

Spatial Intelligence for Wireless Communications and Sensing: Methods to Applications

Speakers:

Prof. Jun Zhang, Dr. Jiajia Guo, Dr. Yumeng Zhang, Dr. Chenghong Bian

**Special thanks to: Chaozheng Wen
Acknowledgement to: Yijie Bian, Ziqi Ke**

About us

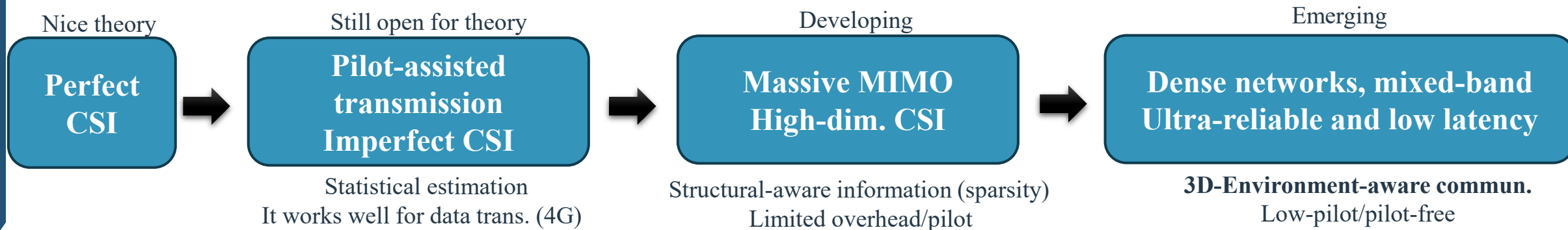
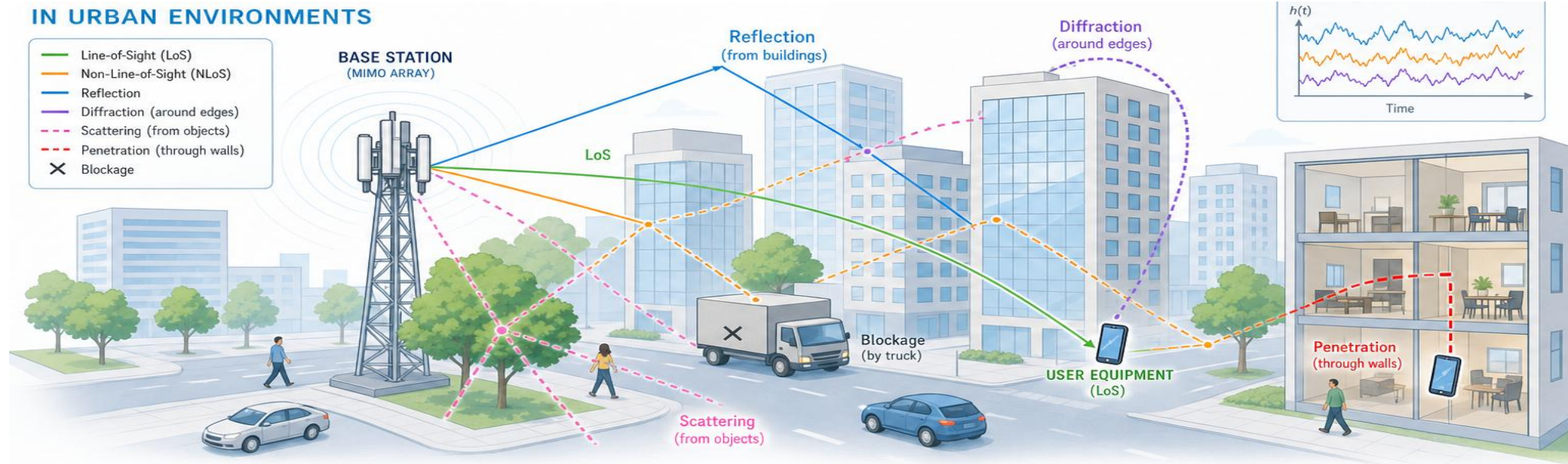
Our group (Wechat card)



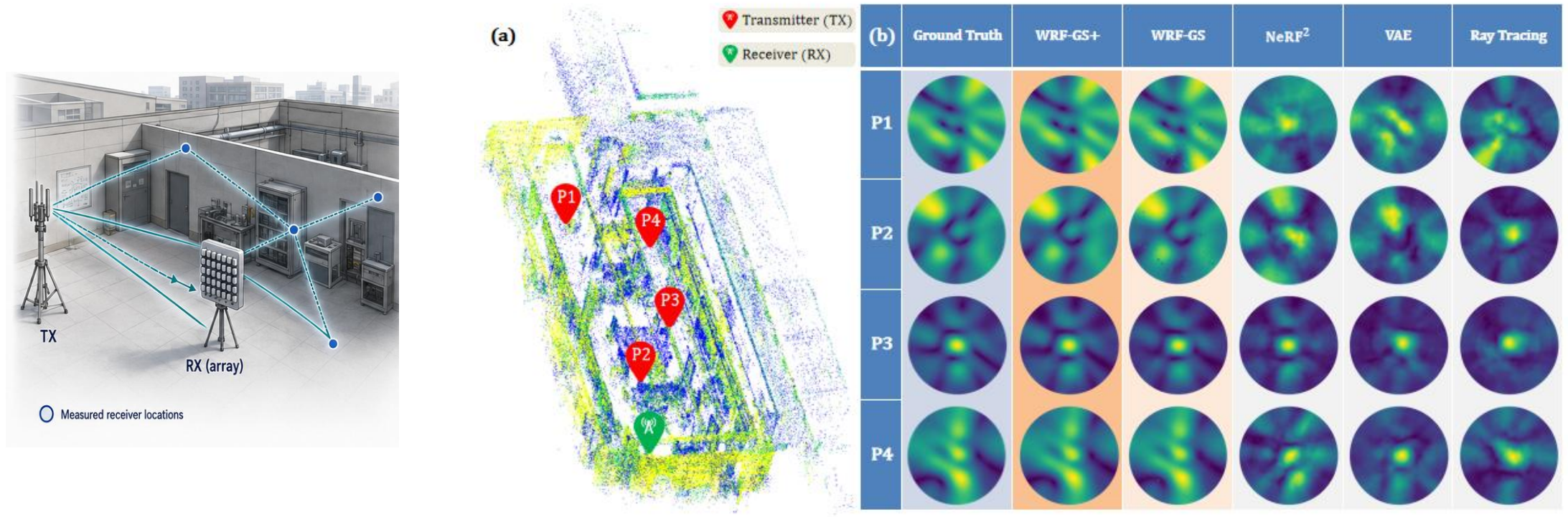
Tutorial Website



The 'LAST' challenge of wireless communications

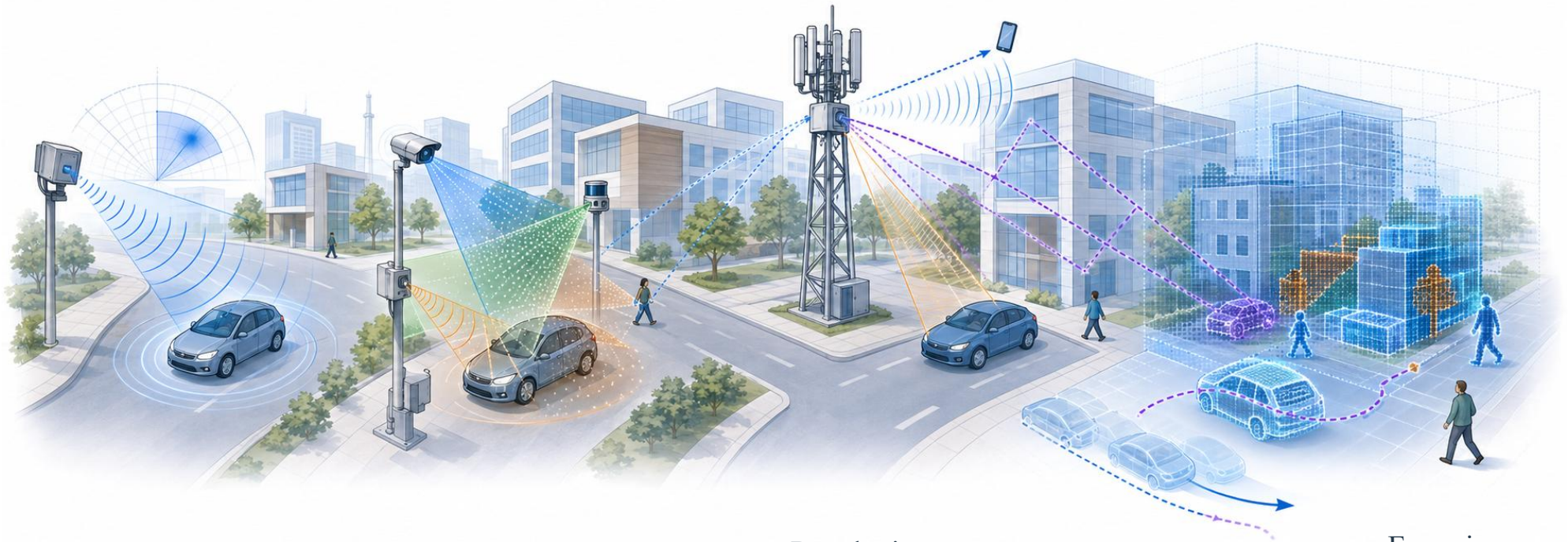


A Comm. example - Learning wireless field, and recover power map at unseen Rx's



Wen C, Tong J, Hu Y, et al. *Neural representation for wireless radiation field reconstruction: A 3D Gaussian splatting approach*[J]. IEEE Transactions on Wireless Communications, 2025.

The 'LAST' challenge of sensing



Nice theory

**Single
target
state est.**

Detection &
Localization

Still open for theory

**Spatial imaging &
Multi-View perception**

SAR, MIMO radar,
multi-static sensing

Developing

**Network-wide multi-
modal sensing**

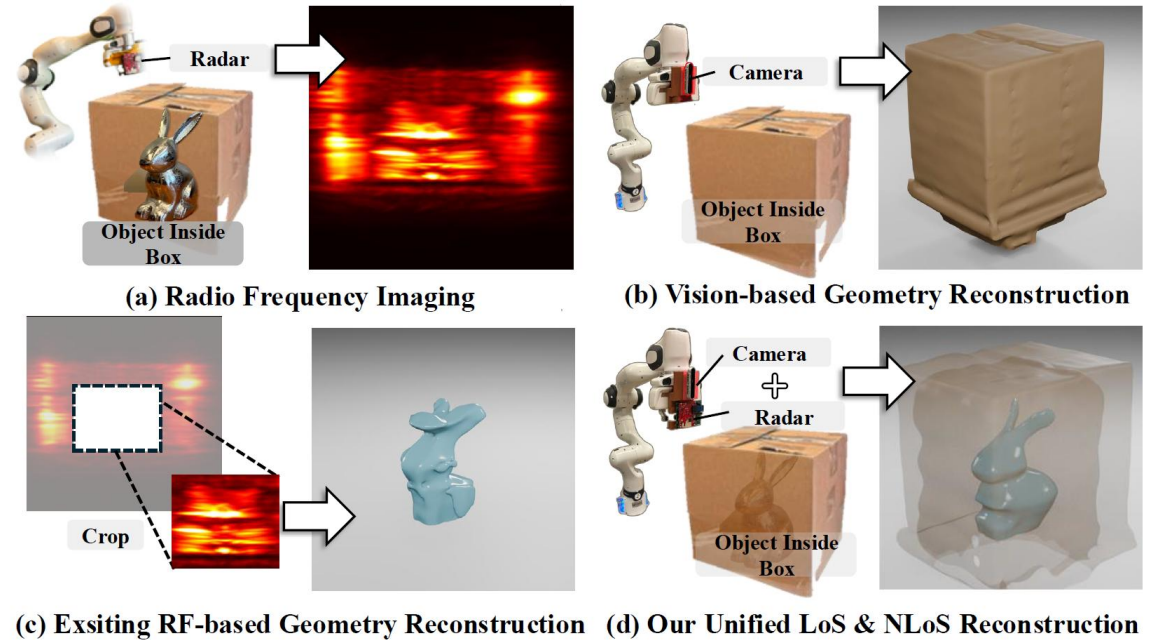
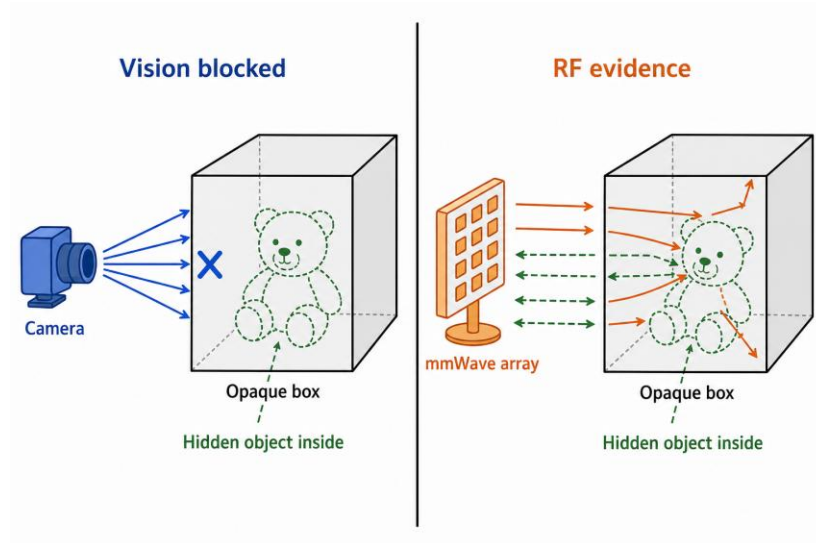
Distributed radar, communication
infrastructure, camera, LiDAR

Emerging

Persistent spatial world model

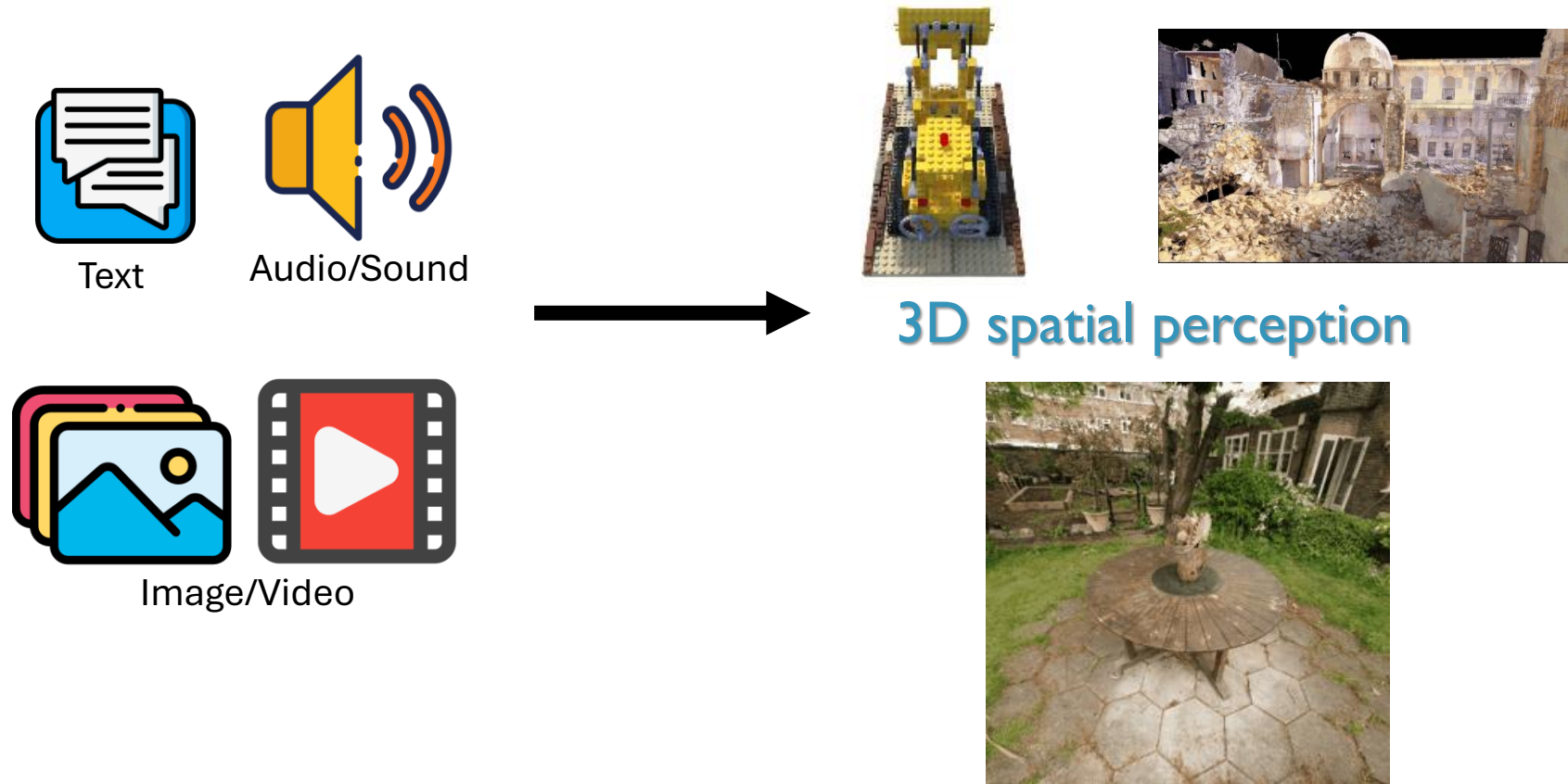
**3D-Environment-aware reconstruction,
Semantics, prediction, reasoning**

A Sensing example: NLoS sensing turns multipath into spatial evidence



Lu J, Shanbhag H, Hassanieh H A. *Seeing through boxes: Non-Line-of-Sight 3D Reconstruction from Radar Signals*[J]. arXiv preprint arXiv:2605.29098, 2026.

3D spatial representation advances in computer science society



Spatial Perception: From text, audio, plane image/video, to 3D multi-view perception

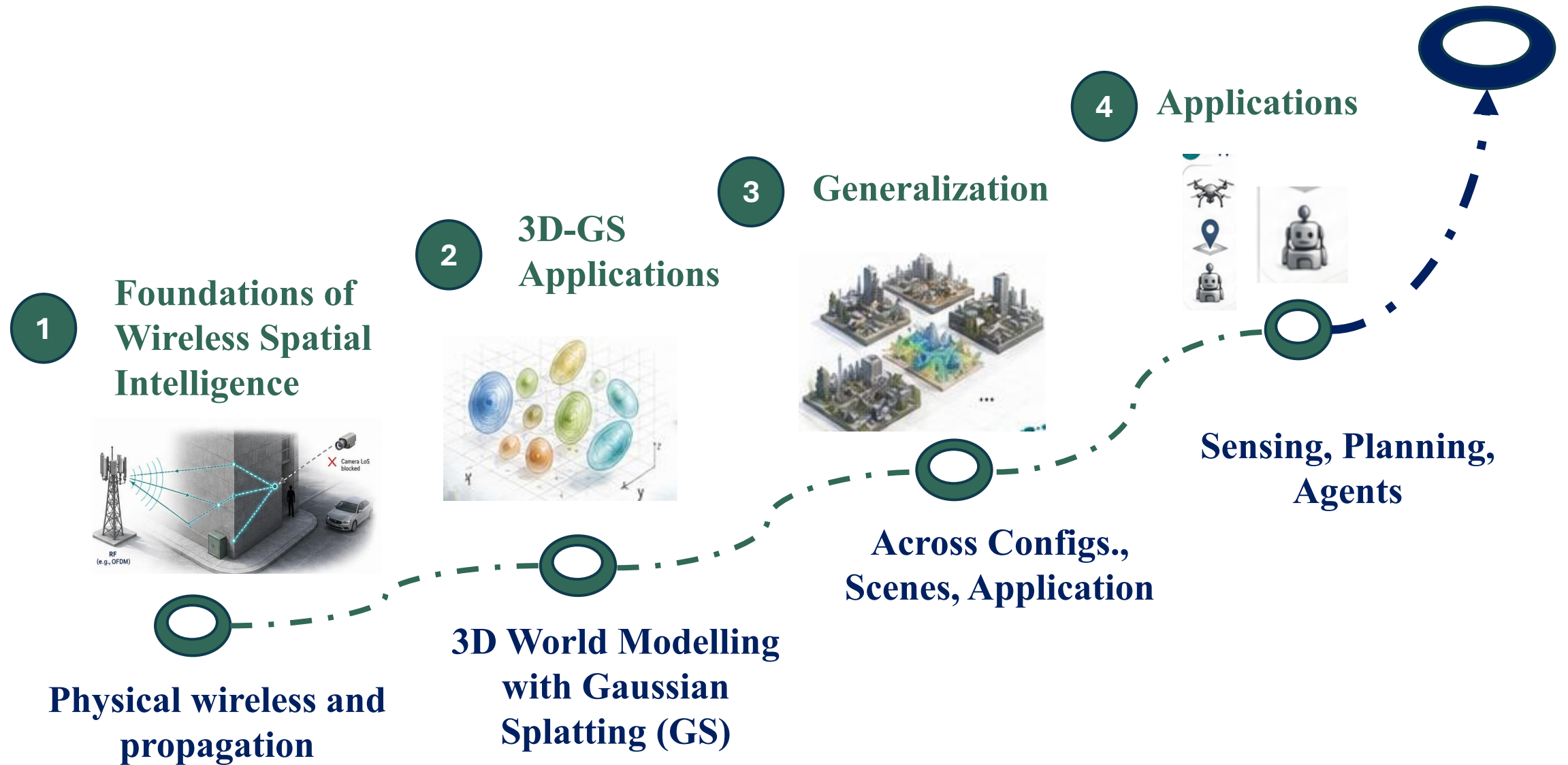
Big perspective: multimodal world model for wireless



Marble: A Multimodal World Model

A world model maintains a queryable and evolving representation of the physical environment.

Roadmap





Part I

Vision for Wireless Spatial Intelligence



Why is reusable spatial knowledge needed?



What is the wireless spatial field?



How should wireless space be represented?

Wireless intelligence begins with knowledge of the physical environment

Two demanding trends

Wireless communications & sensing



- Beam control
- Localization
- Resource Allocation
- Scheduling

AI-native wireless networks



- Prediction
- Optimization
- Planning
- Embodied action

Task-specific knowledge is a projection of shared scene entities and physical attributes—not a collection of independent worlds.



Communications

Uplink/Downlink Channel response (CSI)
UE Loc
Time/Spatial Correlation

SENSING & LOCALIZATION

Anchor Loc
Delay/Angle
Visibility

AI-native

3D Mesh
Uncertainty Fields
Radio Power



SHARED PHYSICAL ENVIRONMENT

Objects & layout

buildings · users · vehicles

Geometry & materials

surfaces · reflectivity · blockage

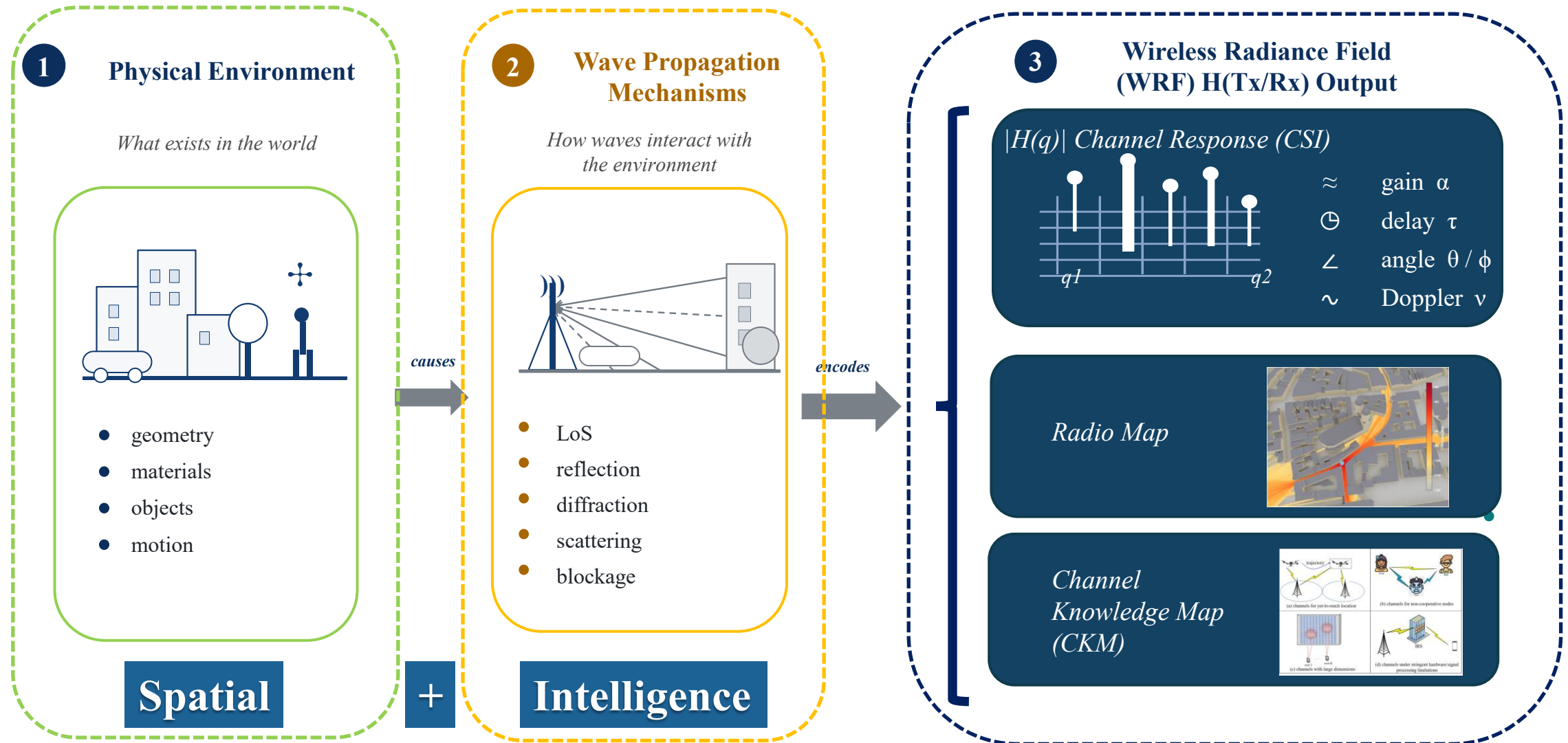
Motion & visibility

trajectories · dynamics · LoS/NLoS

Radio sources & sensors

BS/AP/anchors · arrays · frequency

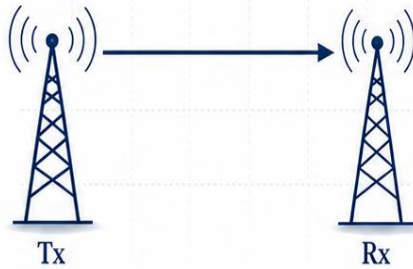
What is spatial intelligence?



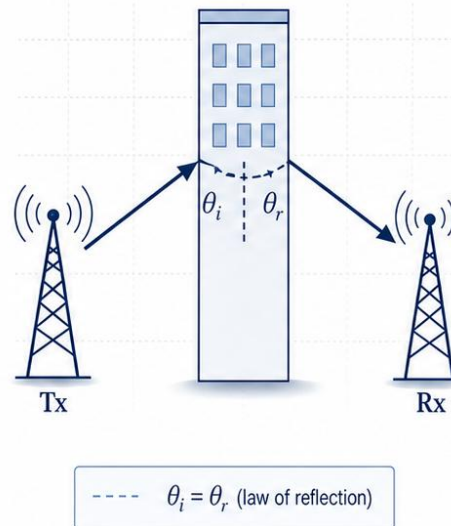
Physics of spatial intelligence: A wireless channel is the coherent sum of multipath components

$$H(f, m, n, t) = \sum_l \alpha(l, t) R_x(\theta_{l,t}, \phi_{l,t}, m) T_x(\theta_{l,t}, \phi_{l,t}, m) \exp\{-j2\pi f \tau_{l,t}\}$$

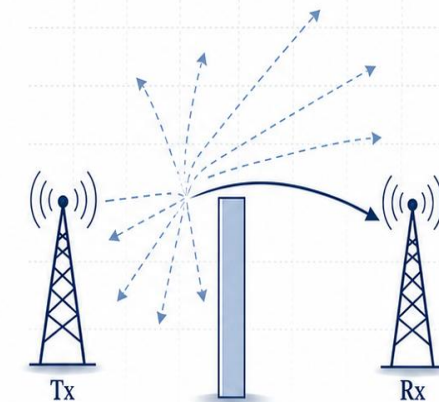
Line of Sight



Reflection



Scattering



A spatial representation should capture the shared path-generating structure—not memorize isolated links.



Why is reusable spatial knowledge needed?

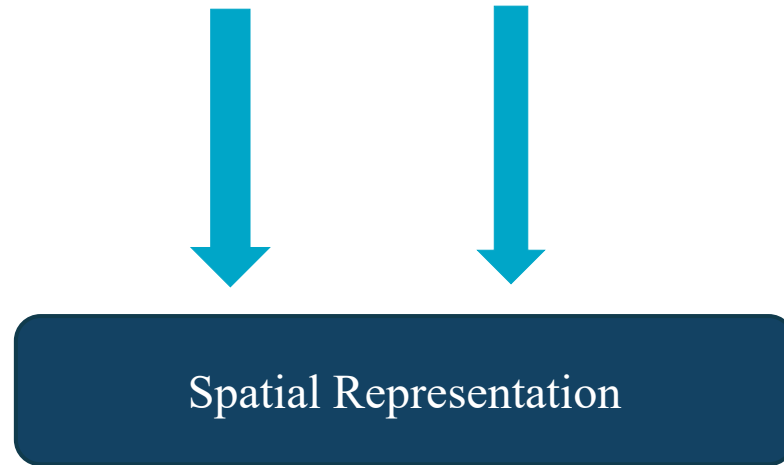
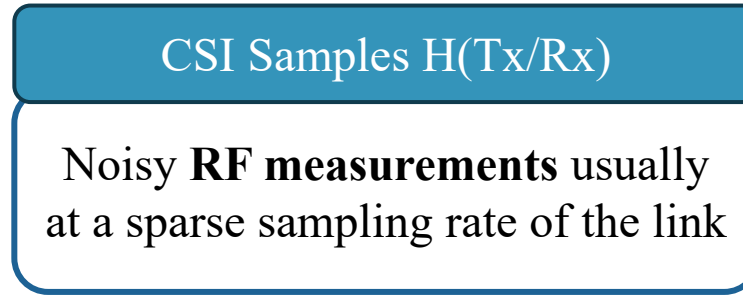
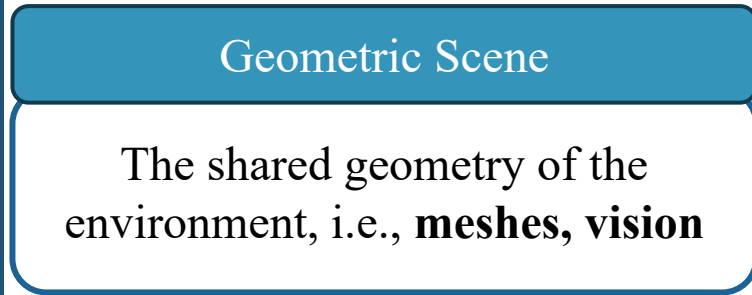


What is the wireless spatial field?



How should wireless space be represented?

Goals: reconstruct once, query and reason across unseen wireless configurations.

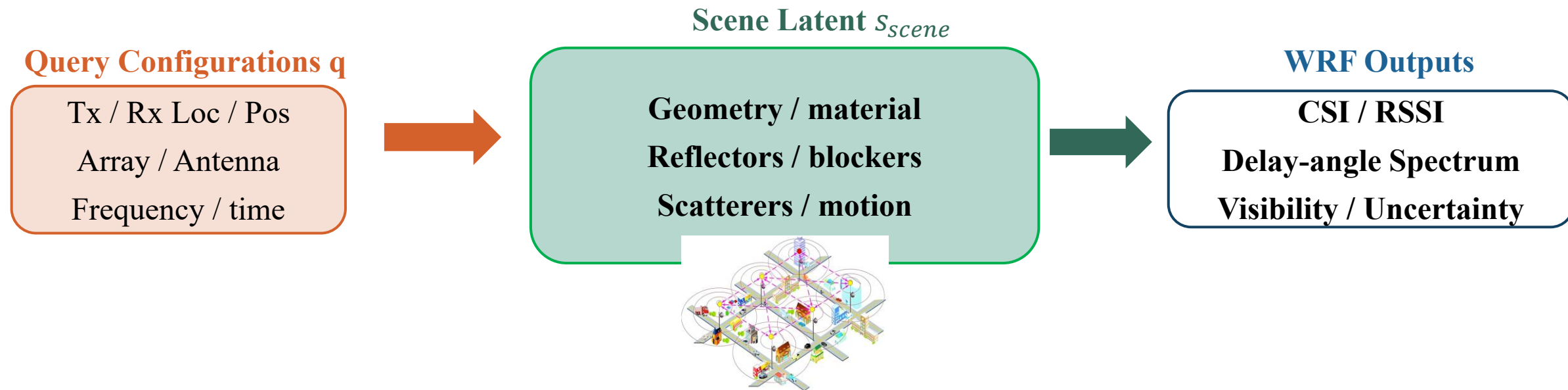


Reconstructed Queryable Field

The shift is from estimating more independent links to learning one field that answers unseen configurations.

Latent representation: A wireless channel is a scene-conditioned function – not a look up table

Learning Objects: $H(q|s_{scene})$



A useful model stores what is shared by configurations and renders what changes with each query.

What is shared?

- **Shared geometry**

The same reflectors and blockers shape nearby links.

- **Smooth evolution**

Delay, angle, and power vary continuously over local motion.

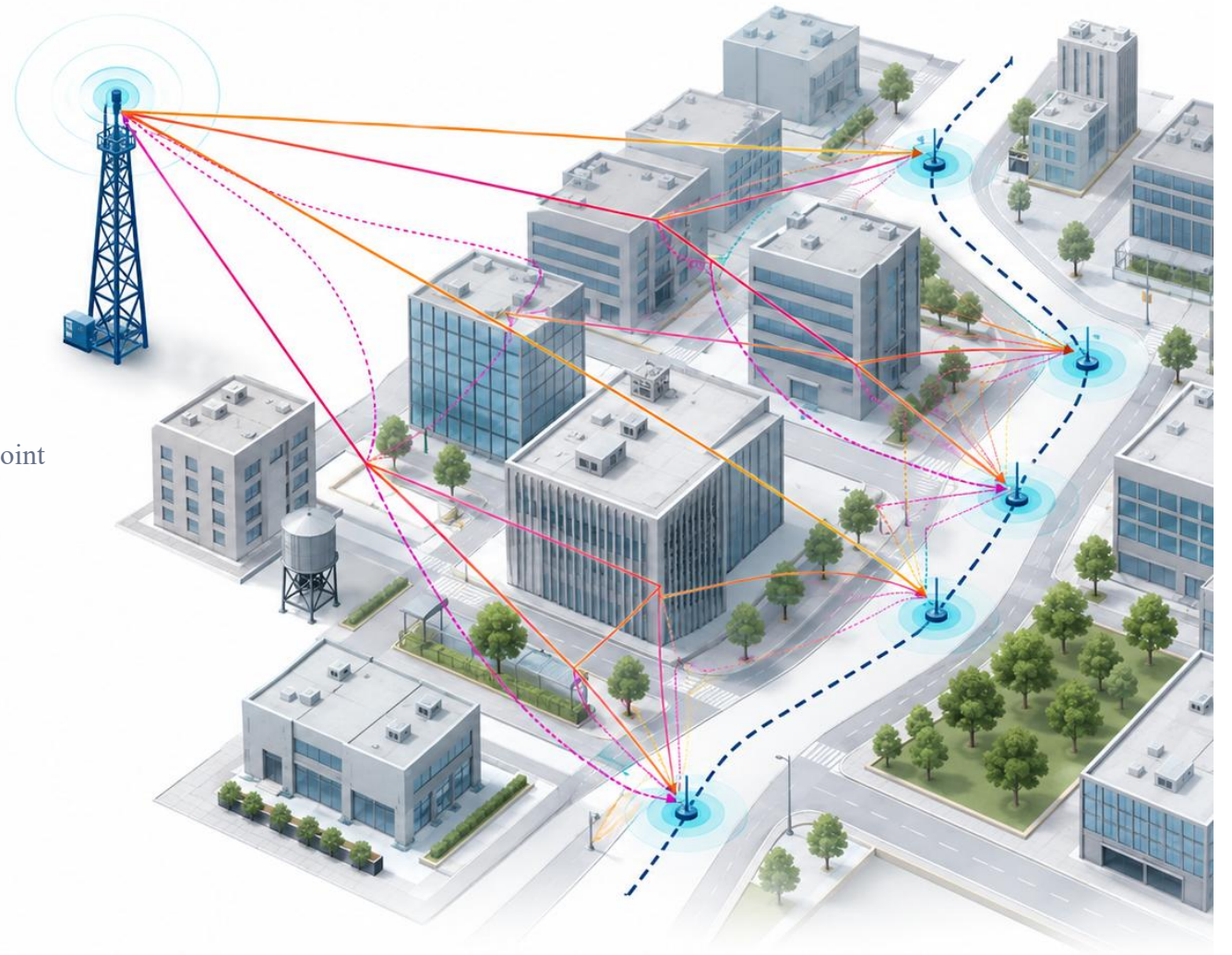
- **Visibility events**

Paths appear or disappear at geometric boundaries.

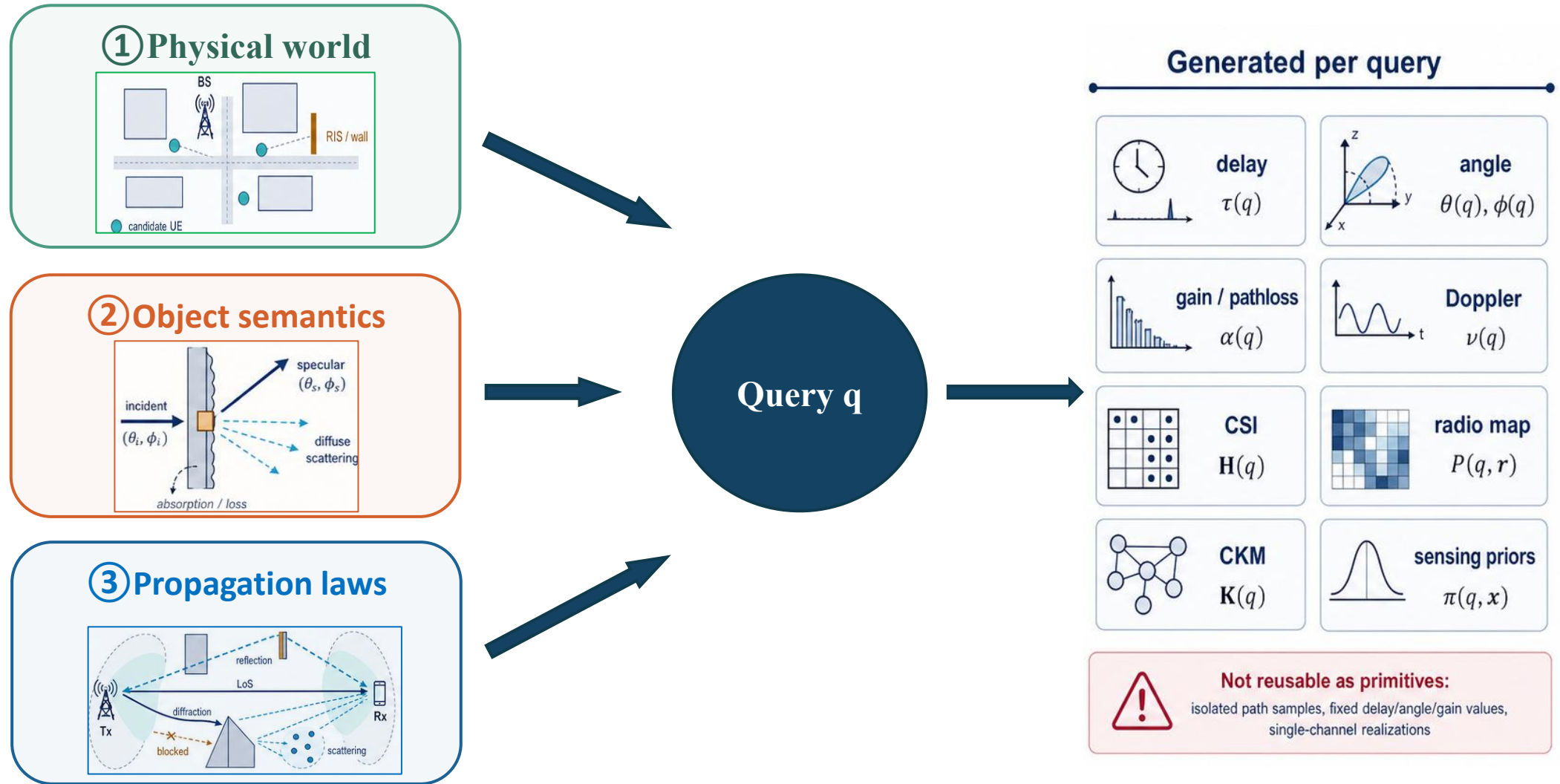
- **Reusable queries**

One scene supports many endpoint and array configurations.

Neighboring links share the same propagation structures

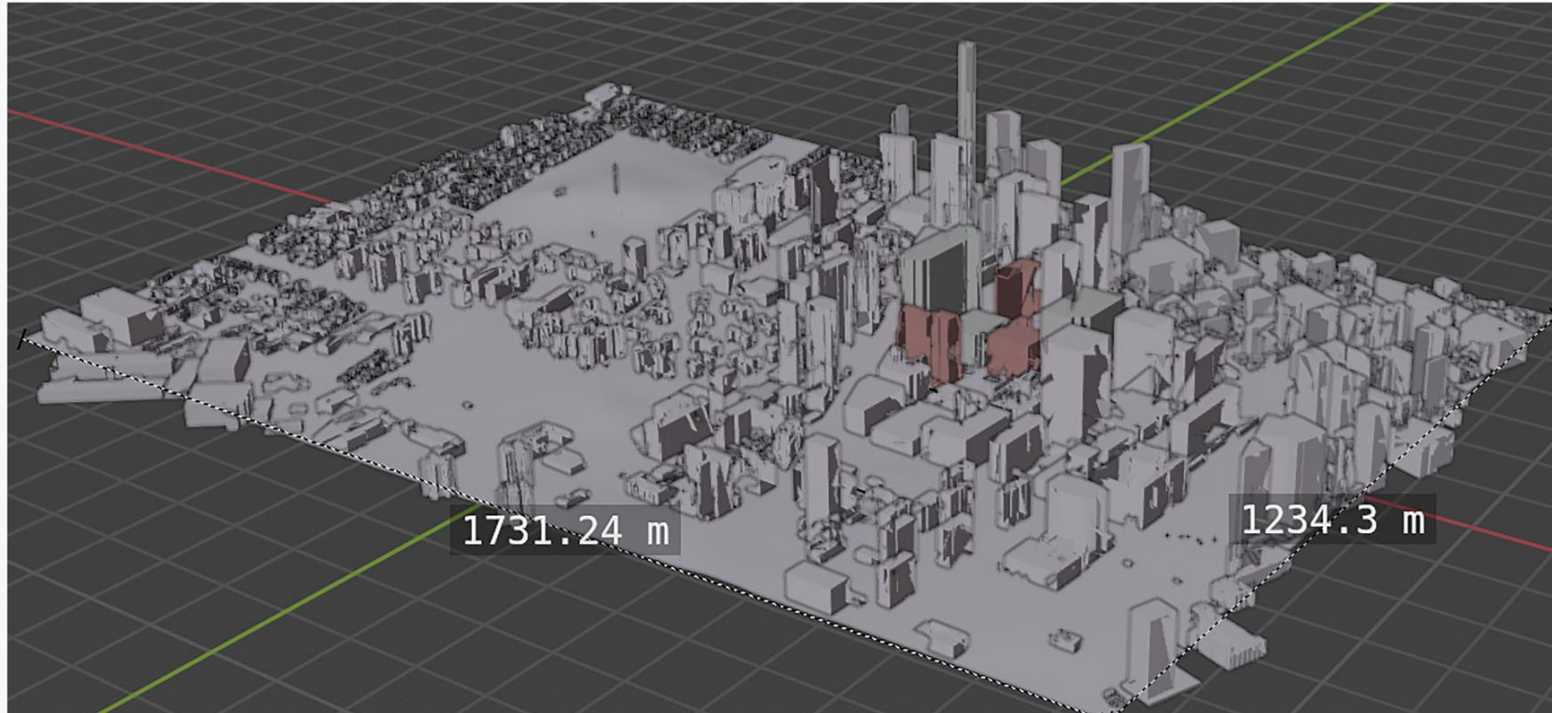


Reusable elements: multipath parameters that evolve coherently across space



What can be re-used? - Geometry of the physical world

3D geometry persists while users, links, and channel realizations change — foundations of ray tracing (RT) simulators



NYURay environment model · Downtown Brooklyn

Geometry is reusable

- Building meshes
- Surface normals
- Object locations
- Spatial topology

REUSABLE CONTEXT

A persistent geometric prior—not a link-specific channel.

M. Ying et al., “Site-specific location calibration and validation of ray-tracing simulator NYURay,” npj Wireless Technology, 2026

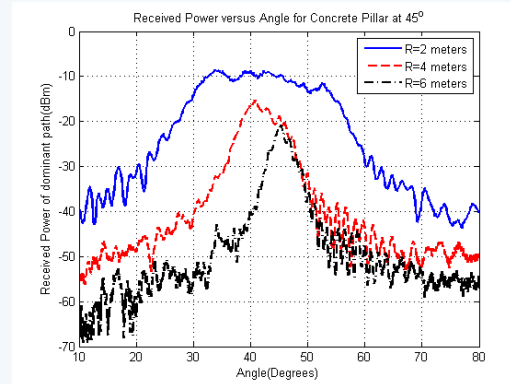
What can be re-used? – Object semantics

The same surface produces a characteristic response across many Tx–Rx configurations.

