be fine-tuned on measurements, and are not differentiable. It additionally limits their usefulness for solving inverse problems.

In contrast, neural models of signal propagation can be evaluated cheaply, can be trained on real measurements in addition to simulation, and are differentiable and thus well-suited for solving inverse problems. However, wireless surrogate modelling faces various challenges. Realistic training data is often scarce, requiring surrogate models to be data efficient. Wireless environments can consist of complex meshes. Finally, input and output data consist of a variety of data types, including the shape of extended 3D objects, point coordinates and spatial orientation of antennas, and information associated with the transmitted signal.

In this work, we present a new approach to modelling wireless signal propagation. It is grounded in the observation that wireless propagation is inherently a geometric problem: a directional signal is transmitted by an oriented transmitting antenna, the signal interacts with surfaces in the environment, and the signal eventually impinges an oriented receiving antenna. We argue that it is critical for neural surrogates to model and flexibly represent geometric aspects (e. g. orientations, shapes) in the propagation environment. We therefore develop surrogate models based on flexible geometric representation and strong geometric inductive biases.

We propose *Wireless Geometric Algebra Transformer* (Wi-GATr), a backbone architecture for wireless signal propagation problems. A key component is a new tokenizer for the diverse, geometric data of wireless scenes. The tokens are processed with a Geometric Algebra Transformer (GATr) network [9]. This architecture is equivariant with respect to the symmetries of wireless channel modelling, but maintains the scalability of a transformer architecture. In addition, we study Wi-GATr models as differentiable, predictive surrogates for the simulator (see Fig. 1). We show how this enables forward modelling, and in addition, inverse problem solving due to Wi-GATr's differentiability.

To enable machine learning development for wireless problems, we finally introduce two new datasets, `Wi3R` and `WiPTR`. Each dataset consists of thousands of indoor scenes of varying complexity and include all the geometric information that characterizes a wireless scene. We demonstrate the predictive models on these datasets Our experiments show that the Wi-GATr approach gives us a higher-fidelity predictions than various baselines, generalizes robustly to unseen settings, and requires up to $20\times$ less data for the same performance than a transformer baseline.

2. Background and related work

Wireless signal propagation. Wireless simulation is often based on ray-tracing approaches [1, 4, 18] which em-

[1]Qualcomm AI Research. (Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.). Correspondence to: Thomas Hehn <thehn@qti.qualcomm.com>.

Published at the 2^{nd} Differentiable Almost Everything Workshop at the 41^{st} International Conference on Machine Learning, Vienna, Austria. July 2024.

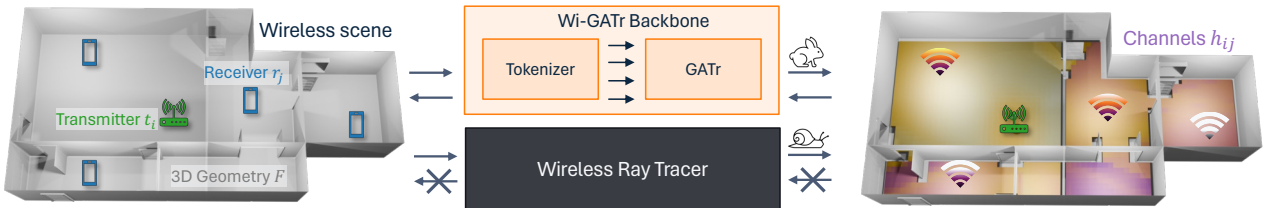

Figure 1: Geometric surrogates for modelling wireless signal propagation.: Predictive modelling of channels from 3D geometry, transmitter, and receiver properties. Wi-GATr is a fast and differentiable surrogate for ray tracers.

ploy the approximations of geometric optics [22]. It approximates the solution to Maxwell's equations as a sum of planar waves propagating in all directions from the transmitter (Tx). Each planar wave is represented as a ray, characterized by various attributes (e. g., power, phase, delay). As a ray hits a surface, the interaction is modelled as reflection, refraction, or diffraction. During such interactions, the power, phase, polarization, and propagation direction of the wave can change in complex, material-dependent ways. After multiple interactions, the rays eventually reach the receiving antenna (Rx). The Tx and Rx are then linked by a connected path p of multiple rays. The effects on the received signal are described by the channel impulse response (CIR) $h(\tau) = \sum_p a_p \delta(\tau - \tau_p)$, where $a_p \in \mathbb{C}$ is the complex gain and τ_p the delay of the incoming rays [38].

Differentiable simulations for wireless propagation. Differentiable simulations has a rich history in physics [29, 34], graphics [14, 21, 26, 30], robotics [17, 36], and numerous other domains [5, 15, 32]. Specific to wireless, approaches can be broadly categorized into analytical, hybrid, or neural techniques. At one end of the spectrum, *analytical* approaches (notably Sionna RT [18]) focus on performing traditional wireless ray tracing computations in a differentiable framework (Tensorflow in case of [18]). While these approaches are differentiable, they largely employ existing physical models (e.g., UTD [24], Antenna radiation patterns) that might be unknown or inaccurate for many scenarios. *Hybrid* approaches [19, 20, 27, 31] lie in the middle of the spectrum. Similar to NeRF [28], they propose combining analytical rendering (e.g., ray tracing, volumetric rendering) with learnable neural components (e.g., learning scene parameters of radio material properties). Although these techniques are fairly new and promising, they are limited to perform simulations in the exact single scene they were trained on. At the other end of the spectrum are neural approaches [16, 25], that make minimal assumptions on propagation physics and almost entirely rely on learning it from data. Our proposed approach falls in this category, where in we simulate wireless signals with a single forward pass through a geometric transformer that is both sample-efficient and generalizes to novel scenes.

