Geometric Wireless Simulation with Equivariant 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 $\mathrm{E}(3)$ 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 inverse problem solving capabilities using differentiable predictive modelling as well as diffusion models.

1. Introduction

Innovations in modern wireless communication systems all 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 often modelled by state-of-the-art ray-tracing simulation software. However, these simulators take substantial time to evaluate for each scene, cannot be fine-tuned on measurements, and are (usually (Hoydis et al., 2022)) not differentiable. This 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. Several such approaches have been proposed recently, often using image-based representations of the inputs and outputs and off-the-shelf vision architectures (Bakirtzis et al., 2022; Gupta et al., 2022; Lee et al., 2019; Levie et al., 2021; Qiu et al., 2022; Ratnam et al., 2020; Sousa et al., 2022; Tian et al., 2021). However, realistic training data is often scarce in practice, requiring surrogate models to be data efficient. Aside from that, wireless environments feature a variety of geometric input/output data types, including the shape of extended 3D objects, point coordinates, spatial orientation of antennas, and information associated with the transmitted signal.

In this work we propose the *Wireless Geometric Algebra Transformer* (Wi-GATr), a backbone architecture for wireless signal propagation problems. 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. 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 (Brehmer et al., 2023). This architecture is equivariant with respect to the symmetries of wireless channel modelling, but maintains the scalability of a transformer architecture. We introduce two new datasets `Wi3R` and `WiPTR` consisting of diverse indoor wireless scenes and find that Wi-GATr gives us 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. The Wireless Geometric Algebra Transformer (Wi-GATr)

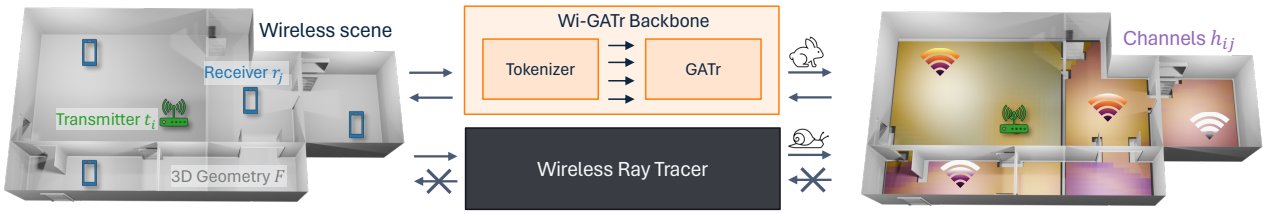
For a short tour of related work, wireless signal propagation, neural surrogates, geometric deep learning, and the GATr architecture, we refer the reader to Appendix A.

Problem formulation. 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:

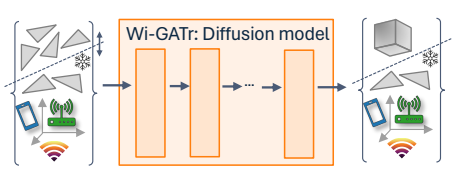
- The 3D geometry F of the environment. We specify it through a triangular mesh with a discrete material

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

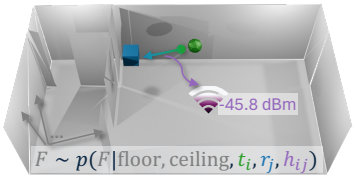
Proceedings of the Geometry-grounded Representation Learning and Generative Modeling at 41st International Conference on Machine Learning, Vienna, Austria. PMLR Vol Number, 2024.



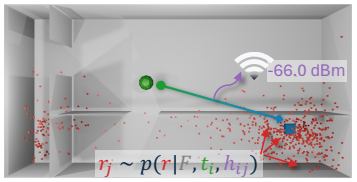
(a) Wi-GATr as a predictive model



(b) Generative model formulation



(c) Conditional generation: 3D geometry



(d) Conditional generation: Receiver location

Figure 1: Geometric surrogates for modelling wireless signal propagation. (a): Predictive modelling of channels from 3D geometry, transmitter, and receiver properties. Wi-GATr is a fast and differentiable surrogate for ray tracers. **(b)**: A probabilistic approach with diffusion models lets us reconstruct 3D environments **(c)** and antenna positions **(d)** from the wireless signal.

class associated with each mesh face.

- 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.
- Analogously, a set of receiving antennas r_i for $i = 1, \dots, n_r$.
- 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 various downstream tasks:

- *Signal prediction* is about predicting the signal received at a single antenna from a single receiver, $p(h|F, t, r)$ with $n_t = n_r = 1$. This is exactly the task that ray-tracing simulators solve. Often, the signal is modelled deterministically as a function $h(F, t, r)$.
- *Receiver localization*: inferring the position and properties of a receiving antenna from one or multiple transmitters, $r \sim p(r|F, \{t_i\}, \{h_i\})$, with $n_r = 1$.
- *Geometry reconstruction* or sensing: reconstructing a 3D environment partially, inferring $p(F_u|F_k, t, r, h)$, where F_u and F_k are the unknown and known subsets of F, respectively.

The latter two problems are examples of *inverse problems*, as they invert the graphical model that simulators are designed for. They are not straightforward to solve with the simulators directly, but we will show how neural surrogates trained on simulator data can solve them.

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

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. A wireless scene consists of the 3D environment mesh F, featuring 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 B. 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. Our tokenization scheme is specified in Tbl. 2 in the appendix.