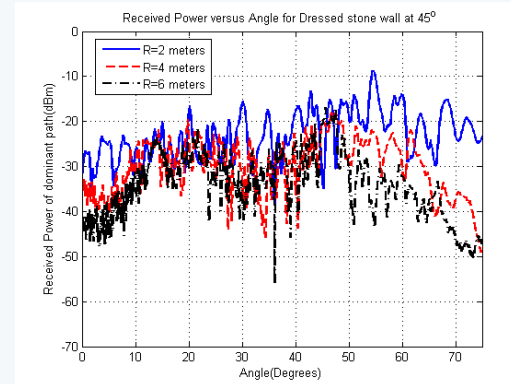


60 GHz measurement

controlled incidence + movable receiver



SMOOTH CONCRETE



ROUGH DRESSED STONE

different angular spreads → different reusable surface responses

material × geometry × frequency

Material

permittivity · conductivity

Surface

roughness · normal · curvature

EM signature

RCS · scattering pattern

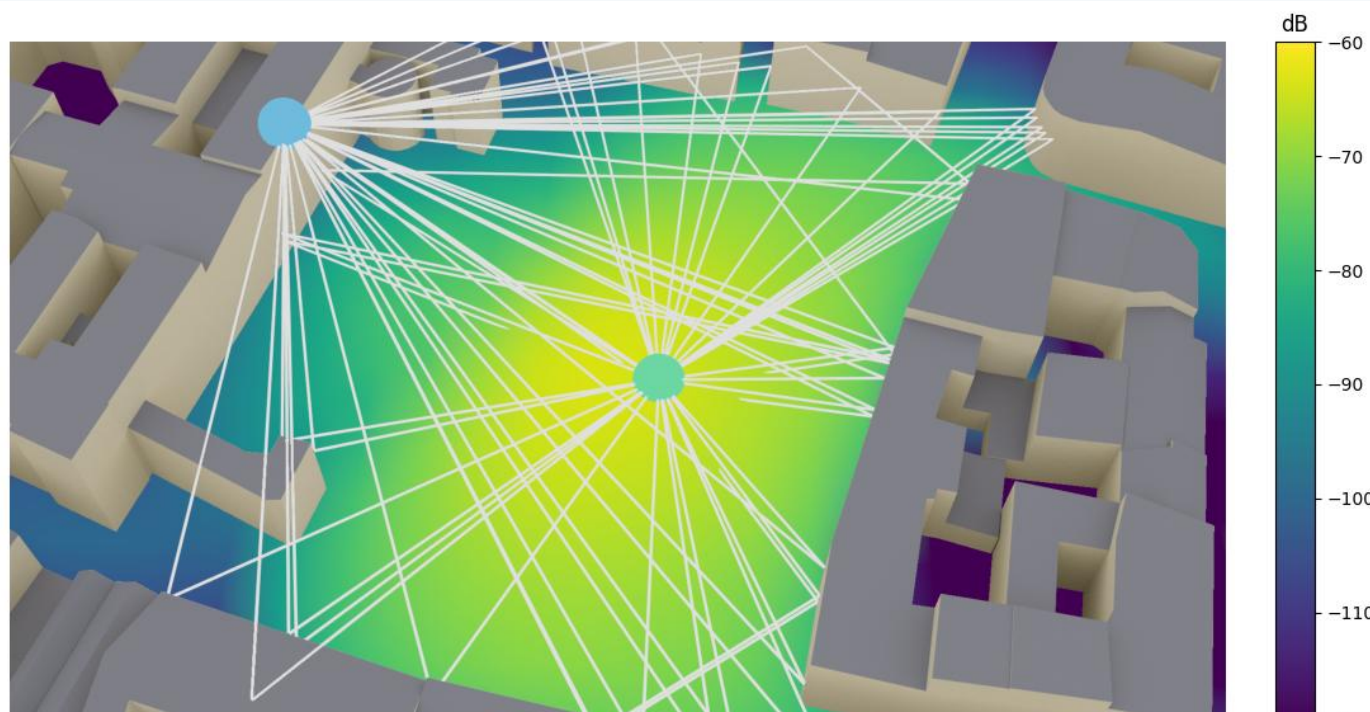
REUSABLE RESPONSE

Object properties—not fixed path gains—govern scattering.

[A. A. Goulios et al., “Measurements and Characterisation of Surface Scattering at 60 GHz,” IEEE VTC-Spring, 2017, Figs. 2, 6–7.]

What can be re-used? – Propagation renderer

These mechanisms persist across scenes, links, and channel realizations.



PROPAGATION LAWS · HOW

Mechanisms persist

- Reflection
- Diffraction
- Scattering
- Attenuation

same laws

different paths and channel realizations

REUSABLE PHYSICS

Propagation laws transform environments into channel responses.

Sionna RT · coverage map + ray-traced paths

J. Hoydis et al., “Sionna RT,” IEEE GLOBECOM Workshops, 2023



Why is reusable spatial knowledge needed?



What is the wireless spatial field?

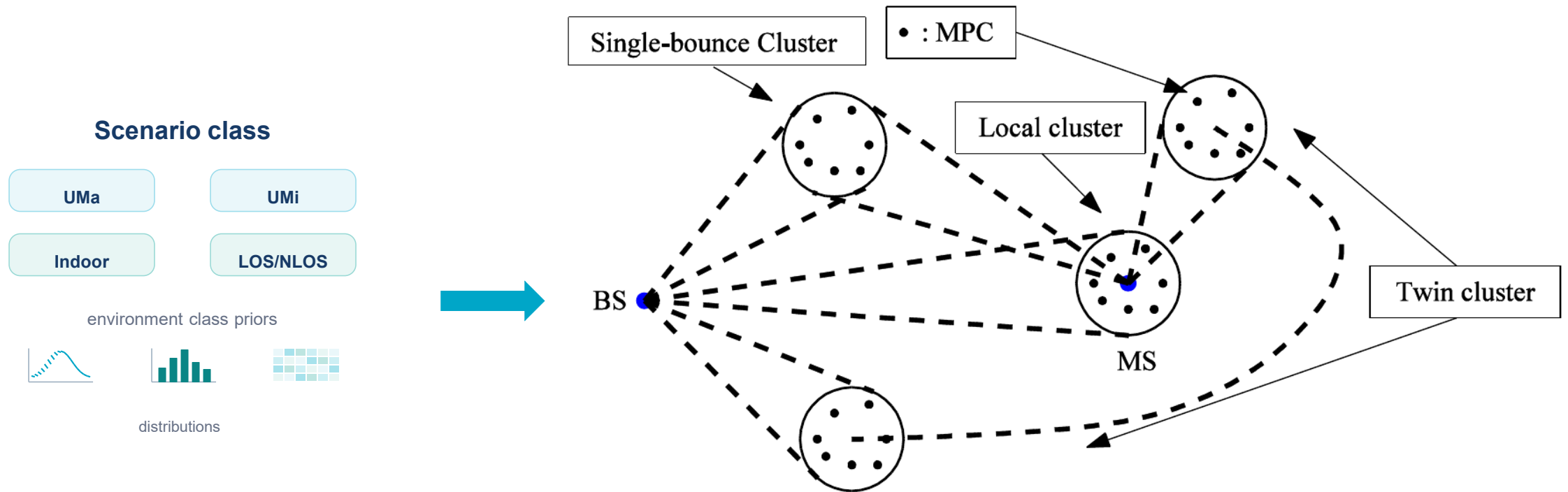


How should wireless space be represented?

Representation families answer different questions

	Representation job	Primary question	Trade-off
Stochastic model	Summarize distributions	What is typical?	Fast, but weak scene specificity
Ray tracing	Enumerate physical paths	Why does this path exist?	Explicit, but model- and compute-heavy
Voxel grid	Store local field samples	What is near this cell?	Queryable, but resolution-bound
Neural field	Fit a continuous function	What is received here?	Continuous, but implicit and fitting-heavy

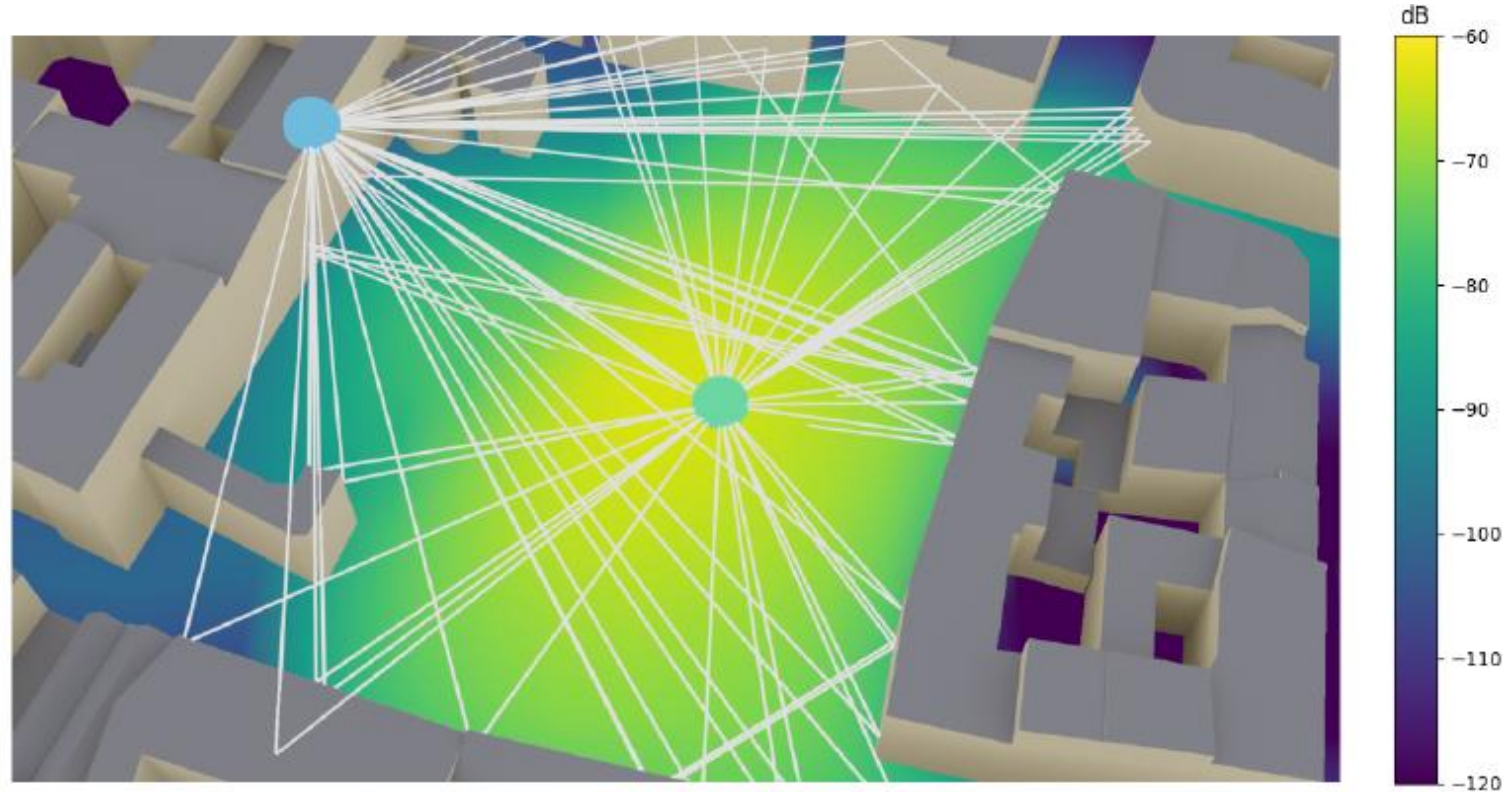
Stochastic models reproduce statistics, not site-specific geometry



Stochastic models reproduce scenario-level channel statistics, not the geometry of a specific site.

[1] L. Liu et al., "The COST 2100 MIMO channel model," IEEE Wireless Commun., vol. 19, no. 6, pp. 92–99, Dec. 2012, doi: 10.1109/MWC.2012.6393523.

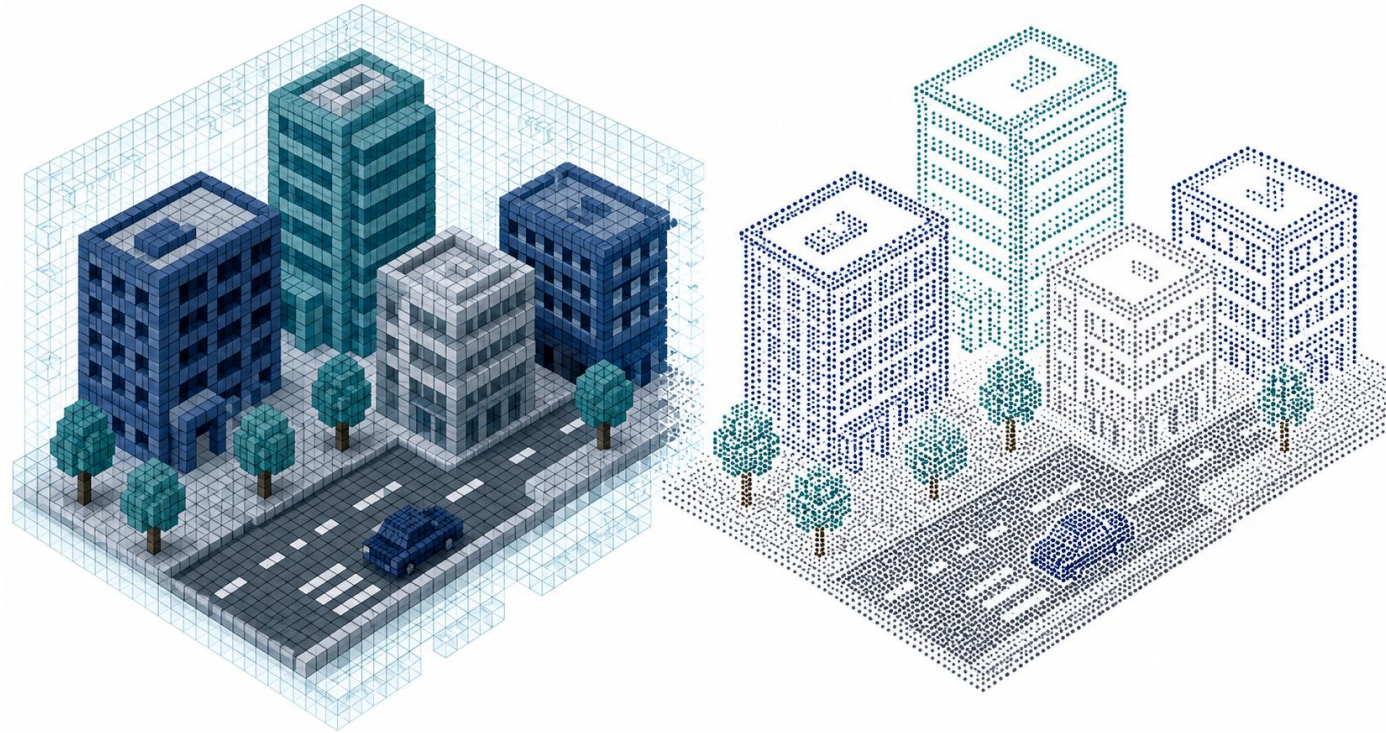
Ray tracing derives channels from explicit scene geometry



Ray tracing explains multipath through explicit geometry, materials, and wave-object interactions.

[1] J. Hoydis et al., “Sionna RT: Differentiable ray tracing for radio propagation modeling,” in Proc. IEEE GLOBECOM Workshops, 2023, pp. 317–321, doi: 10.1109/GCWkshps58843.2023.10465179.

Voxel grids and point clouds store discrete 3-D scene samples



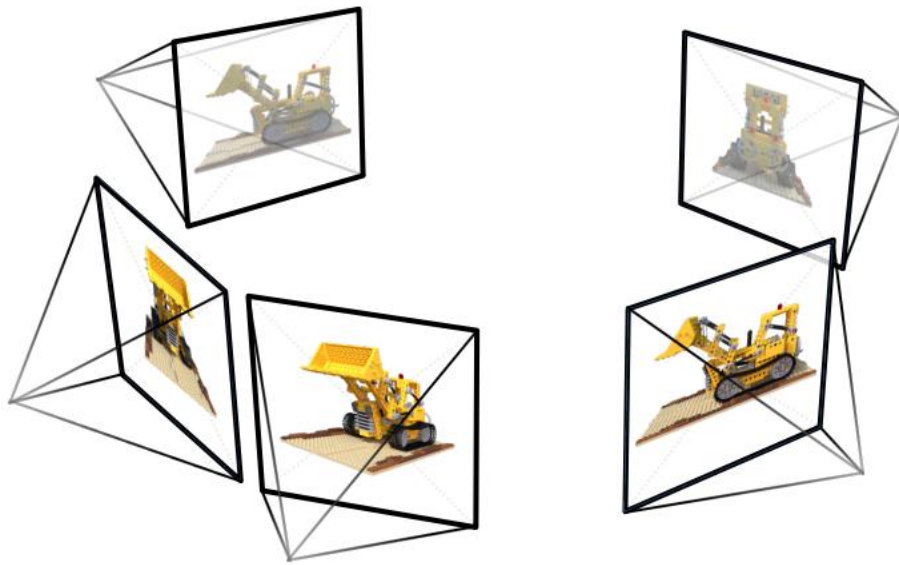
Voxel grids discretize volume; point clouds sample surfaces—both still need a propagation model to produce wireless channels.

[1] D. Maturana and S. Scherer, “VoxNet,” in Proc. IEEE/RSJ IROS, 2015, pp. 922–928, doi: 10.1109/IROS.2015.7353481.

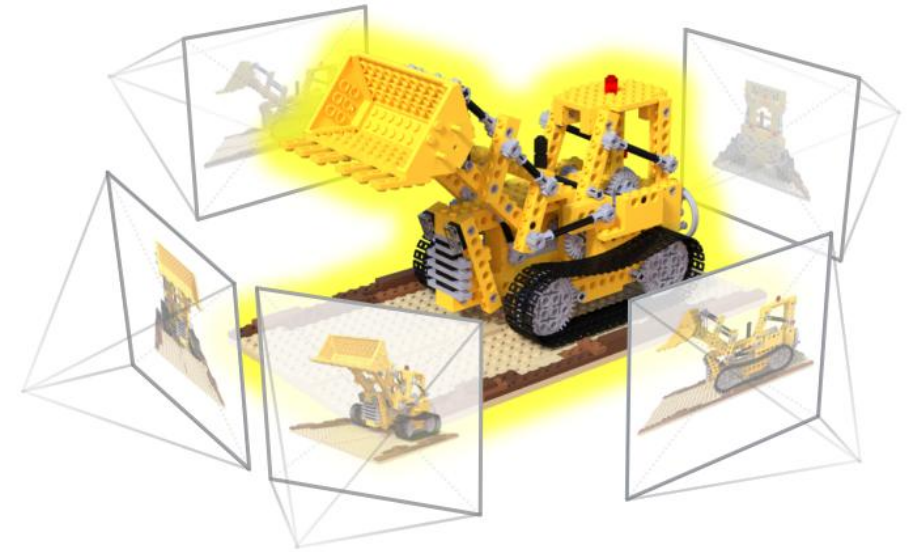
[2] C. R. Qi et al., “PointNet,” in Proc. IEEE CVPR, 2017, pp. 77–85, doi: 10.1109/CVPR.2017.16.

3D-Reconstruction Methods: What is Reconstruction

Create a 3D scene representation from a set of 2D images and their corresponding camera poses.



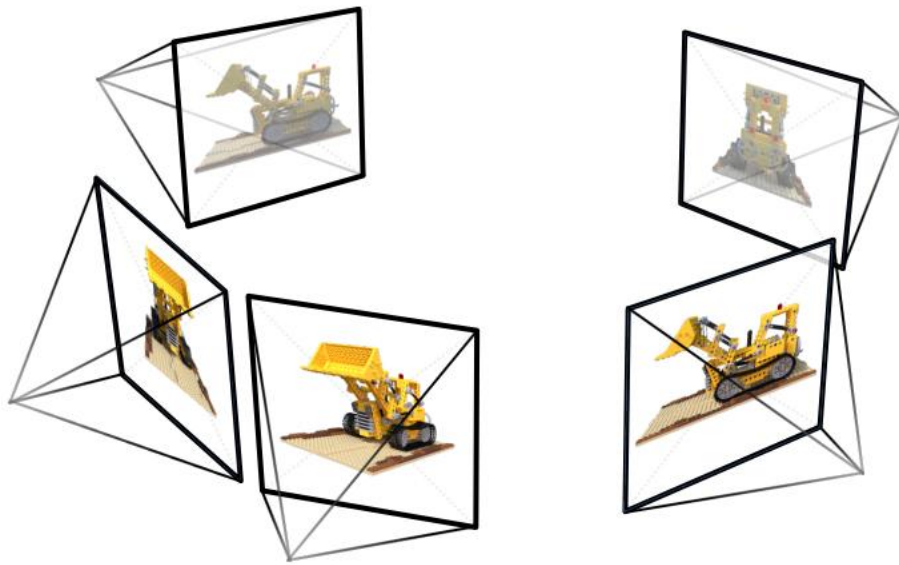
Images & Accurate Camera Poses



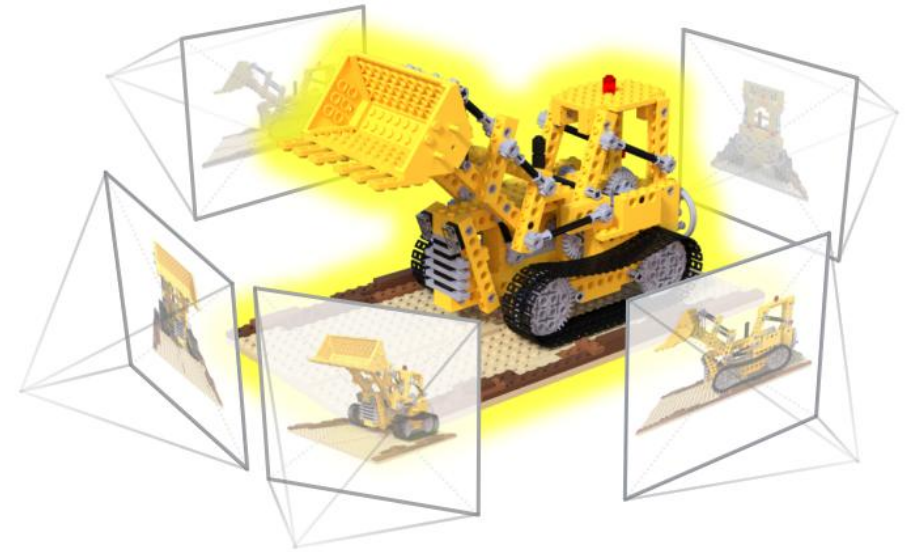
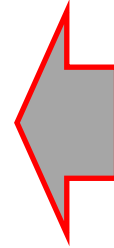
3D Scene Representation

3D-Reconstruction Methods: What is Rendering

Given a 3D scene representation and a camera pose, create a 2D image.



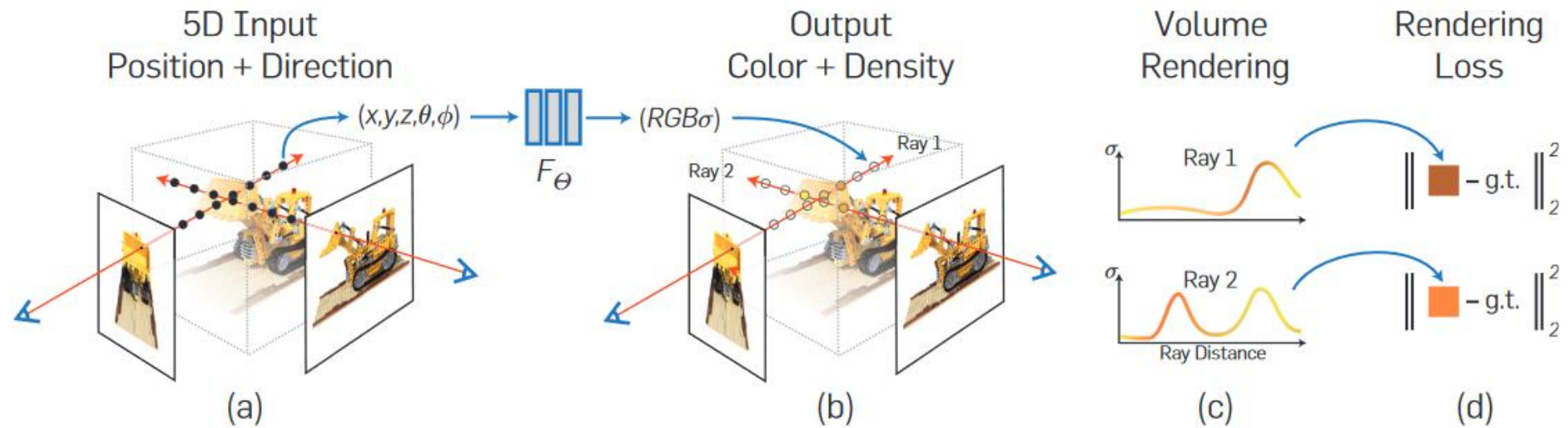
Images & Accurate Camera Poses



3D Scene Representation

3D-Reconstruction Methods: Neural radiance fields (NeRF)

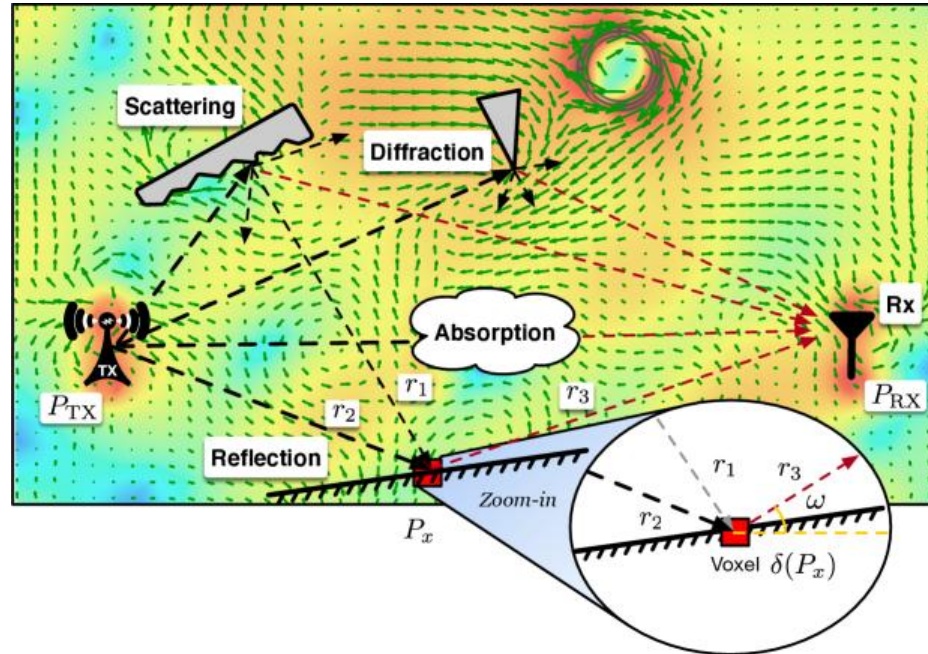
Using a fully connected deep network to learn the underlying continuous volumetric scene function [1].



[1] Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM 65.1 (2021): 99-106.

3D-Reconstruction Methods for Wireless: Neural RF Field (NeRF2)

Extending neural radiance fields from visual scenes to wireless environments



From measurements to spatial fields

INPUT

Wireless measurements
Tx / Rx locations

REPRESENTATION

Continuous neural RF field

OUTPUT

Spatial wireless channel reconstruction

Neural representation of spatially varying wireless propagation characteristics

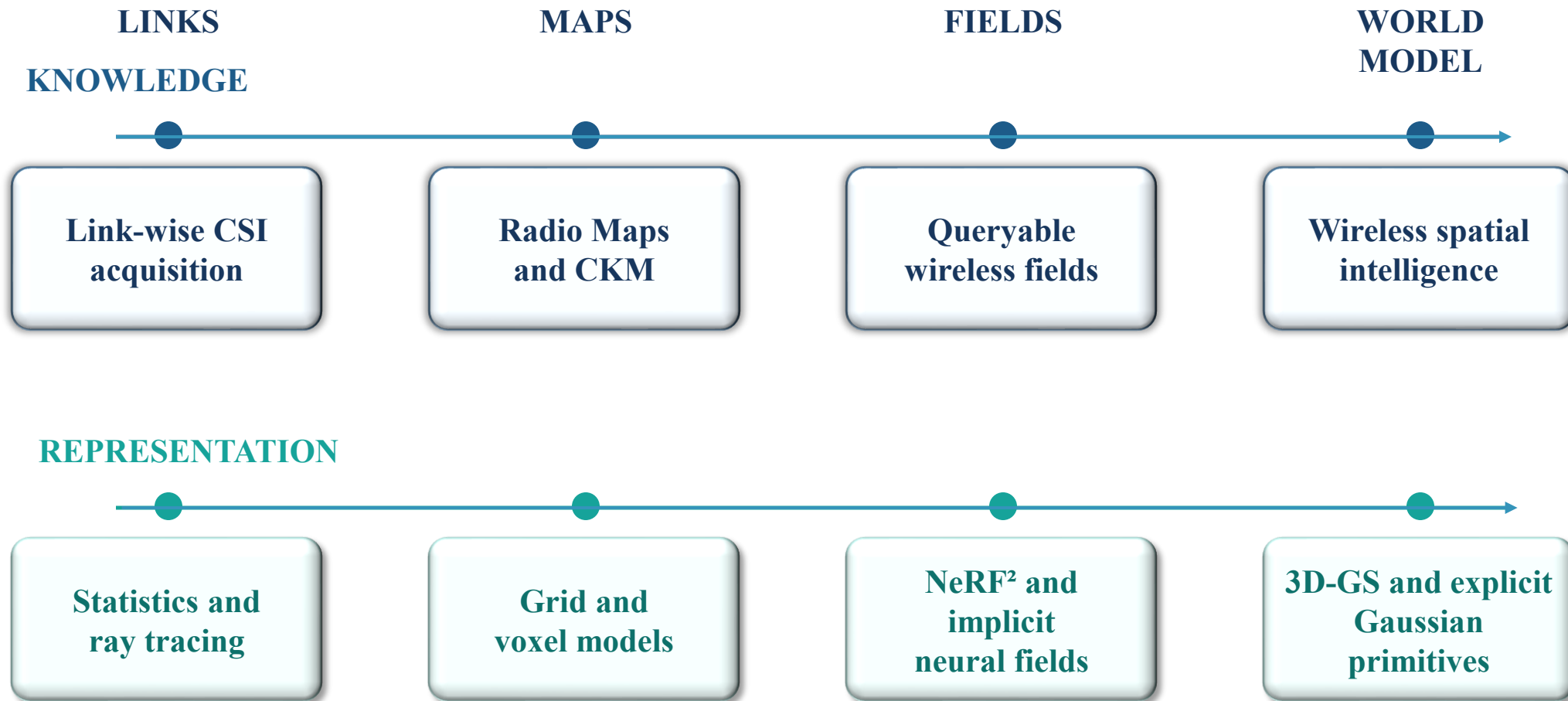
Zhao X, An Z, Pan Q, et al. Nerf2: Neural radio-frequency radiance fields[C]//Proceedings of the 29th Annual International Conference on Mobile Computing and Networking. 2023: 1-15.

No representation dominates every requirement

Method	Scene specificity	Spatial continuity	Physical grounding	Build / update burden	Query speed
Stochastic	Low	Statistical	Low	Low	Very fast
Ray tracing	High	High	High	Very high	Slow
Voxel grid	High	Grid-limited	Low	High memory	Fast
Neural field / NeRF ²	High	High	Implicit	High training	Medium

Design target: scene-specific, continuous, physically grounded, learnable, and fast to query.

Take-Aways: Two lines converge on a wireless spatial world model



Takeaways

- Wireless channels are generated by the **shared physical environment**.
- Reusable knowledge includes **geometry, materials, visibility, motion, and propagation laws**.
- The goal is to **reconstruct once and query many unseen Tx/Rx configurations**.
- Wireless space should be represented as a **scene-conditioned continuous field**.



Part II

3D-GS Methods for Spatially Intelligent Channel Modelling





WRF-GS: wireless field reconstruction



From WRF to WRF-GS+



Temporal non-stationarity and mobility-aware modelling

3D-GS represents a scene with explicit, differentiable Gaussians

Optical 3D reconstruction based on 3D Gaussian Splatting (3D-GS).



Fig. 7: Scenes represented by Gaussian primitives.

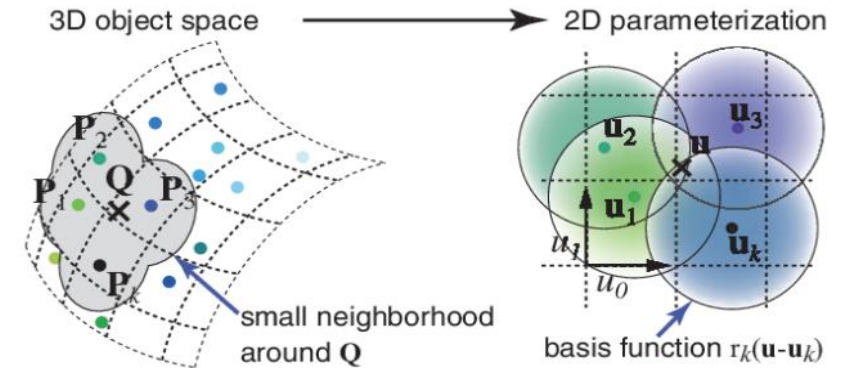
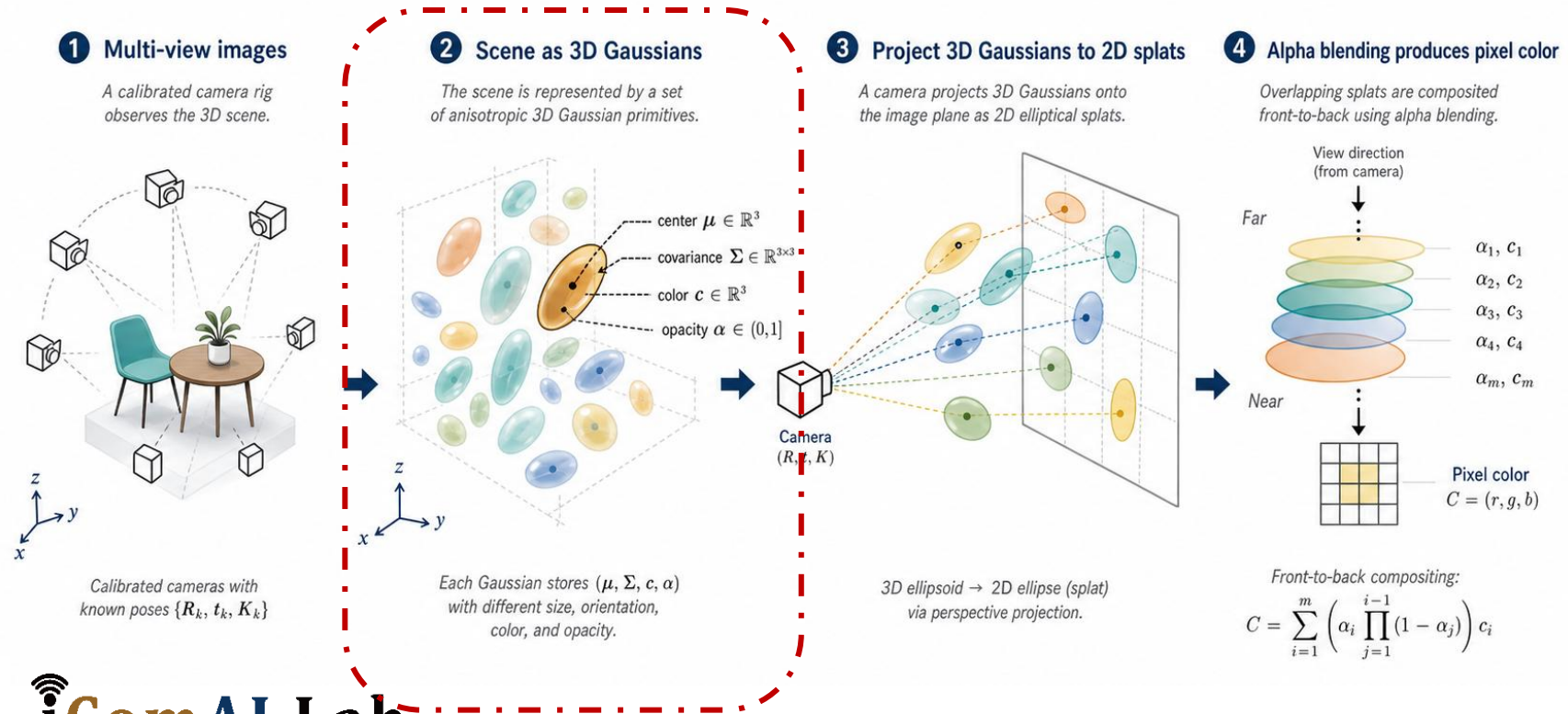


Fig. 8: Rendering process.

Preliminaries: 3D Gaussian Splatting

- Mathematically, each 3D Gaussian distribution is characterized by its own covariance matrix Σ and center position μ vector:

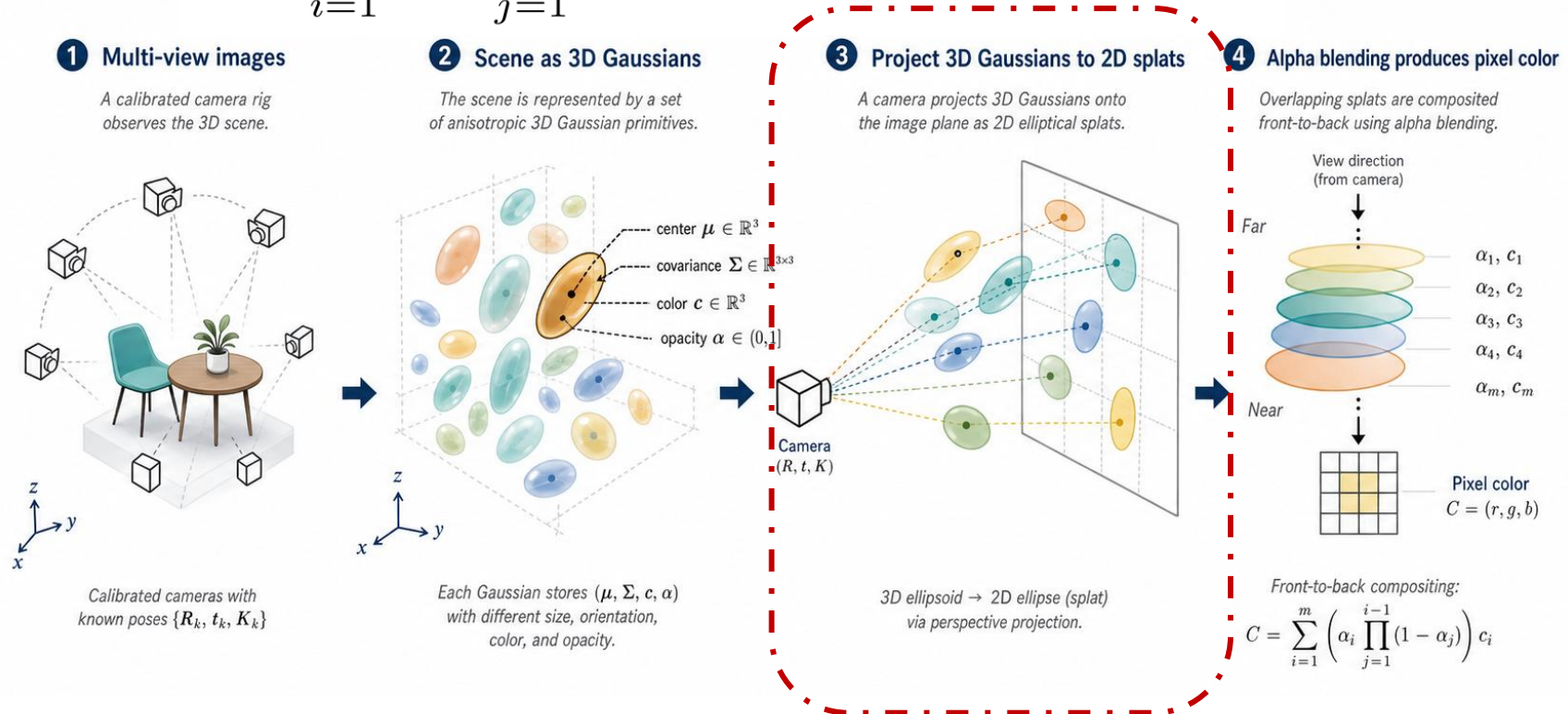
$$G(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}, \quad \Sigma = \mathbf{R} \mathbf{S} \mathbf{S}^T \mathbf{R}^T.$$



Preliminaries: 3D Gaussian Splatting

- Transforming the 3D Gaussians $G(\mathbf{x})$ in space into 2D Gaussians $G'(\mathbf{x}')$ on the 2D image plane, utilizing alpha blending to calculate the values of each pixel.

$$C = \sum_{i=1}^N c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \text{ and } \alpha_i = o_i G'_i(\Delta \mathbf{p}_i),$$





Since optics is also an electromagnetic wave, 3D-GS technology can be used to reconstruct **Wireless Radiation Fields (WRF)** for channel modeling.

Problem Description

$$y = Ae^{j\varphi} \sum_{l=0}^{L-1} \Delta A_l e^{j\Delta\varphi_l},$$

The received signal is a superposition of sub-signals from multiple paths, with each represented as an LoS signal from **a virtual TX (a 3D Gaussian)**.

Spatial Spectrum

$$P(\alpha, \beta) = \left| \frac{1}{K} \sum_{m,n=0}^{\sqrt{K}-1, \sqrt{K}-1} e^{j(\Delta\hat{\theta}_{m,n} - \Delta\theta_{m,n})} \right|^2,$$



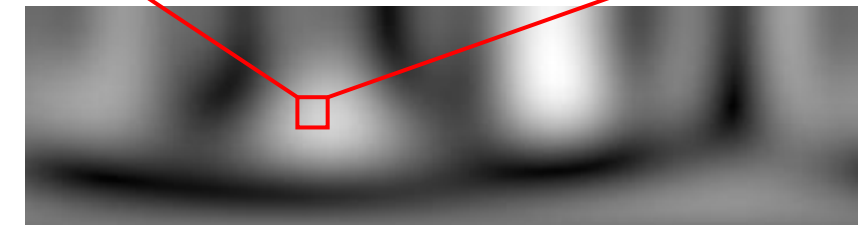
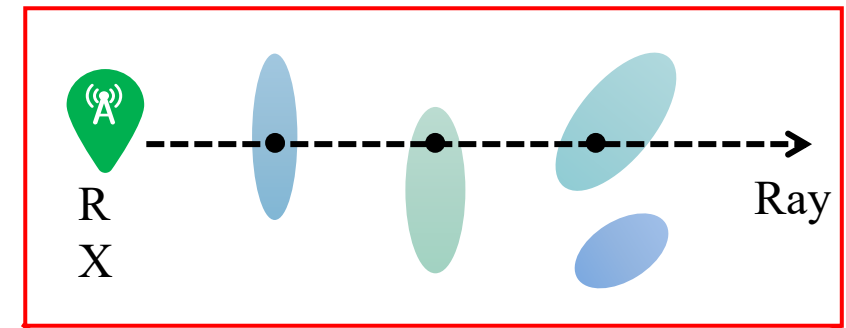
$\alpha \sim (0^\circ, 359^\circ)$ and $\beta \sim (0^\circ, 89^\circ)$

$$\mathbf{P} = \begin{bmatrix} P(0^\circ, 0^\circ) & P(1^\circ, 0^\circ) & \dots & P(359^\circ, 0^\circ) \\ P(0^\circ, 1^\circ) & P(1^\circ, 1^\circ) & \dots & P(359^\circ, 1^\circ) \\ \vdots & \vdots & \ddots & \vdots \\ P(0^\circ, 89^\circ) & P(1^\circ, 89^\circ) & \dots & P(359^\circ, 89^\circ) \end{bmatrix}$$

Convert to image
pixel value



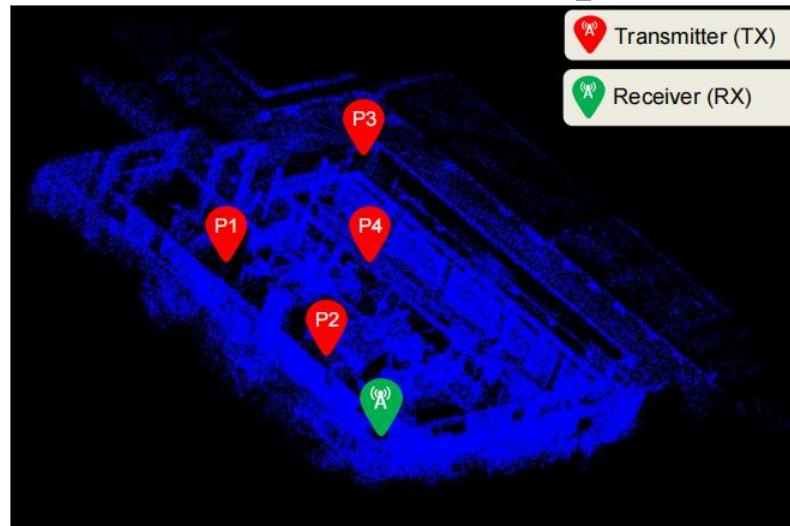
3D Gaussian Splatting



C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10

Problem Description

- **Basic scene settings:** **Fixed** RX position and **random** TX position.
- **Reasonable assumptions:** The main layout in the space remains **unchanged**.
- **Objectives:**
 - Reconstruct the spatial wireless radiation field
 - Given a new TX position, obtain the channel parameters at the RX position.



C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10.

WRF-GS: Framework

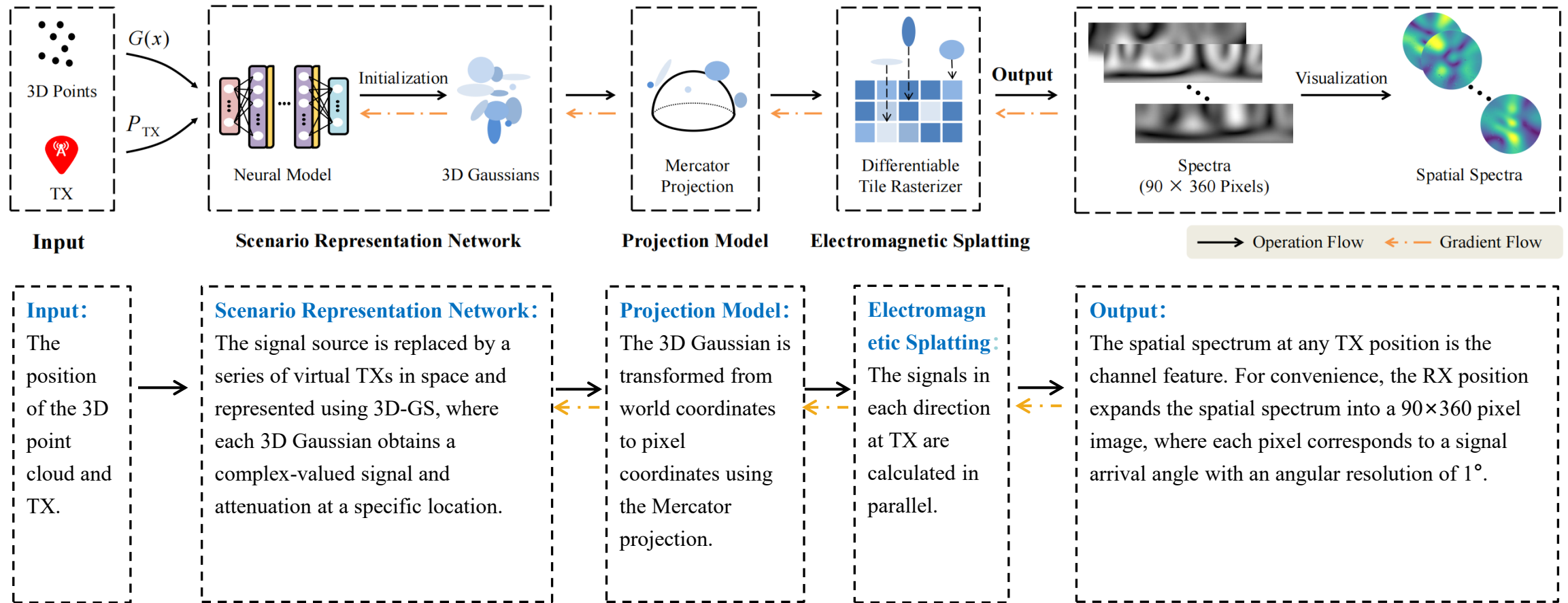


Fig. 10: An overview of the WRF-GS framework.