Geometric deep learning and GATr. The field of geometric deep learning [10] aims to include structural properties of a problem into neural network architectures and algorithms. A central concept is *equivariance* to symmetry groups [11], i. e., outputs of a network transform consistently with a symmetry transformation of the inputs. The Geometric Algebra Transformer (GATr) [9] is an $\mathrm{E}(3)$-equivariant architecture for geometric problems. As we argued above, the physics of wireless signal propagation are invariant under this group.

3. The Wireless Geometric Algebra Transformer (Wi-GATr)

Our goal is to model the interplay between 3D environments, transmitting and receiving antennas, and the resulting transmitted wireless signals. More precisely, we consider *wireless scenes* consisting of: (a) *3D geometry* F of the environment. We specify it through a triangular mesh with a discrete material class associated with each mesh face; (b) A set of *transmitting antennas* t_i for $i = 1, \dots, n_t$. Each t_i is characterized by a 3D position, an orientation, and any antenna characteristics. We will often focus on the case of a single Tx and then omit the index i; (c) Analogously, a set of *receiving antennas* r_i for $i = 1, \dots, n_r$; and (d) *The channel* or signal h_{ij} between each transmitter i and each receiver j, which can be any observable function of the CIR.

In this setting, we consider two downstream tasks: (i) *Signal prediction* is about predicting the signal received at a single antenna from a single receiver $h(F, t, r)$ with $n_t = n_r = 1$. This is exactly the task that ray-tracing simulators solve. (ii) *Receiver localization*: inferring the position and properties of a receiver r from the channel, the geometry and one or multiple transmitters. The latter problem is an example of *inverse problems*, as it inverts the graphical model that simulators are designed for. It is not straightforward to solve with the simulators directly, but we will show how neural surrogates trained on simulator data can solve it.

3.1. Backbone

Core to our approach is the Wireless Geometric Algebra Transformer (Wi-GATr) backbone. It consists of a novel tokenizer and a network architecture.

Wireless GA tokenizer. The tokenizer takes as input some subset of the information characterizing a wireless scene

and outputs a sequence of tokens that can be processed by the network. A key challenge in the neural modelling of wireless problems is the diversity of types of data involved. As we argued above, a wireless scene consists of the 3D environment mesh F, which features three-dimensional objects such as buildings and trees, antennas t and r characterized through a point-like position, an antenna orientation, and additional information about the antenna type, and the characteristics of the channel h.

To support all of these data types, we propose a new tokenizer that outputs a sequence of geometric algebra (GA) tokens. Each token consists of a number of elements (channels) of the projective geometric algebra $\mathbb{G}_{3,0,1}$ in addition to the usual unstructured scalar channels. We define the GA precisely in Appendix A. Its main characteristics are that each element is a 16-dimensional vector and can represent various geometric primitives: 3D points including an absolute position, lines, planes, and so on. This richly structured space is ideally suited to represent the different elements encountered in a wireless problem. Details of our tokenization scheme are specified in Tbl. 1 in the Appendix.

Network. After tokenizing, we process the input data with a Geometric Algebra Transformer (GATr) [9]. This architecture naturally operates on our $\mathbb{G}_{3,0,1}$ parameterization of the scene. It is equivariant with respect to permutations of the input tokens as well as $\mathrm{E}(3)$, the symmetry group of translations, rotations, and reflections. These are exactly the symmetries of wireless signal propagation, with one exception: wireless signals have an additional reciprocity symmetry that specifies that the signal is invariant under an role exchange between transmitter and receiver. We will later show how we can incentivize this additional symmetry property through data augmentation.[1] Finally, because GATr is a transformer, it can process sequences of variable lengths and scales well to systems with many tokens. Both properties are crucial for complex wireless scenes, which can in particular involve a larger number of mesh faces.

3.2. Predictive modelling

We demonstrate predictive modelling of the measured channel information as a function of the complete 3D environment and the information characterizing the transmitter and receiver, $h_\theta(F, t, r)$. This regression model is trained in a supervised way on simulated or measured wireless scenes.

Forward prediction. The network learns a differentiable, deterministic surrogate for the simulator model $h_{\text{sim}}(F, t, r)$. At test time, we can use the network instead of a simulator to predict the signals in unseen, novel scenes. Compared to a simulator based on ray tracing, it has three advantages: it can be evaluated in microseconds rather than seconds or minutes, it can be finetuned on real measurements, and it is differentiable.

Inverse problems. This differentiability makes such a surrogate model well-suited to solve inverse problems. In this paper, we consider the problem of receiver localization [6, 23, 35, 37]: given a 3D environment F, transmitters $\{t_i\}$, and corresponding signals $\{h_i\}$, we can find the most likely receiver position and orientation as $\hat{r} = \arg\min_r \sum_i \|h_\theta(F, t_i, r) - h\|^2$. The minimization can be performed numerically through gradient descent, thanks to the differentiability of the Wi-GATr surrogate.