After tokenizing, we process the input data with a Geometric Algebra Transformer (GATr) (Brehmer et al., 2023). 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 prop-

erty 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.

Predictive modelling. We focus on the prediction of the time-averaged non-coherent received power $h = \sum_p |a_p|^2$, disregarding delay or directional information that may be available in real measurements. We train predictive surrogates $h_\theta(F, t, r)$ that predict the power as a function of the Tx position and orientation t, Rx position and orientation r, and 3D environment mesh F. 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. Compared to a simulator based on ray tracing, h_θ 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.

The differentiability also makes our surrogate model well-suited to solve inverse problems. For instance, we can use it for receiver localization. Given a 3D environment F, transmitters $\{t_i\}$, and corresponding signals $\{h_i\}$, we can find the most likely receiver position and orientation as $\hat{r} = \arg\min_r \sum_i \|h_\theta(F, t_i, r) - h\|^2$, by performing gradient descent through the model.

Probabilistic modelling. Predictive models are deterministic and do not allow us to model stochastic forward processes or express the inherent uncertainty in inverse problems. To overcome this, we draw inspiration from the inverse problem solving capabilities of diffusion models using guidance (Chung et al., 2022). In this case, we formulate the learning problem as a generative modelling task of the joint distribution $p_\theta(F, t, r, h)$ between 3D environment mesh F, transmitter t, receiver r, and channel h, for a single transmitter-receiver pair, using the DDPM framework (Ho et al., 2020). By using an invariant base density and Wi-GATr as an equivariant denoising network, we define an invariant generative model. See Appendix C for a details.

A diffusion model trained to learn the joint density $p_\theta(F, t, r, h)$ does not only allow us to generate unconditional samples of wireless scenes, but also lets us sample from various conditionals: given a partial wireless scene, we can fill in the remaining details using the inpainting approach proposed by Sohl-Dickstein et al. (2015). This lets us solve signal prediction (sampling from $p_\theta(h|F, t, r)$), receiver localization (from $p_\theta(r|F, t, h)$), geometry reconstruction (from $p_\theta(F_u|F_k, t, r, h)$), or any other inference task in wireless scenes. Furthermore, this inference procedure models uncertainty, which is important for inverse problems, where measurements often underspecify the solutions. Finally, to improve the performance of inpainting, we combine training on the unconditional diffusion objective with conditional diffusion objectives by randomly selecting tokens to condition on and evaluate the diffusion loss only on the remaining tokens. See Appendix C for details.

[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.

3. New datasets

While several datasets of wireless simulations and measurements exist (Alkhateeb, 2019; Alkhateeb et al., 2023; Orekondy et al., 2022b; Zhang et al., 2023), they either do not include geometric information, are not diverse, are at a small scale, or the signal predictions are not realistic. 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 (rem) at a frequency of 3.5 GHz.

The first dataset, `Wi3R`, features 5000 floor plans of the same size, the same number of rooms and homogeneous wall materials. Different layouts, Tx and Rx positions are taken from `Wi3Rooms` (Orekondy et al., 2022b). The second dataset, `WiPTR`, is based on the floor layouts in the `ProcTHOR-10k` dataset for embodied AI research (Deitke et al., 2022) and features 12k realistic floor layouts. We extract the 3D mesh information including walls, windows, doors, and door frames and assign 6 different dielectric materials for different groups of objects. For details of the datasets, we refer to the Appendix D.

4. Experiments

Predictive modelling. We compare Wi-GATr against several baselines, including a vanilla transformer (Vaswani et al., 2017) using the same tokenization scheme, the $\mathrm{E}(3)$-equivariant SEGNN (Brandstetter et al., 2022b), though we were only able to fit this model into memory for the `Wi3R` dataset, and an image-based PLViT model, which is a state-of-the-art neural surrogate for wireless scenes (Hehn et al., 2023). Finally, we attempt to compare Wi-GATr also to WiNeRT (Orekondy et al., 2022b), but this architecture was developed to be trained on several measurements on the same floor plan and thus was not able to achieve useful predictions on our diverse datasets. Our experiment setup and the baselines are described in detail in Appendix E.

We study the data efficiency of the surrogates in Fig. 2. Wi-GATr is more data-efficient than any other method with the exception of the $\mathrm{E}(3)$-equivariant SEGNN, which performs

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

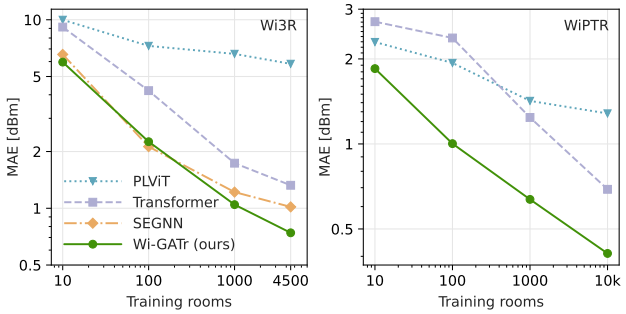


Figure 2: Signal prediction. We show the mean absolute error on the received power as a function of the training data on `Wi3R` (left) and `WiPTR` (right).