C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10

WRF-GS: Scenario Representation Network

- It models the underlying relationships between the environment and radio signals by representing the virtual TXs as 3D Gaussians:

$$F_{\Theta} : (G(\mathbf{x}), P_{\text{TX}}) \Rightarrow (\delta(\mathbf{x}), S(\mathbf{x})),$$

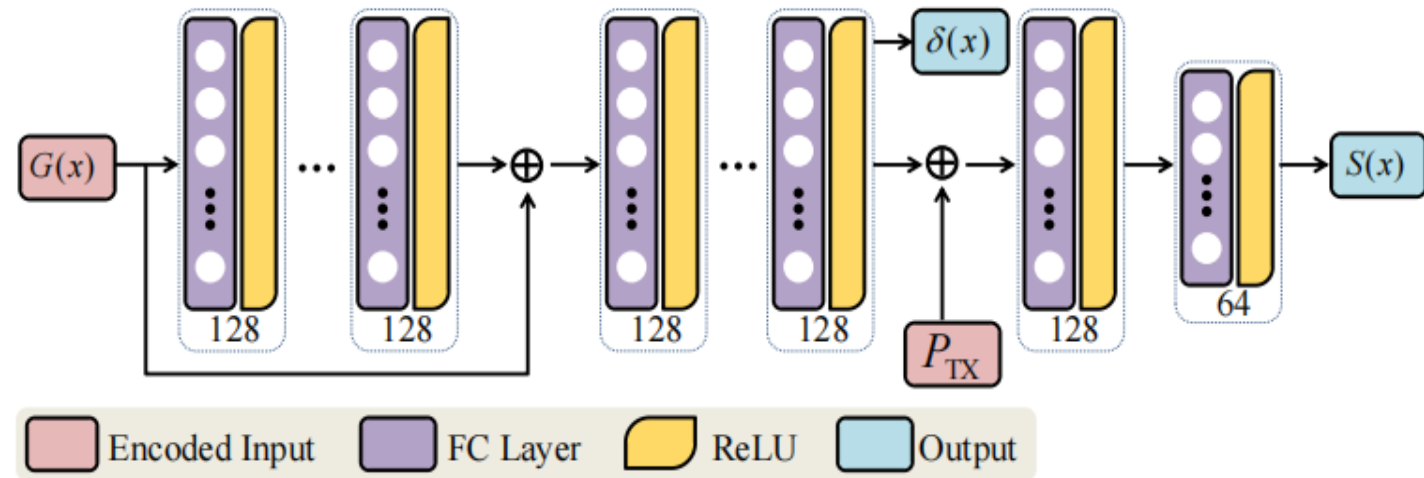


Fig. 11: Architecture of the neural model.

- The two outputs here are denoted by:

$$S(\mathbf{x}) = A(\mathbf{x})e^{j\psi(\mathbf{x})}$$

$$\delta(\mathbf{x}) = \Delta A(\mathbf{x})e^{j\Delta\psi(\mathbf{x})}$$

$$y = Ae^{j\varphi} \cdot \Delta Ae^{j\Delta\varphi},$$

Recall the path loss model

C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10

WRF-GS: Projection Model

- Unlike pinhole cameras, the antenna in our WRF-GS receives signals from a hemispherical direction. Thus, we introduced the **Mercator projection**.

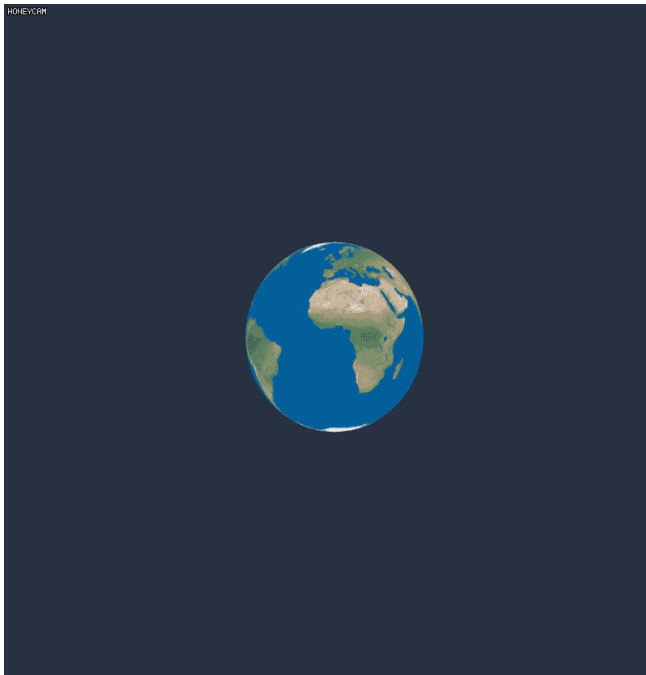
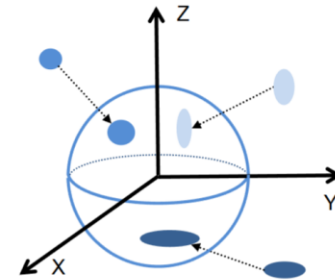
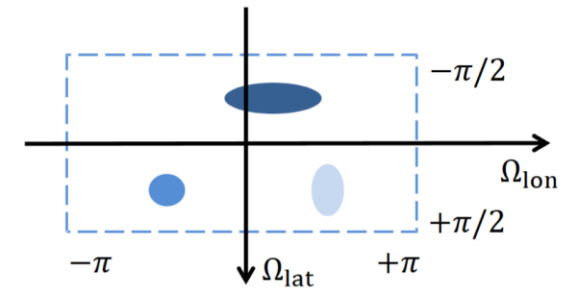


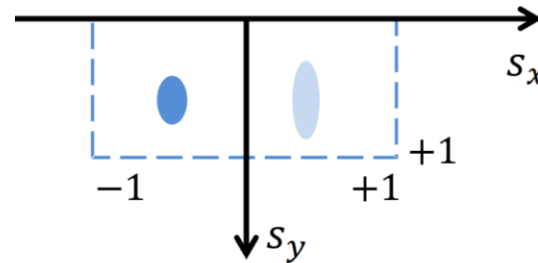
Fig. 12: Mercator projection



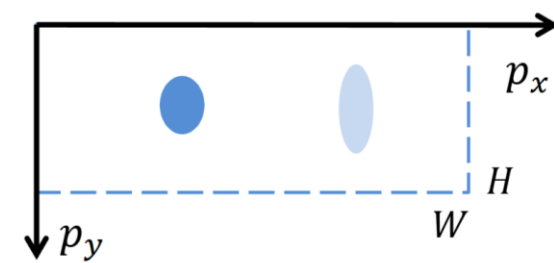
(a) Antenna coordinate.



(b) Latitude-longitude coordinate.



(c) Uniform coordinate.



(d) Image coordinate.

C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10

WRF-GS: Electromagnetic Splatting

- Three steps to adapt electromagnetics 3D-GS to process complex signals:

(a) 3D $\rightarrow$ 2D.

(b) Depth Sorting.

(c) Parallel Computation.

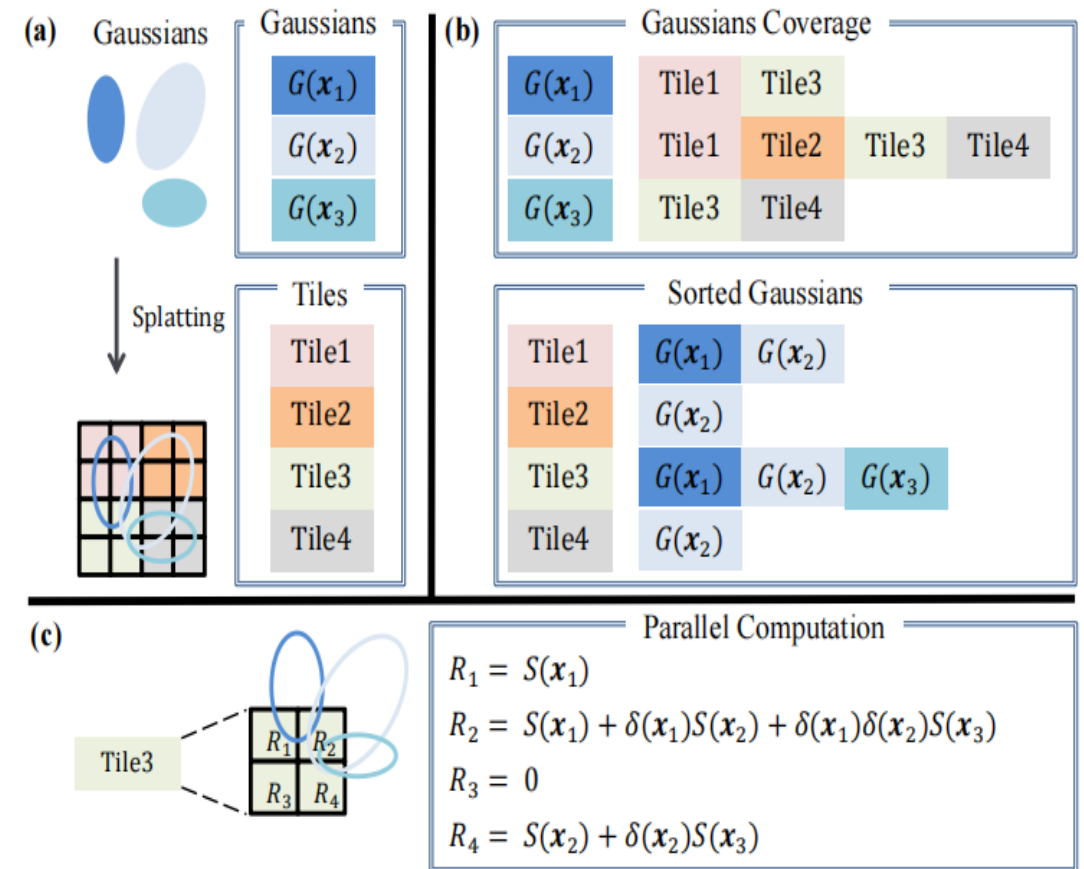


Fig. 13: An illustration of electromagnetic splatting.

C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10.

Experiment

Public Dataset

- The position of the RX is fixed and the TX is varied.
- Each sample consists of the TX position and the corresponding spectrum measured at the RX.

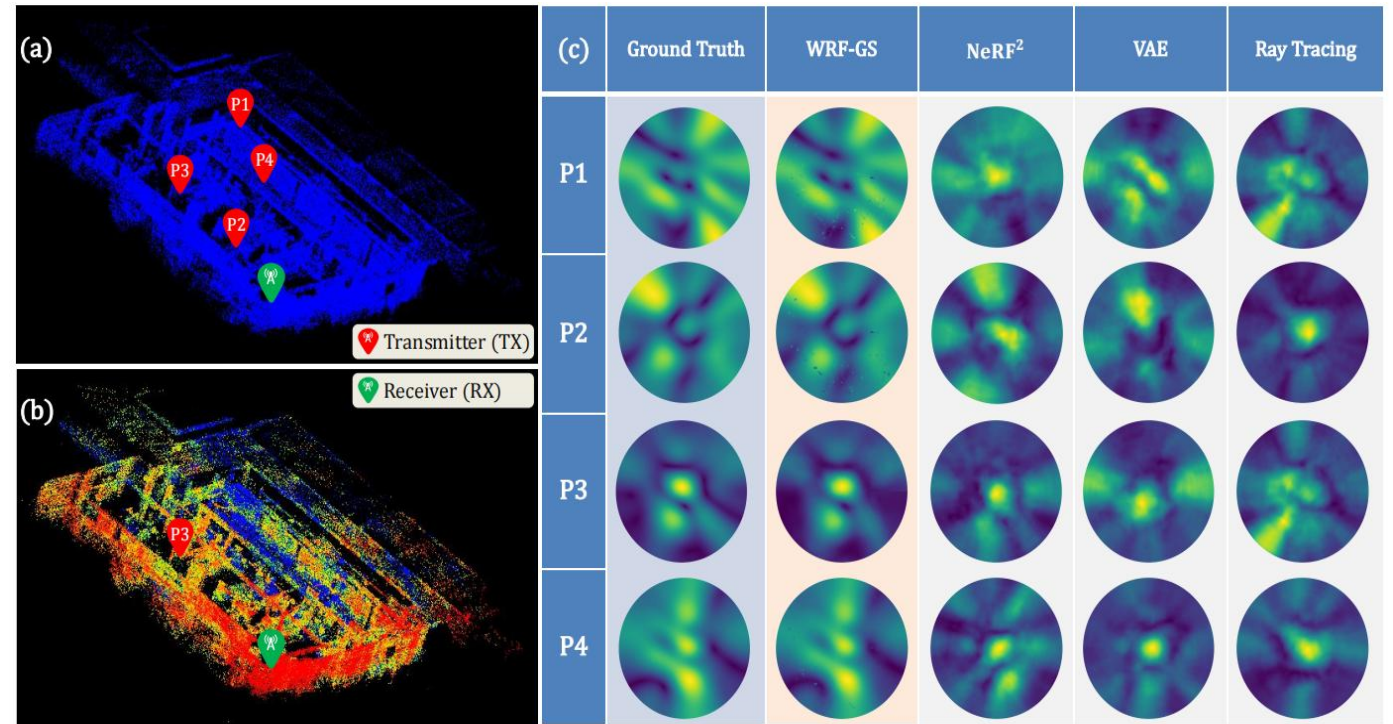


Fig. 14: Synthesized spatial spectrum results.

C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10

Experiment

- We adopt the structural similarity index measure (SSIM) as the performance metric to compare different methods.

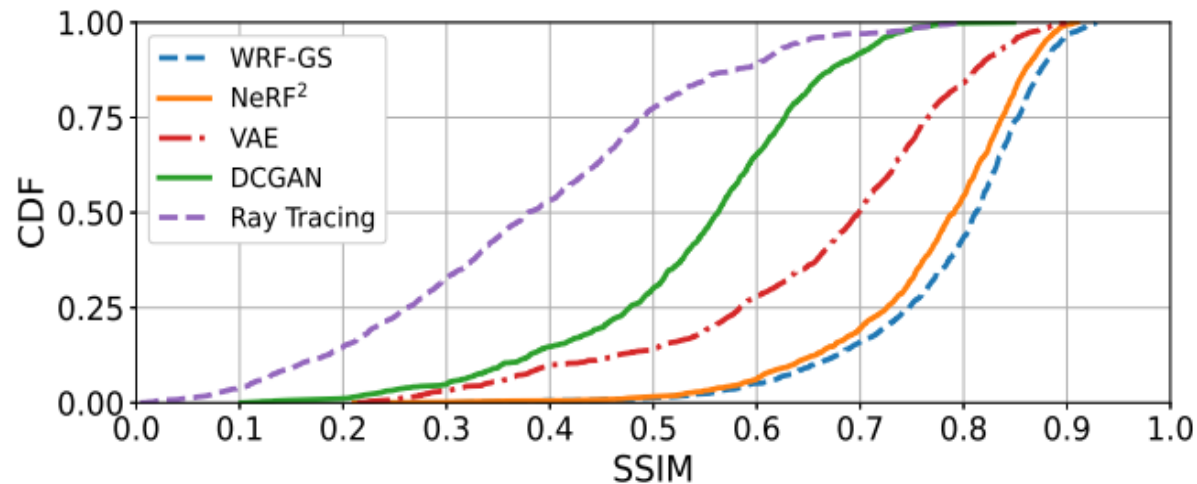


Fig. 15: The SSIM values of the synthesized spatial spectrums generated by different methods.

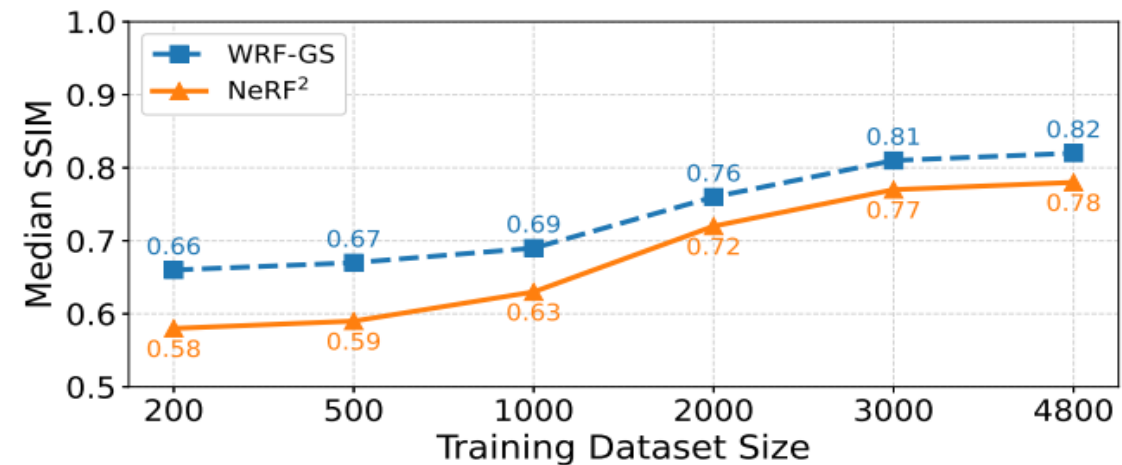


Fig. 16: The median SSIM values of the WRF-GS and NeRF² under different training dataset sizes.

C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10

Experiment

- Fig. 17 illustrates the rendering time required to synthesize a spatial spectrum using different methods under the same settings.
- The processing time of WRF-GS is usually **one to three orders of magnitude less than other methods**.

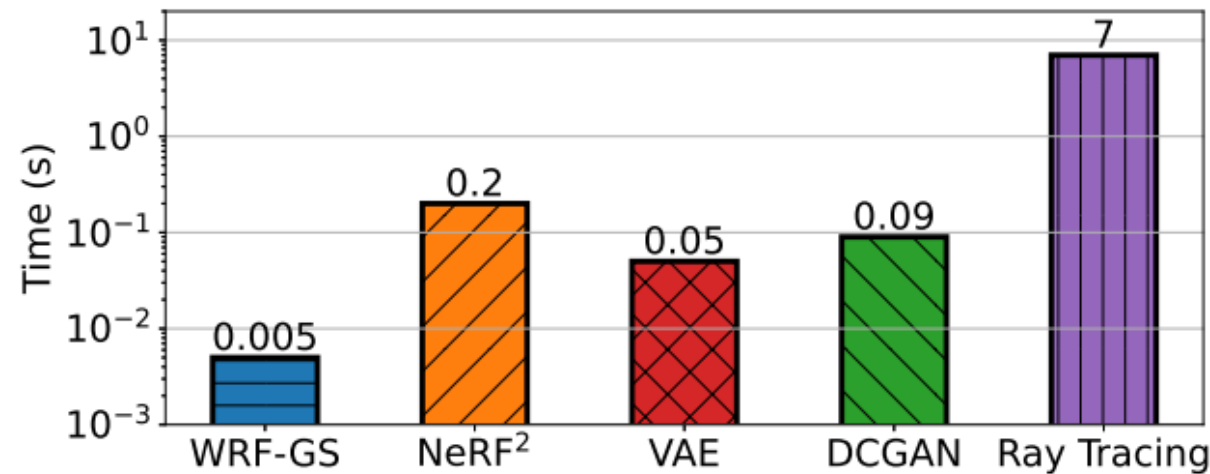


Fig. 17: The rendering time required to synthesize a spatial spectrum using different methods under the same settings.

C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," in Proc. IEEE INFOCOM 2025—IEEE Conf. Comput. Commun., London, U.K., May 2025, pp. 1–10

WRF-GS tracks the dominant CSI structure while reducing prediction error

Simulation Results

- The uplink and downlink transmissions experience the same physical environment, it is reasonable to **use the uplink CSI measured at the BS to infer the downlink CSI**.
- The scenario representation network can be represented as:

$$F_{\Theta} : (G(\mathbf{x}), I_u(\mathbf{x})) \Rightarrow (\delta(\mathbf{x}), S(\mathbf{x})),$$

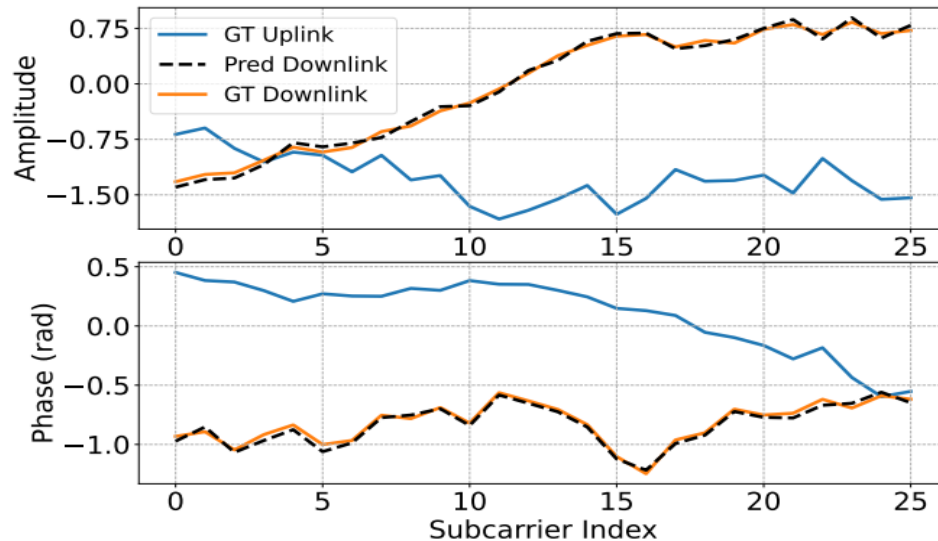


Fig. 19: The comparison of the ground truth CSI and the predicted CSI using WRF-GS.

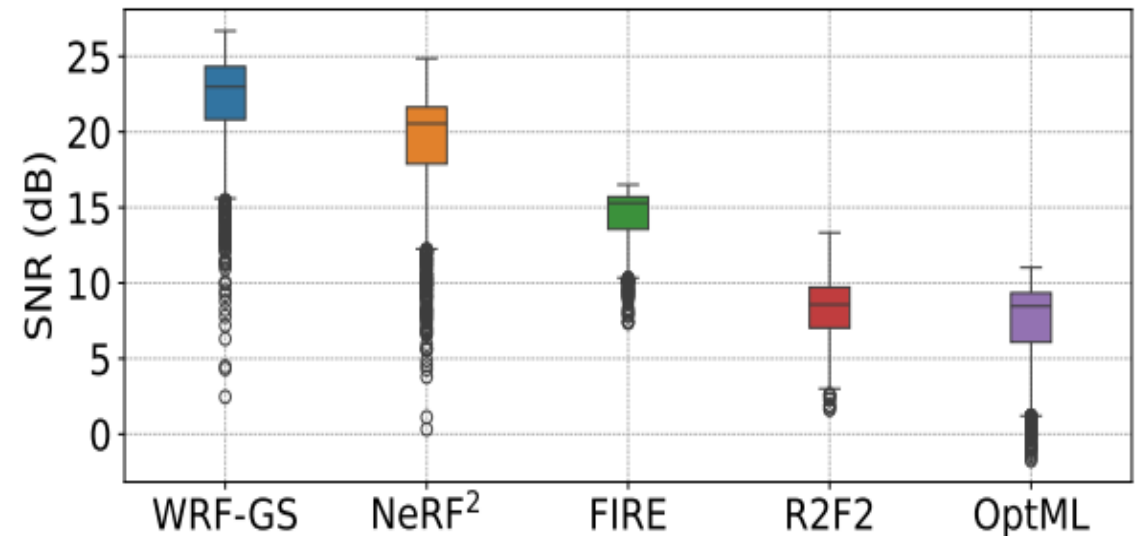


Fig. 20: The prediction accuracy of different methods.



WRF-GS: wireless field reconstruction

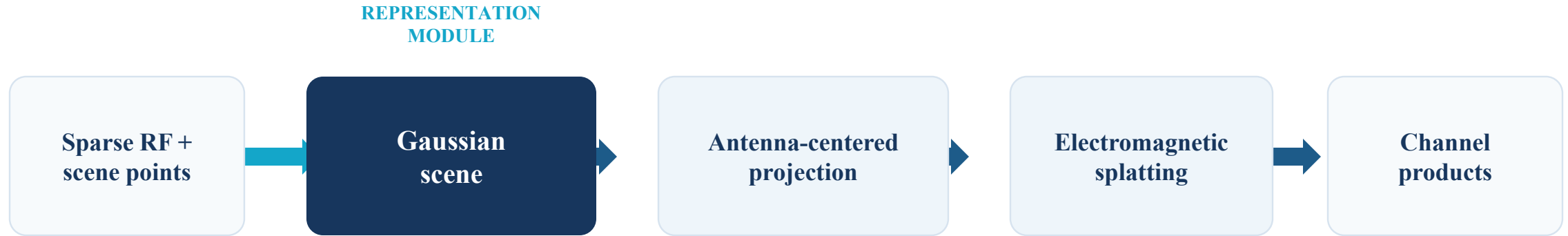


From WRF to WRF-GS+



Temporal non-stationarity and mobility-aware modelling

WRF-GS+ upgrades only the representation module



The projection and splatting stages remain shared; the upgrade increases representation capacity where TX-dependent variation enters.

WRF-GS+: Framework

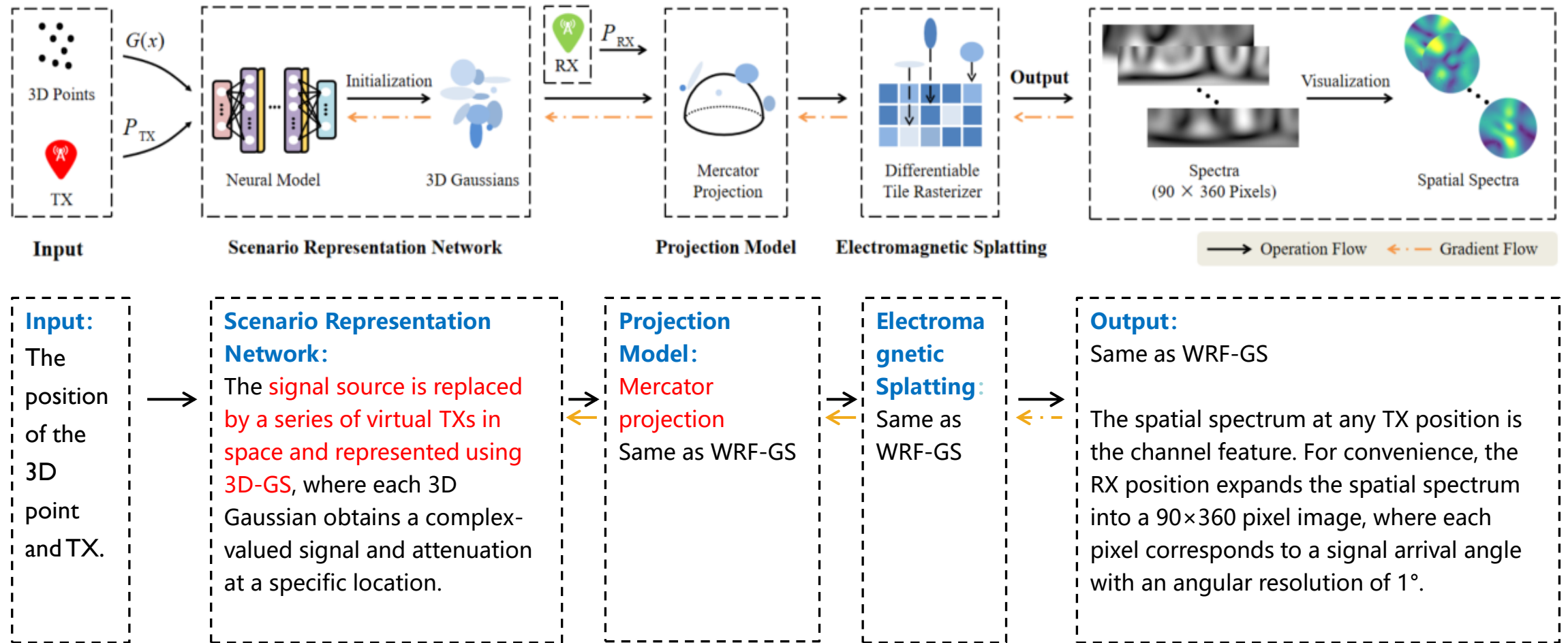


Fig. 5: An overview of the WRF-GS+ framework.

[1] C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "Neural representation for wireless radiation field reconstruction: A 3D Gaussian splatting approach," IEEE Trans. Wireless Commun., vol. 25, pp. 7490–7504, 2026.

WRF-GS+: Scenario Representation Network

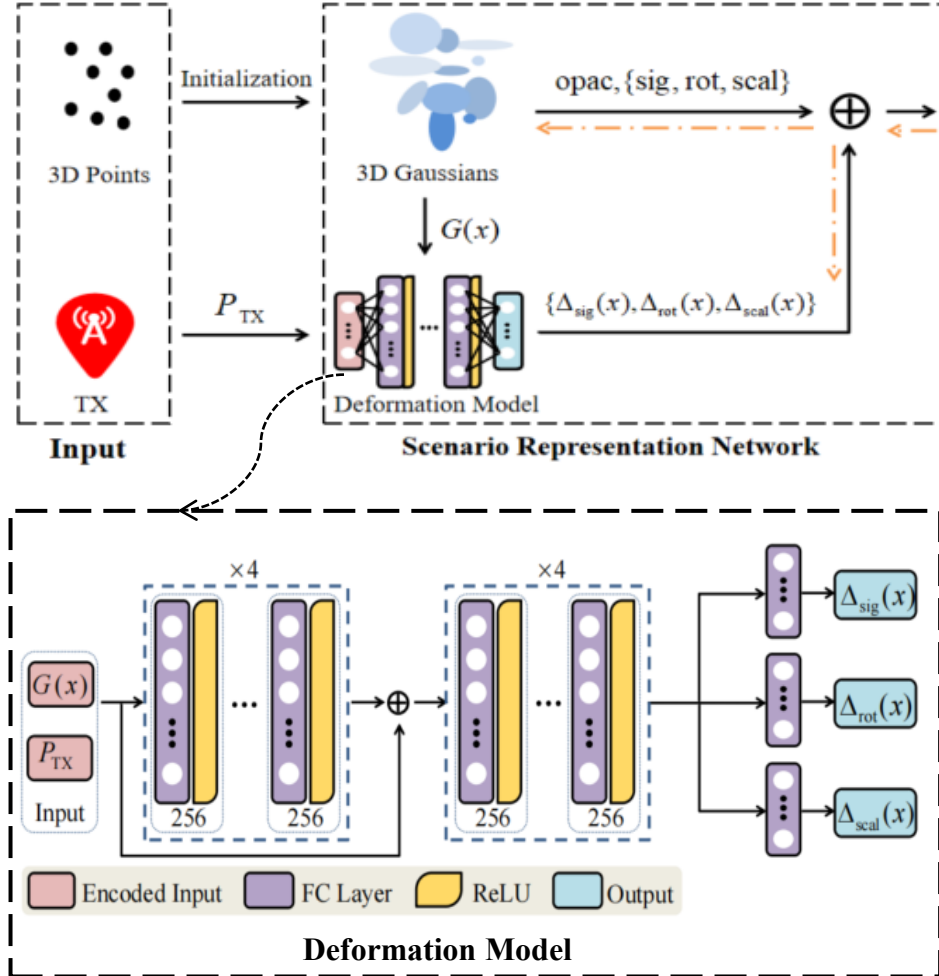


Fig. 6: Architecture of the neural model.

Core idea: Inspired by large-scale and small-scale fading, the 3D Gaussian attributes in the scene are decomposed into **static and dynamic components**.

- **Static component:** analogous to large-scale effects (path loss), determined by the position $\mathbf{G}(\mathbf{x})$ of the 3D Gaussian.
- **Dynamic component:** analogous to the small-scale effect, determined also by the Tx locations, the attribute offsets of each virtual TX (3D Gaussian) are output, including signal strength $\Delta_{sig}(x)$, rotation $\Delta_{rot}(x)$ and scale $\Delta_{scal}(x)$.

[1] C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "Neural representation for wireless radiation field reconstruction: A 3D Gaussian splatting approach," IEEE Trans. Wireless Commun., vol. 25, pp. 7490–7504, 2026.

Experiment: Spatial Spectra

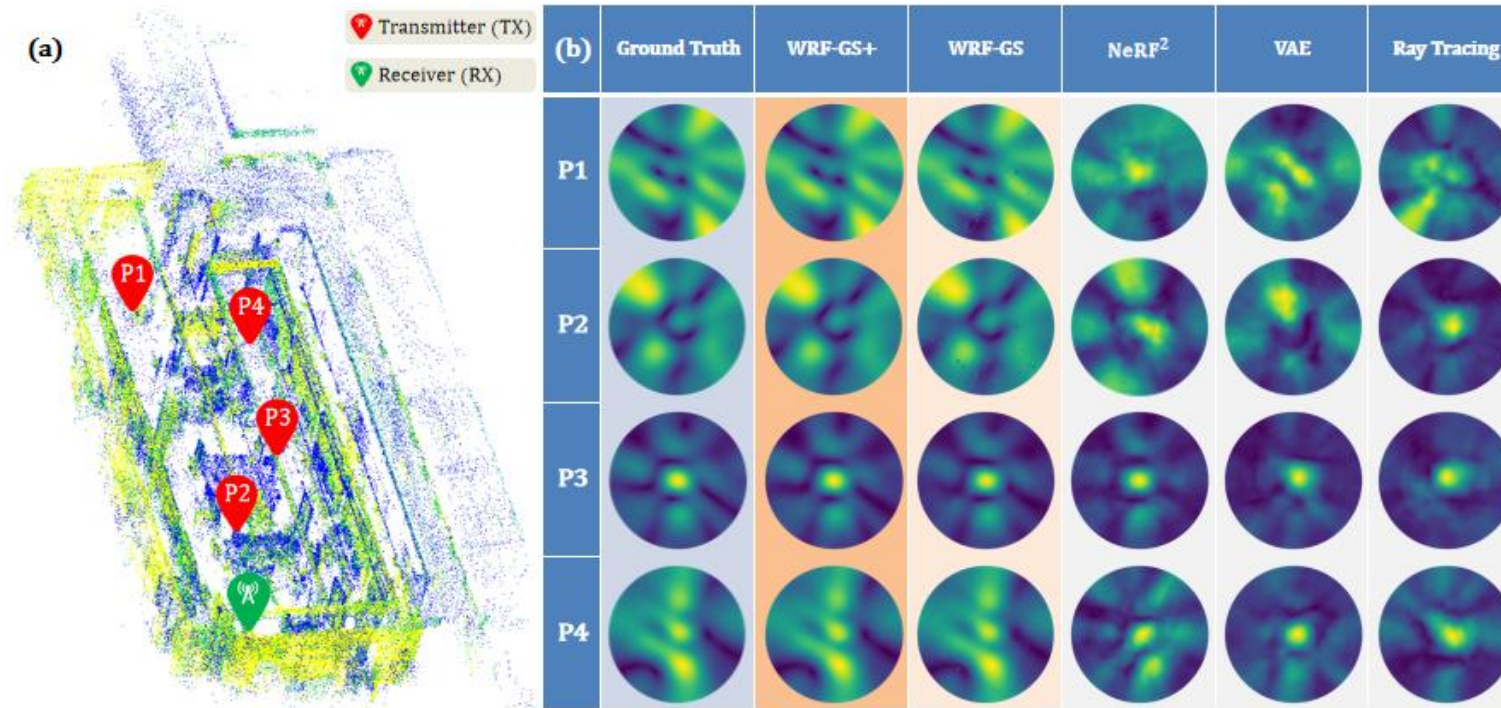


Fig. 10: Spatial spectrum generation results

(a) shows the 3D point clouds of the laboratory created by LiDAR, which are used for the ray tracing algorithm. In this setting, the TX can be located at any position, while the RX is equipped with a 4×4 antenna array and fixed at a corner. (b) compares the synthesized spatial spectra of five algorithms when the TX is located at four different positions, i.e., P1-P4, as shown in (a). The ground truth is obtained using the antenna array.

[1] C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "Neural representation for wireless radiation field reconstruction: A 3D Gaussian splatting approach," IEEE Trans. Wireless Commun., vol. 25, pp. 7490–7504, 2026.

Experiment: Spatial Spectra

- Specifically, the median SSIM of RayTracking, DCGAN, VAE, NeRF2, WRF-GS and **WRF-GS+** are 0.38, 0.56, 0.70, 0.78, 0.82 and **0.90** respectively.
- WRF-GS+** consistently outperforms the WRF-GS and NeRF2 methods for all training dataset sizes, improving the median SSIM by 3% to 13% and 11% to 19%, respectively.

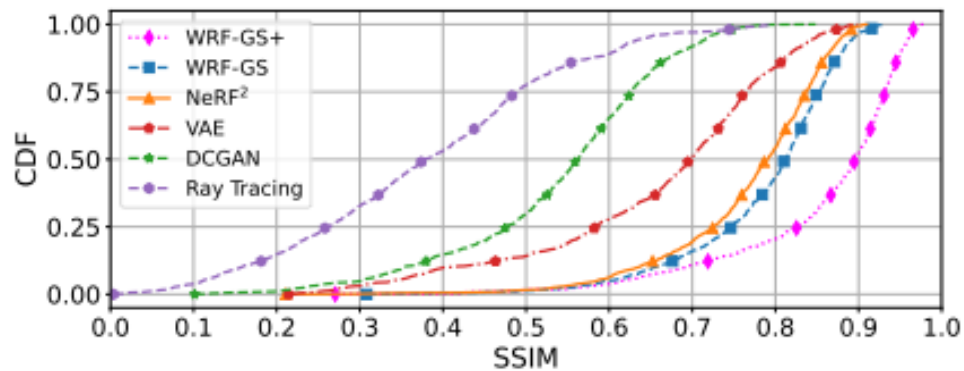


Fig. 11: The SSIM values of the synthesized spatial spectrums generated by different methods.

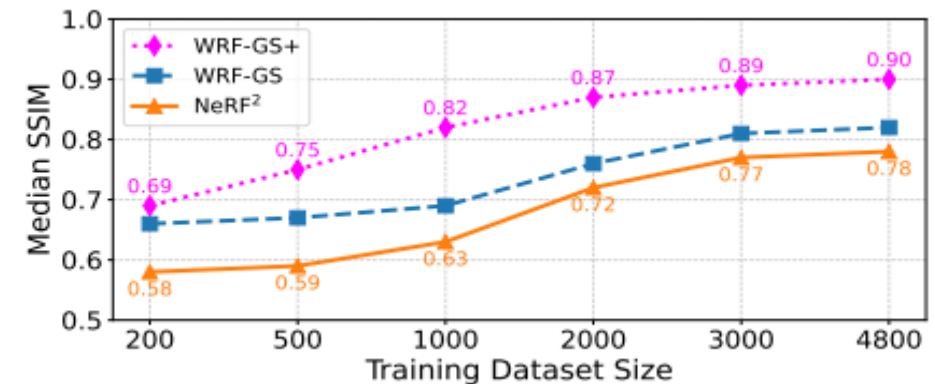


Fig. 12: The median SSIM values of the WRF-GS+, WRF-GS and NeRF2 methods under different training dataset sizes.

[1] C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "Neural representation for wireless radiation field reconstruction: A 3D Gaussian splotting approach," IEEE Trans. Wireless Commun., vol. 25, pp. 7490–7504, 2026.

Experiment: Spatial Spectra

- The processing time of WRF-GS+ and WRF-GS are usually **one to three orders of magnitude less than other methods**.

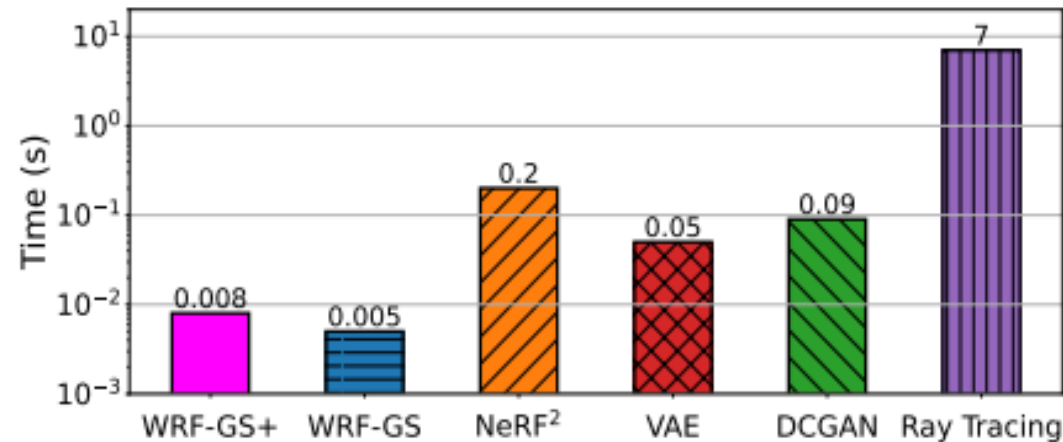


Fig. 13: The rendering time required to synthesize a spatial spectrum using different methods under the same settings.

[1] C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "Neural representation for wireless radiation field reconstruction: A 3D Gaussian splatting approach," IEEE Trans. Wireless Commun., vol. 25, pp. 7490–7504, 2026.