4. New datasets

While several datasets of wireless simulations and measurements exist [2, 3, 31, 40], they lack either geometric information, diversity, or realistic signals. To facilitate the development of machine learning methods with a focus on geometry, we generate two new datasets of simulated wireless scenes.[2] Both feature indoor scenes and channel information generated with a state-of-the-art ray-tracing simulator [1] at a frequency of 3.5 GHz. They provide detailed characteristics for each path between Tx and Rx, such as gain, delay, angle of departure and arrival at Tx/Rx, and the electric field at the receiver itself, which allows users to compute various quantities of interest themselves. We provide a summary of the two datasets here and discuss additional details in Appendix B.

`Wi3R` dataset. Our first dataset focuses on simplicity: each of 5000 floor plans has the same size and number of rooms, and all walls have the same material across layouts. They differ only in their layouts, which we take from `Wi3Rooms` [31], Tx positions, and Rx positions.

`WiPTR` dataset. Next, we generate a more varied, realistic dataset based on the floor layouts in `ProcTHOR-10k` [12]. We extract the 3D mesh information including walls, windows, doors, and door frames and assign 6 different dielectric materials for different groups of objects. `WiPTR` stands out among wireless datasets in terms of its level of detail and scale. Moreover, it is based on `ProcTHOR-10k` is well-suited for integration with embodied AI research.

5. Experiments

We train predictive surrogates $h_\theta(F, t, r)$ that predict the time-averaged non-coherent received power $h = \sum_p |a_p|^2$ as a function of the Tx position and orientation t, Rx position and orientation r, and 3D environment mesh F. We compare the prediction accuracy to several baselines on both the

[1] We also experimented with a reciprocity-equivariant variation of the architecture, but that led to a marginally worse performance without a significant gain in sample efficiency.

[2] We are preparing the publication of the datasets.

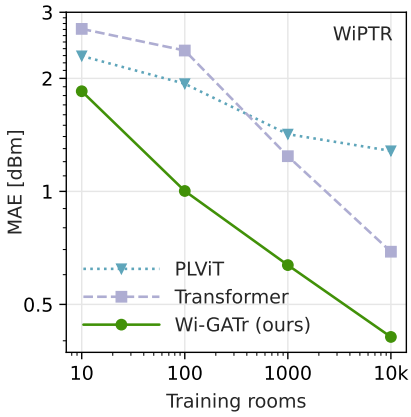

Figure 2: Signal prediction error, on the received power as a function of the training data on `WiPTR`. Wi-GATr outperforms the transformer and PLViT baselines. SEGNN does not scale to `WiPTR`.

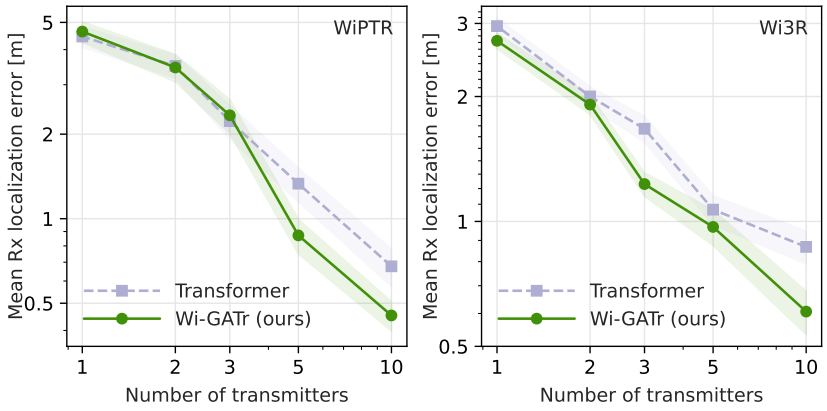

Figure 3: Inverse Rx localization error, as a function of the number of Tx on `Wi3R` (right) and `WiPTR` (left). Lines and error band show mean and its standard error over 240 measurements. Our learned differentiable simulation surrogate models provides gradients to solve this inverse problem.

`Wi3R` and `WiPTR` datasets, and we show the benefits of differentiability by solving the inverse receiver localization problem through gradient descent. All models are trained with reciprocity augmentation, i. e., randomly flipping Tx and Rx labels during training. This improves data efficiency slightly, especially for the transformer baseline.

In addition to our Wi-GATr model, described in Sec. 3, we train several baselines. The first is a vanilla transformer [39], based on the same inputs and tokenization of the wireless scene, but without the geometric inductive biases. Next, we compare to the $\mathrm{E}(3)$-equivariant SEGNN [8], though we were only able to fit this model into memory for the `Wi3R` dataset. In addition, we train a PLViT model, a state-of-the-art neural surrogate for wireless scenes [16] that represent wireless scenes as an image centered around the Tx position. Finally, we attempt to compare Wi-GATr also to WiNeRT [31], a neural ray tracer. However, this architecture, which was developed to be trained on several measurements on the same floor plan, was not able to achieve useful predictions on our diverse datasets with their focus on generalization across floor plans. Our experiment setup and the baselines are described in detail in Appendix C.

Signal prediction. Next, we study the data efficiency and accuracy of the different surrogates for signal prediction in Fig. 2. Wi-GATr is more data-efficient than any other method with the exception of the $\mathrm{E}(3)$-equivariant SEGNN on `Wi3R` (see Appendix C). This confirms that equivariance is a useful inductive bias when data is scarce. But Wi-GATr scales better than SEGNN to larger number of samples, showing that our architecture combines the small-data advantages of strong inductive biases with the large-data advantages of a transformer architecture.