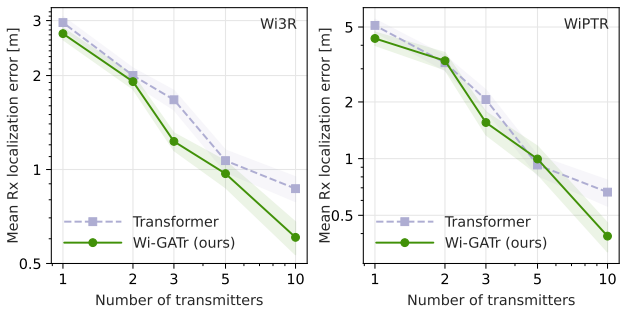


Figure 3: Rx localization error, as a function of the number of Tx, on `Wi3R` (left) and `WiPTR` (right). Lines and error band show mean and its standard error over 240 measurements.

similarly well for a small number of training samples, confirming that equivariance is a useful inductive bias. But Wi-GATr scales better than SEGNN to larger number of samples, showing that our architecture combines small-data advantages of strong inductive biases with the large-data advantages of a transformer architecture. For a visualization of predictions and out of distribution evaluations, we refer to Fig. 5 and Tbl. 5 in the Appendix. Furthermore, we note that both Wi-GATr and a transformer are over a factor of 20 faster than the ground-truth ray tracer (see Appendix E).

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. 2. The performance of our surrogate models is shown in Fig. 3 and Appendix E.[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.

Probabilistic modelling. For the probabilistic approach, we train diffusion models on the `Wi3R` dataset and compare to Wi-GATr to a transformer baseline, as well as a transformer trained on the same data augmented with random rotations. Both models are trained with the DDPM pipeline with 1000 denoising steps and samples from with the DDIM solver (Song et al., 2021a) (see details in Appendix E).

[3]Neither SEGNN nor PLViT are fully differentiable w.r.t. object positions when using the official implementations from Refs. (Brandstetter et al., 2022a; Hehn et al., 2023). We could therefore not use these architectures to infer the transmitter positions.

	Wi-GATr (ours)	Transformer	
		default	data augm.
Canonicalized scenes			
Signal pred.	**1.62**	3.00	15.66
Receiver loc.	**3.64**	8.28	14.42
Geometry reco.	**-3.95**	-3.61	-2.10
Scenes in arbitrary rotations			
Signal pred.	**1.62**	9.57	17.65
Receiver loc.	**3.64**	105.68	14.45
Geometry reco.	**-3.95**	389.34	-2.34

Table 1: Probabilistic modelling results. We show variational upper bounds on the negative log likelihood for different conditional inference tasks (lower is better, best in bold).

For each of the conditional inference tasks described in Sec. 2, we quantitatively evaluate our models through the variational lower bound on the log likelihood of test data under the model. To further analyze the effects of equivariance, we separate testing the model on canonicalized scenes, in which all walls are aligned with the x and y axis, and scenes that are arbitrarily rotated. The results in Tbl. 1 show that Wi-GATr outperforms the transformer baseline across all three tasks, even in the canonicalized setting or when the transformer is trained with data augmentation. The gains of Wi-GATr are particularly clear on the signal prediction and receiver localization problems. Finally, we also show qualitative results for the different inference tasks in Fig. 1, Fig. 7, and Appendix E.2. We observe that Wi-GATr can infer multimodal densities for Rx localization, as well as sample diverse floorplans consistent with the transmitted signals.

5. Conclusion

Wireless signal transmission through electromagnetic wave propagation is an inherently geometric problem. We developed Wi-GATr, a class of neural models grounded in geometric representations and strong inductive biases. In our experiments, we demonstrated the data efficiency of our approach as well as the inverse problem solving capabilities using gradient descent and generative modelling. While deeper analysis in terms of data diversity (e. g., more materials, outdoor scenarios) and realistic target signals (e. g., CIR) is needed, Wi-GATr highlights the benefits of a geometric treatment of wave propagation modelling on two novel geometric datasets of diverse wireless scenes.

Acknowledgements. We thank Suresh Sharma, Pim de Haan, and Jens Petersen for fruitful discussions.

References

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

3GPP TR 38.901. Study on channel model for frequencies from 0.5 to 100 ghz. Standard, 3GPP, Valbonne, FR, March 2022.

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

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

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

Stefanos Bakirtzis, Kehai Qiu, Jie Zhang, and Ian Wassell. DeepRay: Deep Learning Meets Ray-Tracing. In *EuCAP*, 2022.

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

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*, 2022b.

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

Johann Brehmer, Joey Bose, Pim De Haan, and Taco S Cohen. Edgi: Equivariant diffusion for planning with embodied agents. *NeurIPS*, 2024.

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

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv:2209.14687*, 2022.

William Kingdon Clifford. Applications of Grassmann's Extensive Algebra. *Amer. J. Math.*, 1(4):350–358, 1878.

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

Tri Dao, Dan Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. FlashAttention: Fast and memory-efficient exact attention with IO-awareness. *NeurIPS*, 2022.

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.

Sebastian Dörner, Marcus Henninger, Sebastian Cammerer, and Stephan ten Brink. Wgan-based autoencoder training over-the-air. In *SPAWC*, 2020.

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

Yilun Du, Conor Durkan, Robin Strudel, Joshua B Tenenbaum, Sander Dieleman, Rob Fergus, Jascha Sohl-Dickstein, Arnaud Doucet, and Will Sussman Grathwohl. Reduce, reuse, recycle: Compositional generation with energy-based diffusion models and mcmc. In *ICML*, 2023.

Hermann Grassmann. *Die lineale Ausdehnungslehre*. Otto Wigand, Leipzig, 1844.