WRF-GS: wireless field reconstruction

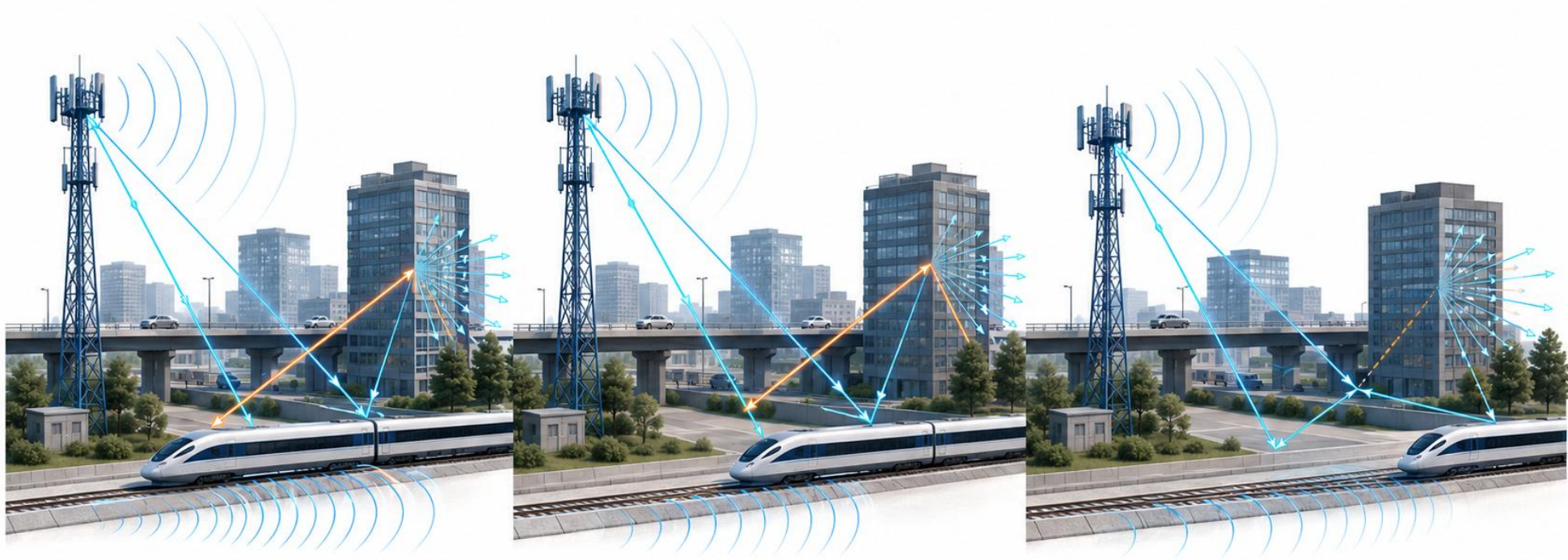


From WRF to WRF-GS+



Temporal non-stationarity and mobility-aware modelling

Towards temporal non-stationary scenarios - High Speed Railways (HSR)



- Dense Users
- Fast aging channels
- Rich Scattering/Speculars paths

Challenges of channel estimation/prediction at HSRs

Challenges 1:

Huge Channel Dimensions

CFR Dim.: Freq (**1024**) * Ant (**100**) * Ues *
Time (**14** per slot)

Challenges 2:

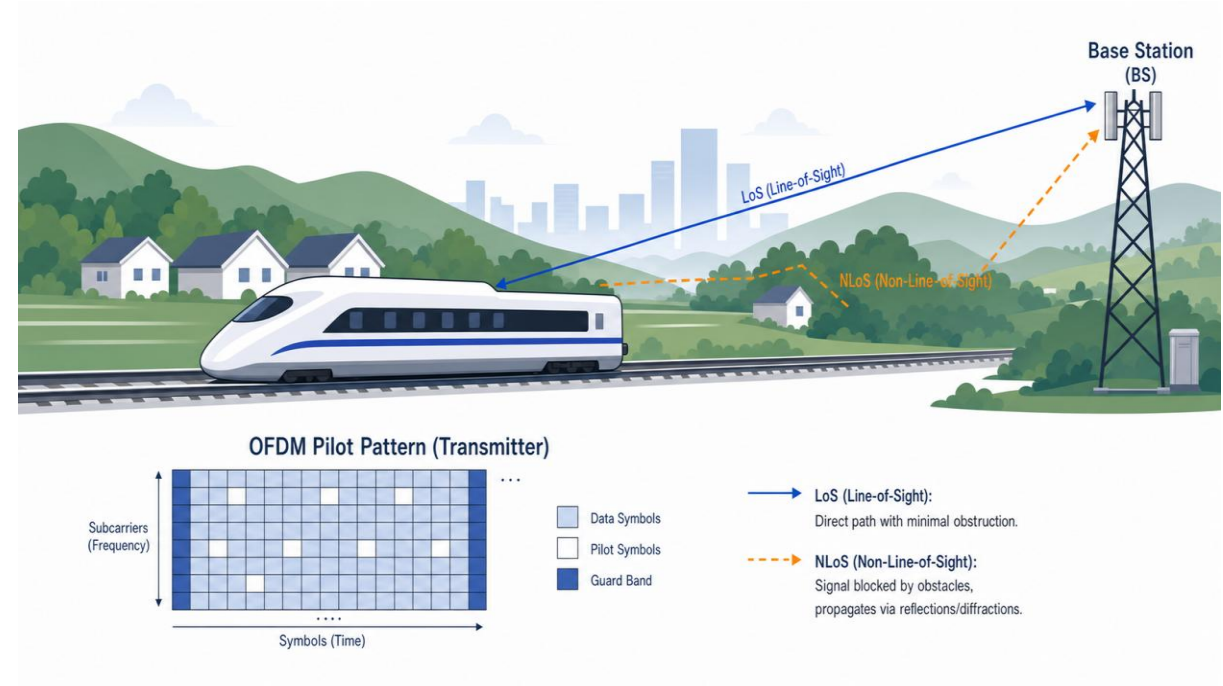
Sparse measurements

Pilot Dim.: Freq (**72**) * Ant (**100**) * Ues *
Time (**1** per slot)

Challenges 3:

Doppler effects

Mobility gives every path **its own**
temporal clock

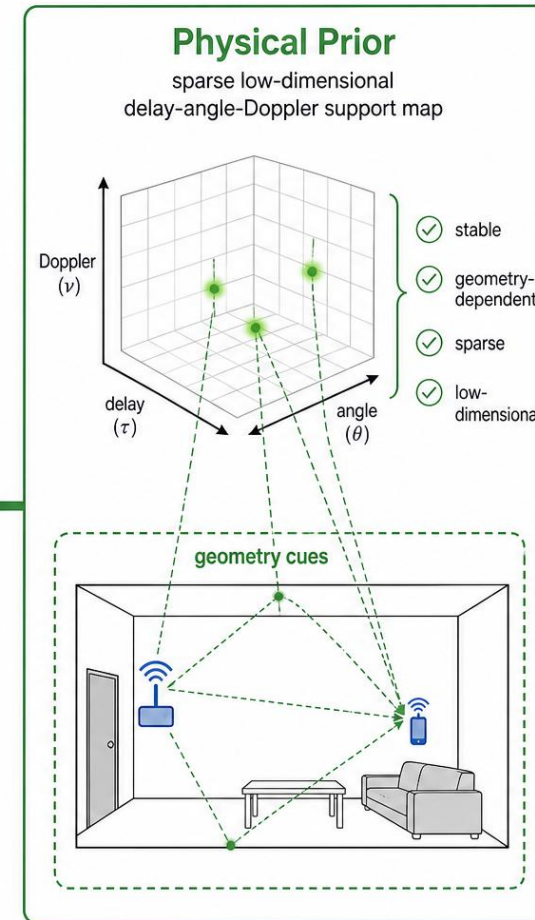
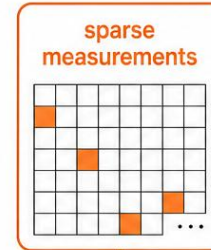
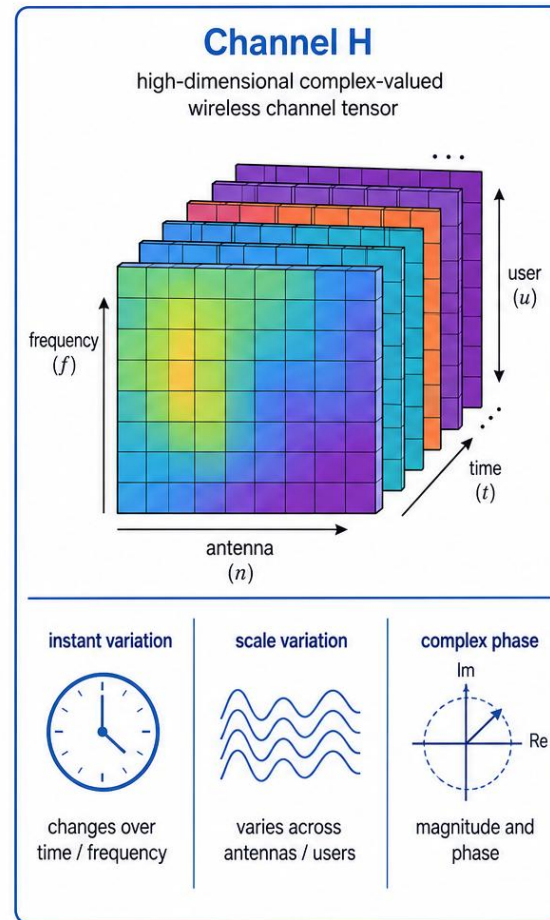


From coherent to non-coherent -- predict the power support, not the phase

FAST / COHERENT

Instantaneous
CFR

H_t :
Dim: 1024×100
(f, n)



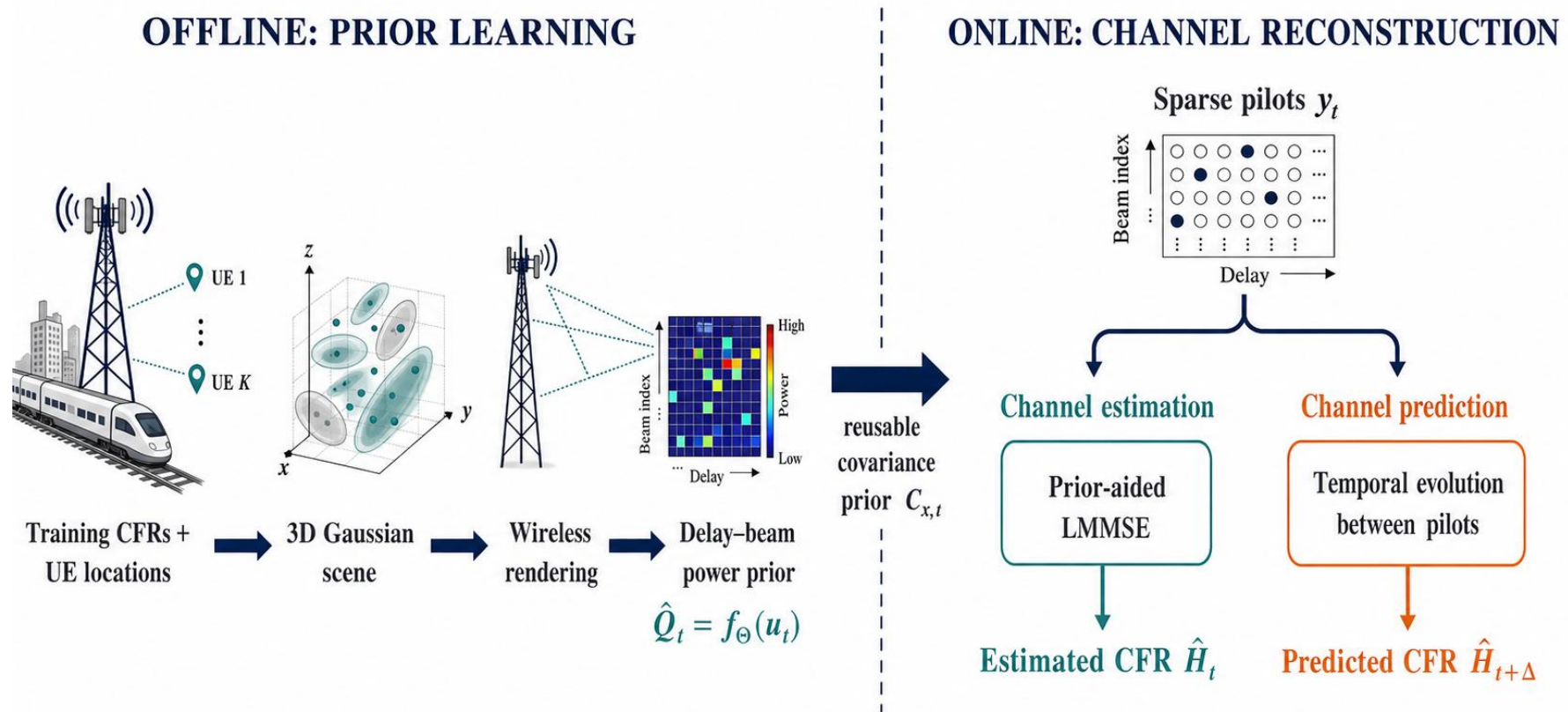
SLOWER /
NON-COHERENT

Delay-beam
power

X_t :
Dim: 1024×100
(τ, θ)

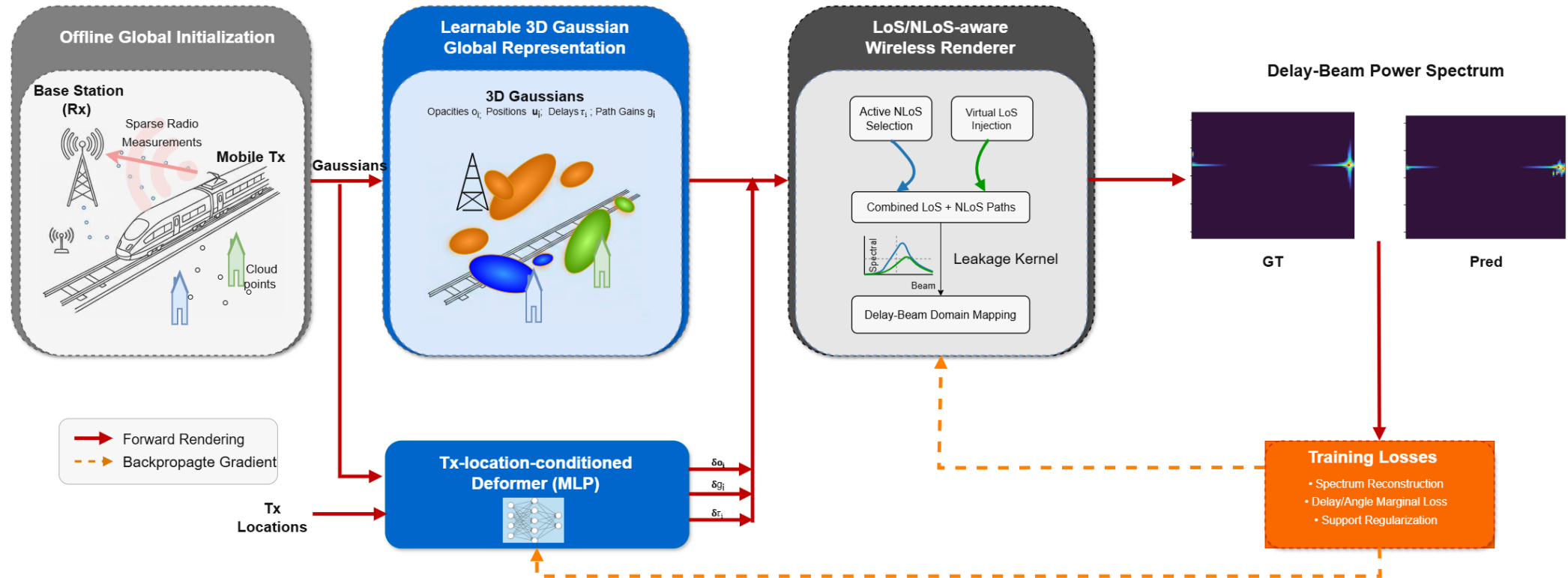
$$\mathbf{X}_t = \mathbf{P}_D \mathbf{X}_t^{\text{FULL}} = \mathbf{P}_D \mathbf{F}_N^H \mathbf{H}_t \mathbf{F}_M$$

Two stage channel estimation and channel learning



Geometry learns the support; pilots recover phase; temporal prediction bridges pilot gaps.

Learning the stable, non-coherent, geometry-dependent delay-beam priors



[1] Y. Zhang, J. Guo, C. Wen, C. Bian, and J. Zhang, "GeoGS-CE: Learning delay-beam channel priors with 3D Gaussians for high-mobility scenarios, ICC 2026

Techniques ① geometry-dependent 3D-Gaussian representatives

Each Gaussian primitives is represented by:

$$\mathcal{G}_{\Theta} = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, o_i, \Delta\tau_i, g_i\}_{i=1}^{N_G}$$

Geometry semantics

$\{\boldsymbol{\mu}_i\}$: central position of Gaussian primitives

$\{\boldsymbol{\Sigma}_i\}$: volumetric support of Gaussian primitives

$\{\Delta\tau_i\}$: the learnable delay residual

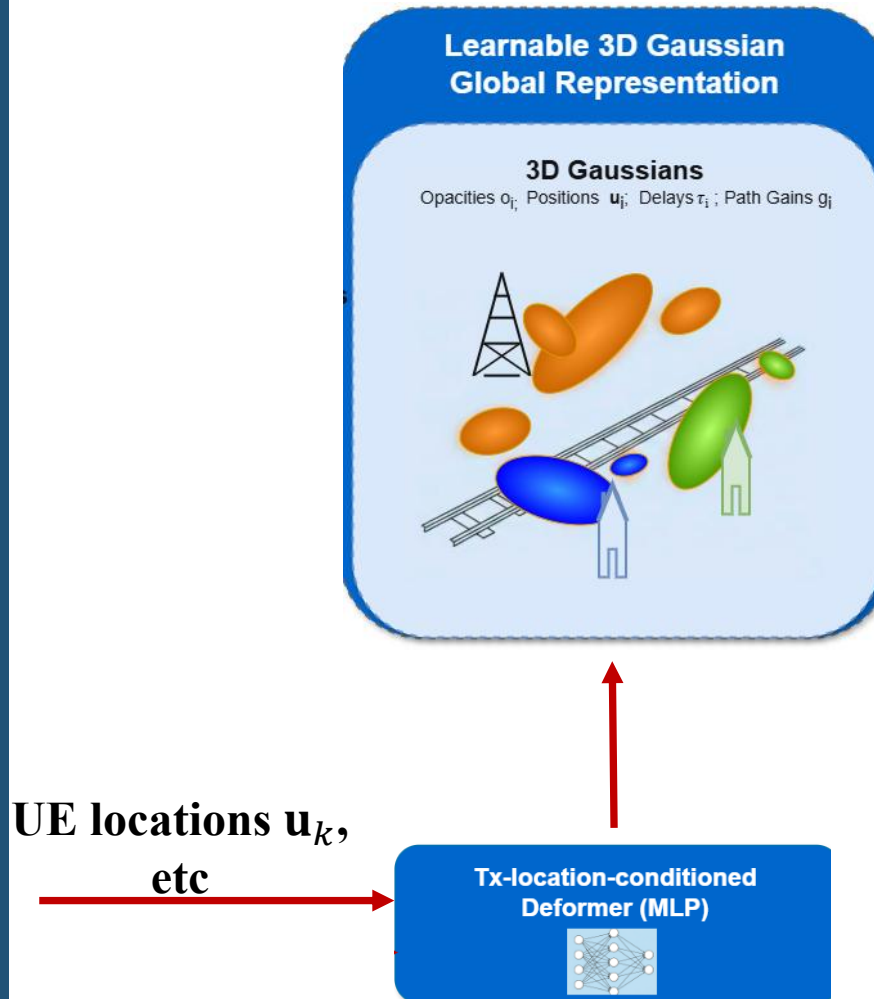
Radio semantics

$\{o_i\}$: the visibility (existence) of the represented NLoS path.

$\{g_i\}$: the learnable base path gain

Learnable features from MLP:

$$(\delta o_{i,k}, \delta\tau_{i,k}, \delta g_{i,k}) = d_{\Phi}(\mathbf{u}_k, \boldsymbol{\mu}_i)$$



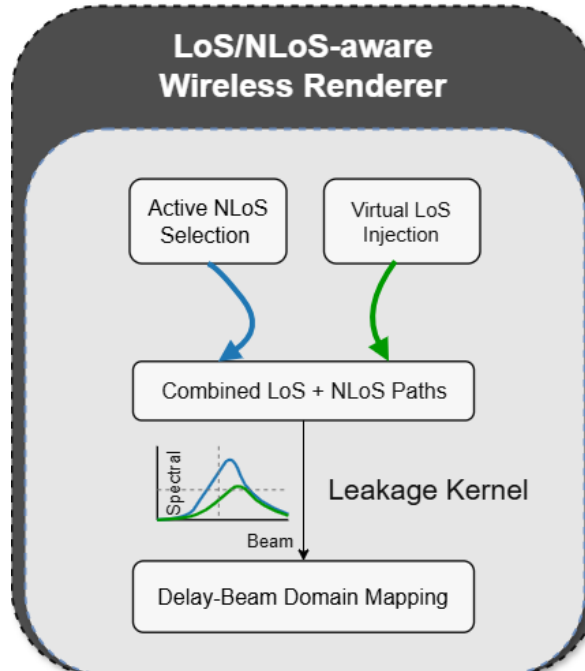
[1] Y. Zhang, J. Guo, C. Wen, C. Bian, and J. Zhang, "GeoGS-CE: Learning delay-beam channel priors with 3D Gaussians for high-mobility scenarios, ICC 2026

Techniques ② physics-embedded wireless renderer

The delay-beam leakage kernel per path:

$$\hat{Q}_k(\ell, b) = \sum_{p \in \mathcal{P}_k} w_{p,k} K_\tau(\tau_\ell - \tau_{p,k}) K_b(v_b - v_{p,k})$$

ℓ : delay index, b : angular index,
 $\tau_{p,k}$: resolution-limited delay bin,
 $v_{p,k}$: resolution-limited angular bin



① FFT windowing frequency leakage:

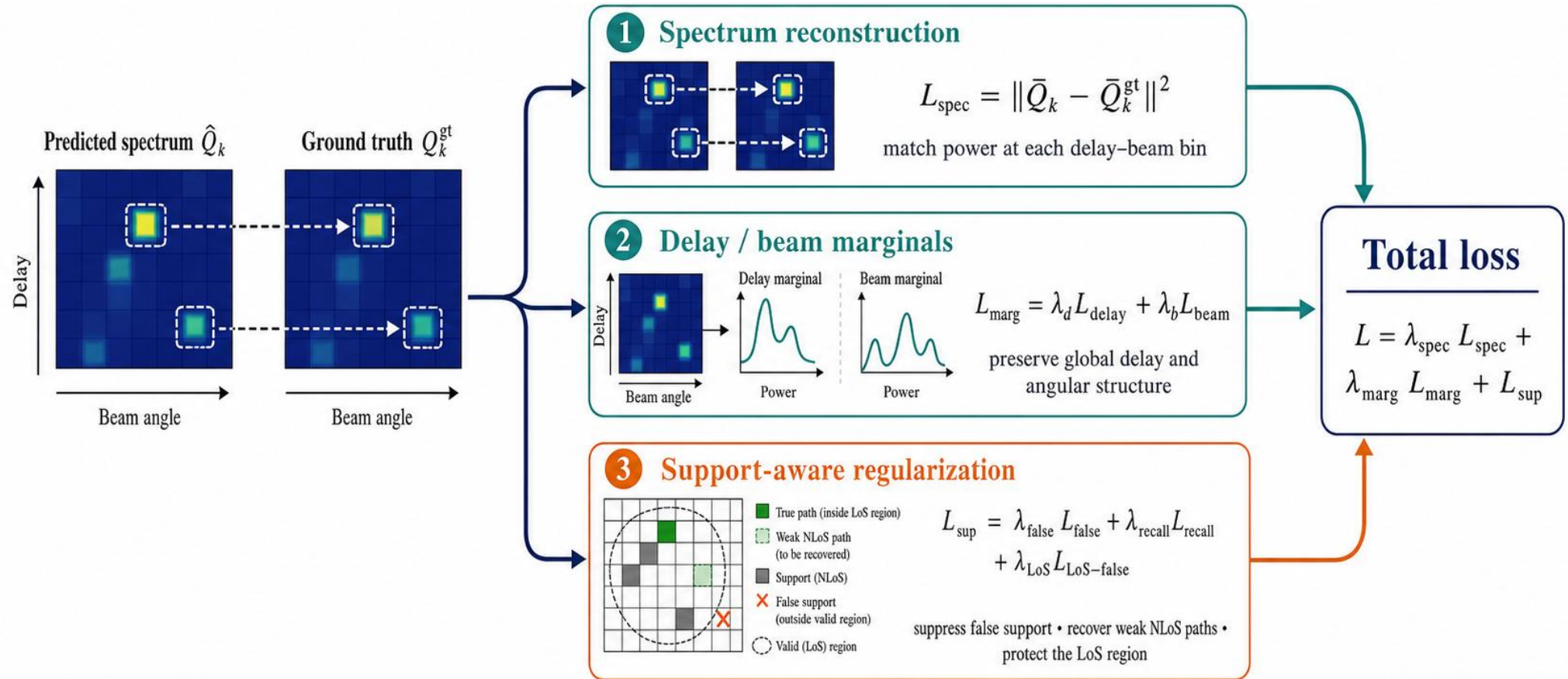
$$K_\tau(\tau_\ell - \tau) = \left| \frac{1}{N} \sum_{n=0}^{N-1} e^{-j2\pi n \Delta f (\tau_\ell - \tau)} \right|^2$$

② IFFT windowing angular leakage:

$$K_b(v_b - v) = \left| \frac{\mathbf{a}(v_b)^H \mathbf{a}(v)}{\|\mathbf{a}(v_b)\|_2 \|\mathbf{a}(v)\|_2} \right|^2$$

$\mathbf{a}(v)$: antenna manifold vector at angle v

Techniques ③ Training Loss

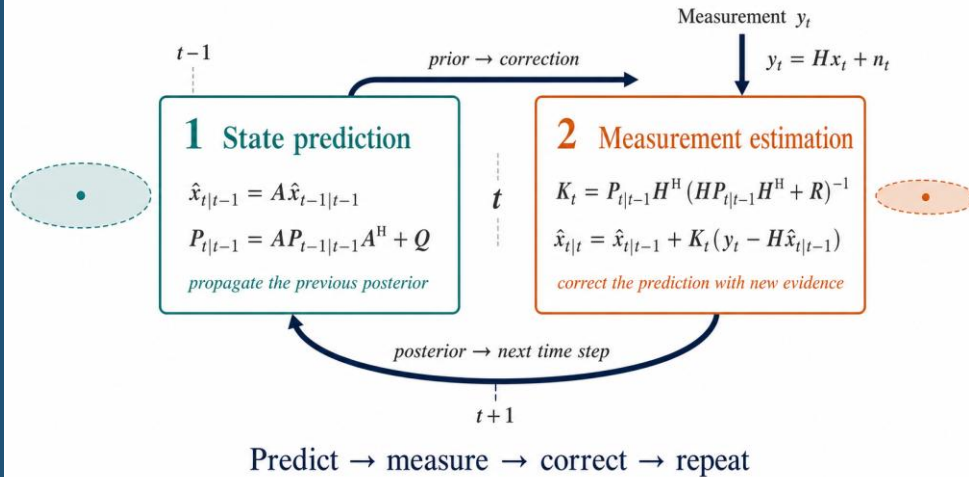


Accurate bins, stable marginals, and physically valid path support.

Online Learning Preliminaries – From Kalman filtering to lightweight causal prediction

Our lightweight causal prediction:

Kalman Filtering: Predict–Correct Iteration



Channel estimation

$$\hat{\mathbf{x}}_t = \mathbf{C}_{x,t} \mathbf{A}_{\mathcal{P}}^H (\mathbf{A}_{\mathcal{P}} \mathbf{C}_{x,t} \mathbf{A}_{\mathcal{P}}^H + \sigma_w^2 \mathbf{I})^{-1} \mathbf{y}_t.$$

$$\mathbf{C}_{x,t} = \text{diag}(\text{vec}(\hat{\mathbf{Q}}_t + \epsilon_q)),$$

$\hat{\mathbf{Q}}_t$: predicted delay-beam prior

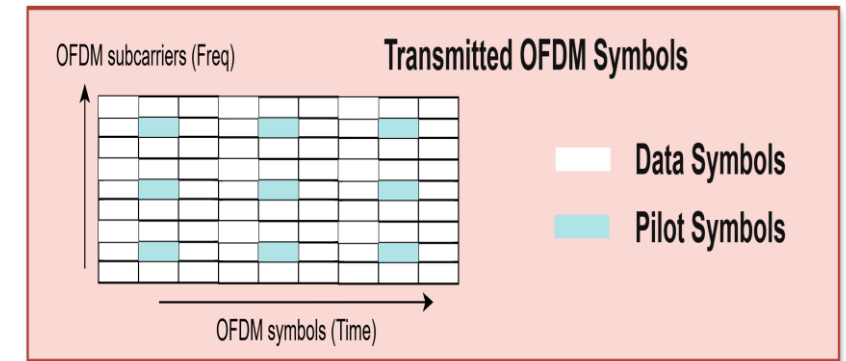
Channel prediction

$$\hat{\mathbf{x}}_t = \hat{\alpha}^{t-t_{\text{ref}}} \hat{\mathbf{x}}_{t_{\text{ref}}}$$

$\hat{\alpha} \in \mathbb{C}$ is estimated from the current and the most recent measurements

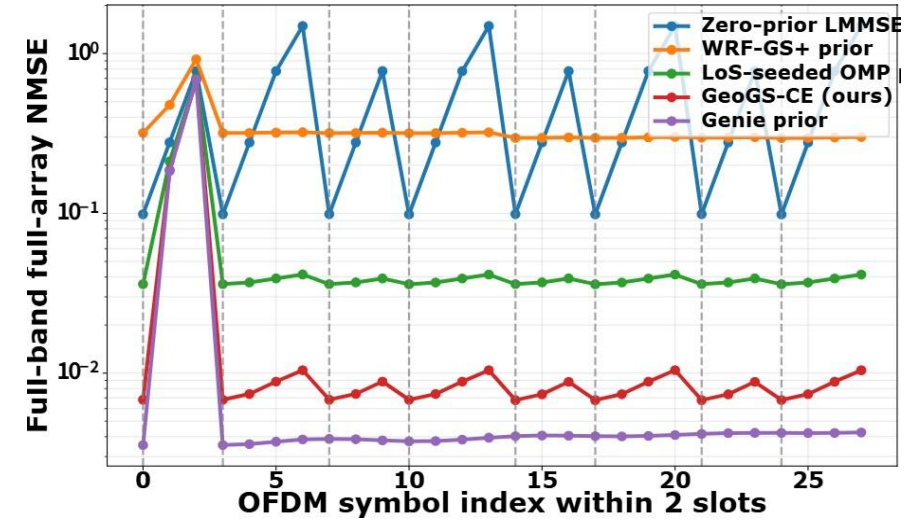
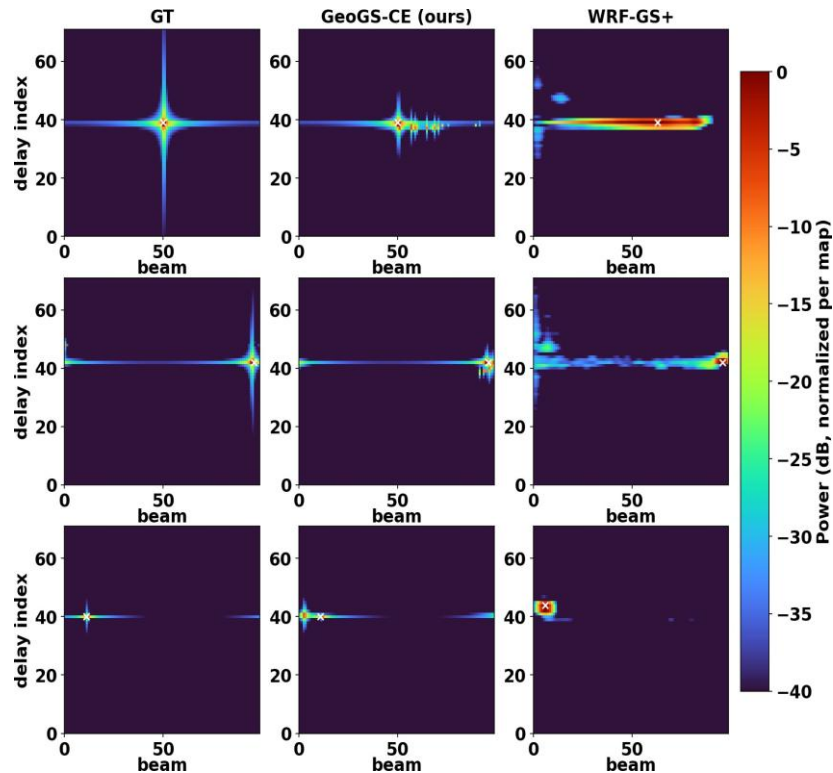
Experiment set-up

- **Low online complexity:** lightweight LMMSE after offline prior learning.
- **Realistic setting:** Guangshen HSR, 350 km/h, 100-antenna BS.
- **Sparse pilots:** 72 (of 1024) pilot tones on 4 of 14 OFDM symbols.
- **Deployment fit:** largest gains appear when pilot resources are tight.



[1] Y. Zhang, J. Guo, C. Wen, C. Bian, and J. Zhang, "GeoGS-CE: Learning delay-beam channel priors with 3D Gaussians for high-mobility scenarios, ICC 2026

Experimental Results:



- Suggests a clean division of labor: **AI for priors, estimators for recovery.**
- Outperforms pilot-only and non-geometric priors.
- Gains are most valuable on non-measurement symbols and sparse-pilot regimes.

Takeaways

- 3D-GS represents the scene with **explicit differentiable Gaussian primitives**.
- WRF-GS adapts **projection, sorting, and splatting** to wireless EM field reconstruction.
- WRF-GS+ separates **static scene structure** from **Tx-dependent dynamic variations**.
- Mobility-aware modelling learns **stable delay-beam priors** for sparse-pilot, time-varying channels.



Part III

Spatial Intelligence Generalization



Few/Zero Shot Generalization in Single Scene



Generalization across Scenes



Indoor Positioning

Motivation

□ Why inverse rendering?

- Inverse rendering aims to estimate the physical attributes of a 3D scene, such as material properties, geometry, and lighting, from captured images.
- It supports multiple downstream tasks such as scene editing and changing lighting conditions.

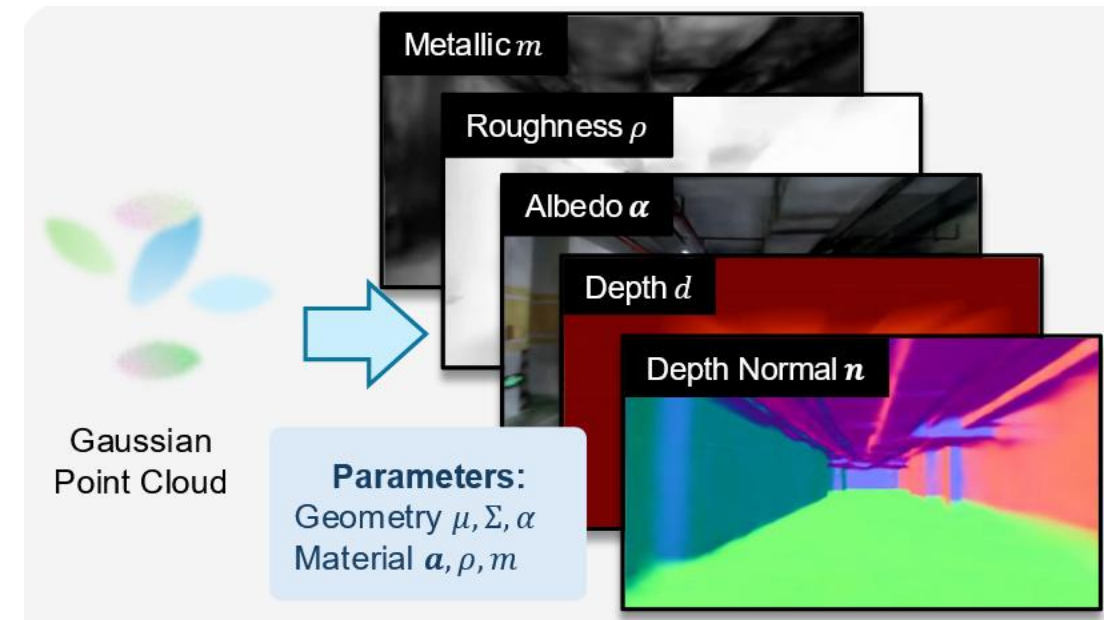


Fig. 6: An illustration of 3D-GS based inverse rendering.



We propose **URF-GS**, a wireless channel **inverse rendering** framework based on **3D-GS**, to support scene **generalization ability**!

Problem Description

- ❑ **Basic scene settings:** **Fixed** TX position and **random** RX position.
- ❑ **Reasonable assumptions:** The main layout in the space remains **unchanged**.
- ❑ **Objectives:**
 - ▶ Reconstruct the spatial wireless radiation field.
 - ▶ Generalization capability: Support changing the position of any TX-RX within the scene:
 - › Zero-shot generalization capability
 - › Few-shot generalization capability

URF-GS: Framework

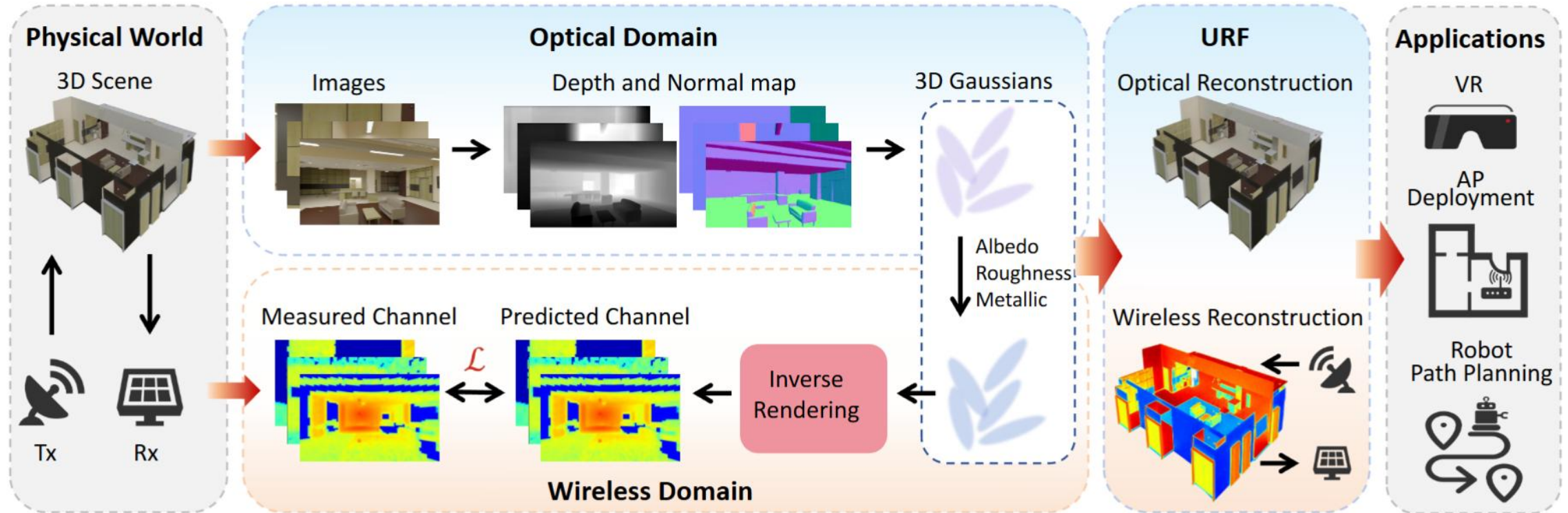


Fig. 7: An overview of the URF-GS framework.

Optical Domain: Utilizing Depth Cues

- We adopt the linear interpolation to compute the depth values of N Gaussians:

$$\hat{D} = \sum_{i=1}^N \hat{w}_i d_i \quad \text{and} \quad \hat{w}_i = \frac{T_i \alpha_i}{\sum_{i=1}^N T_i \alpha_i},$$

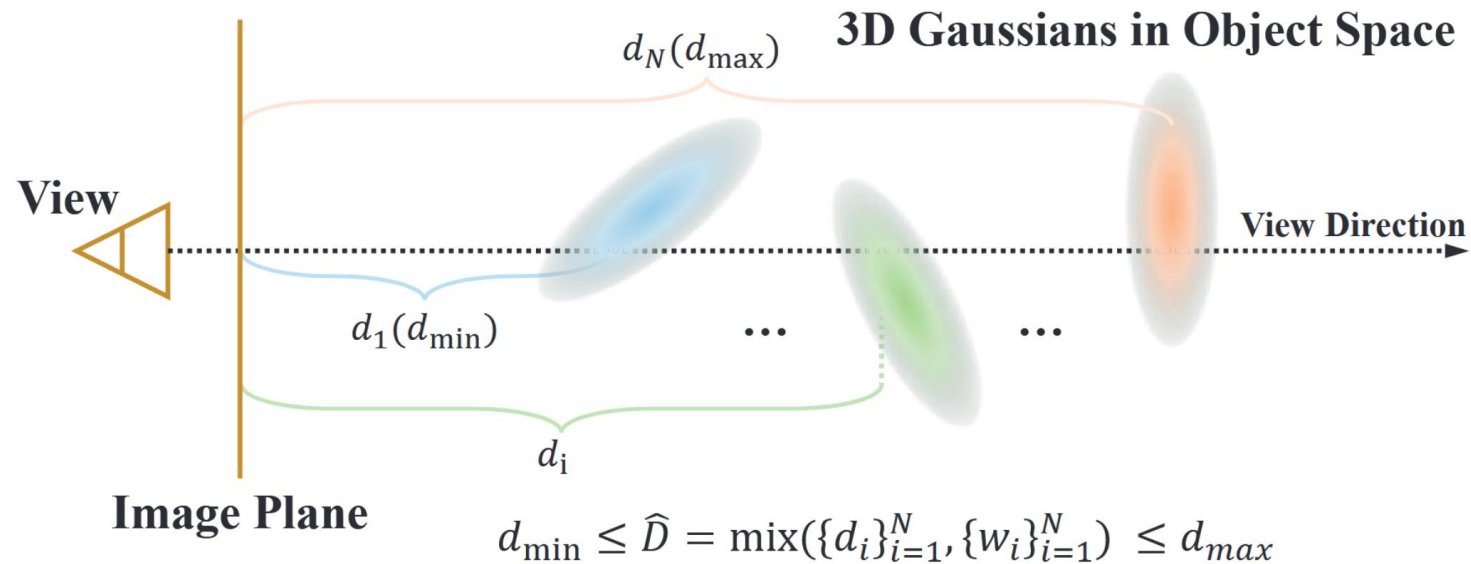


Fig. 8: Depth Illustration.

Optical Domain: Utilizing Depth Cues

- We leverage pre-trained monocular models for depth estimation, such as ZoeDepth and DepthAnything , to obtain dense pixel-wise depth priors.

$$\hat{a}, \hat{b} = \arg \min_{a,b} \sum_{m,n} \|(a \cdot D_{\text{Mono}}^{m,n} + b) - D_{\text{SfM}}^{m,n}\|_2^2,$$

- We define the aligned depth map as:

$$D_{\text{aligned}} \triangleq \hat{a} \cdot D_{\text{Mono}} + \hat{b}.$$

- Then, we employ a gradient-aware loss to effectively balance regularization across various image regions which is expressed as:

$$\mathcal{L}_{\hat{D}} = g_{\text{rgb}} \frac{1}{|\hat{D}|} \sum \log \left(1 + \|\hat{D} - D_{\text{aligned}}\|_1 \right)$$

- where $g_{\text{rgb}} = \exp(-\nabla I)$ and ∇I denoting the gradient of the currently aligned RGB image.

Optical Domain: Utilizing Normal Cues

- The normal map can be obtained by using α -blending:

$$\hat{N} = \sum_{i \in N} \mathbf{n}_i \alpha_i T_i,$$

- We utilize monocular normal estimates from the pre-trained Omnidata model as the GT normal:

$$\mathcal{L}_{\hat{N}} = \frac{1}{|\hat{N}|} \sum \|\hat{N} - N\|_1.$$

- We further apply a prior on the total variation of predicted normals, encouraging smooth normal predictions at neighboring pixels with:

$$\mathcal{L}_{\text{smooth}} = \sum_{m,n} \left(|\nabla_m \hat{N}_{m,n}| + |\nabla_n \hat{N}_{m,n}| \right),$$

- To sum up, the final normal regularization loss is expressed as

$$\mathcal{L}_{\text{normal}} = \mathcal{L}_{\hat{N}} + \mathcal{L}_{\text{smooth}}.$$

Optical Domain: Utilizing Normal Cues

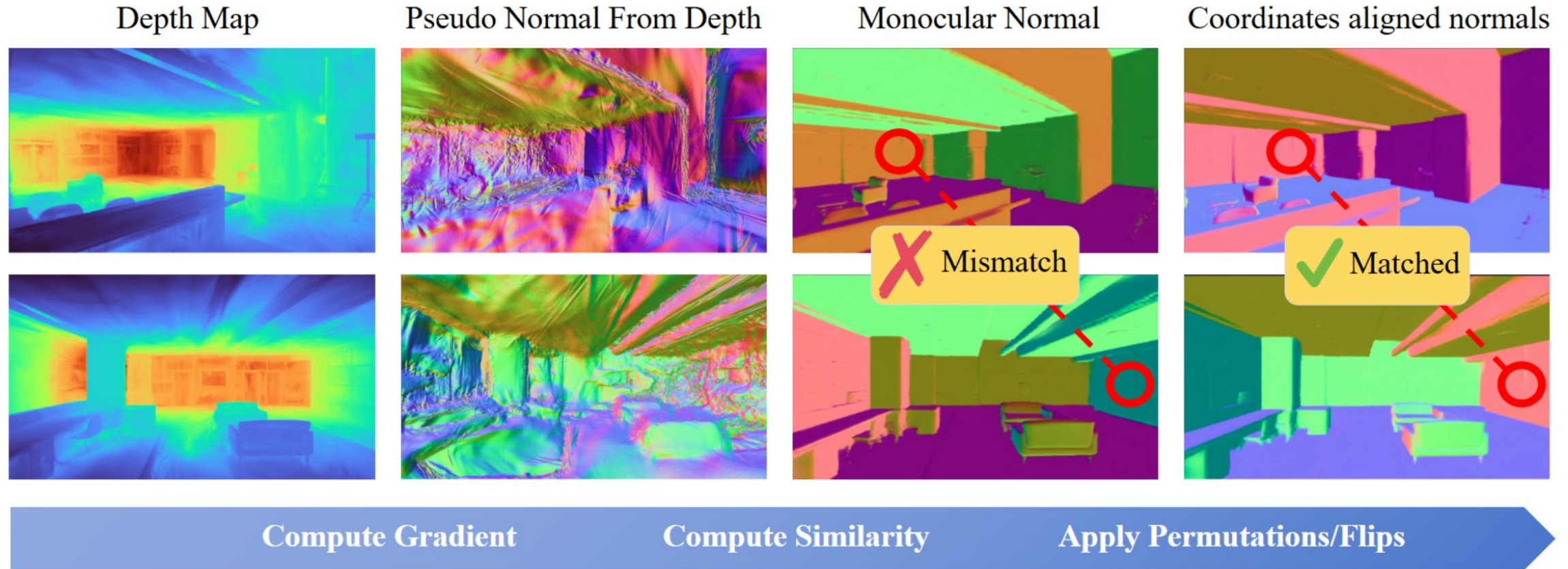


Fig. 9: Coordinate alignment of monocular normal maps using depth-derived pseudo normals.

Wireless Domain

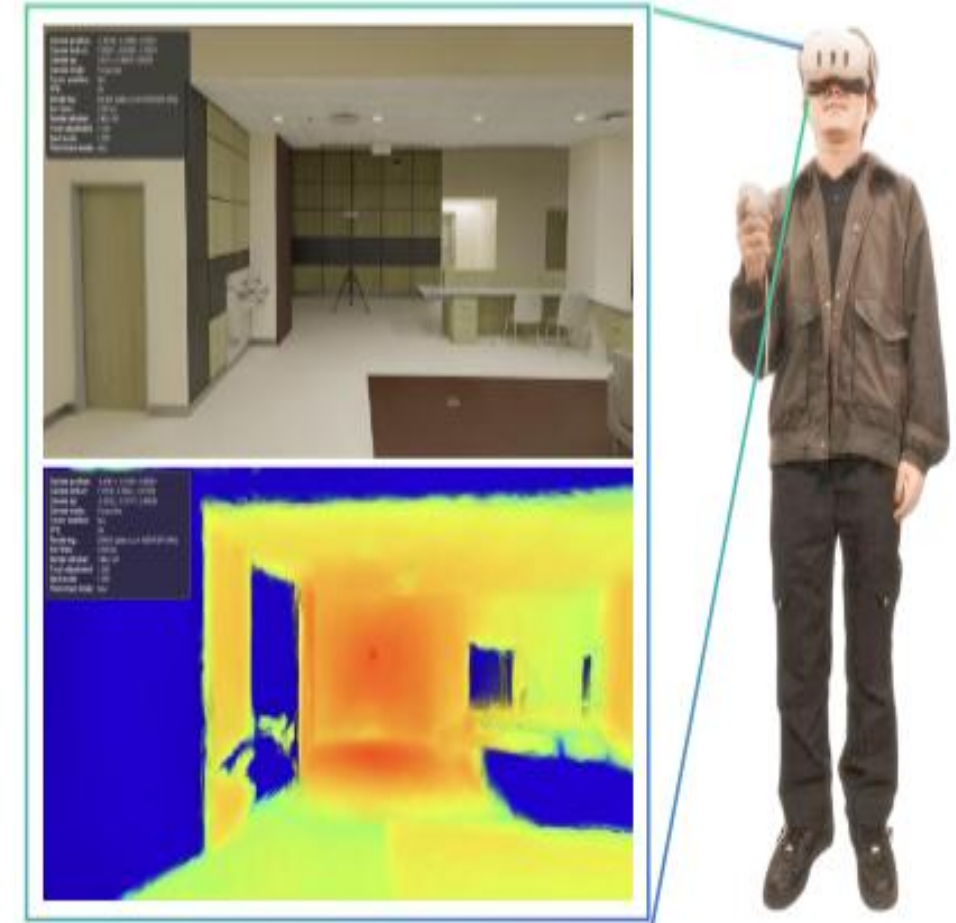
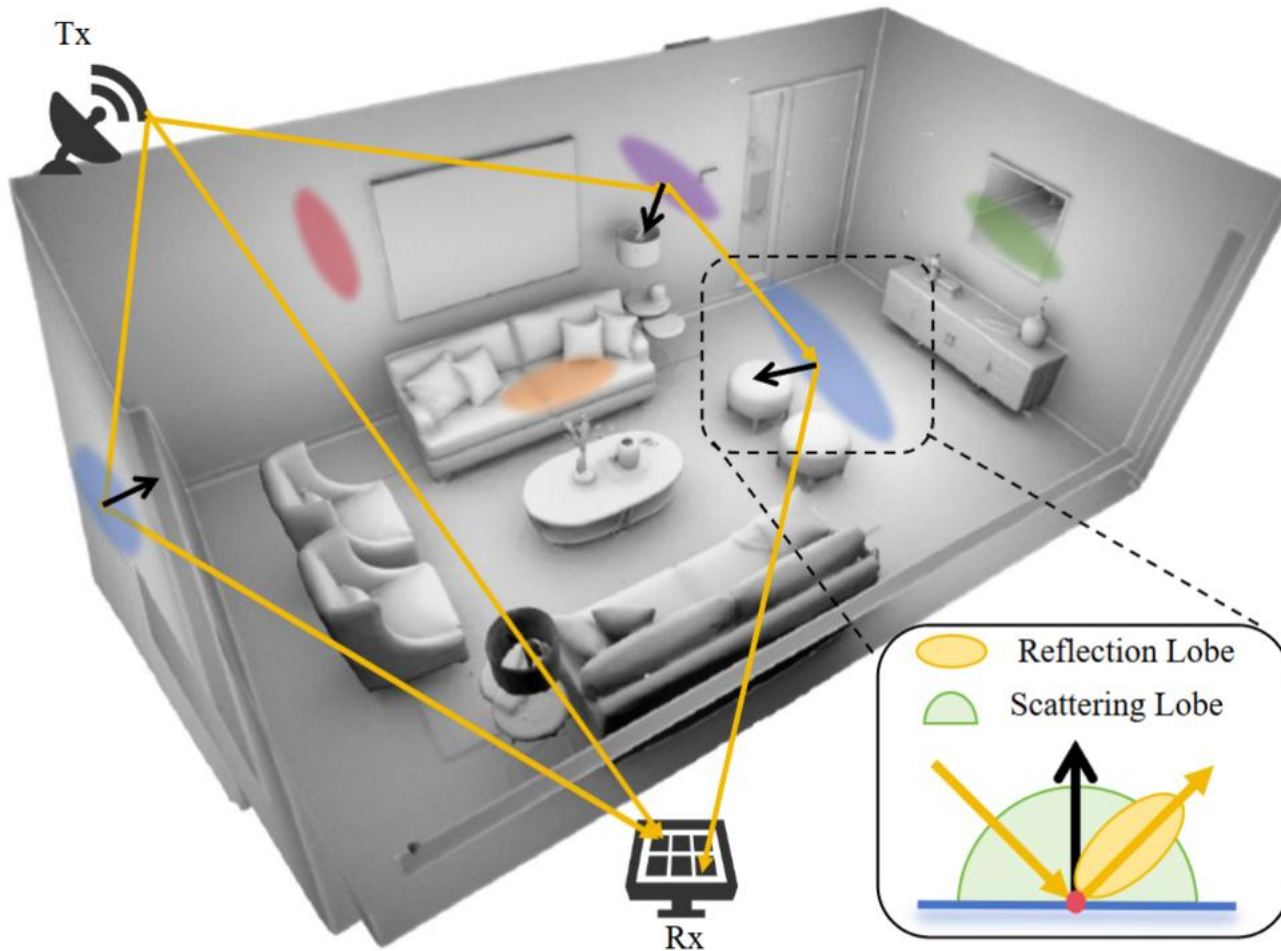


Fig. 10: An illustration of physically-based inverse rendering. Fig. 11: Immersive Experience of URF with VR Devices.