Inference speed. One of the advantages of neural surrogates is their test-time speed. Both Wi-GATr and a transformer are over a factor of 20 faster than the ground-truth ray tracer (see Appendix C).

Receiver localization. Next, we show how differentiable surrogates let us solve inverse problems, focusing on the problem of receiver localization. We infer the Rx position with the predictive surrogate models by optimizing through the neural surrogate of the simulator as discussed in Sec. 3.2. Neither the SEGNN nor PLViT baselines are fully differentiable with respect to object positions when using the official implementations from Refs. [7, 16]. We were therefore not able to accurately infer the transmitter positions with these architectures. The performance of our surrogate models is shown in Fig. 3. The two neural surrogates achieve a similar performance when only one or two transmitters are available, a setting in which the receiver position is highly ambiguous. With more measurements, Wi-GATr lets us localize the transmitter more precisely.

6. Conclusion

Wireless signal transmission through electromagnetic wave propagation is an inherently geometric and symmetric problem. We developed a class of neural surrogates grounded in geometric representations and strong inductive biases. In our experiments, we demonstrated the predictive capabilities and data efficiency of our approach as well as the benefits of its differentiability for solving inverse problems, such as receiver localization. Augmenting or replacing the image-based or general-purpose representations and architectures prevalent in wireless modelling with geometric approaches has the potential of solving inverse problems through differentiability while improving data efficiency, performance, and robustness.

References

[1] Remcom Wireless InSite®. https://www.remcom.com/wireless-insite-propagation-software. [Accessed 10-05-2024].

[2] A. Alkhateeb. DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications. In *ITA*, 2019.

[3] Ahmed Alkhateeb, Gouranga Charan, Tawfik Osman, Andrew Hredzak, Joao Morais, Umut Demirhan, and Nikhil Srinivas. Deepsense 6g: A large-scale real-world multi-modal sensing and communication dataset. *IEEE Communications Magazine*, 2023.

[4] Nicolas Amiot, Mohamed Laaraiedh, and Bernard Uguen. Pylayers: An open source dynamic simulator for indoor propagation and localization. In *ICC*, 2013.

[5] Navid Ansari, Hans-Peter Seidel, Nima Vahidi Ferdowsi, and Vahid Babaei. Autoinverse: Uncertainty aware inversion of neural networks. *NeurIPS*, 2022.

[6] Maximilian Arnold, Jakob Hoydis, and Stephan ten Brink. Novel massive mimo channel sounding data applied to deep learning-based indoor positioning. In *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*, pages 1–6. VDE, 2019.

[7] Johannes Brandstetter, Rianne van den Berg, Max Welling, and Jayesh K Gupta. Clifford neural layers for PDE modeling. *arXiv:2209.04934*, 2022.

[8] Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J Bekkers, and Max Welling. Geometric and physical quantities improve E(3) equivariant message passing. In *ICLR*, 2022.

[9] Johann Brehmer, Pim de Haan, Sönke Behrends, and Taco Cohen. Geometric Algebra Transformer. In *NeurIPS*, 2023.

[10] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. 2021.

[11] Taco Cohen. *Equivariant Convolutional Networks*. PhD thesis, University of Amsterdam, 2021.

[12] Matt Deitke, Eli VanderBilt, Alvaro Herrasti, Luca Weihs, Jordi Salvador, Kiana Ehsani, Winson Han, Eric Kolve, Ali Farhadi, Aniruddha Kembhavi, and Roozbeh Mottaghi. ProcTHOR: Large-Scale Embodied AI Using Procedural Generation. In *NeurIPS*, 2022.

[13] Leo Dorst. A guided tour to the plane-based geometric algebra pga. 2020. URL https://geometricalgebra.org/downloads/PGA4CS.pdf.

[14] Michael Fischer and Tobias Ritschel. Zero grads ever given: Learning local surrogate losses for non-differentiable graphics. In *SIGGRAPH*, 2024.

[15] Josif Grabocka, Randolf Scholz, and Lars Schmidt-Thieme. Learning surrogate losses. *arXiv preprint arXiv:1905.10108*, 2019.

[16] Thomas M Hehn, Tribhuvanesh Orekondy, Ori Shental, Arash Behboodi, Juan Bucheli, Akash Doshi, June Namgoong, Taesang Yoo, Ashwin Sampath, and Joseph B Soriaga. Transformer-based neural surrogate for link-level path loss prediction from variable-sized maps. In *IEEE Globecom*, 2023.

[17] Eric Heiden, David Millard, Erwin Coumans, Yizhou Sheng, and Gaurav S Sukhatme. Neuralsim: Augmenting differentiable simulators with neural networks. In *ICRA*, 2021.

[18] Jakob Hoydis, Sebastian Cammerer, Fayçal Ait Aoudia, Avinash Vem, Nikolaus Binder, Guillermo Marcus, and Alexander Keller. Sionna: An open-source library for next-generation physical layer research. *arXiv*, 2022.

[19] Jakob Hoydis, Fayçal Aït Aoudia, Sebastian Cammerer, Florian Euchner, Merlin Nimier-David, Stephan ten Brink, and Alexander Keller. Learning radio environments by differentiable ray tracing. *arXiv:2311.18558*, 2023.