Ankit Gupta, Jinfeng Du, Dmitry Chizhik, Reinaldo A Valenzuela, and Mathini Sellathurai. Machine learning-based urban canyon path loss prediction using 28 ghz manhattan measurements. *IEEE Transactions on Antennas and Propagation*, 2022.

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.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 2020.

Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J Fleet, et al. Imagen video: High definition video generation with diffusion models. *arXiv:2210.02303*, 2022.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *ICML*, 2022.

Sepidehsadat Sepid Hossieni, Mohammad Amin Shabani, Saghar Irandoust, and Yasutaka Furukawa. Puzzlefusion: Unleashing the power of diffusion models for spatial puzzle solving. *NeurIPS*, 2024.

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.

Michael Janner, Yilun Du, Joshua B Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. *arXiv:2205.09991*, 2022.

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

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *ICLR*, 2014.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: exact likelihood generative learning for symmetric densities. In *ICML*, 2020.

JunG-Yong Lee, Min Young Kang, and Seong-Cheol Kim. Path Loss Exponent Prediction for Outdoor Millimeter Wave Channels through Deep Learning. In *WCNC*, 2019.

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

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022.

T L Marzetta and B M Hochwald. Fast transfer of channel state information in wireless systems. *IEEE Transactions on Signal Processing*, 2006.

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.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *ICML*, 2021.

Tribhuvanesh Orekondy, Arash Behboodi, and Joseph B Soriaga. Mimo-gan: Generative mimo channel modeling. In *ICC*, 2022a.

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*, 2022b.

Timothy J O'Shea, Tamoghna Roy, and Nathan West. Approximating the void: Learning stochastic channel models from observation with variational generative adversarial networks. In *ICNC*, 2019.

William Peebles and Saining Xie. Scalable diffusion models with transformers. *arXiv:2212.09748*, 2022.

Kehai Qiu, Stefanos Bakirtzis, Hui Song, Jie Zhang, and Ian Wassell. Pseudo Ray-Tracing: Deep Leaning Assisted Outdoor mm-Wave Path Loss Prediction. *IEEE Wireless Communications Letters*, 2022.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv:2204.06125*, 1(2):3, 2022.

Vishnu V Ratnam, Hao Chen, Sameer Pawar, Bingwen Zhang, Charlie Jianzhong Zhang, Young-Jin Kim, Soonyoung Lee, Minsung Cho, and Sung-Rok Yoon. Fadenet: Deep learning-based mm-wave large-scale channel fading prediction and its applications. *IEEE Access*, 2020.

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

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *ICLR*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *ICLR*, 2021b.

Marco Sousa, Pedro Vieira, Maria Paula Queluz, and António Rodrigues. An Ubiquitous 2.6 GHz Radio Propagation Model for Wireless Networks using Self-Supervised Learning from Satellite Images. *IEEE Access*, 2022.

Yu Tian, Shuai Yuan, Weisheng Chen, and Naijin Liu. Transformer based radio map prediction model for dense urban environments. In *ISAPE*, 2021.

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

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.

Pascal Vincent. A Connection Between Score Matching and Denoising Autoencoders. *Neural computation*, 23(7): 1661–1674, 2011.

Hao Ye, Geoffrey Ye Li, Biing-Hwang Fred Juang, and Kathiravetpillai Sivanesan. Channel Agnostic End-to-End Learning Based Communication Systems with Conditional GAN. In *IEEE Globecom Workshops*, 2018.

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.

Xiaopeng Zhao, Zhenlin An, Qingrui Pan, and Lei Yang. NeRF2: Neural Radio-Frequency Radiance Fields. In *ACM MobiCom*, 2023.

A. Background and related work

Wireless signal propagation. How do wireless signals propagate from a transmitting antenna (Tx) to a receiver antenna (Rx) in a (static) 3D environment? While the system is fundamentally described by Maxwell's equations, for many realistic problems the ray approximation of geometric optics suffices (Keller). It approximates the solution to Maxwell's equations as a sum of planar waves propagating in all directions from Tx. Each planar wave is represented as a ray, characterized by various attributes (e. g., power, phase, delay) since transmission. As a ray reaches an object—that is, it intersects with its mesh—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. In addition, new rays can emanate from the point of interaction. After multiple interactions, the rays eventually reach the receiving antenna. 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 (Tse and Viswanath, 2005).

Maxwell's equations and in extension ray propagation are highly symmetric. The received signal does not change under rotations, translations, and reflections of the whole scene, as well as the exchange of transmitter and receiver. The latter property is known as reciprocity (Marzetta and Hochwald, 2006).

Wireless simulators. Wireless propagation models play a key role in design and evaluation of communication systems, for instance by characterizing the gain of competitive designs in *realistic* settings or by optimizing systems performance as in base station placement for maximal coverage. Statistical approaches (3GPP TR 38.901) represent propagation as a generative model where the parameters of a probabilistic model are fitted to measurements. On the other hand, wireless ray-tracing approaches (rem; Amiot et al., 2013; Hoydis et al., 2022) are increasingly popular due to their high accuracy and because they do not require expensive field measurement collection campaigns.