Wireless Domain: Surface Interaction Pattern

- In order to avoid solving complex Maxwell's equations and further improve the universality of the model, we simplified the propagation of wireless electromagnetic waves in space:

$$S_o(\omega_o, \mathbf{x}) = \int_{\Omega} f_{\text{brdf}}(\omega_o, \omega_i, \mathbf{x}) S_i(\omega_i, \mathbf{x}) (\omega_i \cdot \mathbf{n}) d\omega_i,$$

- In the wireless domain, the BRDF can be divided into a scattering component f_s and a reflection component f_r according to the Cook-Torrance BRDF model:

$$f_{\text{brdf}}(\omega_i, \omega_o) = f_s + f_r(\omega_i, \omega_o),$$

where $f_s = (1 - m)a/\pi$ and

$$f_r(\omega_i, \omega_o) = \frac{D(\mathbf{h}; \rho) F(\omega_o, \mathbf{h}; \mathbf{a}, m) G(\omega_i, \omega_o, \mathbf{h}; \rho)}{4(\mathbf{n} \cdot \omega_i)(\mathbf{n} \cdot \omega_o)},$$

Wireless Domain: Surface Interaction Pattern

- Consequently, the outgoing signal S_o can be decomposed into scattering component S_s and reflection component S_r as follows:

$$S_o(\omega_o, \mathbf{x}) = S_s(\mathbf{x}) + S_r(\omega_o, \mathbf{x}),$$

where $S_s(\mathbf{x}) = \int_{\Omega} f_s S_i(\omega_i, \mathbf{x}) (\omega_i \cdot \mathbf{n}) d\omega_i$ and

$$S_r(\omega_o, \mathbf{x}) = \int_{\Omega} f_r(\omega_i, \omega_o) S_i(\omega_i, \mathbf{x}) (\omega_i \cdot \mathbf{n}) d\omega_i.$$

- As a result, we can calculate the attenuation coefficient, defined as the ratio of the outgoing signal in a specific direction to the incoming signal in a specific direction at a certain position $\mathbf{x}$ in space, :

$$AC(\omega_o, \omega_i, \mathbf{x}) = S_o(\omega_o, \mathbf{x}) / S_i(\omega_i).$$

Wireless Domain: Free-Space Transmission Pattern

- Regarding the differences between wireless channel and optical modeling, in addition to modeling the attenuation of signals after interaction with object surfaces, the free-space path loss (FSPL) should also be taken into account.
- The resulting power P_{Rx} is given by:

$$P_{Rx} = P_{Tx} G_{Tx}(\theta_D, \phi_D) G_{Rx}(\theta_A, \phi_A) \prod_{n=1}^N AC_n \cdot FSPL,$$

and

$$FSPL = \left(\frac{c}{4\pi f (\sum_{n=1}^{N+1} d_n)} \right)^2,$$

where c represents the speed of light and f is the carrier frequency. Due to considering N interactions with the surface of the object, the number of paths for FSPL is $N + 1$.

Implementation

- **Data Caching.** We set a loading threshold with 500 samples for each retrieval. When a dataset size exceeds this threshold, a portion of this data will be cached on the CPU.
- **Deferred Shading.** It first rasterizes normal, depth, and BRDF maps into a G-buffer. Then, it queries the necessary geometric attributes from the G-buffer and employs a differentiable path tracer to resolve signal–environment interactions when evaluating propagation paths and occlusions.

Experiment

- Dataset: It includes 60 GHz measurements from the National Institute of Standards and Technology (NIST) and a high-fidelity digital replica of a real indoor scene, i.e., a 14 m × 15 m loft with tables, sofas, televisions, and several small rooms.

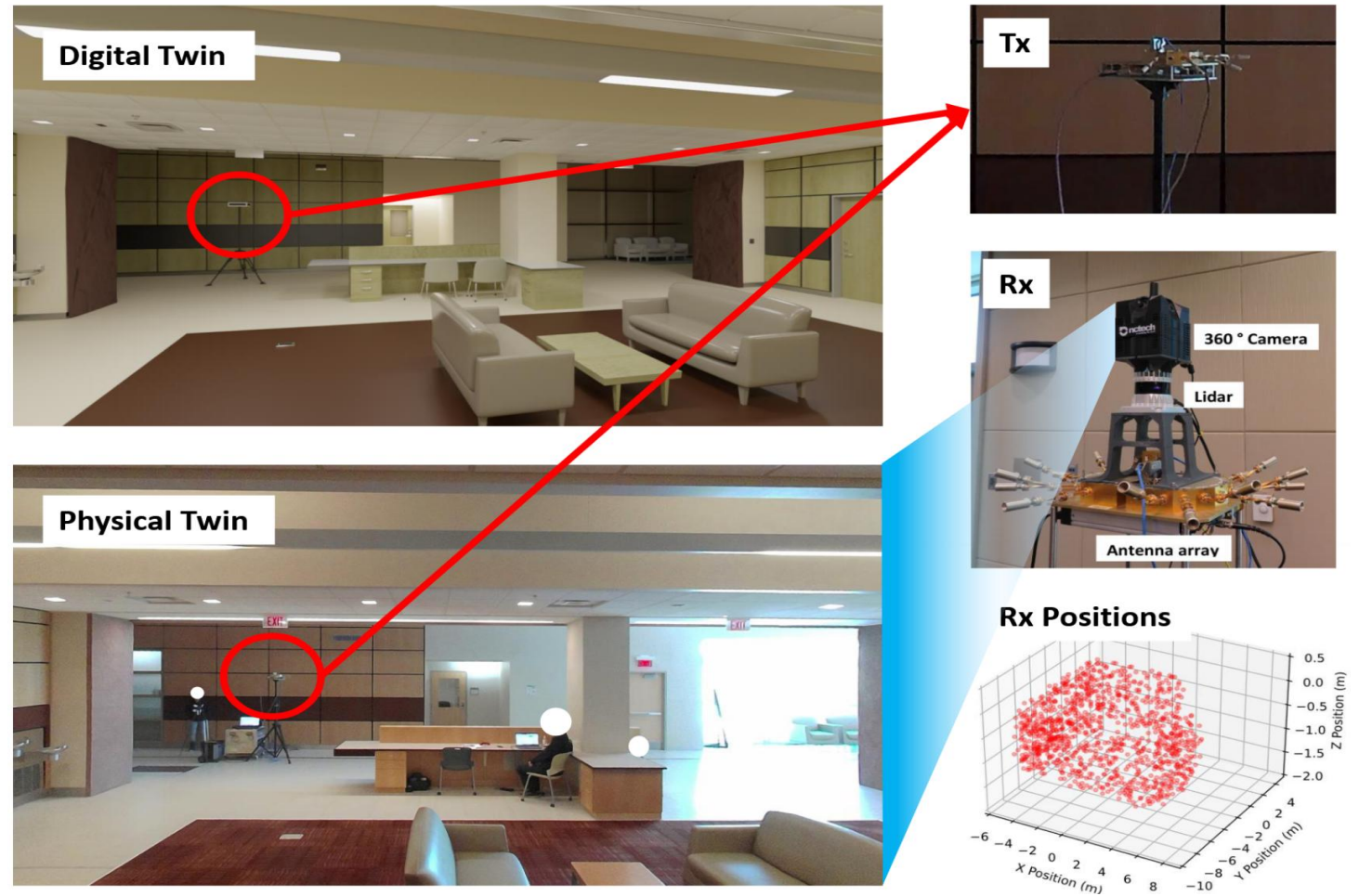


Fig. 12: NIST Dataset.

Experiment

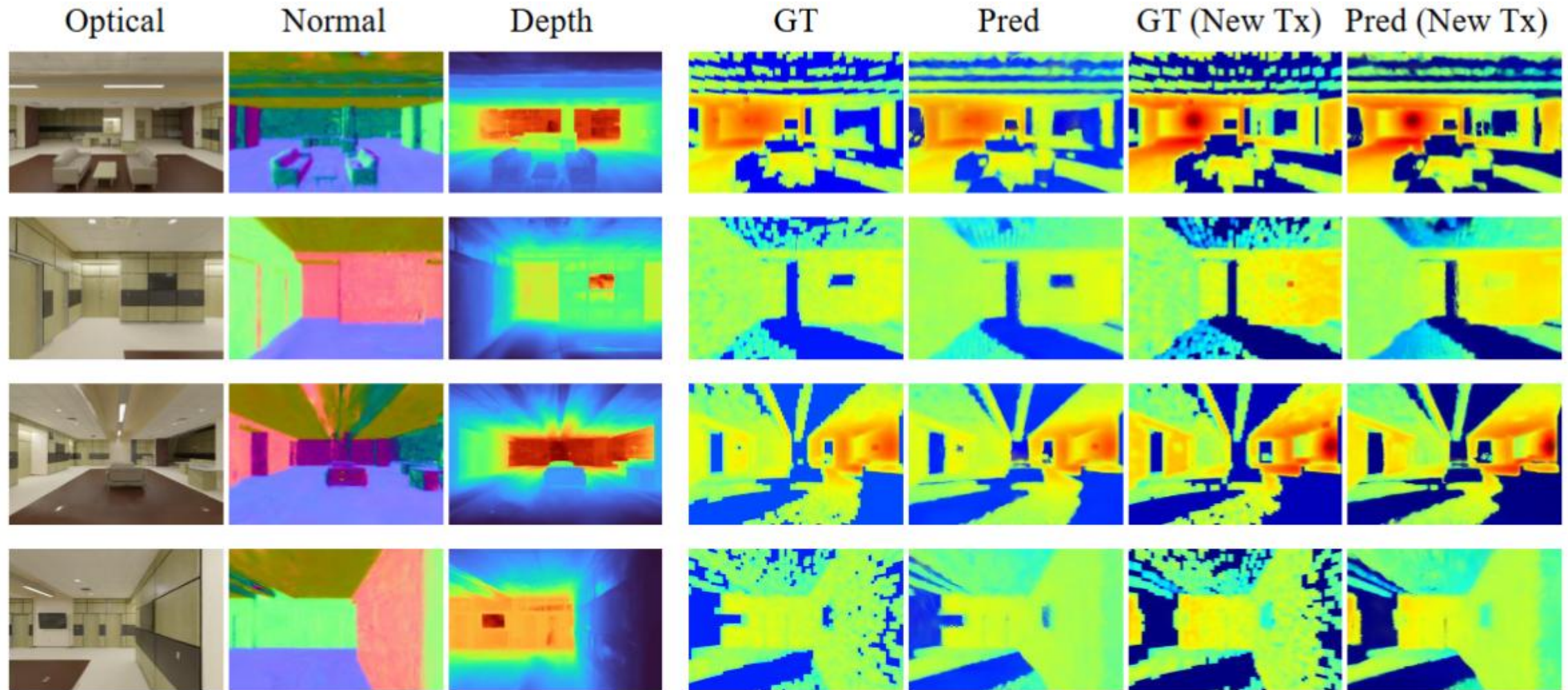


Fig. 13: The reconstructed environment-radio map in a 3D environment using the URF-GS framework.

Experiment: Zero-shot Generalization Capability

Table I compares URF-GS with the RF-3DGS, NeRF2, and WRF-GS+ methods across PSNR, SSIM, and LPIPS. We see that URF-GS owns the **best PSNR (17.3818) and SSIM (0.7012)**, as well as achieves up to **24.7% improvement** in SSIM compared with NeRF2. It matches or surpasses RF-3DGS on LPIPS(0.3340 vs. 0.3345), reflecting comparable perceptual quality.

Table 1: Performance comparison of the URF-GS, RF-3DGS, NeRF², and WRF-GS+ methods.

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
URF-GS	17.3818	0.7012	0.3340
RF-3DGS [43]	16.9949	0.6823	0.3345
NeRF ² [45]	17.0581	0.5623	0.3979
WRF-GS+ [37]	17.1085	0.6917	0.3270

Experiment: Few-shot Generalization Capability

Compares URF-GS with NeRF2, RF-3DGS, and WRF-GS+ across increasing training set sizes using PSNR, SSIM, and LPIPS. Across all data regimes, URF-GS achieves the best performance. It consistently yields the highest PSNR and SSIM, indicating more accurate and structurally faithful reconstructions, while attaining the lowest LPIPS, reflecting superior perceptual quality.

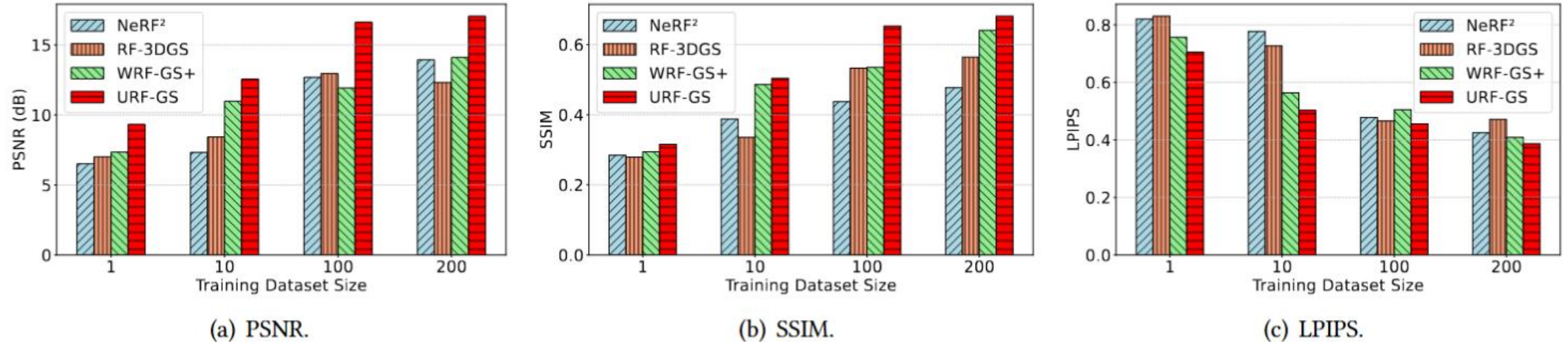


Fig. 14: Performance comparison over different training dataset sizes.

Point2Radio: A Foundation Model for Cross-Scene Radio Fields from Material-Aware Point Clouds

Chaozheng Wen, Chenghong Bian, Hongze Chen and Jun Zhang

Presenter: Chaozheng Wen



香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY



iComAI Lab

INTEGRATED COMMUNICATIONS AND AI LAB

Outline

1. Motivation
2. Problem Gap
3. Idea
4. Method
5. PRISM Dataset
6. Results

Motivation: Why Radio Fields?

Low-altitude aerial networks

Coverage-aware UAV / A2G planning

Embodied AI & multi-robot

Reliable multi-agent communication

Connected autonomy

Environment-aware wireless links

RF as a sensing modality

Complement vision with radio observations

The Challenge

Radio waves undergo reflection, transmission, scattering, and diffraction on complex surfaces and materials.

Classical models

Cheap, but miss fine spatial structure and multipath detail.

Ray tracing

High fidelity, but costly; must be repeated for every scene & TX.

Learning today

Either amortize in 2D layouts, or refit a new model per scene.

Key question: Can one model learn transferable priors of how radio interacts with 3D scenes?

Existing Trade-off

Cross-scene

- ▶ RadioUNet / layout-map networks
- ▶ Feedforward & transferable
- ▶ Usually 2D rasters → fixed path-loss maps
- ▶ Lose fine 3D geometry & materials

Per-scene neural fields

- ▶ NeRF2, WRF-GS, Gaussian RF models
- ▶ High 3D fidelity within one room
- ▶ Need target-scene labels / optimization
- ▶ Do not reuse one model across scenes

Our Idea: Foundation Model for Radio Fields

Shared backbone + task-specific heads — one transferable propagation prior.

$$\mathbf{H} = \mathbf{E}(\mathbf{R}, \mathbf{t})$$

$$\hat{\mathbf{y}} = \mathbf{D}(\mathbf{H}, \mathbf{q})$$

R

Material-aware point cloud
(geometry + EM attributes)

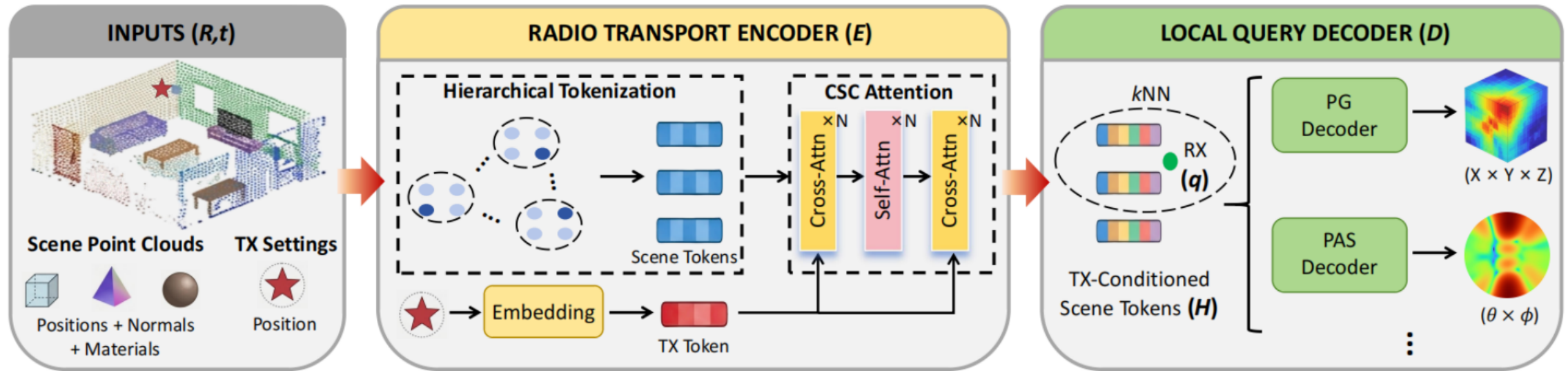
E

Radio transport encoder
→ TX-conditioned codes $\mathbf{H}$

D

Task query decoder
→ PG fields or PAS spectra

Methods — Overview

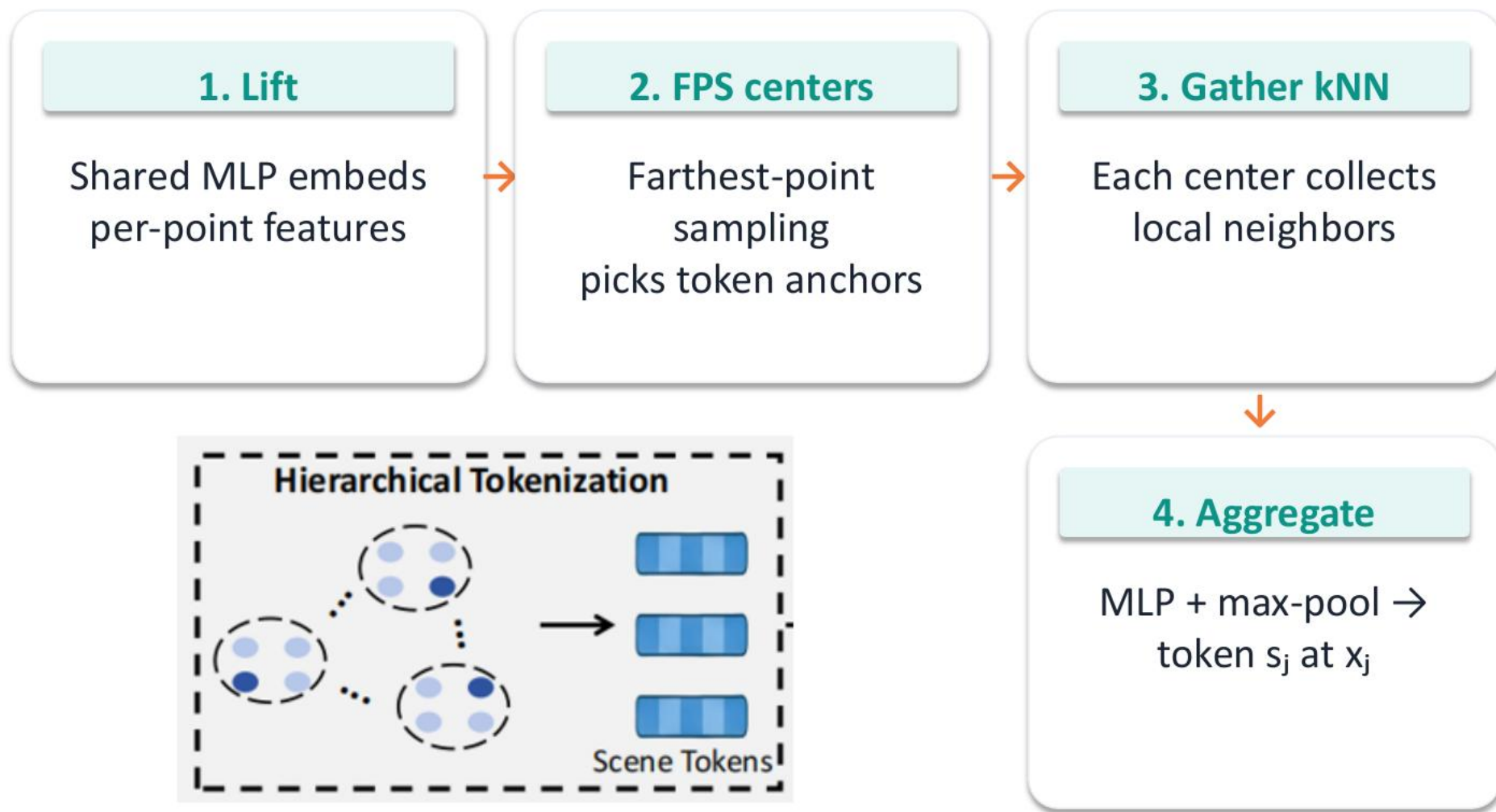


Point features: position, normal, ϵ , $\log_{10}\sigma$, scattering, XPD, emission flag

Attention as a feedforward surrogate for multipath transport (no online ray tracing)

After PG pretraining: freeze encoder $\rightarrow$ train PAS head; optional residual FT per scene

Method — Hierarchical Tokenization



Method — TX-Conditioned CSC Attention

CSC models TX–scene interactions and stores the result in scene tokens for later RX decoding.

Cross-Attn · $\times N$

TX → Scene

Inject TX emission into
scene tokens

Self-Attn · $\times N$

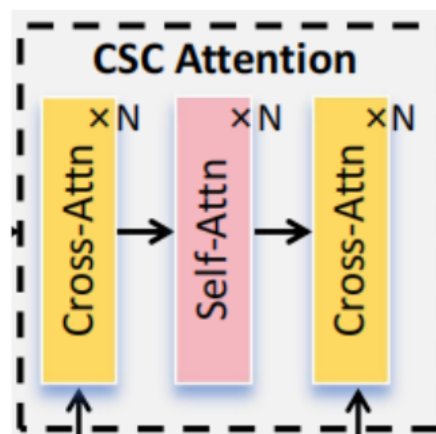
Scene ↔ Scene

Exchange long-range
context among tokens

Cross-Attn · $\times N$

TX → Scene

Re-condition tokens
with TX again



Method — Local Query Decoder

Treat TX-conditioned tokens as virtual scatterers. Decode each RX by aggregating nearby codes.

1. Query embed

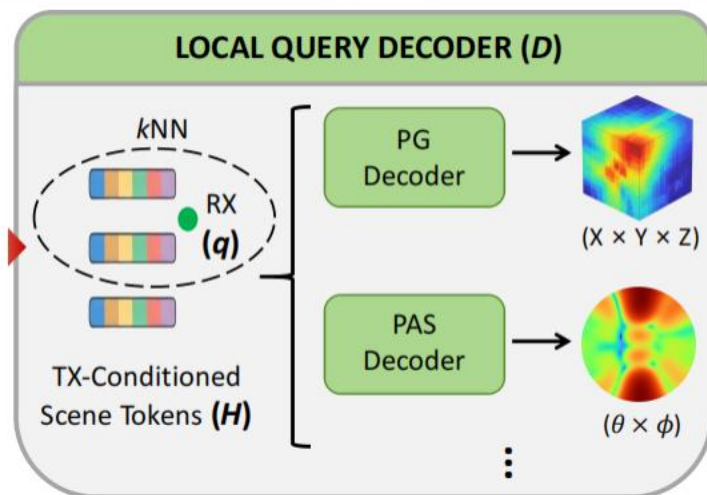
MLP over $[q, q-x_{tx}, \|q-x_{tx}\|]$
encodes RX + link
geometry

2. kNN gather

Select K nearest
codes in H
by anchor distance to
 q

3. Local attend

Same vector cross-attn
with relative RX-anchor
 Δx



4. Task head

Map RX latent $\rightarrow$
scalar PG
or directional PAS

Method — Task Heads: PG & PAS

PG and PAS share the encoder and local readout; only the final head differs.

Path Gain (PG)

- ▶ **Query:** arbitrary 3D RX location q
- ▶ **Head:** MLP $\rightarrow$ scalar $\hat{y}_{pg}(q)$ in dB
- ▶ **Use case:** dense spatial coverage fields
- ▶ **Pretraining:** encoder + PG head with MSE on normalized PG
- ▶ **Output:** full 16^3 volumes at inference
- ▶ **Role:** teaches the shared propagation prior

Power Angular Spectrum (PAS)

- ▶ **Query:** RX q + direction (ϕ, θ)
- ▶ **Direction:** unit vector u + Fourier features $\xi(u)$
- ▶ **Head:** fuse $\xi(u)$ with RX latent $\rightarrow \hat{y}_{pas}$
- ▶ **Transfer:** freeze PG encoder; train PAS head only
- ▶ **Spectrum:** 360×90 az-el from 4×4 array labels
- ▶ **Role:** reuse spatial prior for directional tasks

Method — Residual Adaptation & Training

Target-scene residual adaptation

Zero-shot already transfers, but a held-out scene may have material mismatch, clutter, or sim–reality gaps.

$$\hat{y} = \hat{y}_o + r(g)$$

- ▶ Freeze pretrained encoder + task head
- ▶ r is a tiny MLP on TX–RX geometry g
- ▶ PG: scalar dB correction; PAS: angular residual map
- ▶ Final layer of r is zero-initialized (starts as identity)
- ▶ Only residual params update → cheap, safe adaptation

Staged training

S1

Pretrain encoder + PG head

MSE on normalized PG across many scenes

S2

Attach PAS head

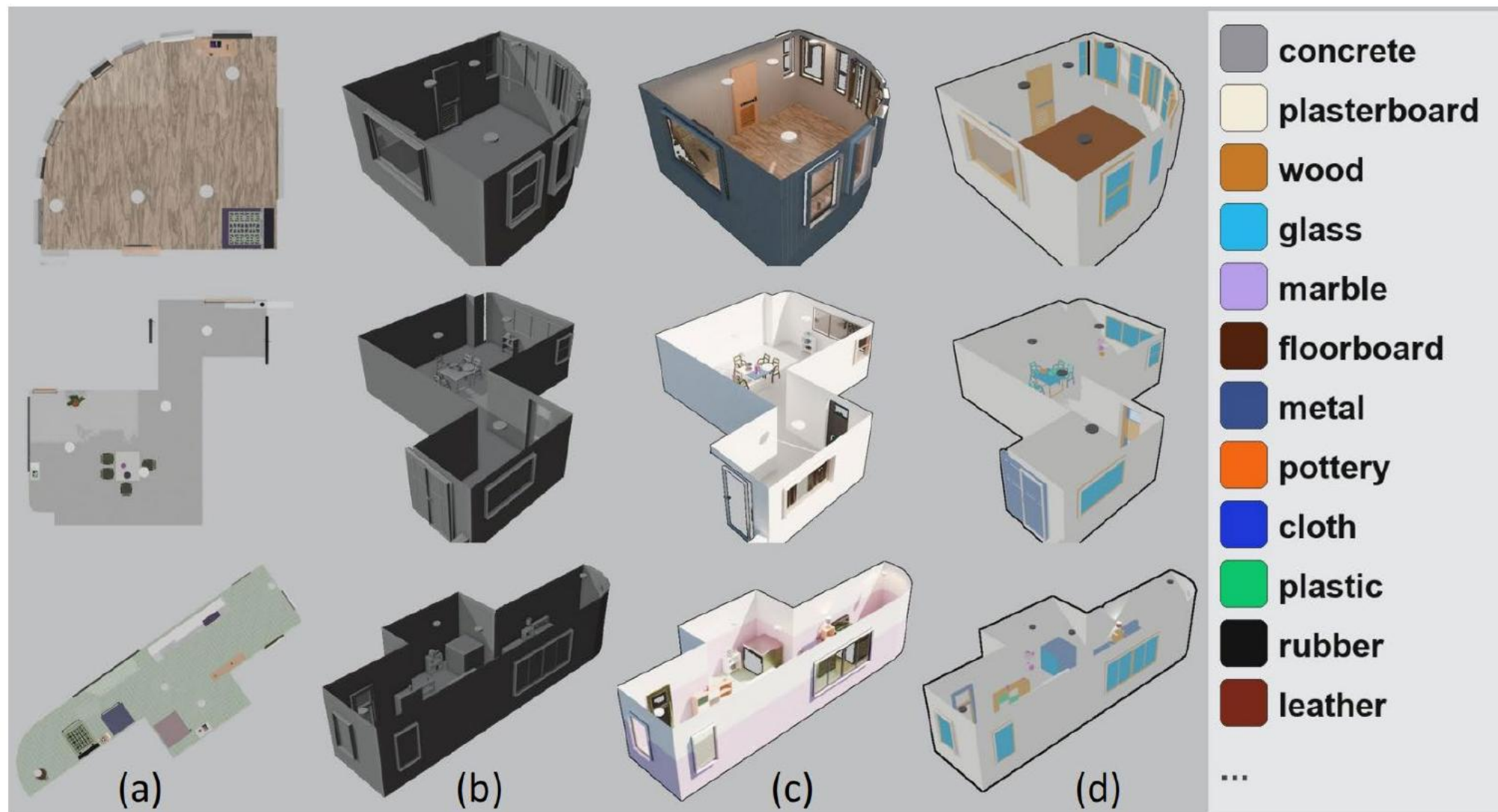
Freeze encoder; train directional head only

S3

Optional residual FT

Freeze encoder + head; fit r on target labels

PRISM Dataset



PRISM Dataset

Procedural Radio–Image Scene Multimodality — Infinigen geometry, Sionna RT labels.

337

Indoor scenes

scene-disjoint split

86,272

TX-conditioned PG fields

3.5 GHz mid-band

15

ITU material classes

ϵ , σ from ITU-R P.2040

269/33/35

Train / Val / Test

held-out rooms

PRISM-PG

Dense 16^3 PG volumes per TX; surface point clouds as model inputs; meshes only for offline labeling.

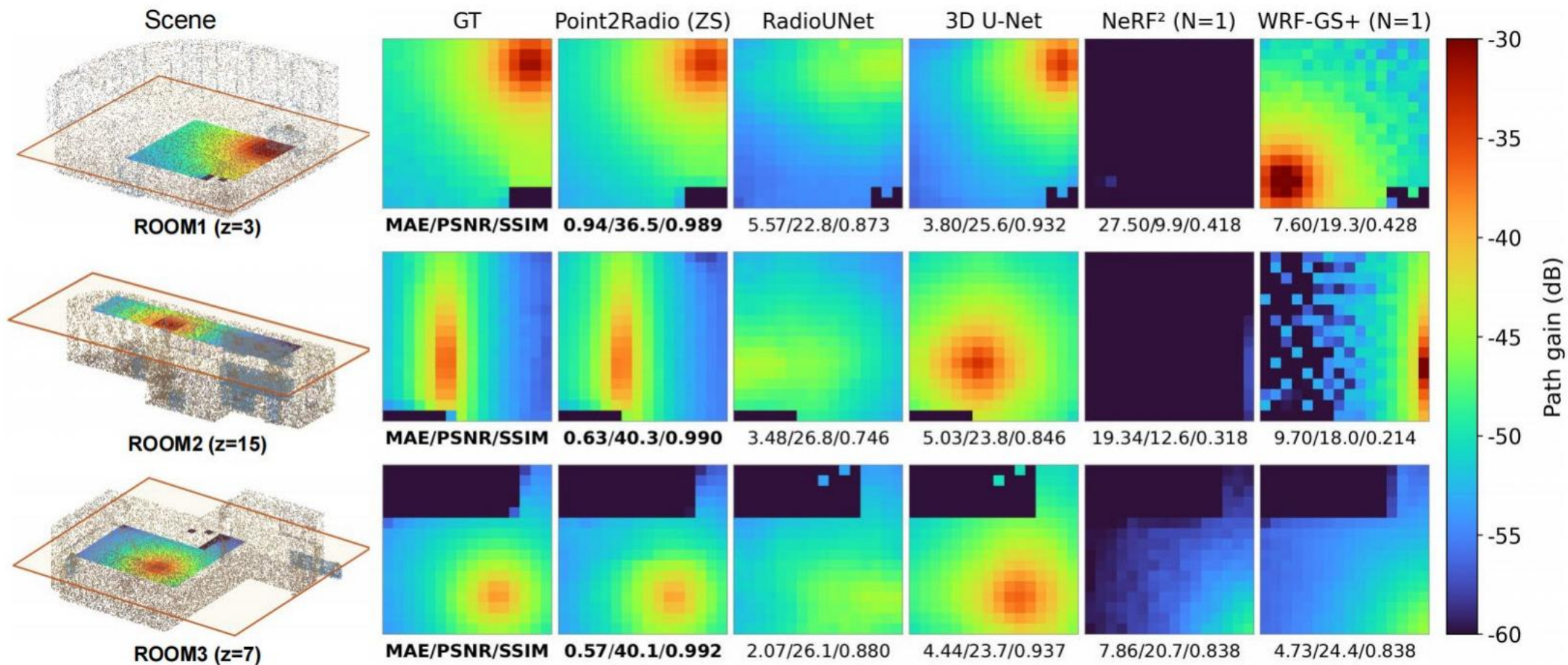
PRISM-PAS

20×20 TX–RX pairs; 4×4 array beamforming → 360×90 az–el spectra; 199 overlapping rooms.

Inference inputs

Material-aware point cloud + transceiver queries only — no mesh, no online path tracing.

Results: Cross-Scene Path Gain



Results: Path Gain

Zero-shot full-volume PG on 35 held-out test scenes.

Method	MAE ↓	NMSE	PSNR ↑	SSIM ↑
RadioUNet (2D)	3.745	0.0159	25.10	0.836
3D U-Net	3.038	0.0152	26.71	0.884
Point2Radio (ZS)	0.871	0.0025	34.83	0.954

76.7%

MAE reduction vs. RadioUNet

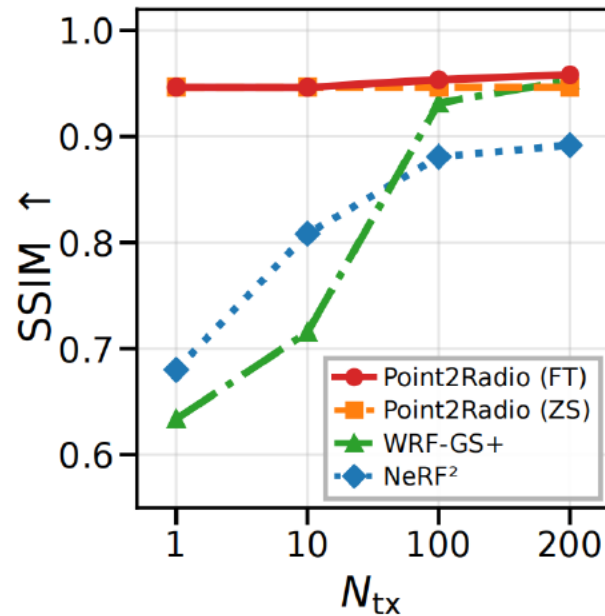
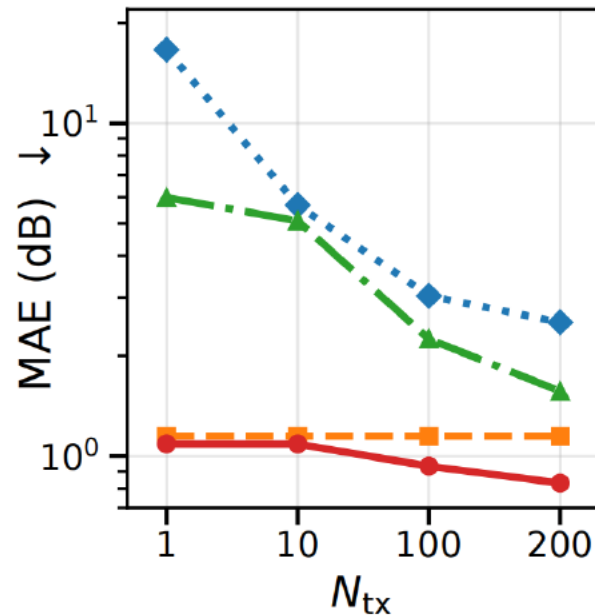
0.871 dB

Zero-shot MAE on held-out rooms

0.954

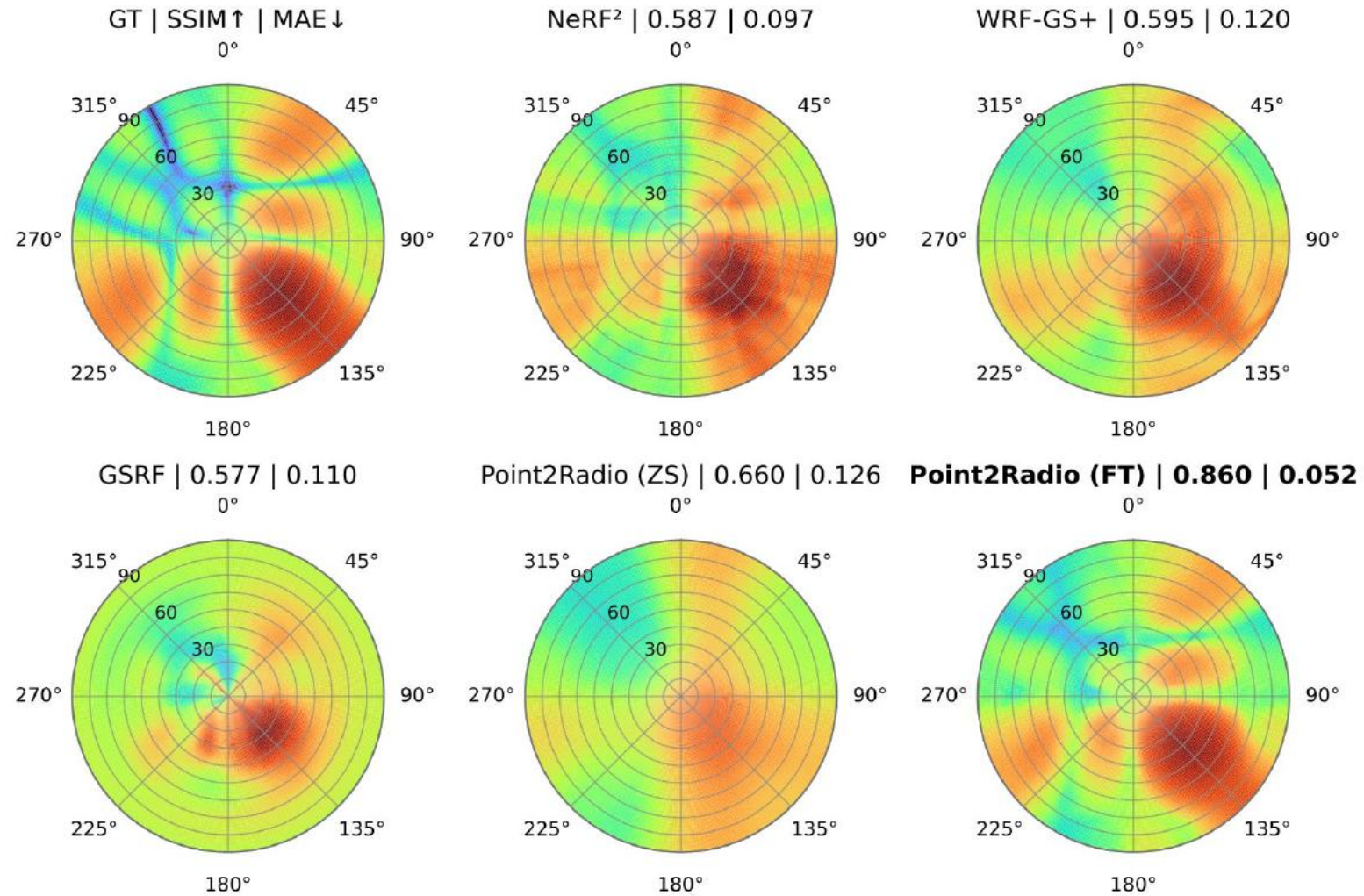
SSIM — strong spatial structure

Results: Path Gain



Variant	MAE (dB)↓	SSIM↑
Full Model	0.871	0.954
w/o ITU materials	1.259	0.943
w/o normals	1.456	0.933
w/o materials & normals	1.685	0.920

Results: PAS



Results: PAS

PAS on a held-out indoor room

Freeze PG-pretrained encoder; train PAS head. Optional residual FT on target TXs.

NeRF2	SSIM 0.678
-------	------------

WRF-GS+	SSIM 0.715
---------	------------

Point2Radio (ZS)	SSIM 0.690
------------------	------------

Point2Radio (FT)	SSIM 0.798
------------------	------------

Latency on RTX 3090

PG volume (16^3)

163 ms vs. 990 ms Sionna · ~6.1× faster

PAS (1 pair)

156 ms vs. 648 ms Sionna · ~4.2× faster

Encode once / TX

~151 ms · Decode: 10 ms / 3 ms

Peak memory (PG)

3.3 GB · vs. 8.3 GB Sionna

Conclusion and Future Work

□ Point2Radio learns a transferable TX-conditioned propagation prior from material-aware point clouds. reusable across scenes and radio tasks, without meshes or online ray tracing.

□ Future Directions:

- Input modality.
- Training and adaptation data.
- ...

GLocFM: A Geometry-Aware Foundation Model for 3D Indoor Wireless Localization

From scene geometry and multi-receiver
WiFi to a single-pass 3D posterior

Anonymous Submission



香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY



iComAI Lab
INTEGRATED COMMUNICATIONS AND AI LAB

Related Work

Radio-only

SpotFi SIGCOMM '15

iARK MobiSys '20

DLoc MobiCom '20

LocGPT MobiSys '24

Learns radio-to-location mapping; geometry remains implicit.

Geometry-aware

MAP-CSI GLOBECOM '21

RayLoc arXiv '25

Wi-GATr ICLR '25

SigMap ICLR '26

Adds spatial priors, often with calibrated maps or iterative inference.

GLocFM

Explicit 3D geometry

Cross-scene training

Candidate likelihood

One forward pass

Geometry-conditioned scoring for unseen indoor scenes.

Venues are taken from the paper's current references.bib.

GLocFM-221: diverse geometry paired with wireless channels

Representative occupancy floor plans from the 221-scene benchmark



1. **Geometry** Infinigen Indoor meshes

2. **Wireless** Sionna RT ray-traced signals

Layout diversity offices, bedrooms, bathrooms, and kitchens
floor area: 3.25 m² to 63.51 m²

221 indoor scenes
independent furnished layouts

88,400 TX–RX links

80 MHz 1024 subcarriers

4 × 4 UPA per receiver

177	22	22
train	validation	test

Infinigen Indoors: Raistrick et al. (2024). Wireless simulation: Sionna RT, Hoydis et al. (2023).