[20] Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, and Anthony Rowe. Dart: Implicit doppler tomography for radar novel view synthesis. In *CVPR*, 2024.

[21] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. Dr.jit: A just-in-time compiler for differentiable rendering. *Transactions on Graphics (Proceedings of SIGGRAPH)*, 2022.

[22] Joseph B Keller. Geometrical theory of diffraction. *J. Opt. Soc. Am., JOSA*.

[23] Ali Khalajmehrabadi, Nikolaos Gatsis, and David Akopian. Modern wlan fingerprinting indoor positioning methods and deployment challenges. *IEEE Communications Surveys & Tutorials*, 19(3):1974–2002, 2017.

[24] Robert G Kouyoumjian and Prabhakar H Pathak. A uniform geometrical theory of diffraction for an edge

in a perfectly conducting surface. *Proceedings of the IEEE*, 1974.

[25] Ron Levie, Çağkan Yapar, Gitta Kutyniok, and Giuseppe Caire. Radiounet: Fast radio map estimation with convolutional neural networks. *IEEE TWC*, 2021.

[26] Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. Differentiable monte carlo ray tracing through edge sampling. *ACM Transactions on Graphics (TOG)*, 2018.

[27] Haofan Lu, Christopher Vattheuer, Baharan Mirzasoleiman, and Omid Abari. Newrf: A deep learning framework for wireless radiation field reconstruction and channel prediction. In *ICML*, 2024.

[28] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

[29] J Krishna Murthy, Miles Macklin, Florian Golemo, Vikram Voleti, Linda Petrini, Martin Weiss, Breandan Considine, Jérôme Parent-Lévesque, Kevin Xie, Kenny Erleben, et al. gradsim: Differentiable simulation for system identification and visuomotor control. In *ICLR*, 2020.

[30] Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: A retargetable forward and inverse renderer. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 2019.

[31] Tribhuvanesh Orekondy, Pratik Kumar, Shreya Kadambi, Hao Ye, Joseph Soriaga, and Arash Behboodi. Winert: Towards neural ray tracing for wireless channel modelling and differentiable simulations. In *ICLR*, 2022.

[32] Simiao Ren, Willie Padilla, and Jordan Malof. Benchmarking deep inverse models over time, and the neural-adjoint method. *NeurIPS*, 2020.

[33] David Ruhe, Jayesh K Gupta, Steven de Keninck, Max Welling, and Johannes Brandstetter. Geometric clifford algebra networks. In *ICLR*, 2023.

[34] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *ICML*, 2020.

[35] Vladimir Savic and Erik G Larsson. Fingerprinting-based positioning in distributed massive mimo systems. In *2015 IEEE 82nd vehicular technology conference (VTC2015-Fall)*, pages 1–5. IEEE, 2015.

[36] Zilin Si, Gu Zhang, Qingwei Ben, Branden Romero, Zhou Xian, Chao Liu, and Chuang Gan. Difftactile: A physics-based differentiable tactile simulator for contact-rich robotic manipulation. *ICLR*, 2024.

[37] Christoph Studer, Saïd Medjkouh, Emre Gonultaş, Tom Goldstein, and Olav Tirkkonen. Channel charting: Locating users within the radio environment using channel state information. *IEEE Access*, 6:47682–47698, 2018.

[38] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *NeurIPS*, 2017.

[40] Lihao Zhang, Haijian Sun, Jin Sun, and Rose Qingyang Hu. WiSegRT: Dataset for Site-specific Indoor Radio Propagation Modeling with 3D Segmentation and Differentiable Ray-Tracing. *arXiv:2312.11245*, 2023.

A. Geometric algebra

As representation, Wi-GATr uses the projective geometric algebra $\mathbb{G}_{3,0,1}$. Here we summarize key aspects of this algebra and define the canonical embedding of geometric primitives in it. For a precise definition and pedagogical introduction, we refer the reader to Dorst [13].

Geometric algebra. A geometric algebra $\mathbb{G}_{p,q,r}$ consists of a vector space together with a bilinear operation, the *geometric product*, that maps two elements of the vector space to another element of the vector space.

The elements of the vector space are known as *multivectors*. Their space is constructed by extending a base vector space $\mathbb{R}^d$ to lower orders (scalars) and higher-orders (bi-vectors, tri-vectors, ...). The algebra combines all of these orders (or *grades*) in one 2^d-dimensional vector space. From a basis for the base space, for instance (e_1, e_2, e_3), one can construct a basis for the multivector space. A multivector expressed in that basis then reads, for instance for $d = 3$, $x = x_\emptyset + x_1 e_1 + x_2 e_2 + x_3 e_3 + x_{12} e_1 e_2 + x_{13} e_1 e_3 + x_{23} e_2 e_3 + x_{123} e_1 e_2 e_3$.

The geometric product is fully defined by bilinearity, associativity, and the condition that the geometric product of a vector with itself is equal to its norm. The geometric product generally maps between different grades. For instance, the geometric product of two vectors will consist of a scalar, the inner product between the vectors, and a bivector, which is related to the cross-product of $\mathbb{R}^3$. In particular, the conventional basis elements of grade $k > 1$ are constructed as the geometric product of the vector basis elements e_i. For instance, $e_{12} = e_1 e_2$ is a basis bivector. From the defining properties of the geometric products it follows that the geometric product between orthogonal basis elements is anti-symmetric, $e_i e_j = -e_j e_i$. Thus, for a d-dimensional basis space, there are $\binom{d}{k}$ independent basis elements at grade k.