Neural wireless simulations. Both statistical and ray-tracing simulation techniques are accompanied by their own shortcomings, subsequently mitigated by their neural counterparts. Neural surrogates for statistical models (Dörner et al., 2020; Orekondy et al., 2022a; O'Shea et al., 2019; Ye et al., 2018) reduce the amount and cost of measurements required. Neural ray tracers (Hoydis et al., 2022; Orekondy et al., 2022b; Zhao et al., 2023) address the non-differentiability of simulators using a NeRF-like strategy (Mildenhall et al., 2020) by parameterizing the scene using a spatial MLP and rendering wireless signals using classic ray-tracing or volumetric techniques. While these techniques are faster than professional ray tracers, they are similarly bottlenecked by expensive bookkeeping and rendering steps (involving thousands of forward passes). In contrast, we propose a framework to 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. The growing field of geometric deep learning (Bronstein et al., 2021) aims to incorporate structural properties of a problem into neural network architectures and algorithms. A central concept is *equivariance* to symmetry groups (Cohen, 2021): a network $f(x)$ is equivariant with respect to a group G if its outputs transform consistently with any symmetry transformation $g \in G$ of the inputs, $f(g \cdot x) = g \cdot f(x)$, where $\cdot$ denotes the group action. Of particular interest to us is the Euclidean group $\mathrm{E}(3)$ of isometries of 3D space, that is, transformations that leave Euclidean distances invariant. This group includes spatial translations, rotations, reflections, and their combinations. As we argued above, the physics of wireless signal propagation are invariant under this group.

GATr. The Geometric Algebra Transformer (GATr) (Brehmer et al., 2023) is an $\mathrm{E}(3)$-equivariant architecture for geometric problems. Among equivariant architectures, it stands out in two ways. First, it uses geometric (or Clifford) algebras (Clifford, 1878; Grassmann, 1844) as representations. For a rigorous introduction to these algebras, we refer the reader to Dorst (2020). From a practical machine learning perspective, these algebras define embeddings for various geometric primitives like 3D points, planes, or $\mathrm{E}(3)$ transformations. We will show that this representation is particularly well-suited for wireless channel modelling. Second, GATr is a transformer architecture (Vaswani et al., 2017). It computes the interactions between multiple tokens through scaled dot-product attention. With efficient backends like FlashAttention (Dao et al., 2022), the architecture is scalable to large systems, without any restrictions on the sparsity of interactions like in message-passing networks.

Diffusion models. Diffusion models (Ho et al., 2020; Sohl-Dickstein et al., 2015; Song et al., 2021b) are a class of generative models that iteratively invert a noising process. They have become the de-facto standard in image and video generation (Ho et al., 2022; Ramesh et al., 2022). Recently, they have also shown to yield promising results in the generation of spatial and sequential data, such as in planning (Janner et al., 2022) and puzzle solving (Hossieni et al., 2024). Aside from their generative modelling capabilities, diffusion models provide a flexible way for solving inverse problems (Chung et al., 2022; Lugmayr et al., 2022) through multiplication with an appropriate likelihood term (Sohl-Dickstein

et al., 2015). Furthermore, by combining an invariant prior distribution with an equivariant denoising network, one obtains equivariant diffusion models (Köhler et al., 2020). These yield a sampling distribution that assigns equal probability to all symmetry transformations of an object, which can improve performance and data efficiency in symmetry problems like molecule generation (Hoogeboom et al., 2022) and planning (Brehmer et al., 2024). We will demonstrate similar benefits in modelling wireless signal propagation.

B. 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 (2020).

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.[4] 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. (2023), or to Dorst (2020).

C. Probabilistic model

Formally, we employ the standard DDPM framework (Song et al., 2021b) to train a latent variable model $p_\theta(\mathbf{x}_0) = \int p_\theta(\mathbf{x}_{0:T}) d_{\mathbf{x}_{1:T}}$, where $\mathbf{x}_0 = [rsrp, \mathbf{tx}, \mathbf{rx}, \mathbf{mesh}]$ denotes the joint vector of variables following the dataset distribution $p_{data}(\mathbf{x}_0)$. In DDPM, the latent variables $\mathbf{x}_{1:T}$ are noisy versions of the original data, defined by a discrete forward noise process $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}\right)$ and $\beta_i > 0$. We approximate the reverse distribution $q(\mathbf{x}_{t-1}\mathbf{x}_t)$ with $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \sum_{\hat{\mathbf{x}}_0} q(\mathbf{x}_{t-1}|\mathbf{x}_t, \hat{\mathbf{x}}_0) p_\theta(\hat{\mathbf{x}}_0|\mathbf{x}_t, t)$, where $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ is a normal distribution with closed-form parameters (Ho et al., 2020). The forward and backward distributions q and p form a variational auto-encoder (Kingma and Welling, 2014) which can be trained with a variational lower bound loss. Using the above parametrization of $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$, however, allows for a simple approximation of this lower bound by training on an MSE objective $\mathcal{L} = \mathbb{E}_{\mathbf{x}_t, \mathbf{x}_0}\left[||f_\theta(\mathbf{x}_t, t) - \mathbf{x}_0||^2\right]$ which resembles denoising score matching (Vincent, 2011).

To parametrize $p_\theta(\hat{\mathbf{x}}_0|\mathbf{x}_t, t)$, we pass the raw representation of $\mathbf{x}_t$ through the wireless GA tokenizer of Wi-GATr and, additionally, we embed the scalar t through a learned timestep embedding (Peebles and Xie, 2022). The embedded timesteps can then be concatenated along the scalar channels in the GA representation in a straightforward manner. Similar to GATr (Brehmer et al., 2023), the neural network outputs a prediction in the GA representation, which is subsequently converted to the original latent space. Note

[4] 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. (2023); Dorst (2020); Ruhe et al. (2023) 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		• Normalized power (scalar)
	• Phase, delay, ...		• Additional data (scalars)

Table 2: 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 B.

that this possibly simplifies the learning problem, as the GA representation is inherently higher dimensional than our diffusion space with the same dimensionality as $\mathbf{x}_0$.