From continuous maximum likelihood to learned candidate scoring

CONTINUOUS MAXIMUM LIKELIHOOD

$$\hat{\mathbf{x}}_{\text{ML}} = \arg \max_{\mathbf{x} \in \Omega_{\text{free}}} \log p(\{\mathbf{Y}_r\}_{r=1}^{N_R} \mid \mathbf{x}, \mathcal{S})$$

CANDIDATE POSTERIOR

$$P_m = \frac{\exp S_m}{\sum_{m'=1}^M \exp S_{m'}}$$

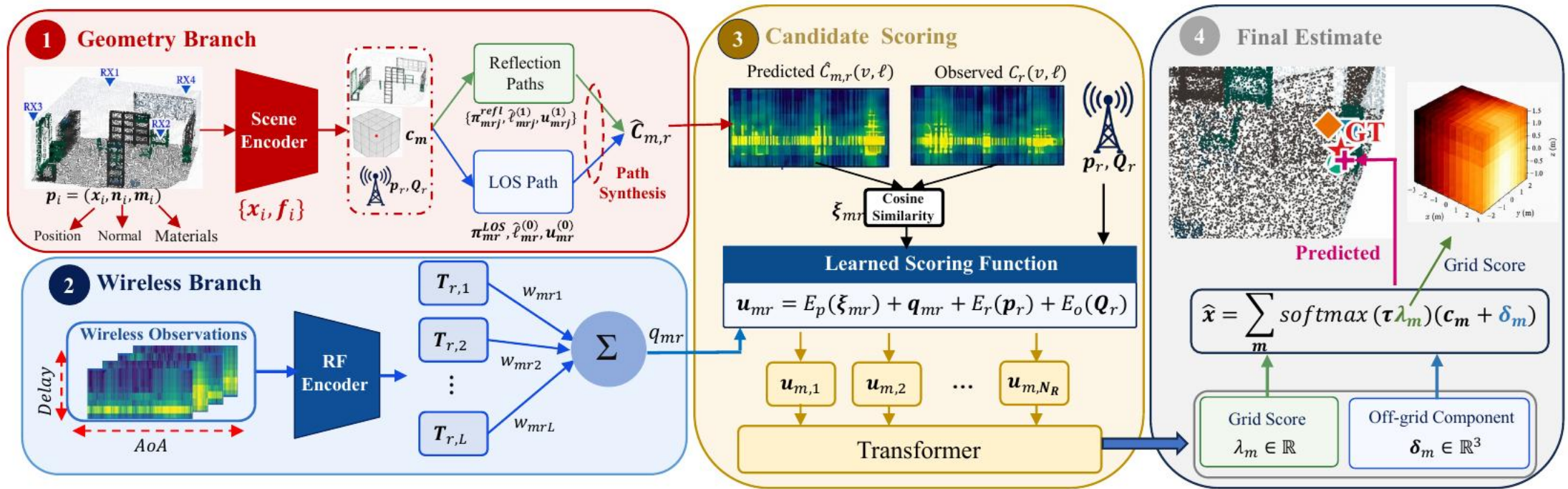
S_m : learned scoring function for candidate $\mathbf{c}_m$

Ω_{free} is the admissible indoor volume not occupied by object meshes.

Geometry defines the hypotheses; the network learns how strongly each hypothesis explains the RF evidence.

A finite candidate set makes the search tractable; the final position remains continuous.

One model, two encoders, one candidate posterior



1 GEOMETRY 2 PATH SYNTHESIS 3 CANDIDATE SCORING 4 FINAL ESTIMATE

Geometry generates candidate-specific path hypotheses

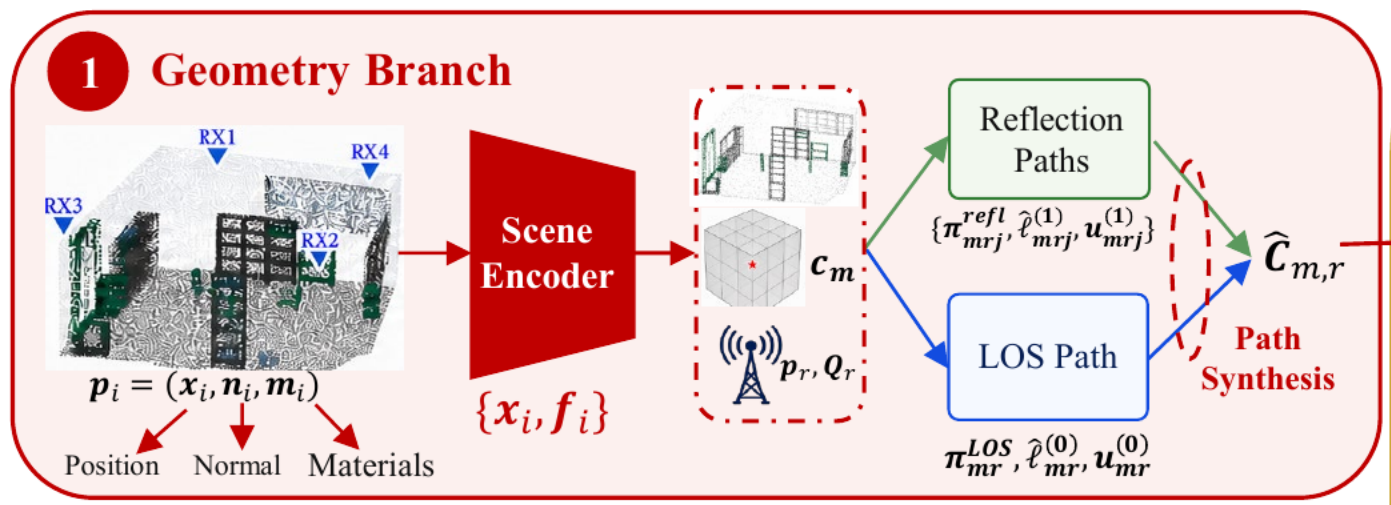
1. Discretize free space into an $N_x \times N_y \times N_z$ grid:

$$N_x = N_y = N, \quad N_z = 4,$$

$$M = N^2 N_z.$$

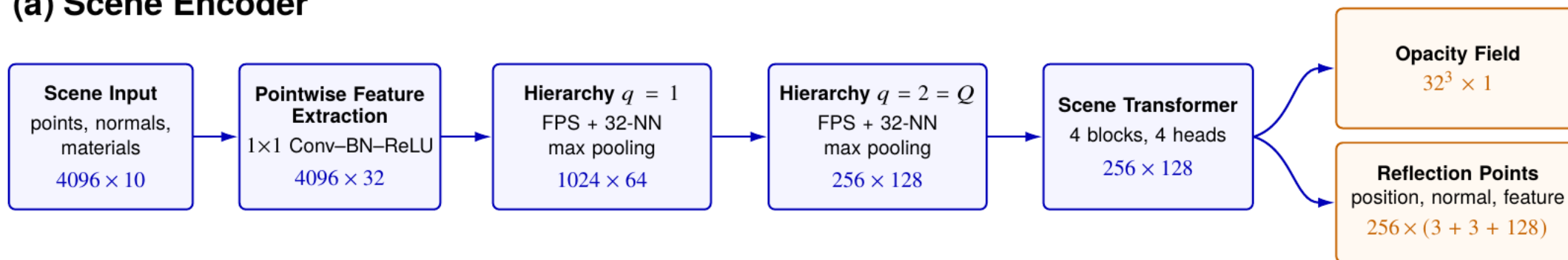
2. For each candidate–RX pair $(\mathbf{c}_m, \mathbf{p}_r)$, generate:
 - LoS path
 - one-bounce alternatives
3. Predict delay, receiver-local AoA, and attenuation priors.

Reflections turn NLoS energy into a geometric constraint.



The scene encoder predicts soft propagation structure

(a) Scene Encoder



Input

4096×10 point features:
position, normal, and EM
attributes.

Hierarchy

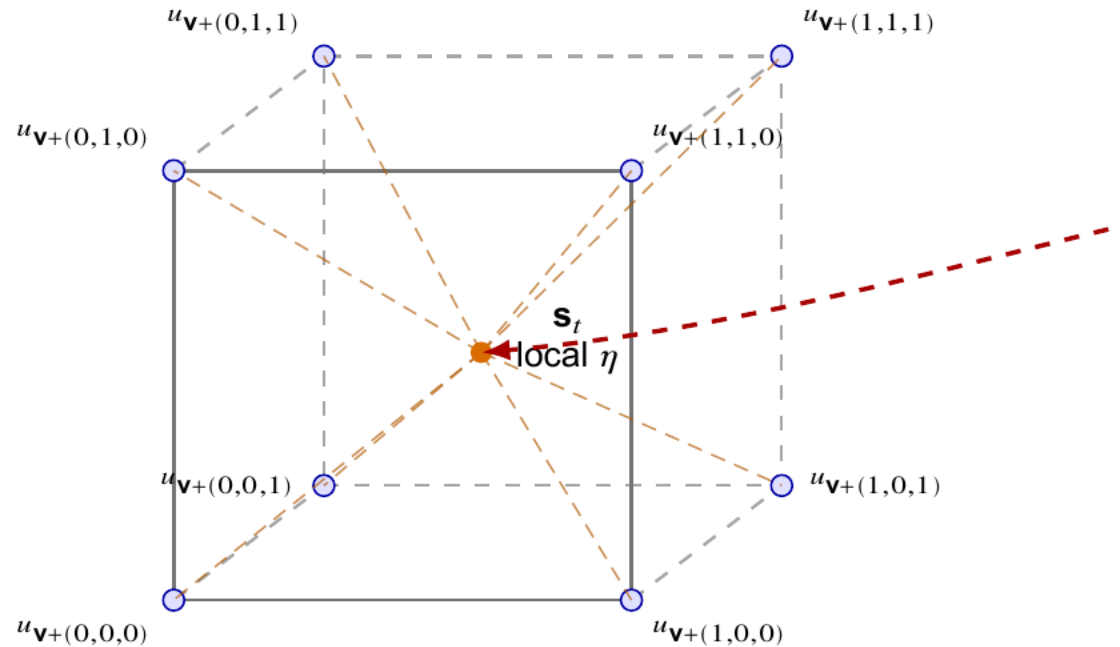
$4096 \rightarrow 1024 \rightarrow 256$ points,
then four Transformer
blocks.

Outputs

A 32^3 opacity field and 256
reflection points with
features.

Visibility field: query each point by trilinear interpolation

(a) Eight-corner trilinear interpolation



The query point $\mathbf{s}_t$ interpolates the eight opacity-grid vertices of its containing voxel using separable trilinear weights.

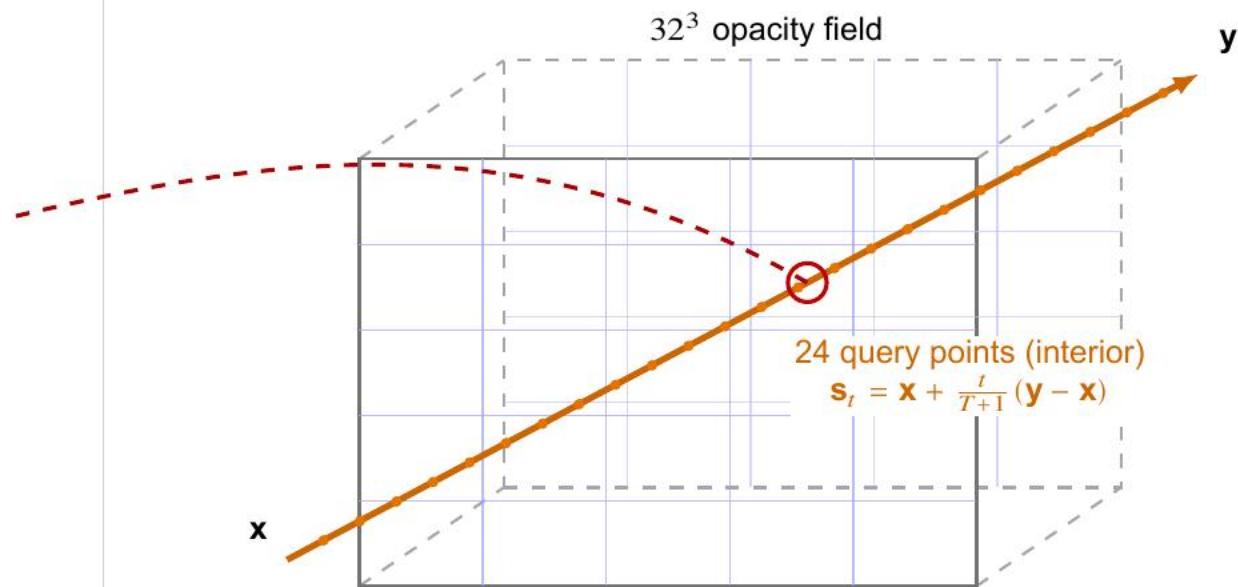
Why it matters

The learned opacity field remains continuous and differentiable between grid vertices.

$$\omega_{\delta}(\eta) = \prod_{r \in \{x,y,z\}} \eta_r^{\delta_r} (1 - \eta_r)^{1-\delta_r}$$
$$\rho(\mathbf{s}_t) = \sum_{\delta \in \{0,1\}^3} \omega_{\delta}(\eta) u_{\mathbf{v}+\eta}$$

Visibility field: integrate occupancy along each path segment

(b) Segment querying in the opacity field



$$\log V(\mathbf{x}, \mathbf{y}) = -\text{softplus}(\gamma) \frac{\|\mathbf{y} - \mathbf{x}\|_2}{T} \sum_{t=1}^T \rho_t, T = 24$$

The supplement evaluates the learned 32³ field at exactly $T = 24$ interior samples along every candidate–receiver or reflection segment.

$$\log V(\mathbf{x}, \mathbf{y}) = -\text{softplus}(\gamma) \frac{\|\mathbf{y} - \mathbf{x}\|_2}{T} \sum_{t=1}^T \rho(\mathbf{s}_t)$$

LoS and one-bounce paths synthesize a candidate spectrum

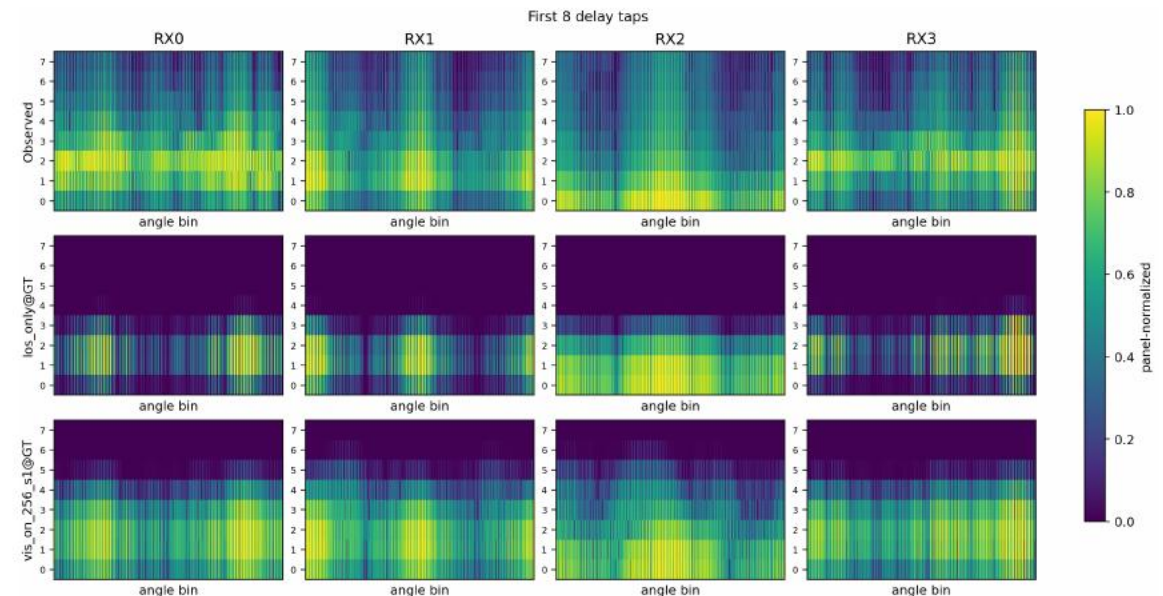
Geometry-predicted complex spectrum

$$\hat{h}_{m,r,a,\ell} = \hat{h}_{m,r,a,\ell}^{\text{LoS}} + \sum_j \hat{h}_{m,r,a,\ell}^{\text{refl}}$$

Direct and reflected components are summed in the complex delay–array domain before forming the predicted delay–AoA spectrum. **Spectrum match**

$$s_{mr} = \frac{\langle \mathbf{C}_r, \hat{\mathbf{C}}_{m,r} \rangle_F}{\|\mathbf{C}_r\|_F \|\hat{\mathbf{C}}_{m,r}\|_F + \epsilon}$$

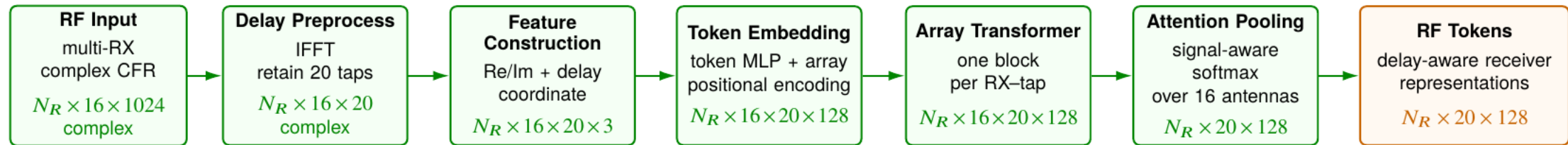
Cosine similarity in delay–AoA space.



Supplement audit: LoS+reflection improves cosine similarity from 0.6974 to 0.8697.

The RF encoder preserves both delay and array structure

(b) RF Encoder



Input

Multi-receiver complex CFR: $N_R \times 16 \times 1024$.

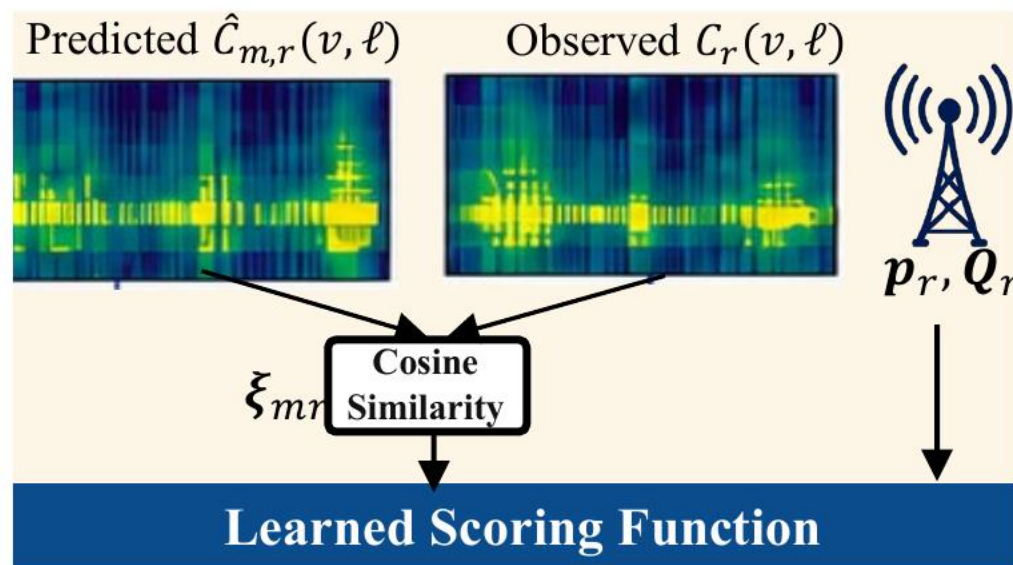
Processing

IFFT, 20 retained taps, token embedding, and array Transformer.

Output

Delay-aware RF tokens: $N_R \times 20 \times 128$.

A learned scorer calibrates the physics-based match



Candidate–receiver evidence

$\xi_{mr} = \text{MLP}([s_{mr}, P_r])$ combines spectrum agreement and receiver strength.

Candidate-specific RF readout

$\mathbf{q}_{mr} = \sum_{\ell} w_{mr\ell} \mathbf{T}_{r,\ell}$ reads a geometry-conditioned combination of delay tokens.

Cross-receiver fusion

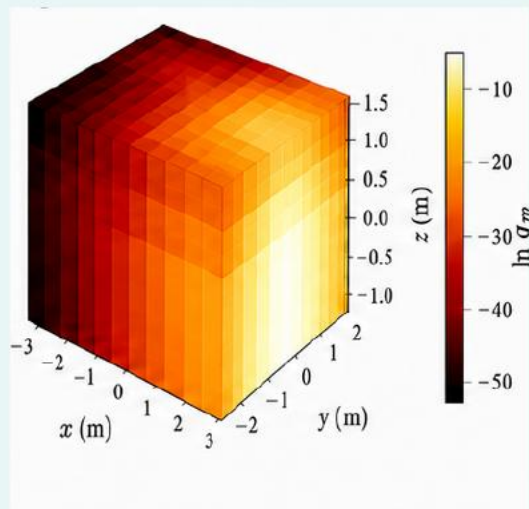
Self-attention fuses all receiver tokens into candidate feature $\mathbf{f}_m$.

Two output heads

Grid logit λ_m and bounded off-grid residual δ_m .

Posterior weighting yields a continuous 3D estimate

Grid score



Each cell receives
a learned logit λ_m .



Posterior-weighted estimation

$$\hat{\mathbf{x}} = \sum_{m=1}^M \text{softmax}_m(\tau \lambda_m) (\mathbf{c}_m + \boldsymbol{\delta}_m)$$

$\mathbf{c}_m$: grid candidate

λ_m : learned grid score

$\boldsymbol{\delta}_m$: bounded off-grid offset

**Soft posterior + off-grid refinement
avoids hard grid quantization.**

Timing robustness is learned by exposure, not test-time search

Synchronized model

$$\mathcal{L}_{\text{train}} = \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2$$

Assumes absolute delay is aligned.



ToF-robust model

$$\mathcal{L}_{\text{robust}} = \mathbb{E}_{b \sim \mathcal{N}(0, \sigma_b^2)} \|\hat{\mathbf{x}}_i(b) - \mathbf{x}_i\|_2^2$$

One common transmitter-side offset per sample.

CFR phase ramp



Delay shift



Augmented training



Single-pass inference

Simulation configuration

Component	Configuration
Scene Transformer	4 blocks, $d = 128$, 4 heads
Array Transformer	1 block, $d = 128$, 4 heads
Candidate–RX Transformer	4 blocks, $d = 128$, 4 heads
Dropout	0.1
Visibility / reflection anchors	512 / 64
Segment samples	24
Candidate grid	$12 \times 12 \times 4$
RF input	4 RX / 16 antennas / 20 taps

Training

Adam, 10^{-4} learning rate, cosine schedule, 200 epochs, 8 NVIDIA A6000 GPUs.

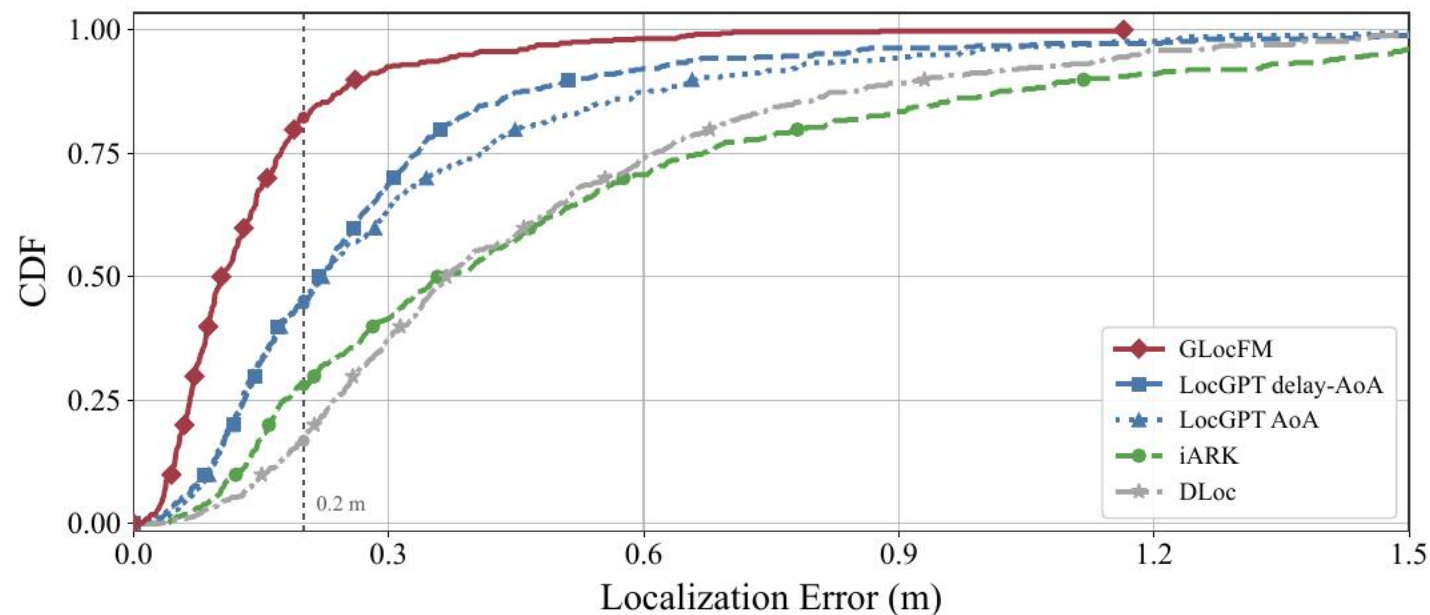
Channel

3.5 GHz carrier, 80 MHz bandwidth, 1024 OFDM subcarriers.

Inference

Scene and RF encodings are cached; all candidates are scored in one feed-forward pass.

GLocFM roughly halves synthetic 3D localization error



The gain is distribution-wide, not driven by a few easy samples.

0.1429 m

mean error, 4 RX

vs. 0.2827 m LocGPT

delay-AoA

82.2%

$e \leq 0.2$ m, 4 RX

vs. 45.2% strongest

neural baseline

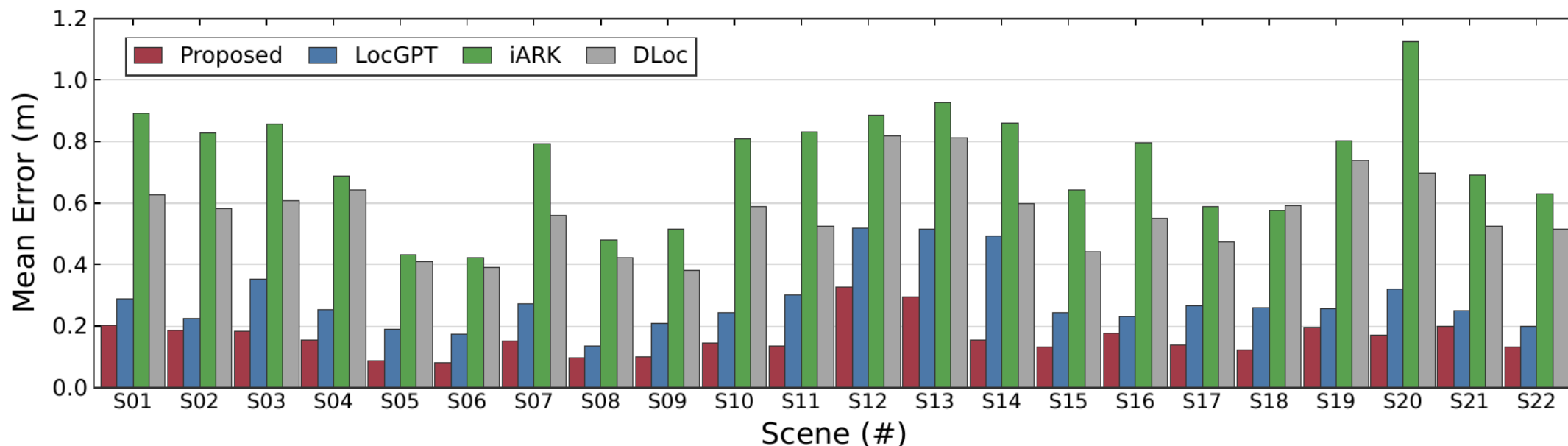
0.1746 m

mean error, 3 RX

reflections stabilize fewer

receivers

The improvement persists across all 22 held-out scenes



22/22

test scenes

GLocFM is lowest in
every scene

Geometry-conditioned scoring is especially useful when radio-only methods encounter scene-specific blockage and multipath. The advantage remains visible across small and large rooms rather than collapsing to an average-case win.

The representation adapts to measured NeRF² data

Measured setting

Fixed 4×4 UPA receiver; 6,123 angular spectra;
no subcarrier-resolved delay information.

Method	Train share	Mean (m)	F_e (0.1)
iARK	100%	0.2849	6.9%
LocGPT	100%	0.1244	63.5%
GLocFM	100%	0.0637	86.6%
GLocFM	50%	0.0854	78.8%
GLocFM	25%	0.1304	67.3%

48.8%

mean-error reduction

relative to LocGPT

25%

training data still

usable

0.1304 m mean error

The geometry-aware prior remains useful
after removing delay.

Scene-specific AoA adaptation; this is not claimed as zero-shot cross-scene

transfer.

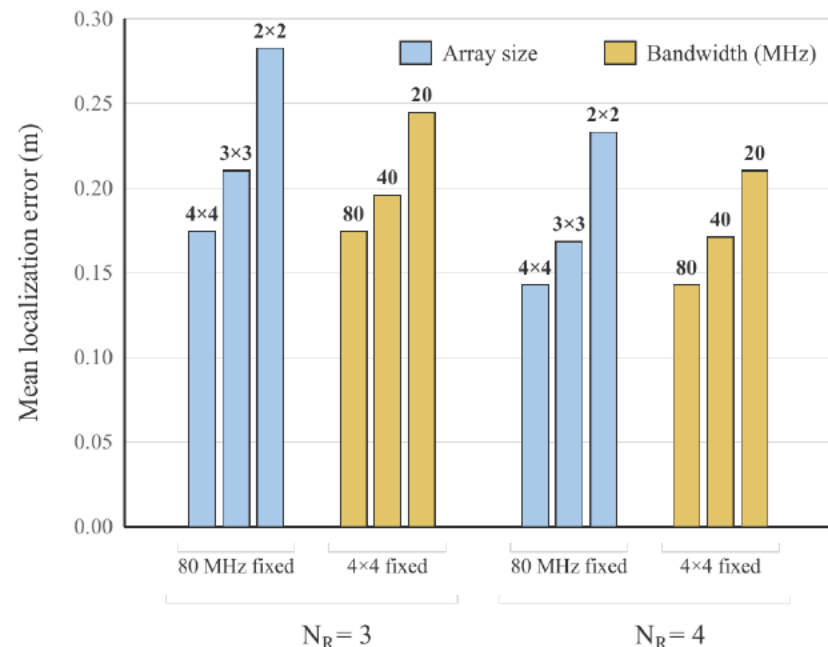
Measured dataset: NeRF² (Zhou et al. 2023).

Ablations identify where the accuracy comes from

Variant	Mean (m)	Δ
Reference ($12 \times 12 \times 4$)	0.1429	—
Coarse grid ($3 \times 3 \times 4$)	0.4494	+0.3065
Fine grid ($16 \times 16 \times 4$)	0.1360	-0.0069
Cosine only	0.3914	+0.2485
LoS only	0.2285	+0.0856

Interpretation

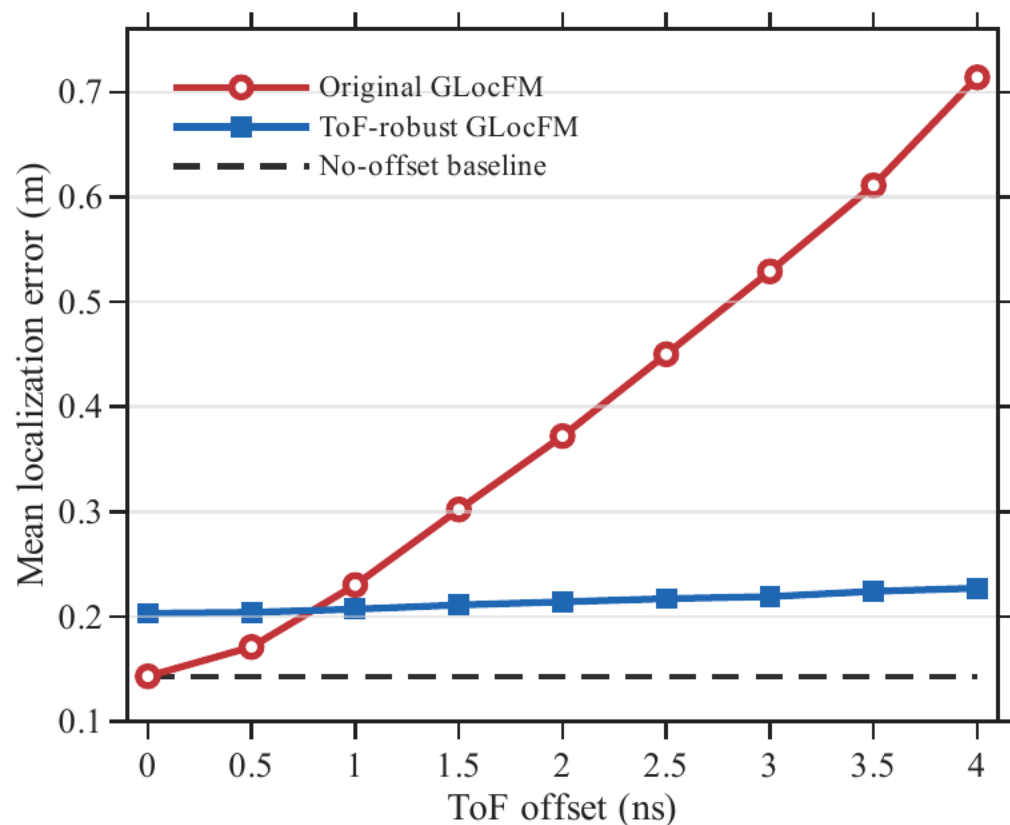
Learned scoring and reflected paths deliver most of the gain.



Fewer

antennas weaken angular resolution; lower bandwidth weakens range resolution. One-epoch adaptation remains effective.

Bias augmentation prevents ToF-offset failure



Original / mismatch

0.7141 m at 4 ns

ToF-robust

0.2270 m at 4 ns

Robustness trades a small loss at perfect synchronization for a nearly flat error curve across 0–4 ns.

Backup: optical materials are mapped to EM attributes

Mesh token	Radio label	Index
concrete	concrete	0
brick	brick	1
plaster	plasterboard	2
wood / shelves	wood	3
glass	glass	4
metal	metal	5
tile	pottery	6

Point material feature

$$\mathbf{m} = [\varepsilon_r, \sigma, S, K_x]$$

Relative permittivity, conductivity, scattering coefficient, and cross-polarization coefficient are stored per point.

Optical color and texture are not passed to the radio simulator. Unmatched tokens default to plasterboard.

Current evidence is strong—but deployment gaps remain

Simulation gap

No diffraction or refraction

Perfect geometry registration

Static transmitter and scene

Measurement scope

One fixed-receiver scene

AoA-only adaptation

Antenna mismatch remains

Next steps

Joint pose calibration

More measured environments

Geometry uncertainty

Privacy-aware deployment

Backup: selected references

Wireless localization

Kotaru et al. SpotFi. SIGCOMM 2015.

Ayyalasomayajula et al. DLoc. CoNEXT 2020.

An et al. iARK. IPSN 2020.

Zhao et al. LocGPT. 2024.

Geometry-aware localization

Vuckovic et al. MAP-CSI. 2021.

Han et al. RayLoc. 2025.

Hehn et al. Wi-GATr. ICLR 2025.

Chu et al. SigMap. ICLR 2026.

Data and simulation

Raistrick et al. Infinigen Indoors. CVPR 2024.

Hoydis et al. Sionna. 2023.

Zhao et al. NeRF². 2023.

Geometric representation

Peng et al. Convolutional Occupancy Networks. ECCV 2020.

Full bibliography

See references.bib in the paper source.



Part IV

Spatial-Intelligent Enabled Applications



Intelligent Agents

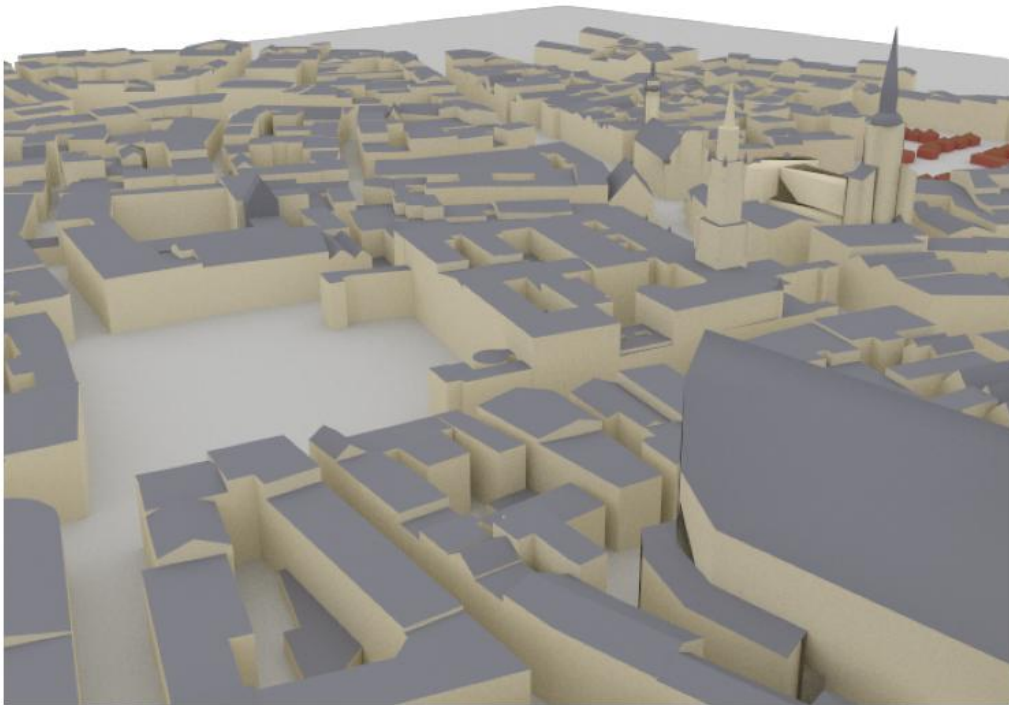


Beamforming

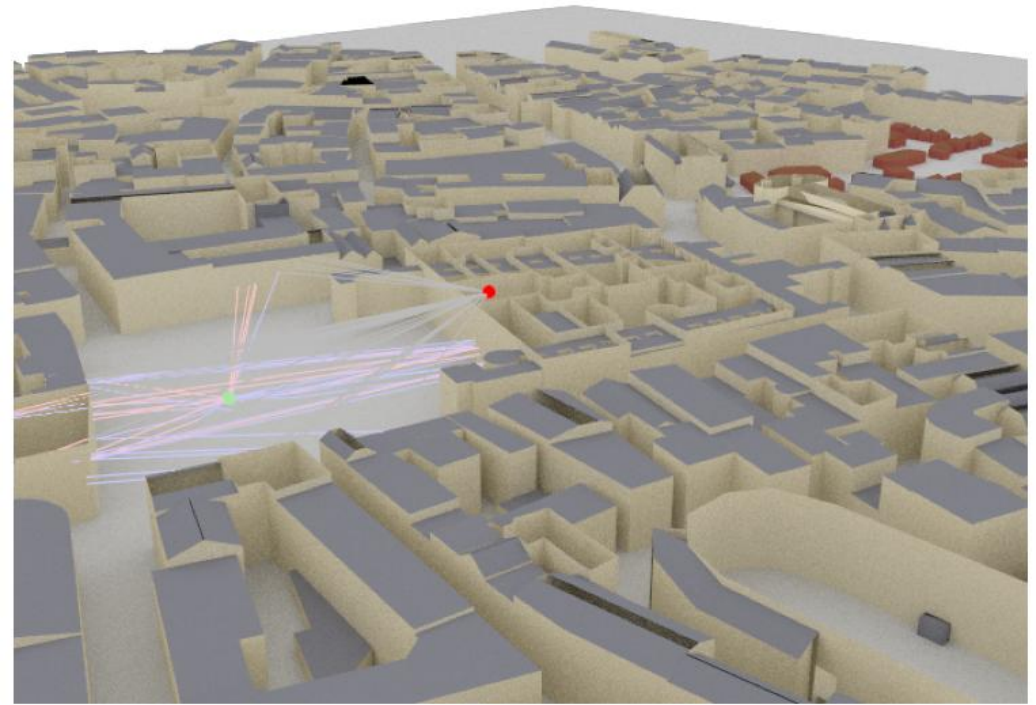


AP Deployment and Low-altitude Spatial Intelligence

Ray tracing connects a 3D environment to its wireless channel



3D scene · geometry and radio materials

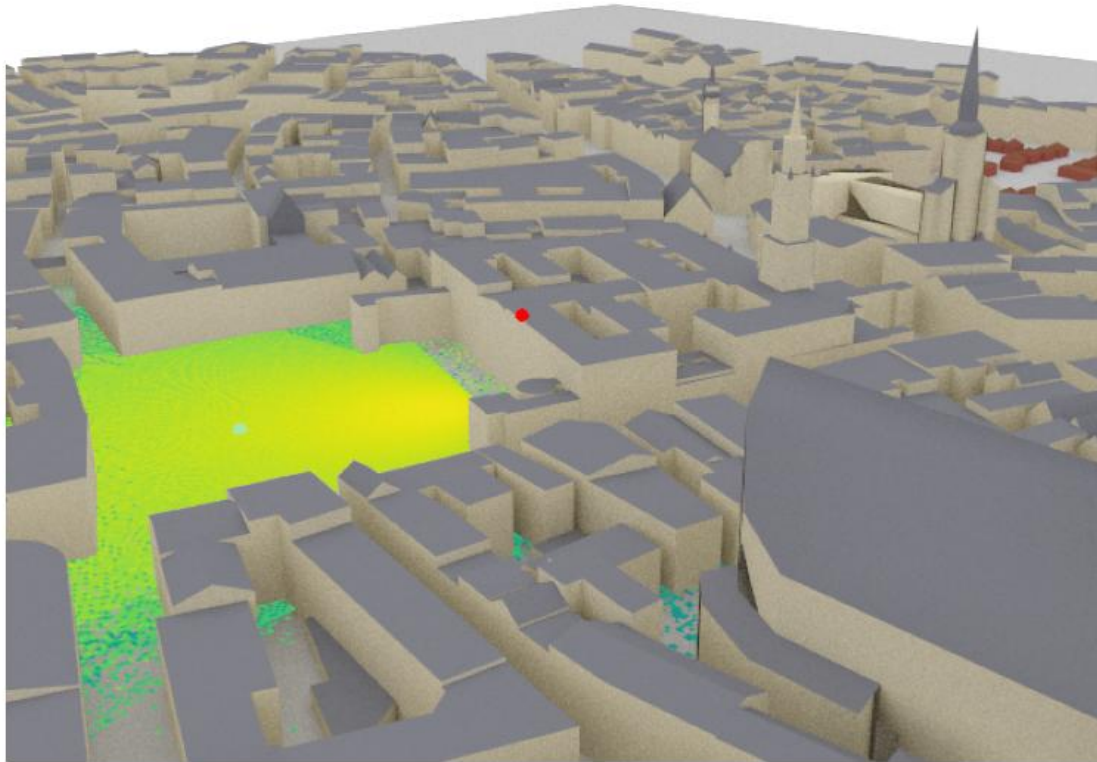


Computed paths · gain, delay, angles and LoS/NLoS

Scene + devices + radio configuration → multipath → CIR / CFR

Images: NVIDIA Sionna RT 2.0.1 tutorial

The environment shapes where radio energy can—and cannot—propagate



Sionna RT tutorial radio map · path-gain visualization

Image: NVIDIA Sionna RT 2.0.1 tutorial ·

OFFICIAL SIONNA RT EXAMPLE

Geometry

Controls visibility, blockage, path length, and multipath topology.

Radio materials

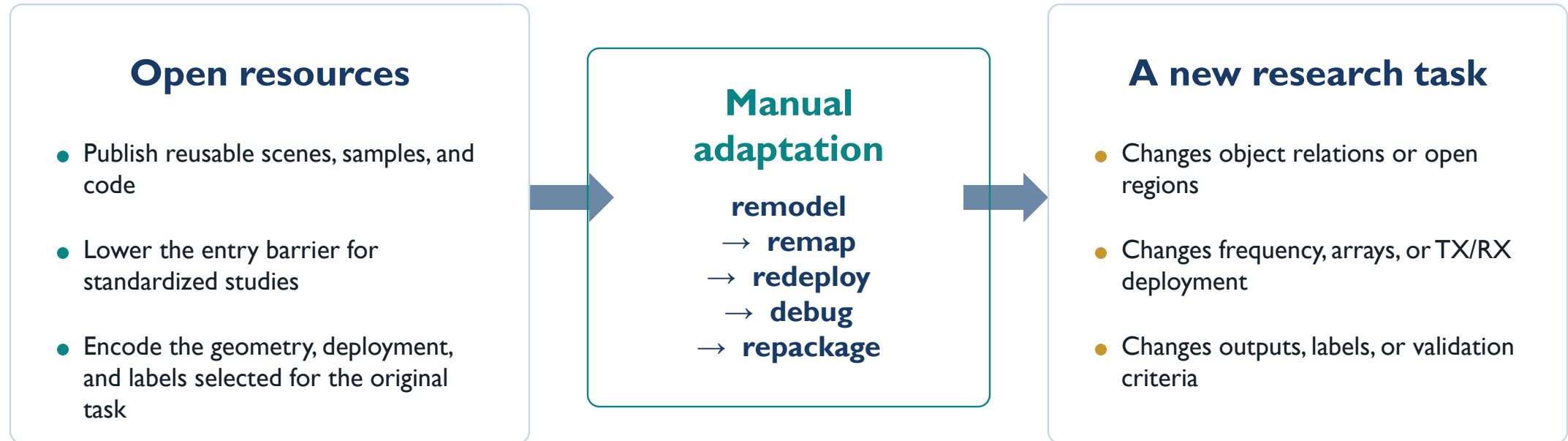
Control reflection, scattering, refraction, and attenuation.

Deployment

Determines which TX/RX links are feasible inside the frozen scene.

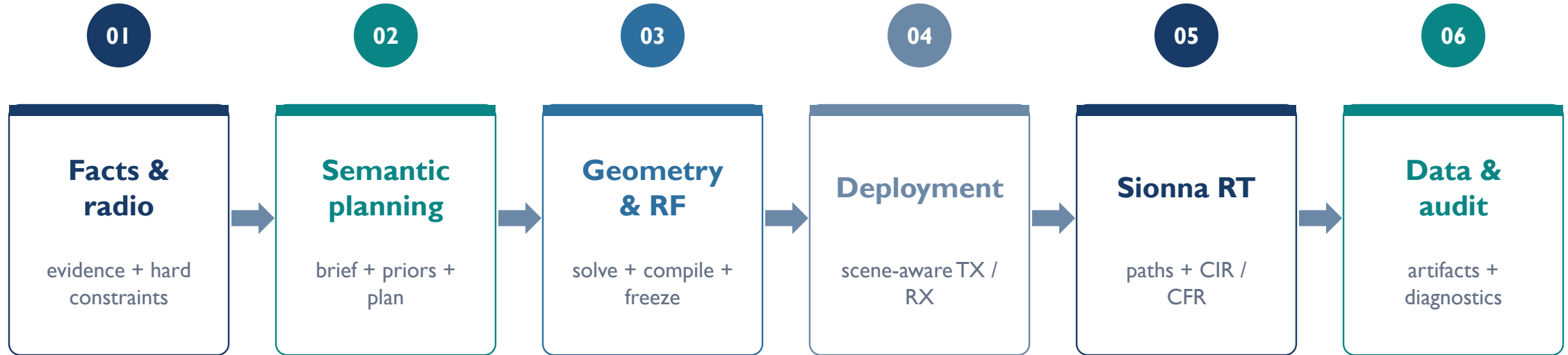
A scene can look plausible—and still fail at runtime.