Projective geometric algebra. To represent three-dimensional objects including absolute positions, we use a geometric algebra based on a base space with $d = 4$, adding a *homogeneous coordinate* to the 3D space.[3] We use a basis (e_0, e_1, e_2, e_3) with a metric such that $e_0^2 = 0$ and $e_i^2 = 1$ for $i = 1, 2, 3$. The multivector space is thus $2^4 = 16$-dimensional. This algebra is known as the projective geometric algebra $\mathbb{G}_{3,0,1}$.

Canonical embedding of geometric primitives. In $\mathbb{G}_{3,0,1}$, we can represent geometric primitives as follows:

- Scalars (data that do not transform under translation, rotations, and reflections) are represented as the scalars of the multivectors (grade $k = 0$).
- Oriented planes are represented as vectors ($k = 1$), encoding the plane normal as well as the distance from the origin.
- Lines or directions are represented as bivectors ($k = 2$), encoding the direction as well as the shift from the origin.
- Points or positions are represented as trivectors ($k = 3$).

For more details, we refer the reader to Tbl. 1 in Brehmer et al. [9], or to Dorst [13].

Wi-GATr tokenization details. Table 1 lists details of the tokenization schemes proposed for wireless propagation modelling.

B. Datasets

Table 2 summarizes major characteristics of the two datasets. In the following we explain more details on data splits and generation.

`Wi3R` dataset. Based on the layouts of the Wi3Rooms dataset by Orekondy et al. [31], we run simulations for 5000 floor layouts that are split into training (4500), validation (250), and test (250). These validation and test splits thus represent generalization across unseen layouts, transmitter, and receiver locations. From the training set, we keep 10 Rx locations as additional test set to evaluate generalization only across unseen Rx locations. To evaluate the generalization performance, we also introduce an out-of-distribution (OOD) set that features four rooms in each of the 250 floor layouts. In all layouts, the interior walls are made of brick while exterior walls are made of concrete. The The Tx and Rx locations are sampled uniformly within the bounds of the floor layouts (10m $\times$ 5m $\times$ 3m).

`WiPTR` dataset. Based on the floor layouts in the ProcTHOR-10k dataset for embodied AI research [12], we extract the 3D mesh information including walls, windows, doors, and door frames. The layouts comprise between 1 to 10 rooms and can cover up to 600 m^2. We assign 6 different dielectric materials for different groups of objects (see Tbl. 3). The 3D Tx and Rx locations are randomly sampled within the bounds of the layout. The training data comprises 10k floor layouts, while test and validation sets each contain 1k unseen layouts, Tx, and Rx locations. Again, we introduce an OOD validation set with 5 layouts where we manually remove parts of the walls such that two rooms become connected. While the multi-modality in combination with the ProcTHOR dataset enables further research for joint sensing and communication in wireless, our dataset set is also, to the best of our knowledge, the first large-scale 3D

[3] A three-dimensional base space is not sufficient to represent absolute positions and translations acting on them in a convenient form. See Brehmer et al. [9], Dorst [13], Ruhe et al. [33] for an in-depth discussion.

Data type	Input parameterization	Tokenization	Channels ($\mathbb{G}_{3,0,1}$ embedding)
3D environment F	• Triangular mesh	1 token per mesh face	• Mesh face center (point) • Vertices (points) • Mesh face plane (oriented plane)
	• Material classes		• One-hot material emb. (scalars)
Antenna t_i / r_i	• Position • Orientation • Receiving / transmitting • Additional characteristics	1 token per antenna	• Position (point) • Orientation (direction) • One-hot type embedding (scalars) • Characteristics (scalars)
Channel h_{ij}	• Antennas	1 token per link	• Tx position (point) • Rx position (point) • Tx-Rx vector (direction)
	• Received power • Phase, delay, ...		• Normalized power (scalar) • Additional data (scalars)

Table 1: Wireless GA tokenizer. We describe how the mesh parameterizing the 3D environment and the information about antennas and their links are represented as a sequence of geometric algebra tokens. The mathematical representation of $\mathbb{G}_{3,0,1}$ primitives like points or orientated planes is described in Appendix A.

	Wi3R	WiPTR
Total Channels	5M	>5.5M
Materials	2	6
Transmitters per layout	5	1-15
Receivers per layout	200	Up to 200
Floor layouts	5k	12k
Simulated frequency	3.5 GHz	3.5 GHz
Reflections	3	6
Transmissions	1	3
Diffractions	1	1
Strongest paths retained	25	25
Antennas	Isotropic	Isotropic
Waveform	Sinusoid	Sinusoid

Table 2: Dataset details and simulation settings for dataset generation.

Object	Material name
Ceiling	ITU Ceiling Board
Floor	ITU Floor Board
Exterior walls	Concrete
Interior walls	ITU Layered Drywall
Doors and door frames	ITU Wood
Windows	ITU Glass

Table 3: Dielectric material properties of objects in WiPTR.

wireless indoor datasets suitable for embodied AI research.