Equivariant generative modelling. A diffusion model with an invariant base density and an equivariant denoising network defines an invariant density, but equivariant generative modelling has some subtleties (Köhler et al., 2020). Because the group of translations is not compact, we cannot define a translation-invariant base density. Previous works have circumvented this issue by performing diffusion in the zero center of gravity subspace of euclidean space (Hoogeboom et al., 2022). However, we found that directly providing the origin as an additional input to the denoising network also resulted in good performance, at the cost of full E(3) equivariance. We also choose to generate samples in the convention where the z-axis represents the direction of gravity and positive z is "up"; we therefore provide this direction of gravity as an additional input to our network.

Masking strategies. To improve the performance of conditional sampling, we randomly sample conditioning masks during training which act as an input to the model, as well as a mask on the loss terms. Namely, we sample masks from a discrete distribution with probabilities $p = (0.2, 0.3, 0.2, 0.3)$ corresponding to masks for unconditional, signal, receiver and mesh prediction respectively. If we denote this distribution over masks as $p(m)$, the modified loss function then reads as $\mathcal{L} = \mathbb{E}_{\mathbf{m}\sim p(\mathbf{m}),\mathbf{x}_t,\mathbf{x}_0}\left[||\mathbf{m} \odot f_\theta(\mathbf{x}_t^{\mathbf{m}}, t, \mathbf{m}) - \mathbf{m} \odot \mathbf{x}_0||^2\right]$, where $\mathbf{x}_t^{\mathbf{m}}$ is equal to $\mathbf{x}_0$ along the masked tokens according to $\mathbf{m}$.

D. Datasets

Table 3 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. (2022b), 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 (Deitke et al., 2022), 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. 4). 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 wireless indoor datasets suitable for embodied AI research.

	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 3: 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 4: Dielectric material properties of objects in WiPTR.

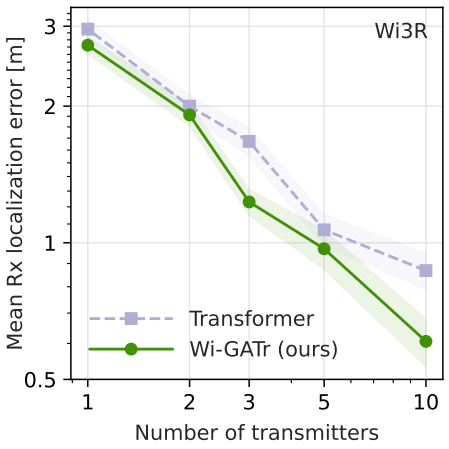


Figure 4: Rx localization error, as a function of the number of Tx. Lines and error band show mean and its standard error over 240 measurements.

E. Experiments

E.1. Predictive modelling

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. (2022b).

The PLViT model is based on the approach introduced by Hehn et al. (2023). 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.

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 6 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

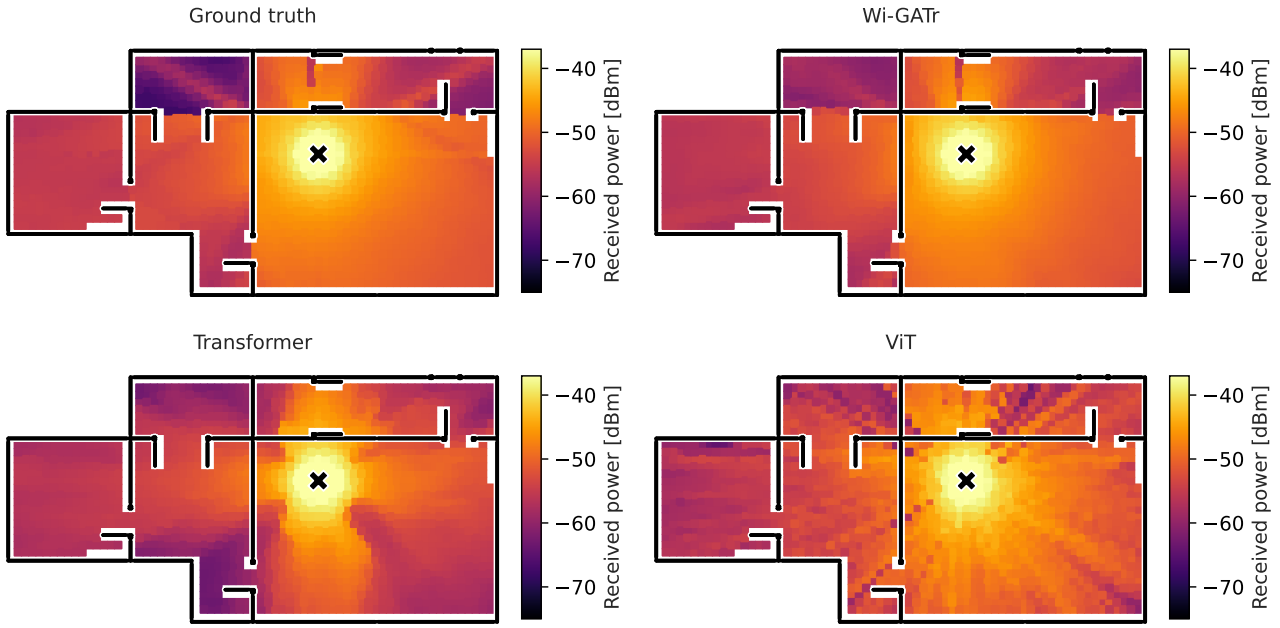


Figure 5: Qualitative signal prediction results. We show a single floor plan from the `WiPTR` test set. The black lines indicate the walls and doors, the colors show the received power as a function of the transmitter location (brighter colours mean a stronger signal). The transmitting antenna is shown 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 5: 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.