Open datasets help—but new tasks still require substantial engineering



The missing layer is a compiler from intent to verified wireless channels.

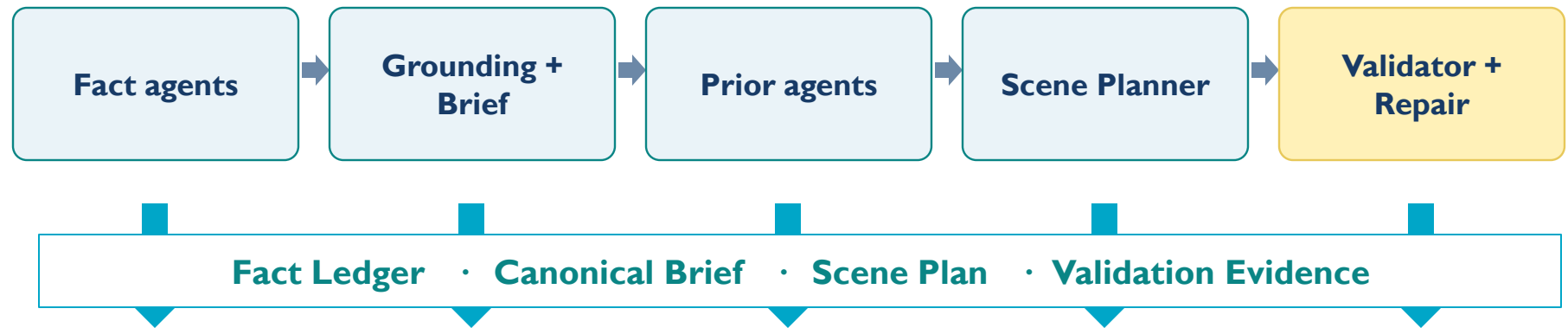
ChannelAgent is an agentic compiler—not a sequence of tools



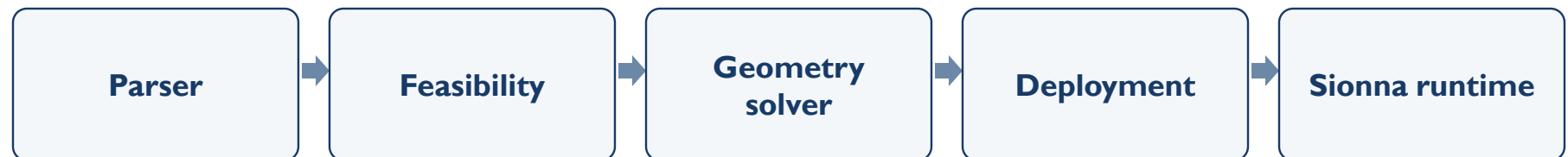
Typed artifacts carry intent through deterministic execution, validation, and runtime proof.

Specialized agents collaborate through explicit artifacts

LLM planning lane



Deterministic lane



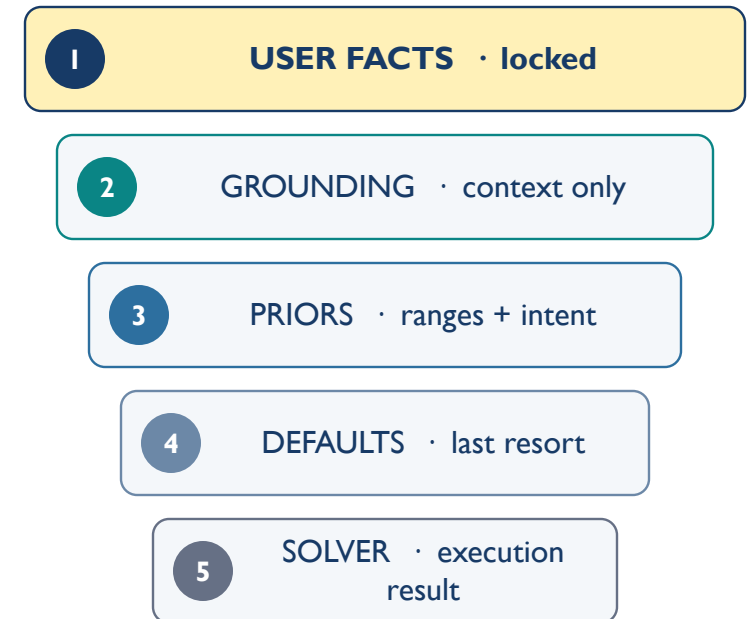
Agents plan and critique. Executors compute, compile, and simulate.

User facts remain immutable and traceable

User Fact Ledger

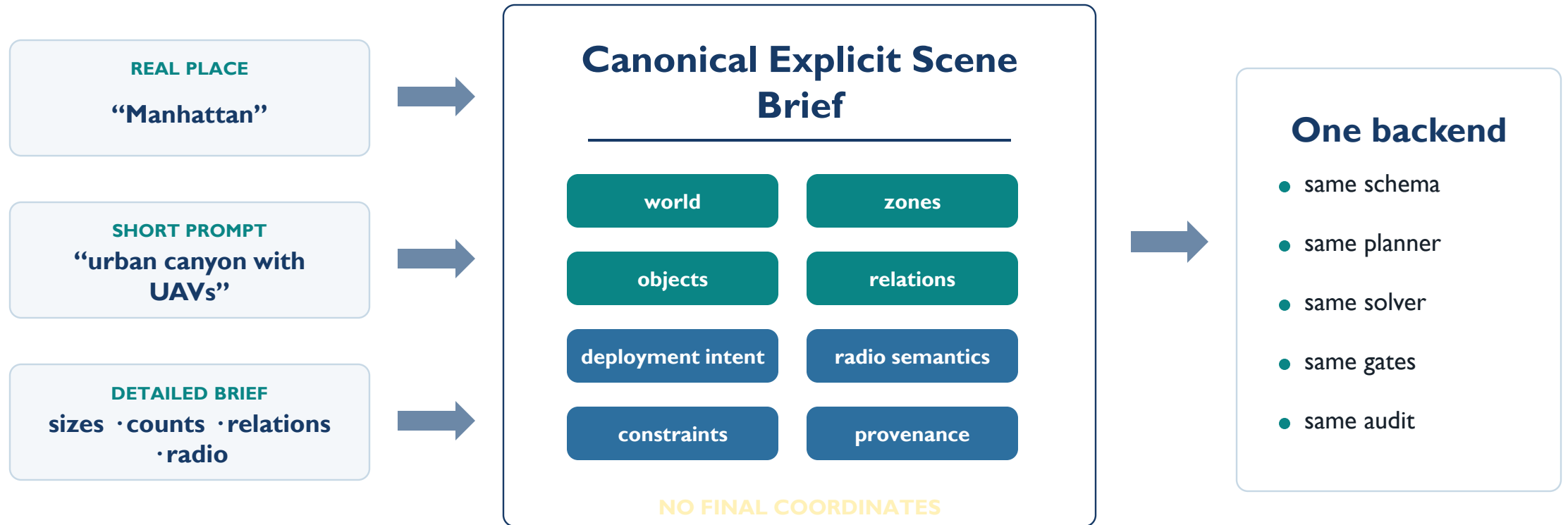
ID	Predicate	Value	Origin	Strength
f1	world.size	500 × 500 m	user	HARD
f2	building.count	20	user	HARD
f3	relation	around open core	user	HARD
p1	clearance.preferred	12–18 m	prior	SOFT
s1	building.centers	solved	solver	RESULT

Precedence is explicit



Grounding may enrich context; it may never overwrite explicit user evidence.

One coordinate-free contract unifies every request



After the contract, downstream modules no longer need the raw prompt or its source mode.

LLM plans the scene; deterministic code solves it

PLANNING RIGHTS

- scene archetype and propagation intent
- morphology and topology relations
- minimum / preferred / maximum ranges
- importance, prohibitions, repair scope

Scene Plan

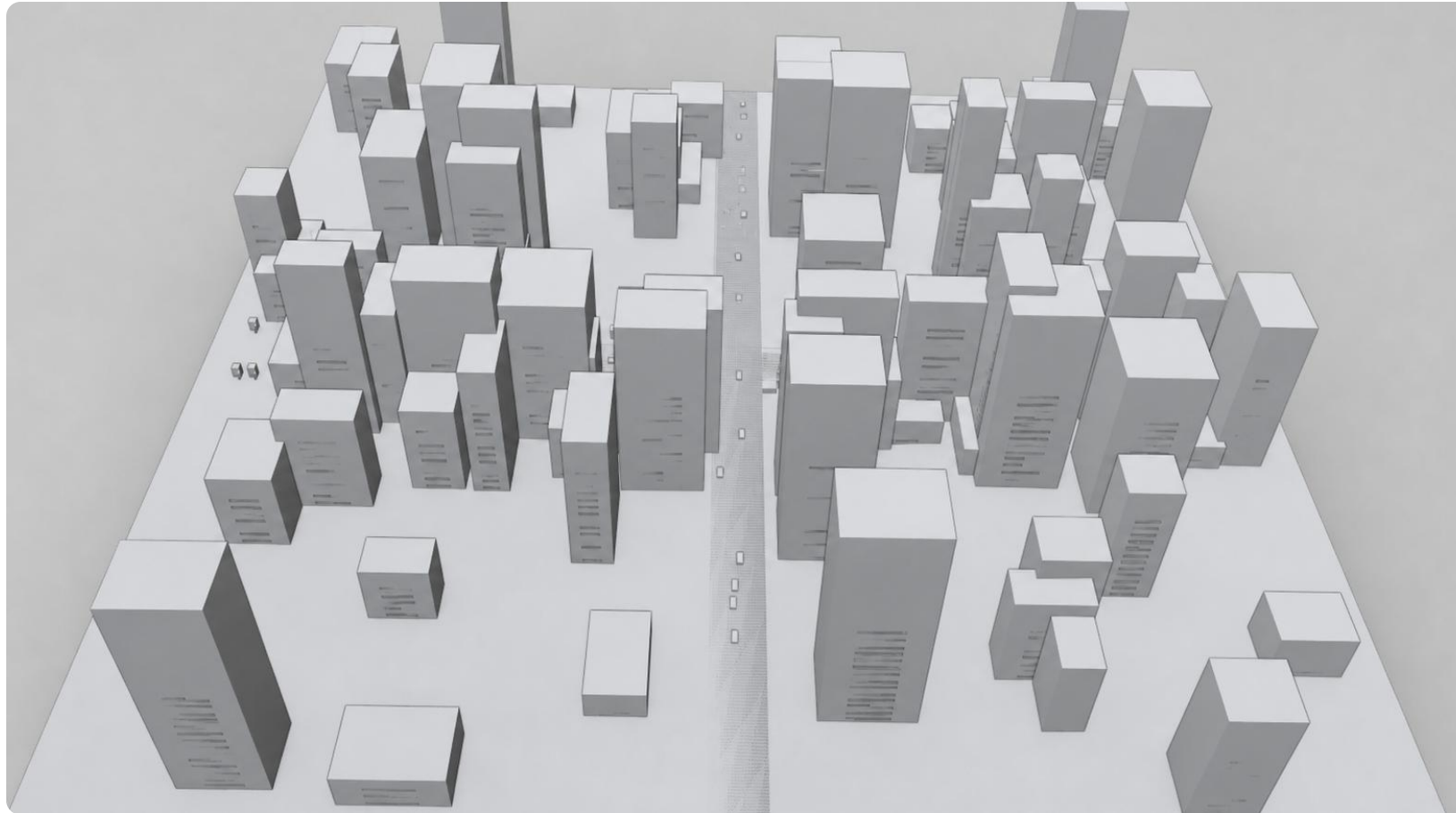


COORDINATE RIGHTS

- region polygons and exact coordinates
- footprints, heights, mesh compilation
- overlap, boundary and clearance checks
- Sionna assets and numerical validation

The solver may reject an infeasible plan; it may not silently redesign it.

Example: “Xinjiekou” becomes a solver-ready urban scene



CASE STUDY

Urban morphology

Dense high-rise blocks create multiple street-canyon conditions.

Spatial structure

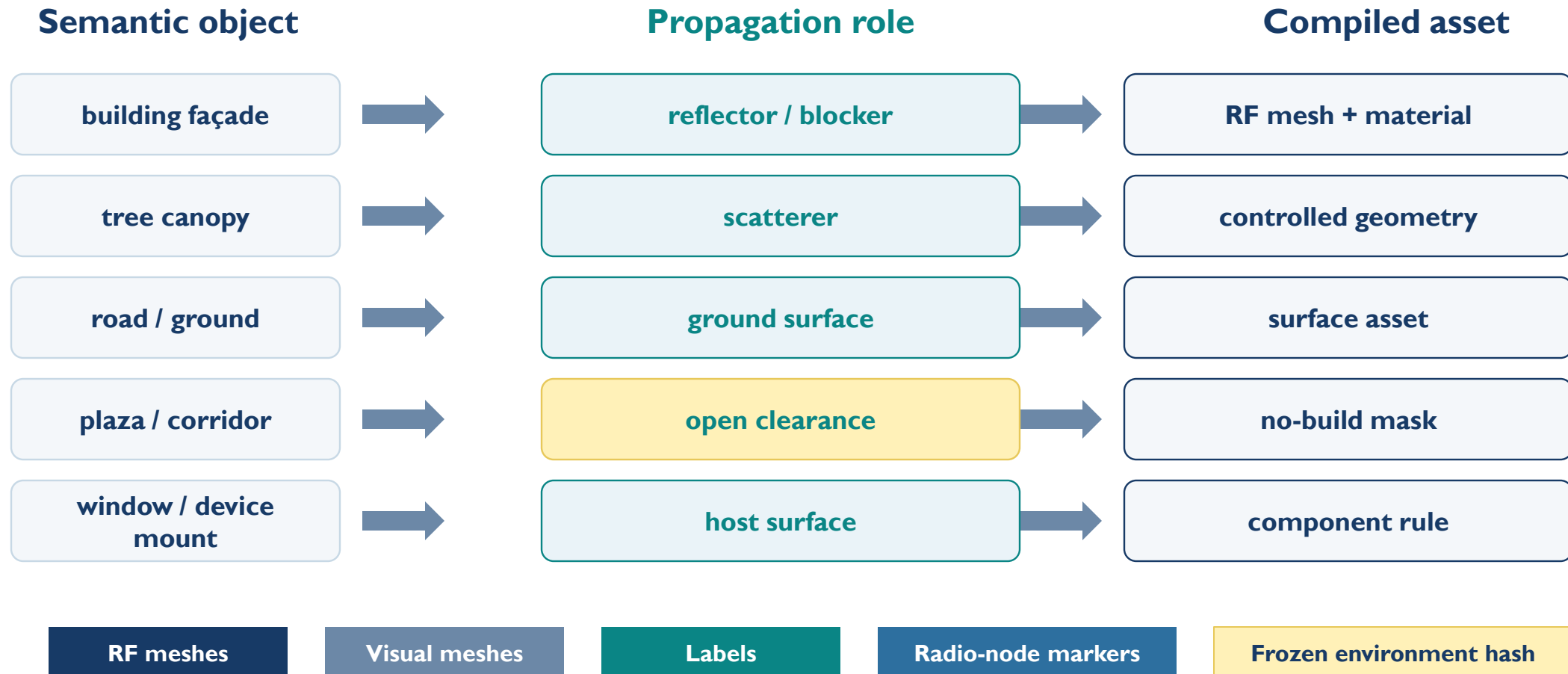
A dominant road axis becomes open space and a deployment corridor.

Next stage

Compile RF roles and materials, then place TX/RX and run quality gates.

Illustrative local RT abstraction—not a verified GIS replica.

The compiled world must be an RF environment



Radio nodes are deployed only after the environment is frozen



Quality gates build a proof chain from semantics to channels

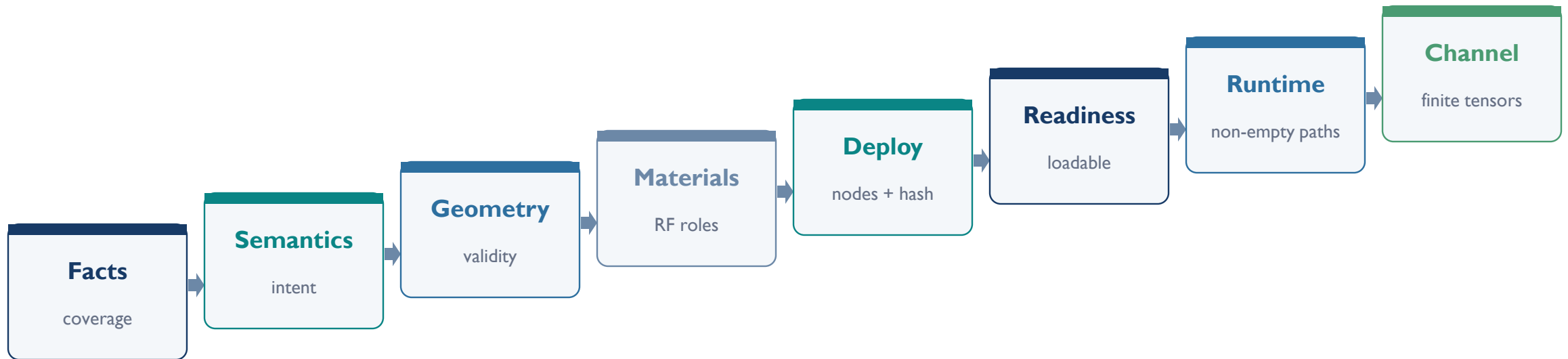
PASS

WARN

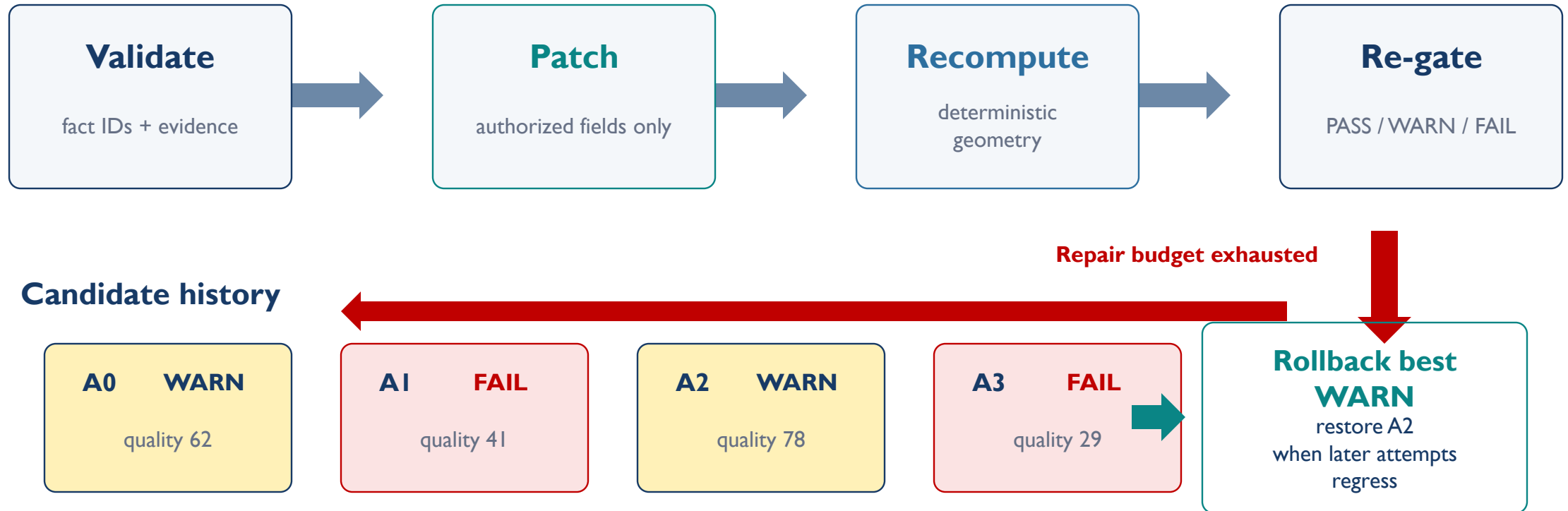
FAIL

SKIP · dependency missing

Static readiness is necessary—but only runtime and numerical gates prove that a usable channel exists.

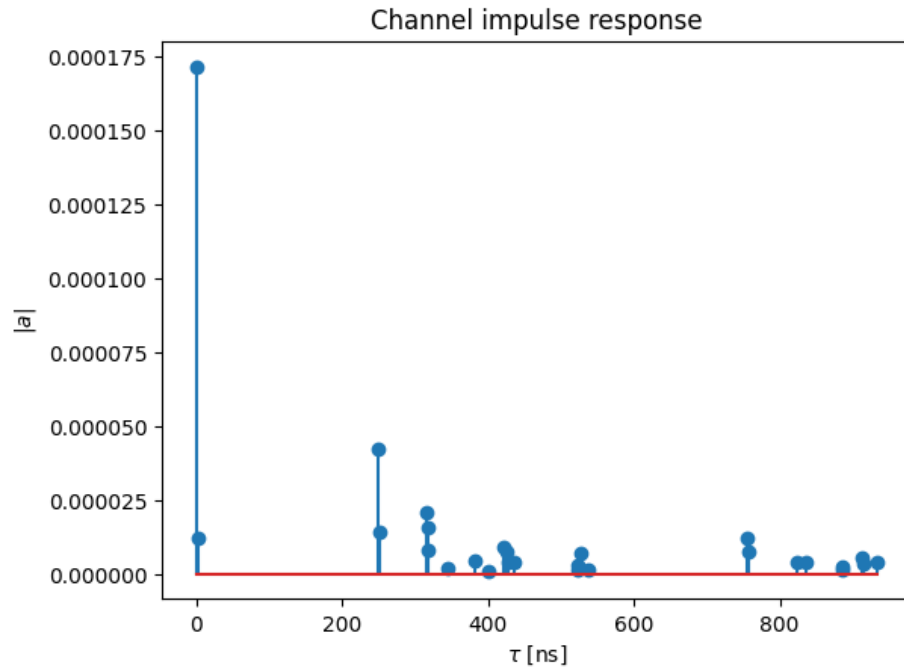


Scoped repair and rollback make failure safe

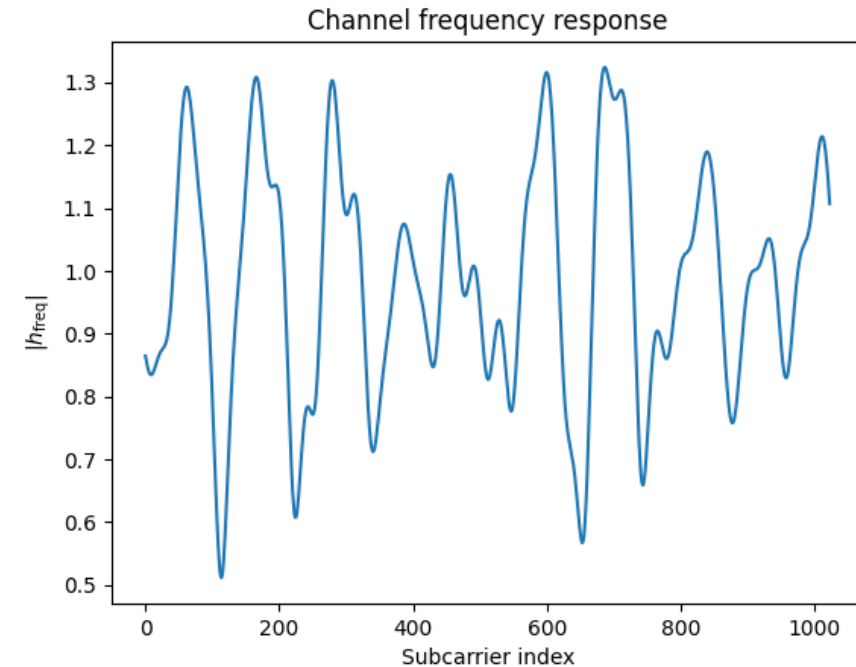


Unsatisfiable hard constraints fail explicitly; optional context alone may degrade.

Sionna turns propagation paths into inspectable channel evidence



CIR · multipath gain versus normalized delay



CFR · frequency selectivity across subcarriers

Export: channels.npz · labels.npz · scene assets · final report · manifest + checksum

Auditable evidence does not imply exact GIS or measurement-calibrated truth.



Intelligent Agents

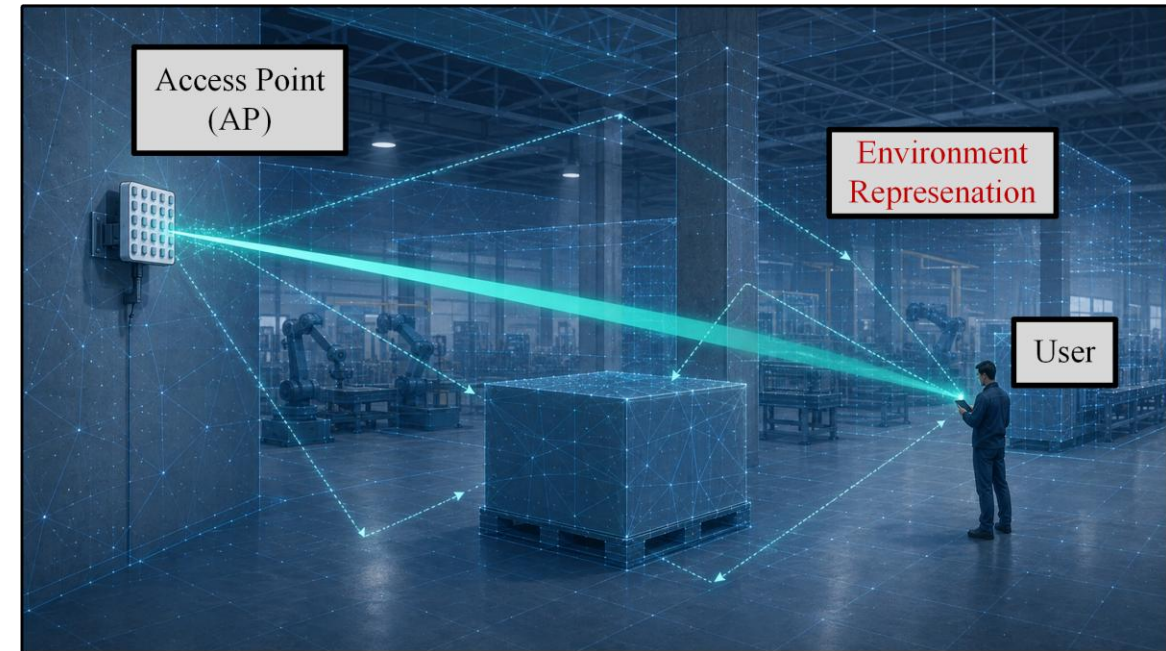
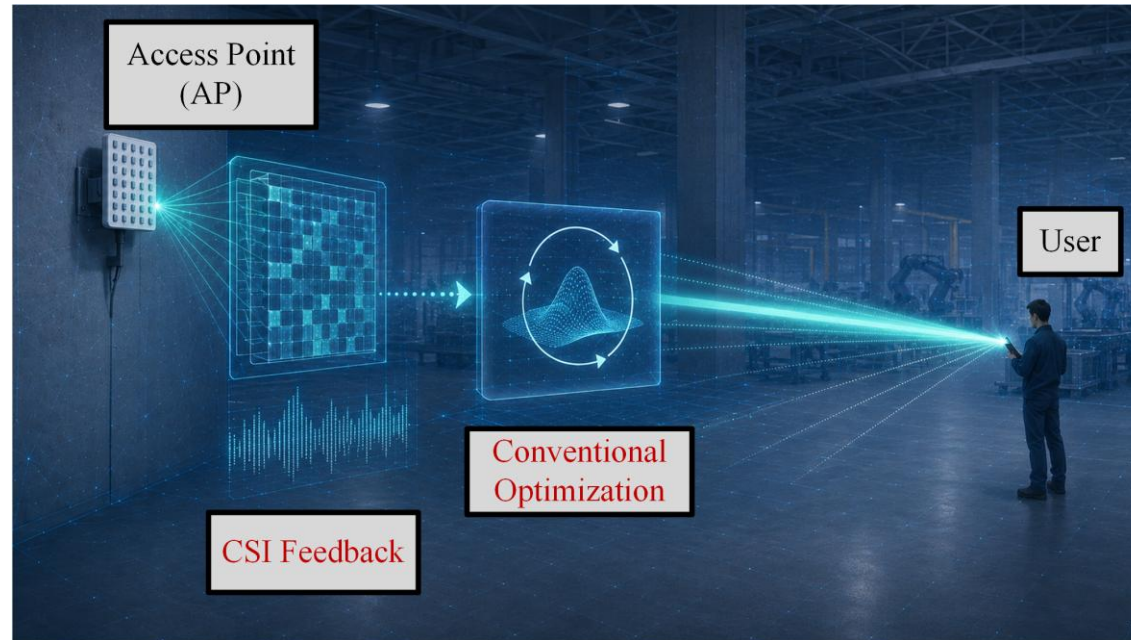


Beamforming



AP Deployment and Low-altitude Spatial Intelligence

Towards Environment-Aware Beamforming



$$\begin{aligned} \max_{\mathbf{f}_t} \quad & \log_2 \left(1 + \frac{P_t |\mathbf{h}_t \mathbf{f}_t|^2}{N_0 \cdot \text{BW}} \right), \\ \text{s.t.} \quad & |\{\mathbf{f}_t\}_i| = \frac{1}{\sqrt{N}}, \quad \forall i = 1, 2, \dots, N. \end{aligned}$$

Conventional optimization



Model the environment

Towards Environment-Aware Beamforming

- **Frequent Pilot Transmission**

- Conventional beamforming requires frequent pilot transmission and instantaneous CSI

- **High Computational Complexity and Latency**

- Conventional methods rely on iterative optimization for each channel realization,

- **Codebook Quantization Loss**

- Codebook-based methods suffer from limited resolution and increasingly costly beam sweeping as the antenna array grows

- **No Online CSI Acquisition**

- Generates beams directly from the AP–user geometry and the learned environment

- **One-Shot Beam Generation**

- Directly synthesizes the beam without exhaustive search within multiple frames

- **Grid-Free Beam Synthesis**

- Provides fine-grained steering through an over-complete angular dictionary and scalable with different antenna sizes

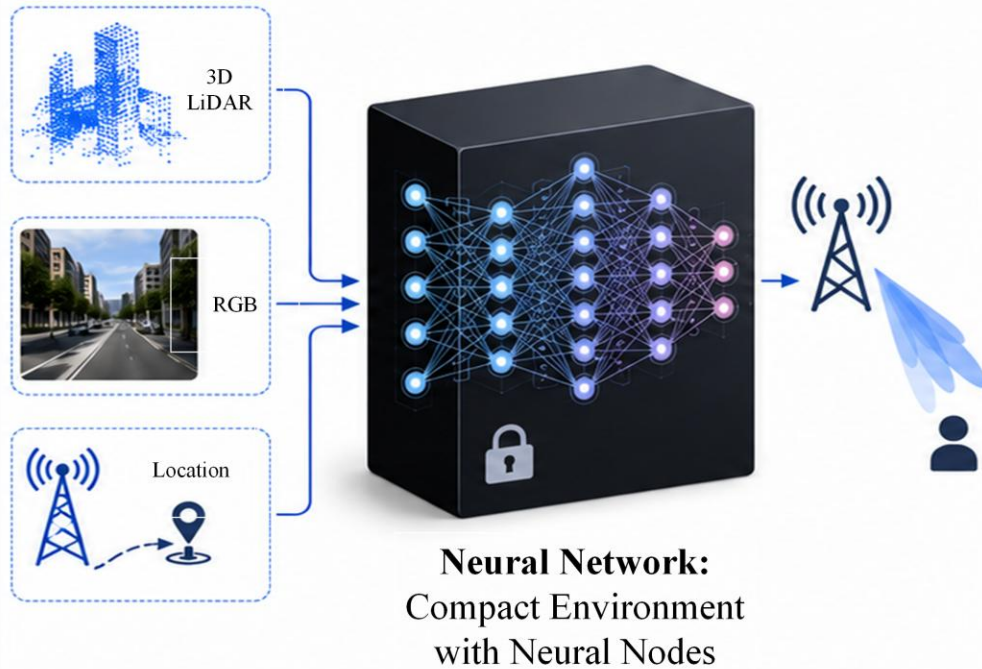
Conventional optimization



Model the environment

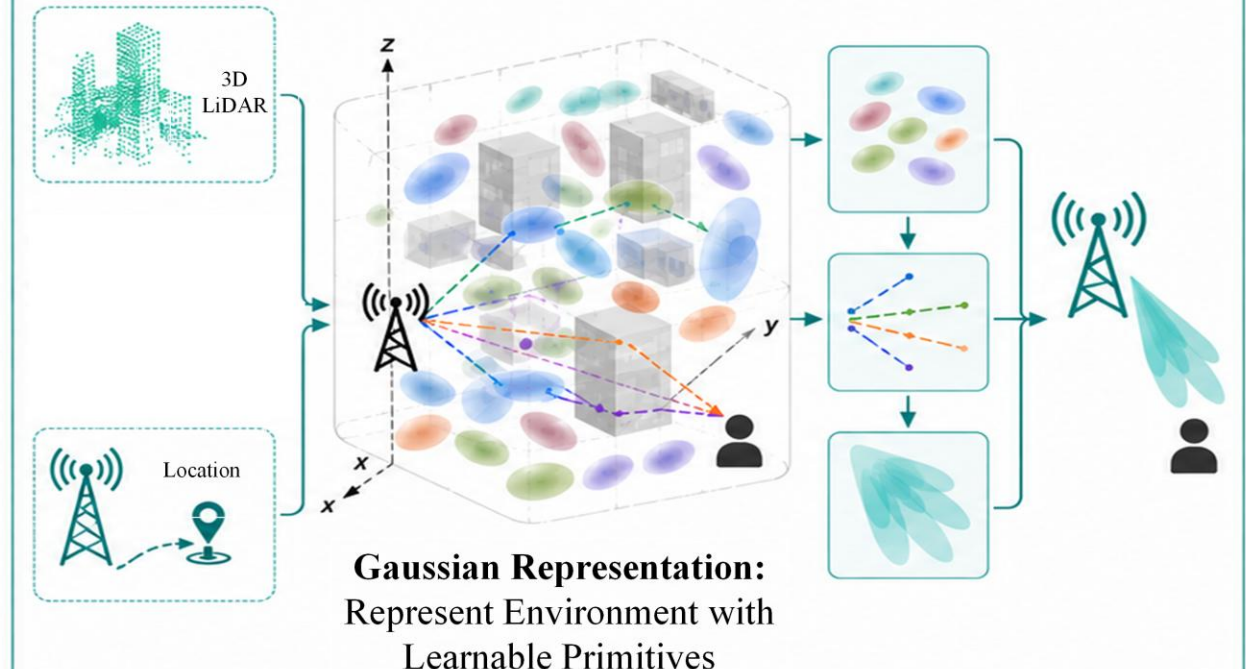
How to Model the Environment? Neural Network (NN) vs. Gaussians

Multi-modal and NN Assisted Beamforming



Implicit Modeling

Gaussian Splatting Based Beamforming

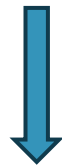
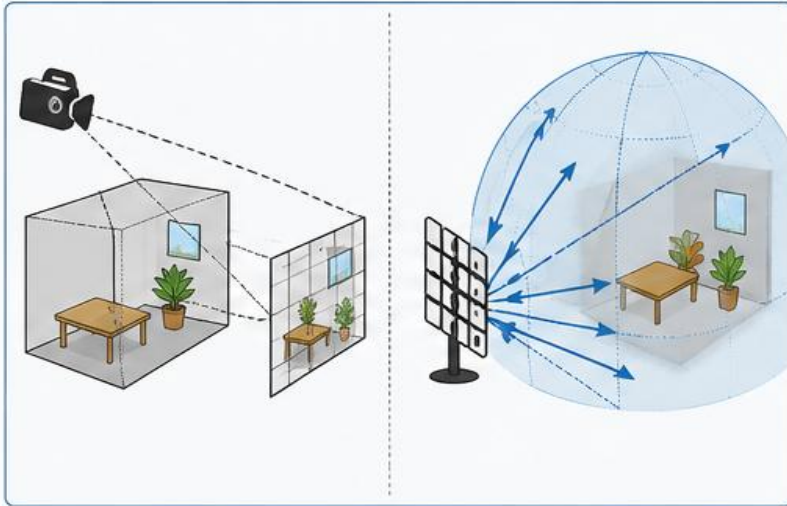


Explicit Modeling

Discrepancy: Vanilla 3DGS vs. Wireless Domain

Propagation Mismatch

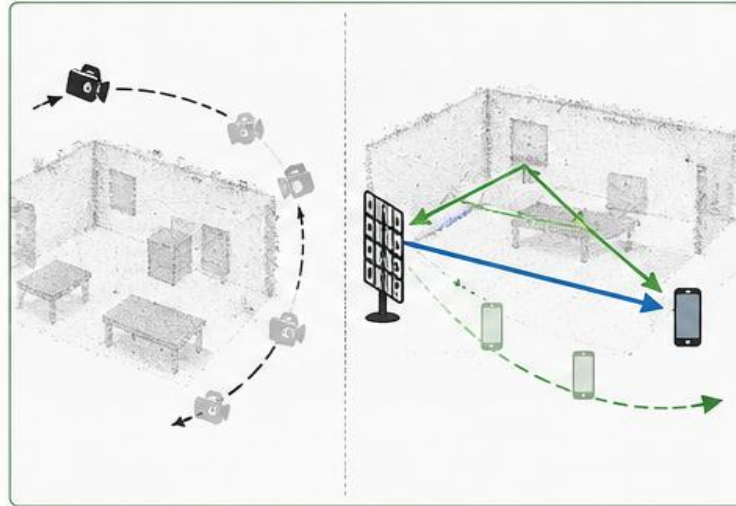
Pinhole camera view
vs.
Hemispherical wireless reception



Change the projection model

Viewpoint Mismatch

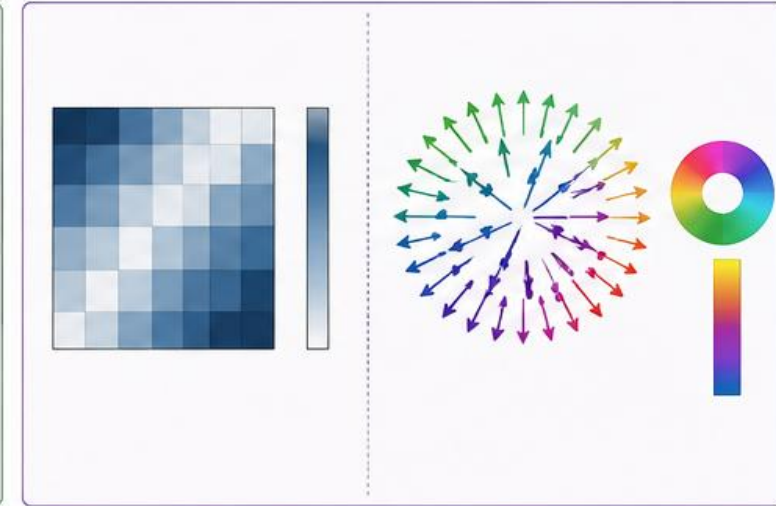
Static scene + moving camera, rendering at camera
vs.
Static AP + moving user, rendering at AP



Change the kernel function

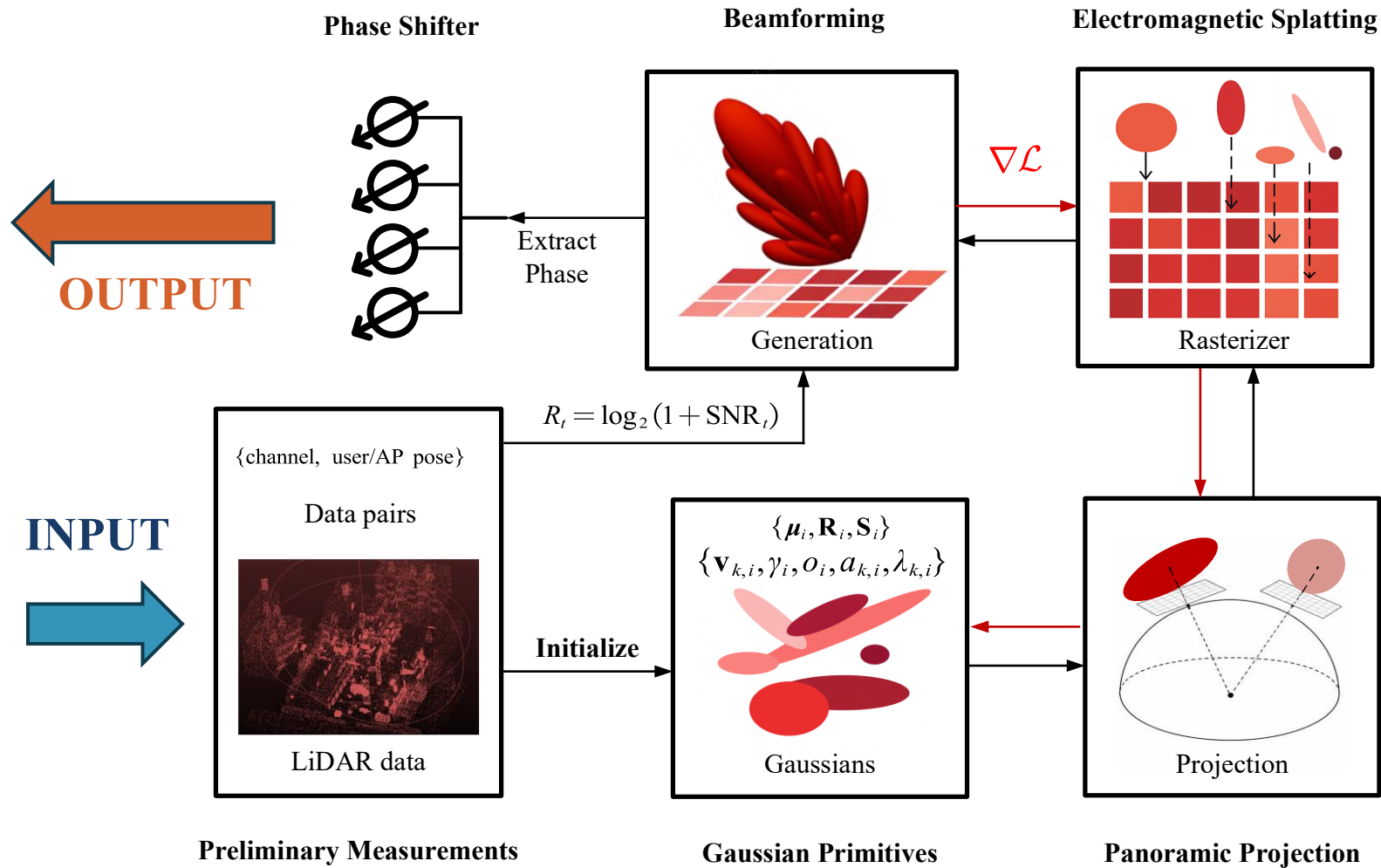
Signal Mismatch

Real-valued radiance / color
vs.
Complex-value signal



Change the rendering process

GSBF Pipeline and Overview



- I. Gaussian primitives initialization
- II. Project 3D scene to 2D plane
- III. Compute angular propagator map
- IV. Beamforming Generation

Bi-SG Scattering Kernel Design: Change the Core

Spherical Gaussian (SG)

$$S_i(\omega_{\text{in}}, \omega_{\text{out}}) = \gamma_i + \sum_{k=1}^K a_{k,i} \exp[\lambda_{k,i}(\kappa_{k,i} - 1)]$$

- **Learnable Complex-Valued Parameters:**

- $\gamma_i \in \mathbb{C}$: Complex-valued albedo
- $a_{k,i} \in \mathbb{C}$: Lobe amplitude

- **Scattering Lobe Characteristics (k -th lobe):**

- $\mathbf{v}_{k,i} \in \mathbb{R}^{3 \times 1}$: Unit-normalized learnable lobe axis
- $\lambda_{k,i} \in \mathbb{R}^+$: Lobe sharpness parameter

- **Directional Response Alignment Metric:**

$$\kappa_{k,i} = \beta(\omega_{\text{in}}^\top \mathbf{v}_{k,i})(\omega_{\text{out}}^\top \mathbf{v}_{k,i}) + (1 - \beta)\mathbf{e}^\top \mathbf{v}_{k,i}$$

- **Key Components:**

- **Near-collinear term:** $(\omega_{\text{in}}^\top \mathbf{v}_{k,i})(\omega_{\text{out}}^\top \mathbf{v}_{k,i})$
- **Redirected-scattering term:** $\mathbf{e}^\top \mathbf{v}_{k,i}$
- **Blending factor:** $\beta = \sigma[\nu(\omega_{\text{in}}^\top \omega_{\text{out}} - \xi)]$

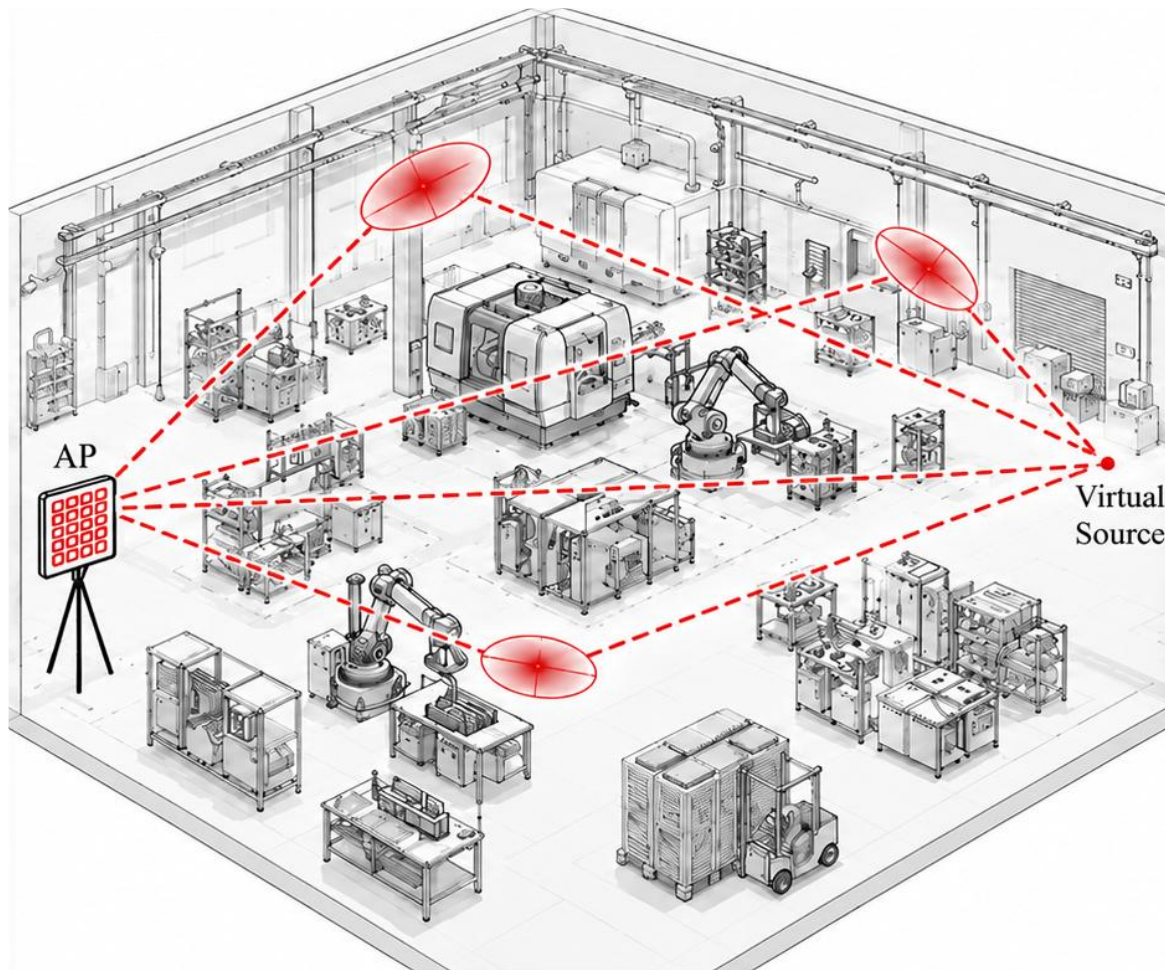
$f(\omega)$
Unidirectional



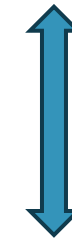
$f(\omega_{\text{in}}, \omega_{\text{out}})$
Bidirectional (Bi)

$$S_i(\omega_{\text{in}}, \omega_{\text{out}}) = S_i(-\omega_{\text{out}}, -\omega_{\text{in}})$$

Two-side Electromagnetic Rasterization: Change the Rasterization



Source-to-Primitive Transmittance

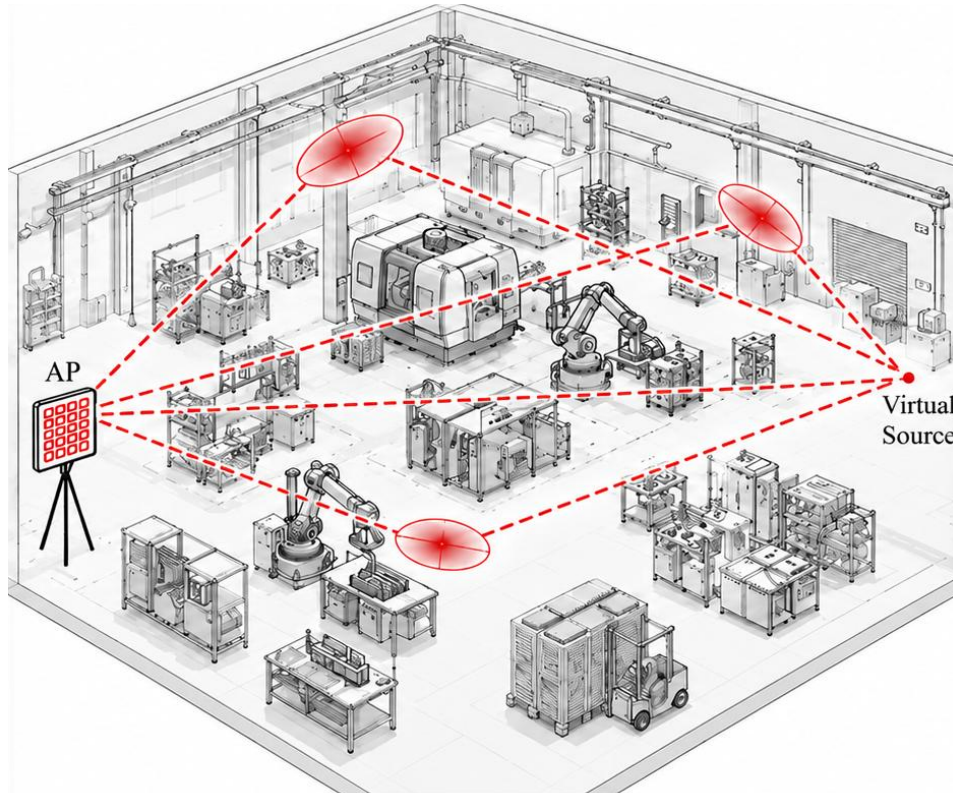


Panoramic Projection

Propagator Accumulation

Two-side Electromagnetic Rasterization: Change the Rasterization

Panoramic Projection



- **Panoramic Equirectangular Projection:**

- Unlike standard 3DGS (pinhole model), AP receives signals from the entire **forward hemisphere**.
- Projects 3D Gaussians onto a $W \times H$ angular map, then cropped to the antenna coverage.