C. Experiments

Models. We use an Wi-GATr model that is 32 blocks deep and 16 multivector channels in addition to 32 additional scalar channels wide. We use 8 attention heads and multi-query attention. Overall, the model has $1.6 \cdot 10^7$ parameters. These settings were selected by comparing five differently sized networks on an earlier version of the Wi3R dataset, though somewhat smaller and bigger networks achieved a similar performance.

Our Transformer model has the same width (translating to 288 channels) and depth as the Wi-GATr model, totalling $16.7 \cdot 10^6$ parameters. These hyperparameters were independently selected by comparing five differently sized networks on an earlier version of the Wi3R dataset.

For SEGNN, we use representations of up to $\ell_{\text{max}} = 3$, 8 layers, and 128 hidden features. The model has $2.6 \cdot 10^5$ parameters. We selected these parameters in a scan over all three parameters, within the ranges used in Brandstetter et al. [8].

The PLViT model is based on the approach introduced by Hehn et al. [16]. We employ the same centering and rotation strategy as in the original approach around the Tx. Further, we extend the original approach to 3 dimensions by providing the difference in z-direction concatenated with the 2D x-y-distance as one token. Since training from scratch resulted in poor performance, we finetuned a ViT-B-16 model pretrained on ImageNet and keeping only the red channel. This resulted in a model with $85.4 \cdot 10^7$ parameters and also required us to use a fixed image size for each dataset that ensures the entire floor layout is visible in the image data.

Optimization. All models are trained on the mean squared error between the model output and the total received power in dBm. We use a batch size of 64 (unless for SEGNN, where we use a smaller batch size due to memory limitations), the Adam optimizer, an initial learning rate of 10^{-3}, and a cosine annealing scheduler. Models are trained for $5 \cdot 10^5$ steps on the Wi3R dataset and for $2 \cdot 10^5$steps on the WiPTR dataset.

Data efficiency and accuracy. In addition to the result

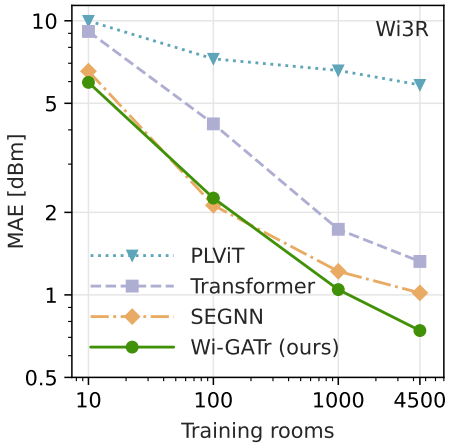

Figure 4: Signal prediction. We show the mean absolute error on the received power as a function of the training data on `Wi3R`. Wi-GATr outperforms the transformer and PLViT baselines at any amount of training data, and scales better to large data or many tokens than SEGNN.

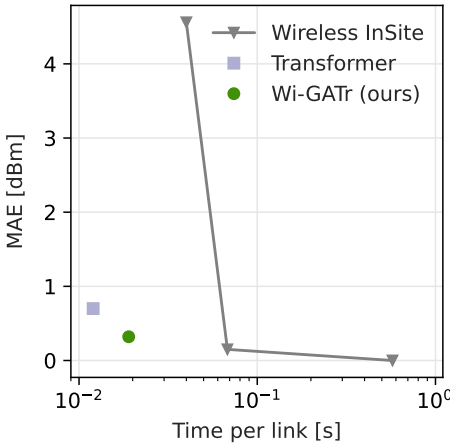

Figure 5: Inference wall time vs signal prediction error per Tx/Rx prediction on the first room of the `WiPTR` validation set.

shown in the paper on `WiPTR`, Fig. 4 shows the prediction results on `Wi3R`. Notably, the SEGNN baseline performs almost as well as our Wi-GATr in the small data regime. As discussed in the paper, however, SEGNN does not scale to large scenes. In Fig. 6 we illustrate the prediction task on a `WiPTR` floor plan. We show signal predictions for the simulator as well as for surrogate models trained on only 100 floor plans. Despite this floor plan not being part of the training set, Wi-GATr is able to capture the propagation pattern well, while the transformer and ViT show memorization artifacts.

In Tbl. 4 we compare surrogate models trained on the full `Wi3R` and `WiPTR` datasets. Both when interpolating Rx positions on the training floor plans as well as when evaluating on new scenes unseen during training, Wi-GATr offers the highest-fidelity approximation of the simulator. Wi-GATr as well as the equivariant baselines are by construction robust to symmetry transformations, while the performance of a vanilla transformer degrades substantially. All methods but SEGNN struggle to generalize to an OOD setting on the `Wi3R` dataset. This is not surprising given that the training samples are so similar to each other. On the more diverse `WiPTR` dataset, Wi-GATr is robust under domain shift.