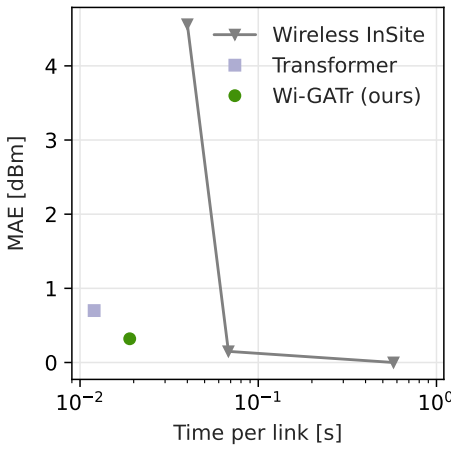


Figure 6: Inference wall time vs signal prediction error per Tx/Rx prediction on the first room of the `WiPTR` validation set.

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.

E.2. Probabilistic modelling

Experiment setup. For all conditional samples involving $p(F_u|F_k, t, r, h)$, we always choose to set F_k to be the floor and ceiling mesh faces only and F_u to be the remaining geometry. This amounts to completely predicting the exterior walls, as well as the separating walls/doors of the three rooms, whereas the conditioning on F_k acts only as a mean to break equivariance. Since F is always canonicalized in the non-augmented training dataset, this allows for direct comparison of variational lower bounds in Tbl. 1 with the non-equivariant transformer baseline.

Models. For both Wi-GATr and the transformer baseline, we follow similar architecture choices as for the predictive models, using an equal amount of attention layers. To make the models timestep-dependent, we additionally employ a standard learnable timestep embedding commonly used in diffusion transformers (Peebles and Xie, 2022) and concatenate it to the scalar channel dimension.

Optimization. We use the Adam optimizer with a learning rate of 10^{-3} for the Wi-GATr models. The transformer models required a smaller learning rate for training stability, and thus we chose $3 \cdot 10^{-4}$. In both cases, we linearly anneal the learning rate and train for $7 \cdot 10^5$ steps with a batchsize of 64 and gradient norm clipping set to 100.

Evaluation. We use the DDIM sampler using 100 timesteps for visualizations in Fig. 7 and for the error analysis in Fig. 8. To evaluate the variational lower bound in Tbl. 1, we follow (Nichol and Dhariwal, 2021) and evaluate $L_{vlb} := L_0 + L_1 + \dots L_T$, where $L_0 := -\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)$, $L_{t-1} := D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))$ and $L_T := D_{KL}(q(\mathbf{x}_T|\mathbf{x}_0), p(\mathbf{x}_t))$. To be precise, for each sample $\mathbf{x}_0$ on the test set, we get a single sample $\mathbf{x}_t$ from q and evaluate L_{vlb} accordingly. Table 1 reports the mean of all L_{vlb} evaluations over the test set.

Additional results. Fig. 8, shows the quality of samples from $p_\theta(h|F, t, r)$ as a function of the amount of available training data, where we average over 3 samples for each conditioning input. It is worth noting that diffusion samples have a slightly higher error than the predictive models. This shows that the joint probabilistic modelling of the whole scene is a more challenging learning task than a deterministic forward model.

To further evaluate the quality of generated rooms, we analyze how often the model generates walls between the receiver and transmitter, compared to the ground truth. Precisely, we plot the distribution of received power versus the distance of transmitter and receiver in Fig. 9 and color each point according to a line of sight test. We can see that, overall, Wi-GATr has an intersection error of 0.26, meaning that in 26% of the generated geometries, line of sight was occluded, while the true geometry did not block line of sight between receiver and transmitter. This confirms that the diffusion model correctly correlates the received power and receiver/transmitter positions with physically plausible geometries. While an error of 26% is non-negligible, we note that this task involves generating the whole geometry given only a single measurement of received power, making the problem heavily underspecified. Techniques such as compositional sampling (Du et al., 2023) could overcome this limitation by allowing to condition on multiple receiver and received power measurements.