- **World-to-Local Transformation:**

$$\mathbf{p} = \mathbf{V}\mathbf{p}_w = [p_x, p_y, p_z, 1]^\top$$

- $\mathbf{p}_w$: Primitive center in world coordinates.
- $\mathbf{V} \in \mathbb{R}^{4 \times 4}$: Viewing transformation matrix (Rotation $\mathbf{W}$ and Translation $\mathbf{t}$).

- **Azimuth (ϕ) & Elevation (θ) Mapping:**

$$\theta = \arctan2(p_y, \rho), \quad \phi = \arctan2(p_x, p_z), \quad \rho = \sqrt{p_x^2 + p_z^2}$$

- **2D Pixel Coordinate Transformation $\mathbf{p}' = [u, v]^\top$:**

$$u = W \left(\frac{\phi}{2\pi} + \frac{1}{2} \right), \quad v = H \left(\frac{\theta}{\pi} + \frac{1}{2} \right)$$

Two-side Electromagnetic Rasterization: Change the Rasterization

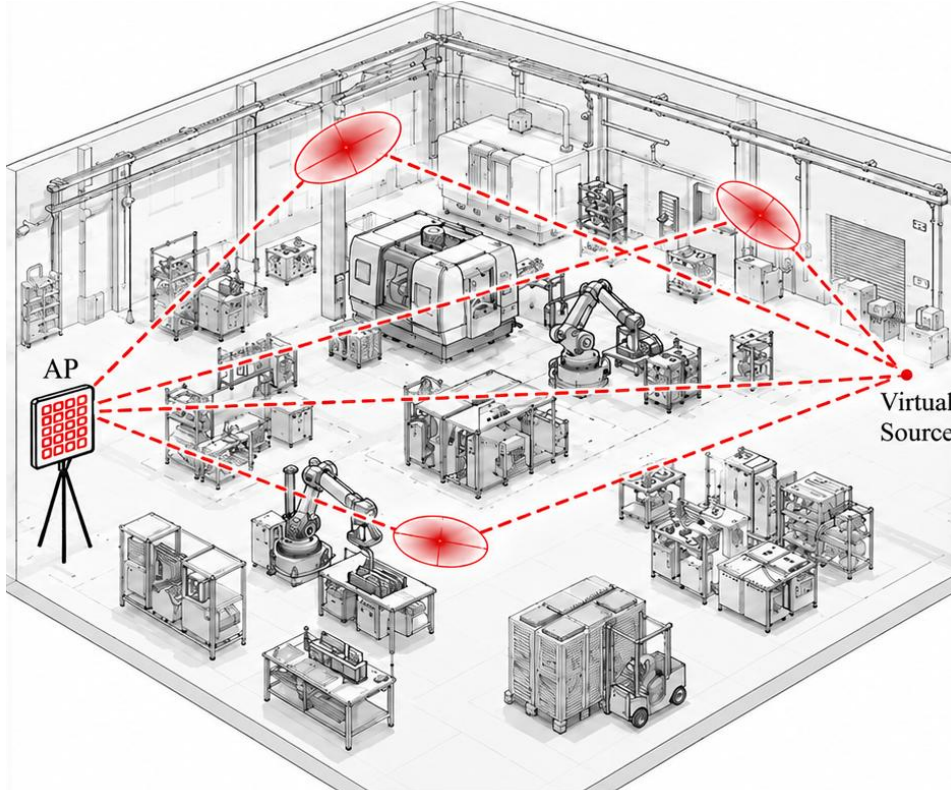
Source-to-Primitive Transmittance

$$\tau_i^{\text{UE}} = \prod_{j \in \mathcal{R}_i} (1 - \delta_{ij}^{\text{UE}})$$

- **Occluder Set ($\mathcal{R}_i$):**
 - Denotes the set of occluder primitives located between the user and target primitive i .
- **User-Side Attenuation Contribution (δ_{ij}^{UE}):**

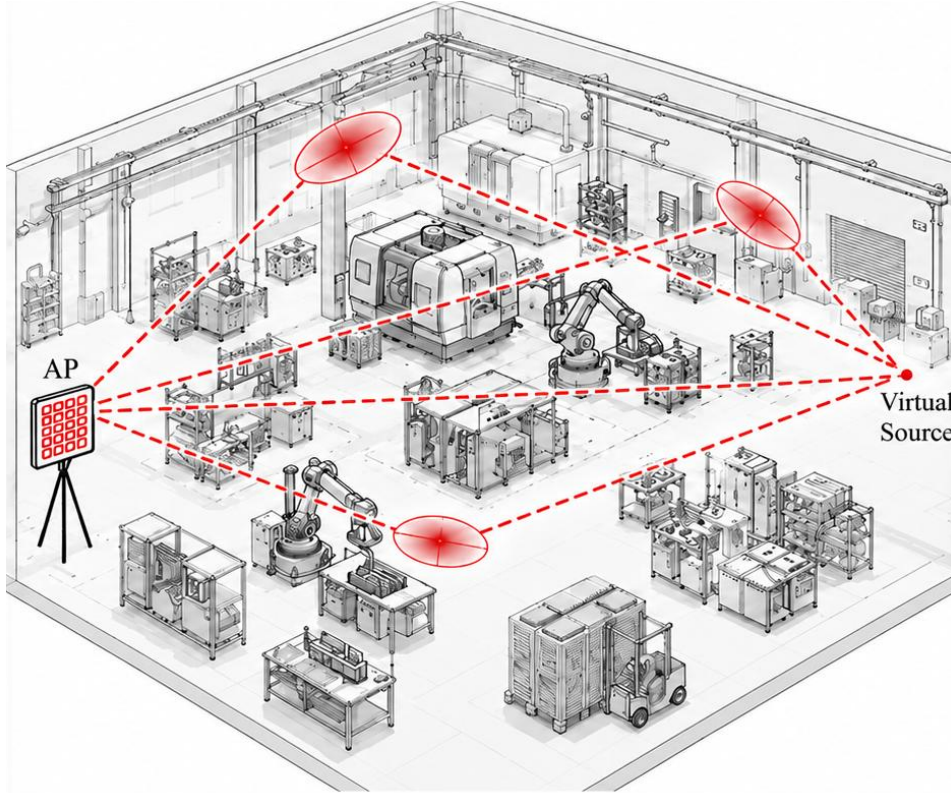
$$\delta_{ij}^{\text{UE}} = o_j G'(\mathbf{p}'_i; \mathbf{p}'_j, \Sigma_j^{\text{UE}'})$$

- Attenuation contribution of occluder j evaluated at the projected center of primitive i .
- $\mathbf{p}'_i$: Projected position of Gaussian primitive i in user-side view.



Two-side Electromagnetic Rasterization: Change the Rasterization

Propagator Accumulation



$$b_i(u, v) = o_i \tau_i^{\text{UE}} \tau_i^{\text{AP}}(u, v) S_i(\omega_{\text{in}}, \omega_{\text{out}}) G'(\mathbf{q}'; \mathbf{q}'_i, \Sigma_i^{\text{AP}'})$$

- **AP-Side Transmittance:**

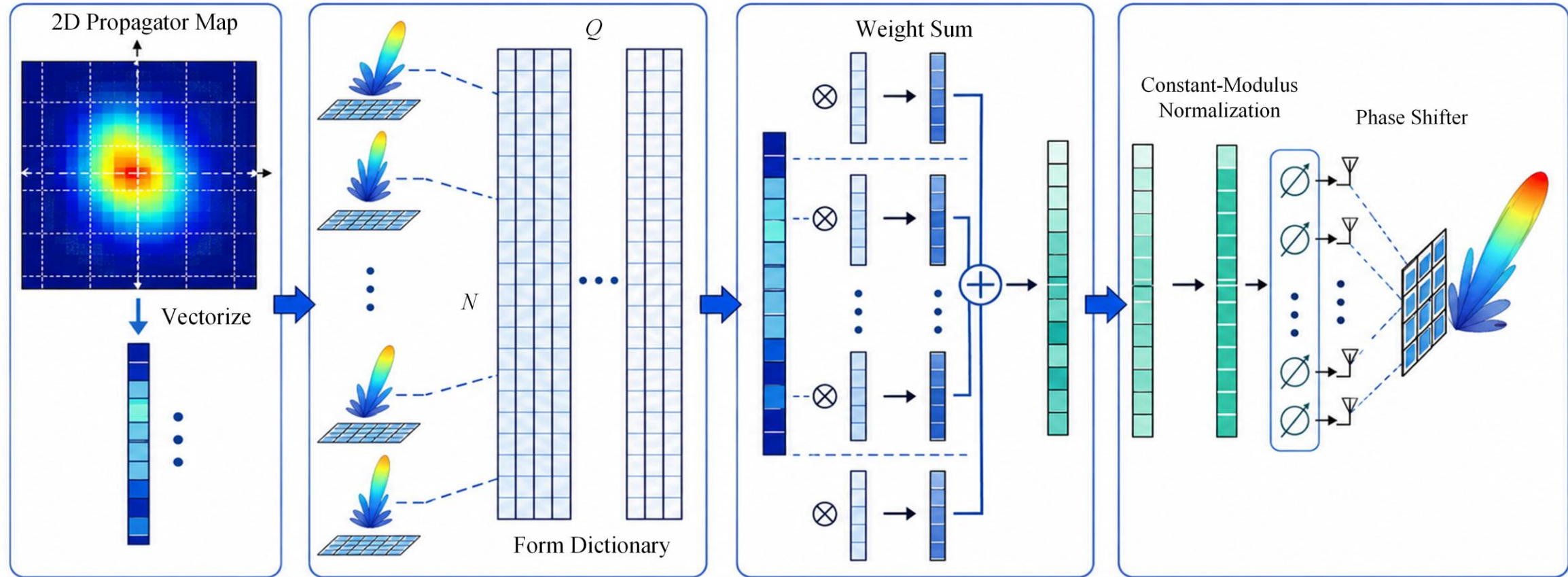
$$\tau_i^{\text{AP}}(u, v) = \prod_{j \in \mathcal{T}_{i,u,v}} (1 - \delta_{ij}^{\text{AP}}(u, v))$$

- Accumulates attenuation $\delta_{ij}^{\text{AP}}(u, v) = o_j G'(\mathbf{q}'; \mathbf{q}'_j, \Sigma_j^{\text{AP}'})$ from overlapping primitives $j \in \mathcal{T}_{i,u,v}$ at pixel $\mathbf{q}'$.
- Transformed covariance $\Sigma_j^{\text{AP}'}$ relies on fixed AP rotation matrix $\mathbf{W}^{\text{AP}}$.

- **Geometric Scattering Directions:**

- Directions ω_{in} and ω_{out} are set by relative geometry among user, primitive center μ_i , and AP.
- Feeds into scattering kernel $S_i(\omega_{\text{in}}, \omega_{\text{out}})$.

Beamforming Generation: Change the Rendering



Angular Propagator Map

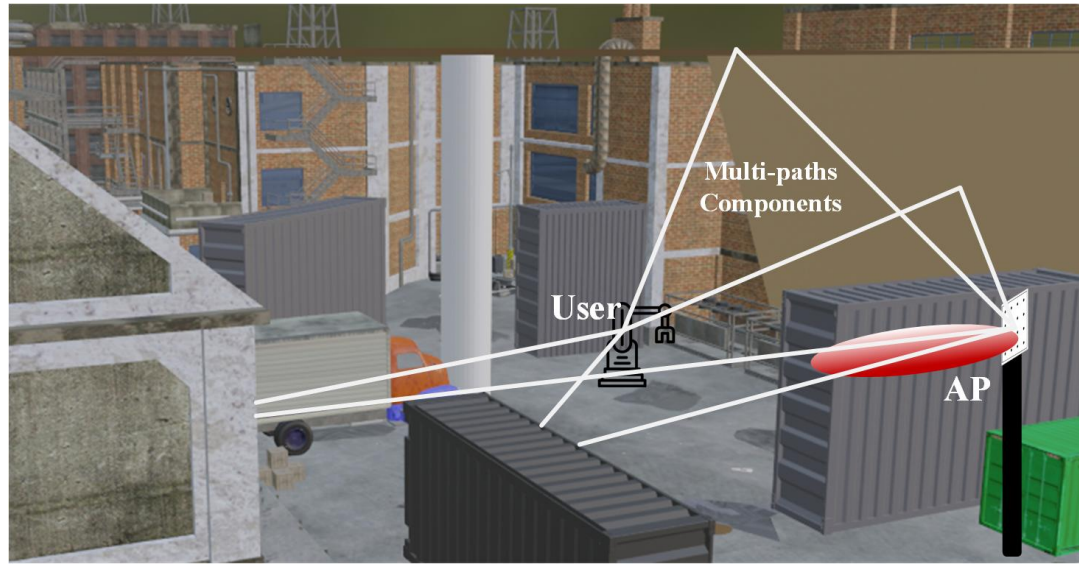
Steering Vector Dictionary

Beamforming Synthesis

Normalized Beamformer

$$\tilde{\mathbf{f}}_t = \mathbf{A}\mathbf{b}_t \quad \longrightarrow \quad \{\hat{\mathbf{f}}_t\}_i = \frac{\{\tilde{\mathbf{f}}_t\}_i}{\sqrt{N}|\{\tilde{\mathbf{f}}_t\}_i|}$$

Simulation Setup



Created Blender Scene

- **EBA:** a discrete Fourier transform (DFT) codebook-based exhaustive beam search that always selects the beam with the largest virtual gain
- **LocBF:** a purely geometry-based beamforming baseline
- **DBF:** the fully digital architecture with the optimal beamformer

Baseline and Benchmark

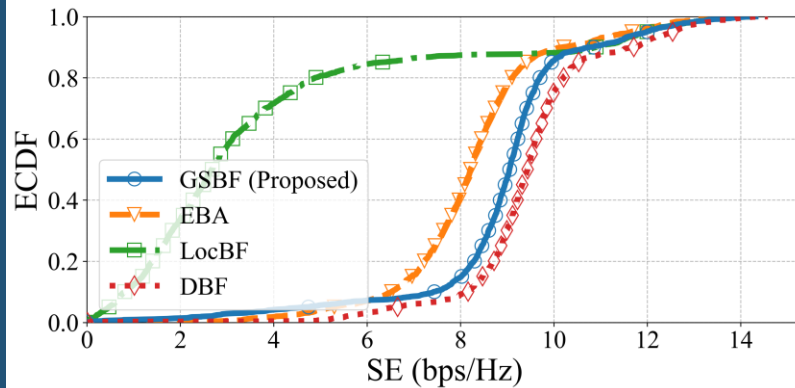
Parameters	Values
Central frequency	28 GHz
bandwidth	100 MHz
Transmit power	15 dBm
Noise power density	-174 dBm/Hz
Angular propagator map resolution	[45, 45]
Lobe number	4
Training: test samples	5600: 1400 (8:2)

Simulation Configurations

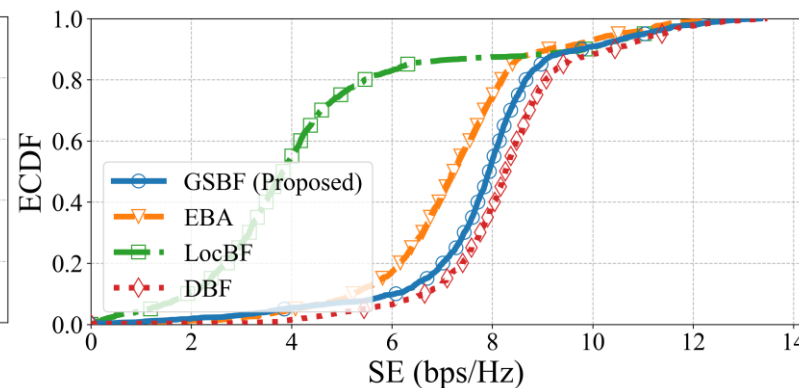
Simulation Results

Antenna Size	Mean / Median SE (bps/Hz)			
	GSBF (Ours)	EBA	LocBF	DBF (Upper Bound)
4×4	7.798 / 7.932	7.203 / 7.227	4.485 / 3.824	8.281 / 8.261
6×6	8.874 / 9.051	8.238 / 8.216	3.771 / 2.671	9.490 / 9.437
8×8	9.625 / 9.812	9.066 / 9.071	3.648 / 2.415	10.345 / 10.252

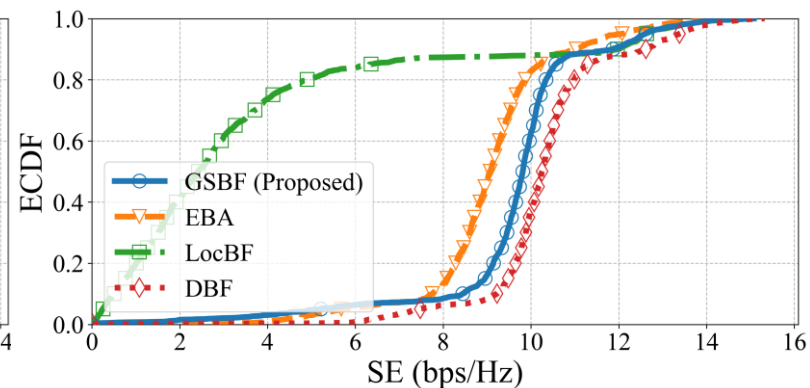
Mean & Median Report



AP: 4×4



AP: 6×6



AP: 8×8



Intelligent Agents



Beamforming



AP Deployment and Low-altitude Spatial Intelligence

AP deployment tests whether a field model supports unseen candidate locations

- Wi-Fi AP deployment involves choosing the number, placement, and configuration of access points to meet coverage and capacity goals while accounting for propagation, interference, and user distribution.



Fig. 15: BISTRO dataset.

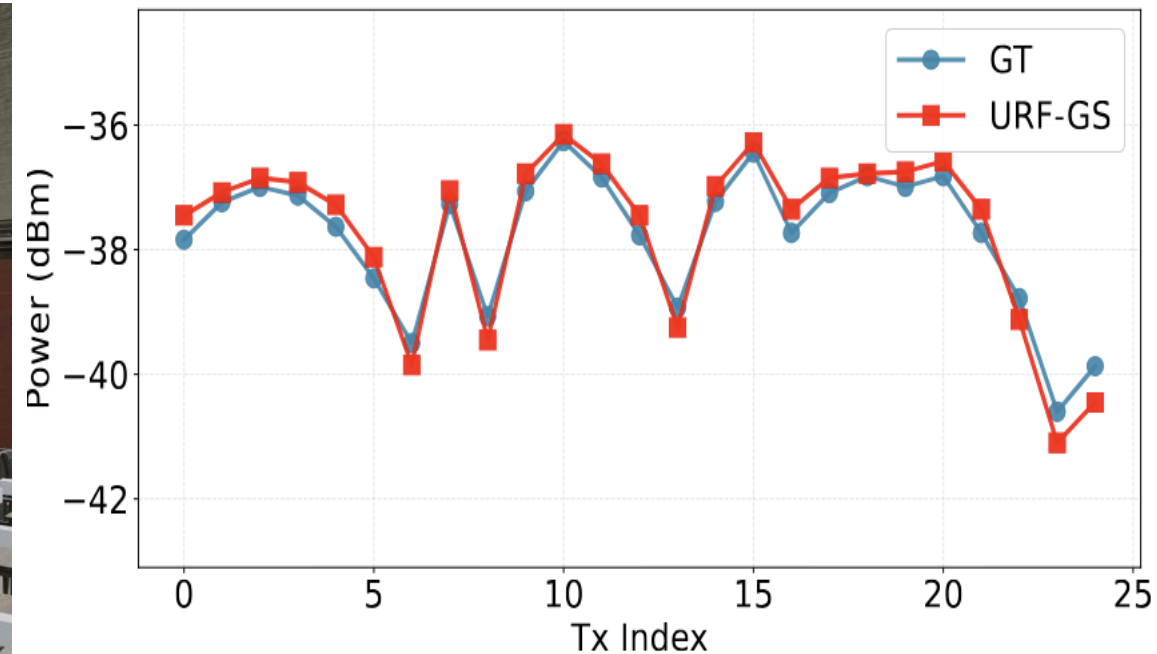
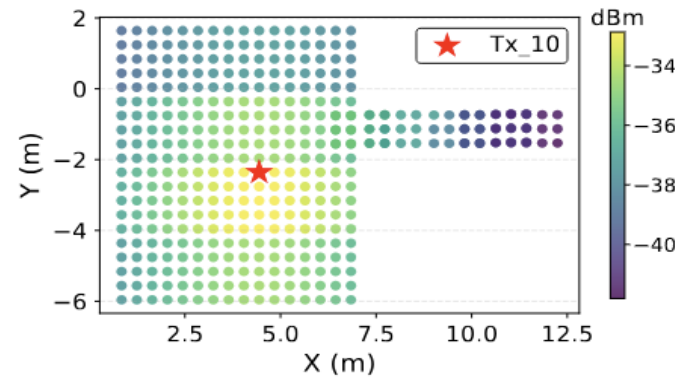


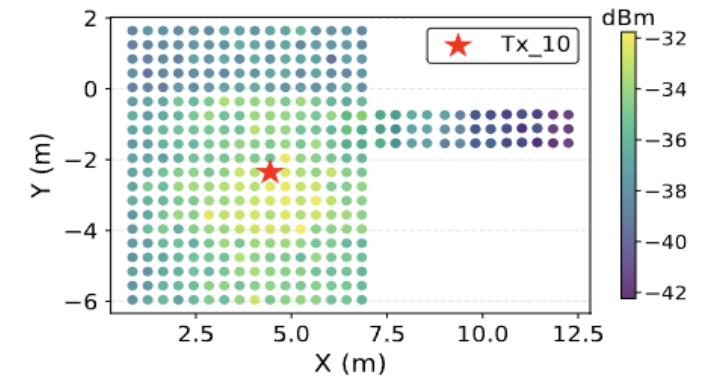
Fig. 16: The average received power of all Rx at different Tx positions.

Candidate-position queries turn one field into a deployment decision surface

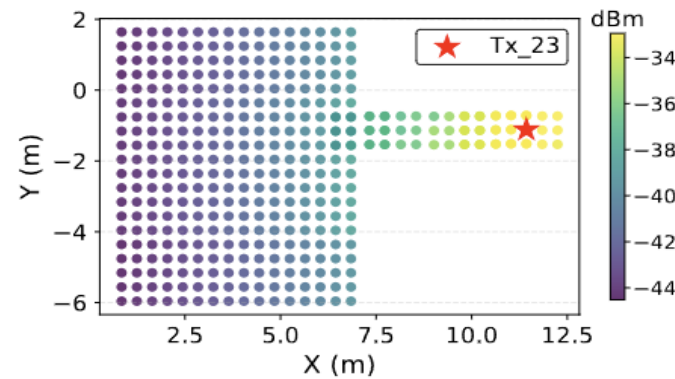
- The training set includes 5 randomly placed transmitters; the testing set uses 25 unseen transmitters mounted along the ceiling to simulate practical AP placement.
- For each Tx, there are 386 receivers uniformly sampled at a height of 1.2 m, which is typical for a user who holds devices, to emulate a dense user population.



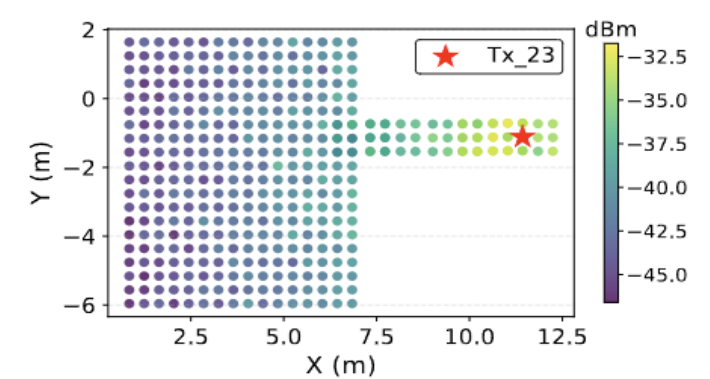
(a) Best GT Tx position.



(b) Predict Tx position..



(c) Worst GT Tx position.



(d) Predict Tx position.

Fig. 17: A visualization of the GT and predicted Tx positions.

A radio map steers the robot away from geometrically valid but unreliable paths

- Classical planners primarily exploit optical geometry from LiDAR/cameras to build maps and avoid obstacles, but typically ignore radio propagation.
- We produce a dataset based on the VVi3room layout by removing the wall between the left room and the upper right room.
- At each grid, the robot can stop or move in four directions, where the locations of the start, goal, and Tx are randomly placed.

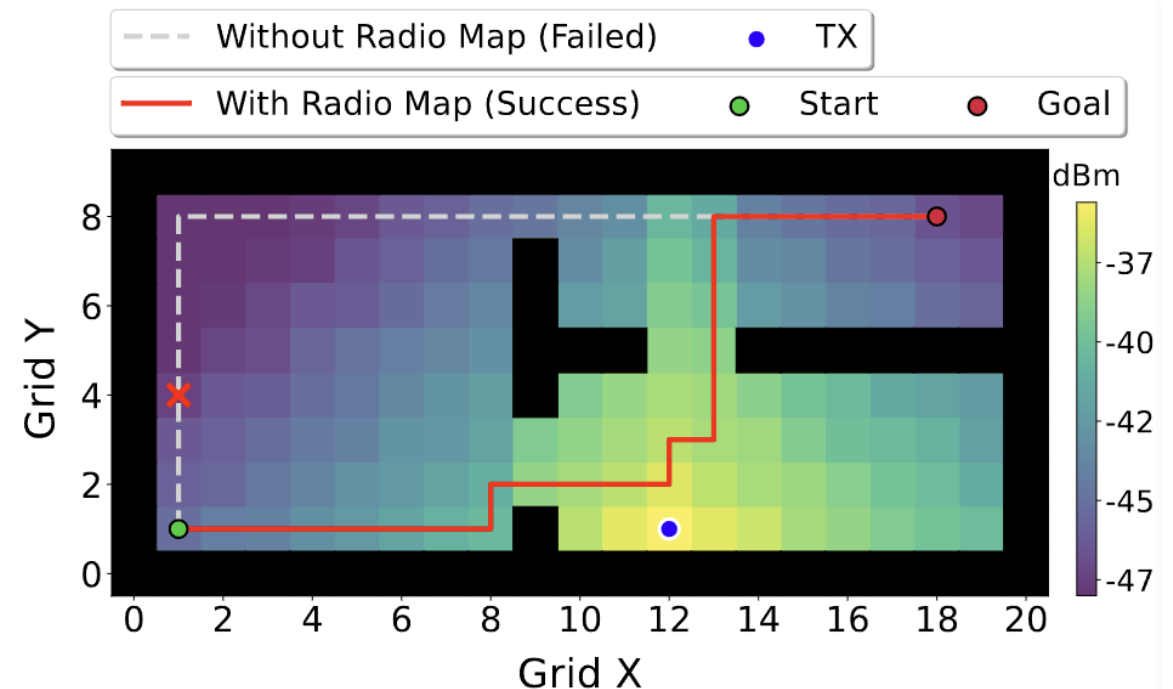


Fig. 18: Comparison of path planning results with and without radio map assistance.

Planning quality should be measured by mission failure—not map error alone

- We define that the robot reaches a grid where the average signal power is below the threshold; this event carries a certain probability of resulting in mission failure.

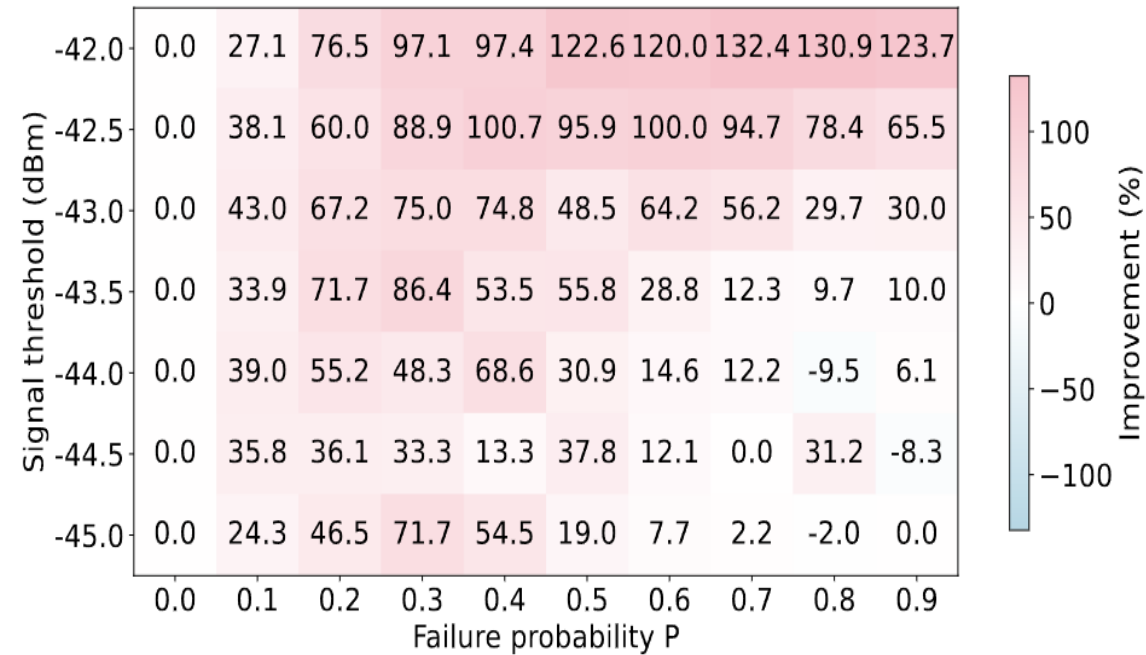
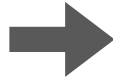


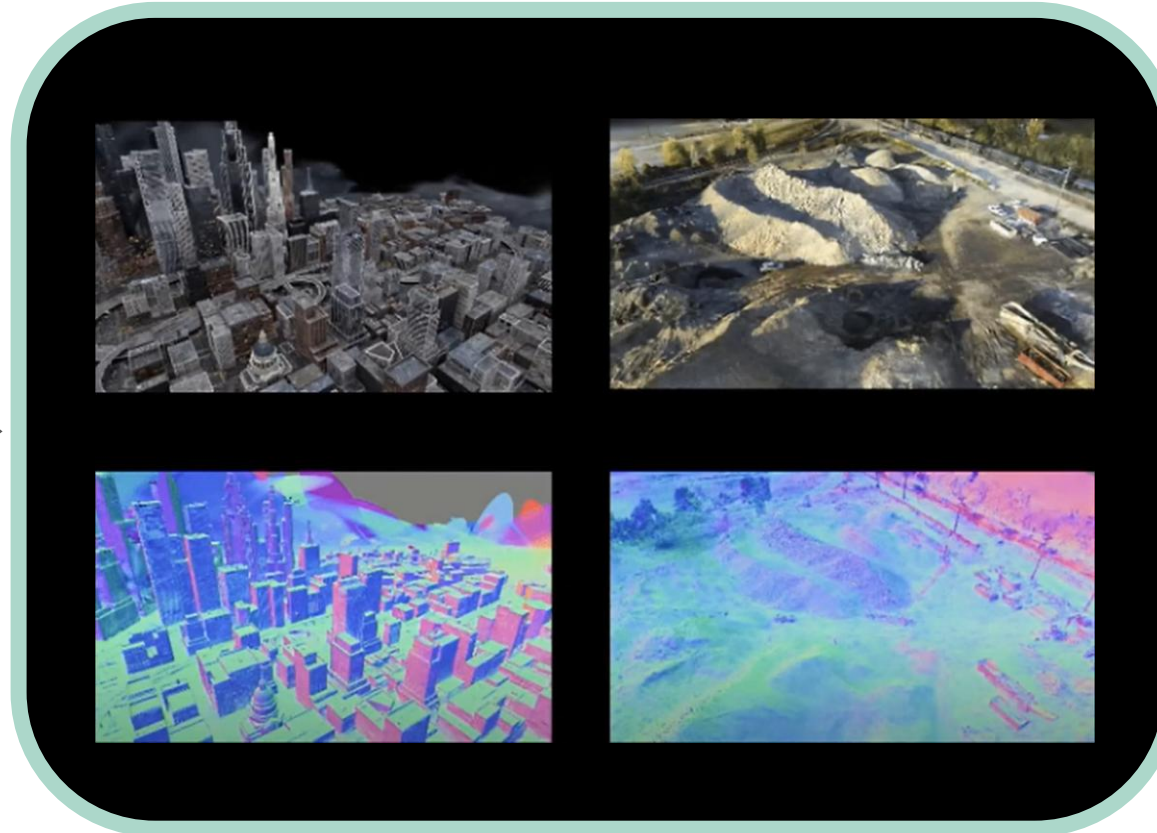
Fig. 19: Statistics on the success rate improved by introducing the radio map method.

Urban 3D modeling using low-altitude UAV photogrammetry

UAV images



Gaussian-splatting rendering



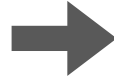
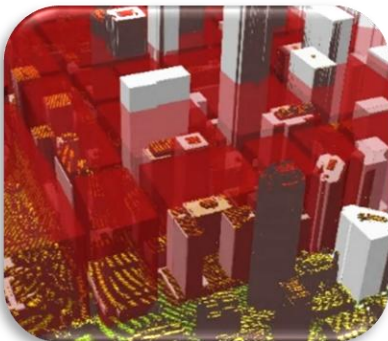
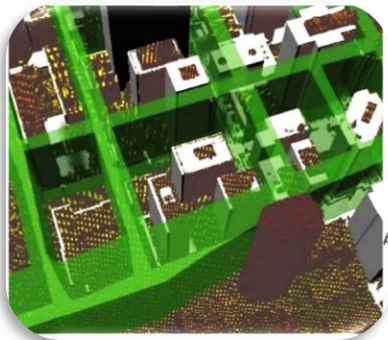
RGB rendering

Normal rendering

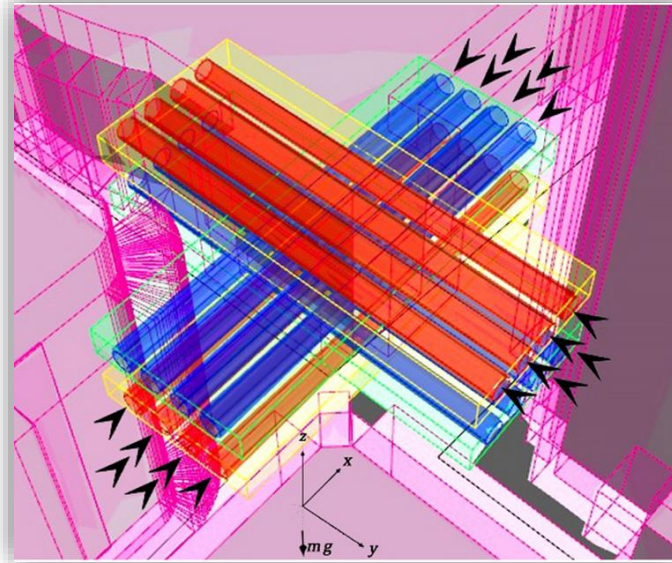
A divide-and-conquer framework enabling real-time urban reconstruction from UAV images.

Spatial intelligence for UAV trajectory optimization

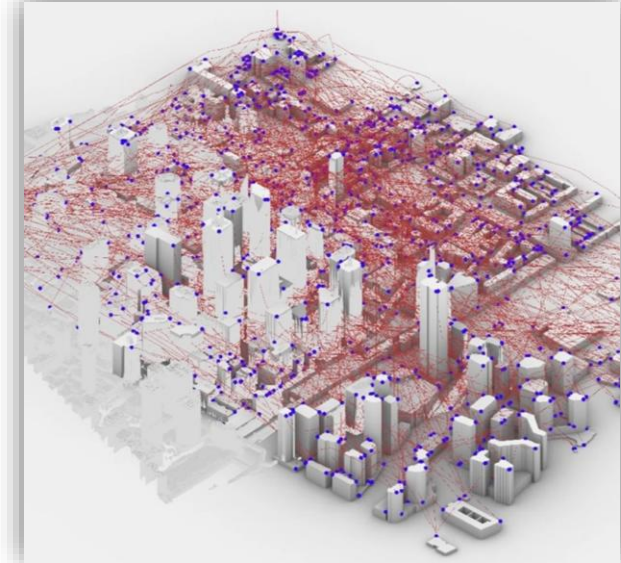
Digital-twin airspace



Trajectory optimization under airspace control



(a) hybrid layered, zonal, and tubed discretization.

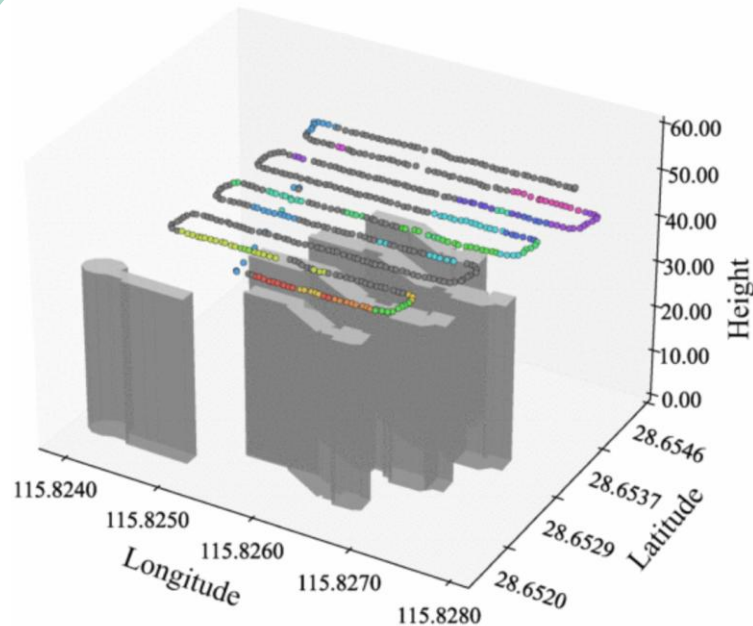


(b) UAV lane trajectory options under control.

An advanced aerial mobility system that dynamically discretizes the viable airspace into UAV trajectories.

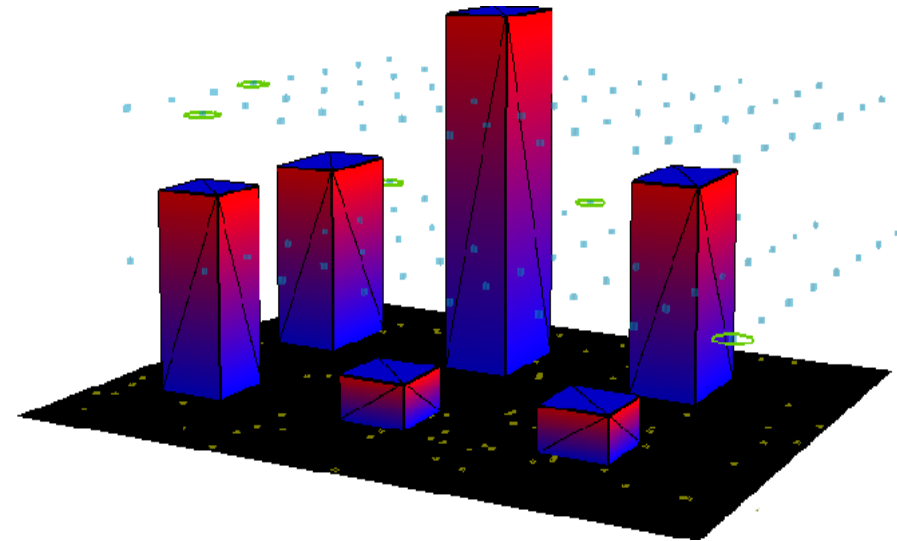
Spatial intelligence for aerial base station placement

UAV-based Radio Map Estimation



(a) Radio Map based on sparse UAV measurements [1].

Aerial base station placement



(b) Optimization of aerial base station for wider connectivity [2].

As a type of spatial prior knowledge, the Radio Map can also be used for the optimization of the layout of airborne base stations.



Part V

Conclusions & Outlook

Conclusions

- Wireless spatial intelligence shifts wireless modelling from independent links to reusable, queryable, and scene-conditioned world models.
- 3D-GS based methods provide an efficient path to reconstruct and render wireless radiation fields with explicit spatial primitives.
- Generalizable radio-field models further extend wireless spatial intelligence from one scene to unseen configurations, new scenes, and downstream tasks.
- Applications show that radio fields are not only prediction targets, but also decision surfaces for deployment, planning, localization, and agentic data generation.

Our work turns 3D environments into reusable, queryable, and task-aware wireless spatial representations.

Summarize of our work

① Explicit Representatio

WRF-GS, WRF-GS+

Use 3D Gaussian primitives/cloud points to **encode** geometry, materials, visibility and radio interactions.

② Efficient Query

WRF-GS, WRF-GS+

Enable **fast** radio-map, PAS, CSI and spatial spectrum prediction

③ Strong Generalization

URF-GS, Point2Radio

Enable **unseen** Tx/Rx data generation and **cross-scene** transfer

④ Physical-aware learning

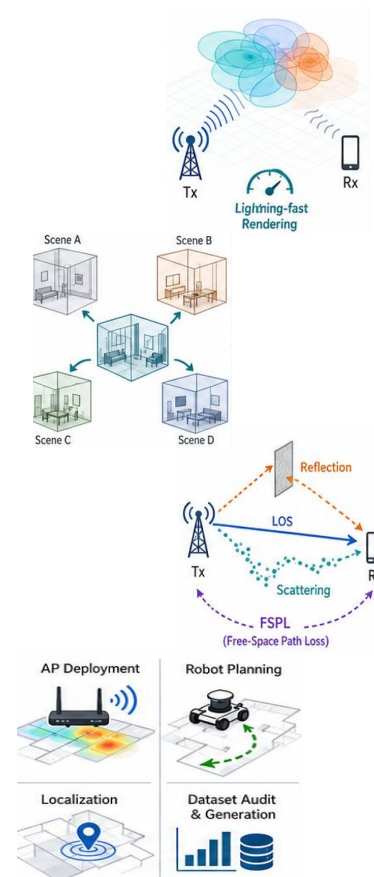
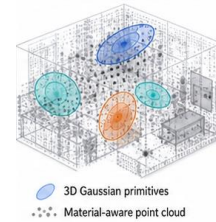
GeoGS-CE

Embed propagation and signal processing priors for **structural rendering**.

⑤ Task-level Value

GLoCFM

Support real deployment, planning, localization and towards **intelligent agent generation**.



Outlook

- Unified multimodal wireless world models.
- Cross-scene, cross-frequency, and cross-device generalization.
- Online adaptation to mobility and dynamic environments.
- Trustworthy wireless agents with validation, repair, and auditable channel evidence.

Toward wireless world models: unify geometry, semantics, mobility, and RF measurements into an evolving, queryable representation for environment-aware wireless intelligence.



Questions & Discussion