Inference speed. To quantify the trade-off between inference speed and accuracy of signal prediction, we compare the ray tracing simulation with our machine learning approaches. For this purpose, we evaluate the methods on a single room of the validation set with 2 different Tx locations and two equidistant grids at $z \in \{2.3, 0.3\}$ with each 1637 Rx locations. Figure 5 summarizes the average inference times per link with the corresponding standard deviation. While Wireless InSite ($6/3/1$, i.e., 6 reflections/3 transmissions/1 diffraction) represents our method that was used to generate the ground truth data, it is also by far the slowest approach. Note that we only measure the inference speed of Wireless InSite for each Tx individually without the preprocessing of the geometry. By reducing the complexity, e.g., reducing the number of allowed reflections or transmissions, of the ray tracing simulation the inference time can be reduced significantly. For example, the configuration $3/2/1$ shows a significant increase in inference speed, but at the same time we can already see that the simulation results do not match the ground truth anymore. This effect is even more pronounced for the case of Wireless InSite $3/1/1$. Our machine learning solutions outperform all tested configurations of Wireless InSite in terms of inference speed, while at the same time keeping competitive performance in terms of prediction accuracy (MAE) compared to the data generation simulation itself in a simpler configuration setting.

In addition, the differentiability of ML approches enables them to solve inverse problems and such as finetuning to real-world measurement data. Finetuning, often referred to as calibration, remains challenging for simulation software and will likely lead to increased MAE as the ground truth is not given by Wireless InSite itself anymore.

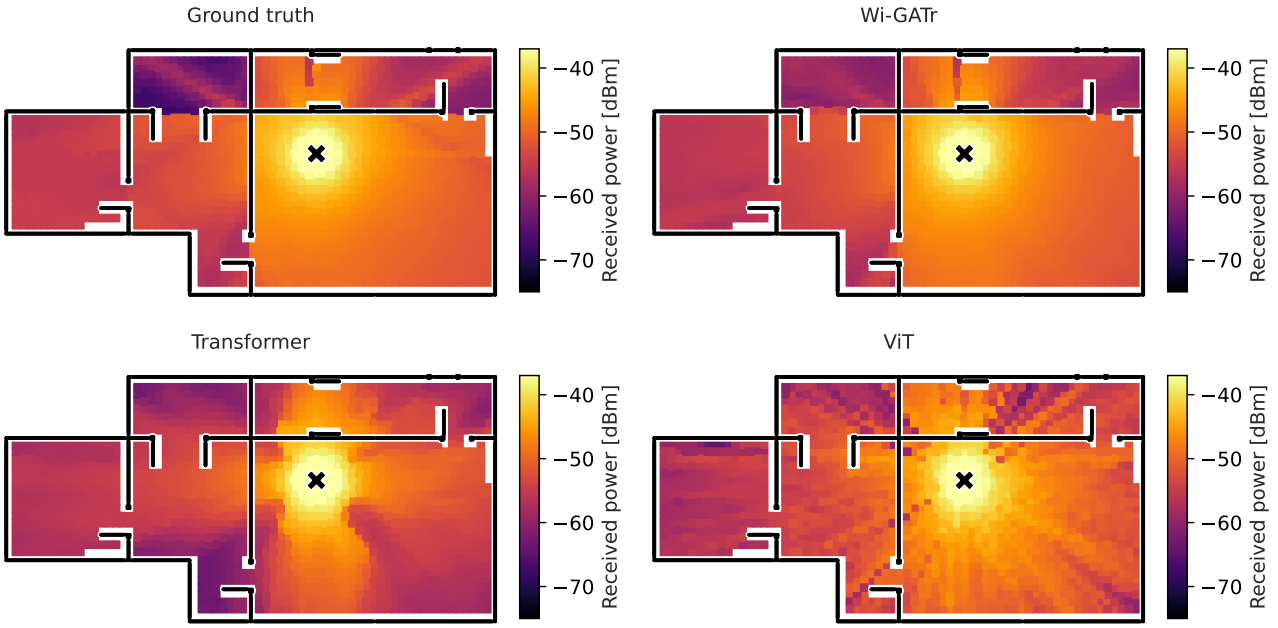

Figure 6: Qualitative signal prediction results. We show a single floor plan from the `WiPTR` test set. The black lines indicate the walls and doors, the colors show the received power as a function of the transmitter location (brighter colours mean a stronger signal). The transmitting antenna is shown as a black cross. The z coordinates of transmitter and receiver are all fixed to the same height. We compare the ground-truth predictions (top left) to the predictions from different predictive models, each trained on only 100 `WiPTR` floor plans. Wi-GATr is able to generalize to this unseen floor plan even with such a small training set.

	`Wi3R` dataset				`WiPTR` dataset		
	Wi-GATr (ours)	Transf.	SEGNN	PLViT	Wi-GATr (ours)	Transf.	PLViT
In distribution							
Rx interpolation	**0.63**	1.14	0.92	4.52	**0.39**	0.62	1.27
Unseen floor plans	**0.74**	1.32	1.02	4.81	**0.41**	0.69	1.28
Symmetry transformations							
Rotation	**0.74**	78.68	1.02	4.81	**0.41**	38.51	1.28
Translation	**0.74**	64.05	1.02	4.81	**0.41**	4.96	1.28
Permutation	**0.74**	1.32	1.02	4.81	**0.41**	0.69	1.28
Reciprocity	**0.80**	1.32	1.01	10.15	**0.41**	0.69	1.28
Out of distribution							
OOD layout	7.03	14.06	**2.34**	5.89	**0.43**	0.86	1.23

Table 4: Signal prediction results. We show the mean absolute error on the received power in dBm (lower is better, best in bold). **Top**: In-distribution performance. **Middle**: Generalization under symmetry transformations. **Bottom**: Generalization to out-of-distribution settings. In almost all settings, Wi-GATr is the highest-fidelity surrogate model.