F. Discussion

Progress in wireless channel modelling is likely to lead to societal impact. Not all of it is positive. The ability to reconstruct details about the propagation environment may have privacy implications. Wireless networks are ubiquitous and could quite literally allow to see through walls. At the same time, we believe that progress in the development of wireless channel models may help to reduce radiation ex-

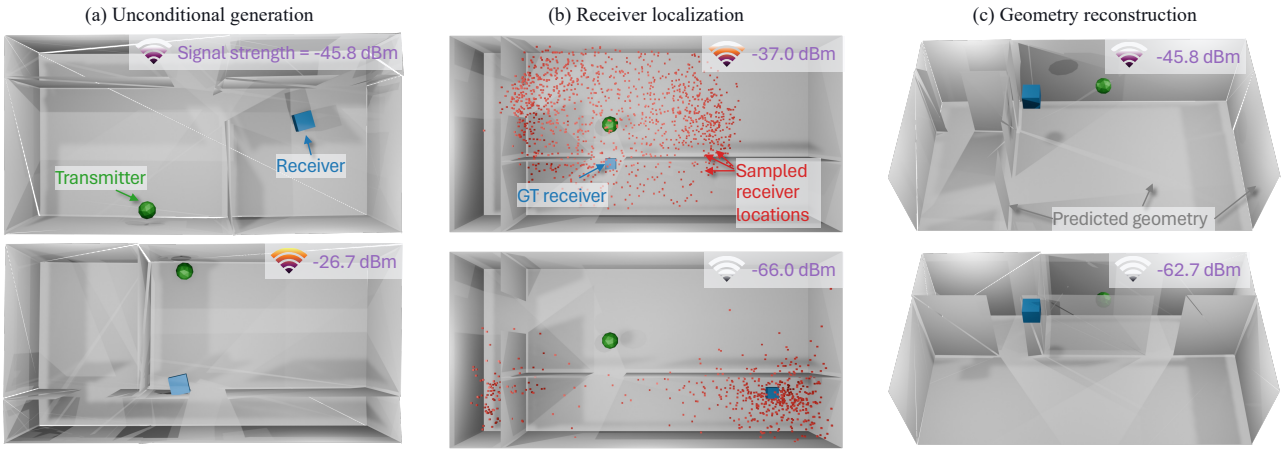


Figure 7: Probabilistic modelling. We formulate various tasks as sampling from the unconditional or conditional densities of a single diffusion model. **(a)**: Unconditional sampling of wireless scenes $p(F, t, r, h)$. **(b)**: Receiver localization as conditional sampling from $p(r|F, t, h)$ for two different values of h and r. **(c)**: Geometry reconstruction as conditional sampling from $p(F_u|F_k, t, r, h)$ for two different values of h, keeping t, r, F_k fixed.

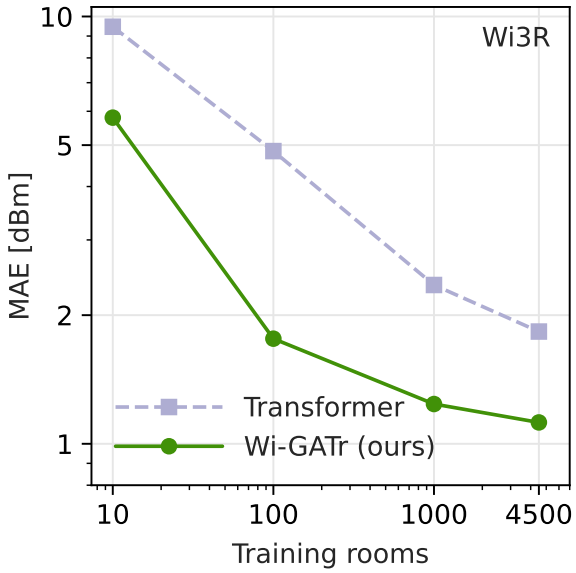


Figure 8: Mean absolute errors of received power as a function of number of training rooms for conditional diffusion model samples.

posure and power consumption of wireless communication systems, and generally contribute to better and more accessible means of communication.

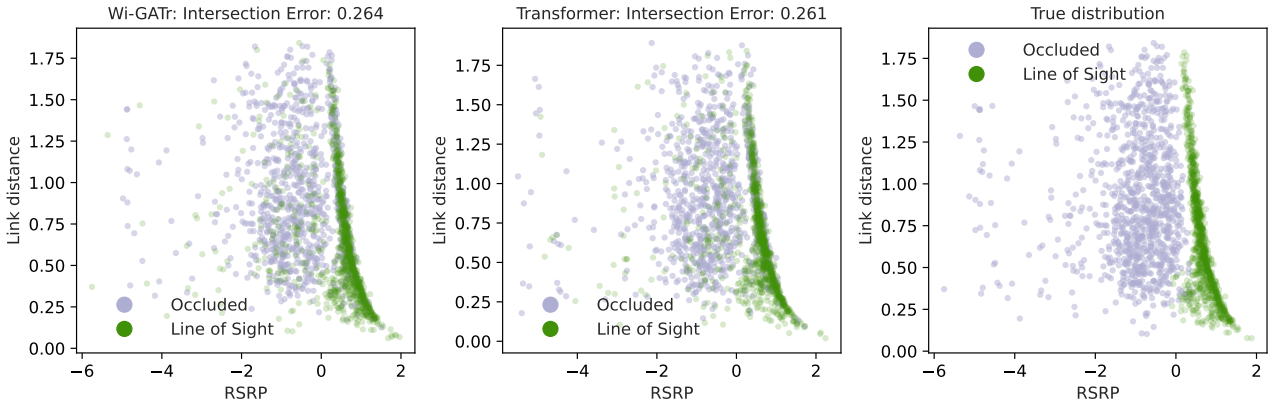


Figure 9: A scatter plot of normalized received power versus normalized distance between receiver and transmitter. Each point is colored depending on having line of sight between the receiver and transmitter given the room geometry. Left: The geometry used for calculating line of sight is given by conditional diffusion samples using Wi-GATr. Middle: The geometry used for calculating line of sight is given by transformer samples. Right: The geometry used for calculating line of sight is taken from the test data distribution.