# BeamMix: 3D Gaussian Mixture-of-Experts for Element-Space Wireless Channel Modeling

William Bjorndahl
Southern Methodist University
Dallas, TX, USA
wbjorndahl@smu.edu

Joseph Camp
Southern Methodist University
Dallas, TX, USA
camp@smu.edu

## Abstract

Many-antenna base stations must rapidly select beams and compute multi-user precoders from uplink measurements. This requires accurate, decision-ready models of the downlink channel in element space. Recent work has ported radiance field graphics techniques to the RF domain to learn how the propagation environment shapes the wireless channel. These scene-centric methods first learn a volumetric RF field or a cloud of 3D Gaussian primitives that encodes this propagation. They then ray-march or splat through this field to render channel state information (CSI) or spatial spectra at the array. For scheduling and multi-user precoding, however, this workflow is indirect. The models are designed to learn the 3D RF field, and the downlink channel information needed for resource allocation (CSI or spatial spectra) only becomes available after the rendering step. To bypass rendering we propose BeamMix, a map-free, element-space model that represents the channel as a mixture of physics-guided 3D Gaussian primitives. Each primitive has structure inspired by geometry-based stochastic channel models, with a learnable spatial footprint, direction, delay, and frequency response. An uplink-conditioned mixture-of-experts gate selects a top set of active primitives per observation. Their complex-valued contributions are splatted directly onto the antenna manifold. This produces element-space CSI for schedulers and precoders without additional rendering. We evaluate BeamMix on: (i) site-level uplink-to-downlink CSI prediction, (ii) full-array uplink-to-downlink CSI prediction, and (iii) RF spectrum synthesis. When directly compared, BeamMix improves accuracy over wireless radiance field baselines on site-level CSI and spectrum synthesis.

## CCS Concepts

• **Networks → Mobile networks**; • **Computing methodologies → Machine learning**; **Modeling methodologies**.

## Keywords

3D Gaussians, Mixture-of-Experts, Element Space, Multiple Input Multiple Output

**ACM Reference Format:**
William Bjorndahl and Joseph Camp. 2026. BeamMix: 3D Gaussian Mixture-of-Experts for Element-Space Wireless Channel Modeling. In *The 24th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '26), June 21–25, 2026, Cambridge, United Kingdom.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3745756.3809247


ACM ISBN 979-8-4007-2027-7/2026/06
https://doi.org/10.1145/3745756.3809247

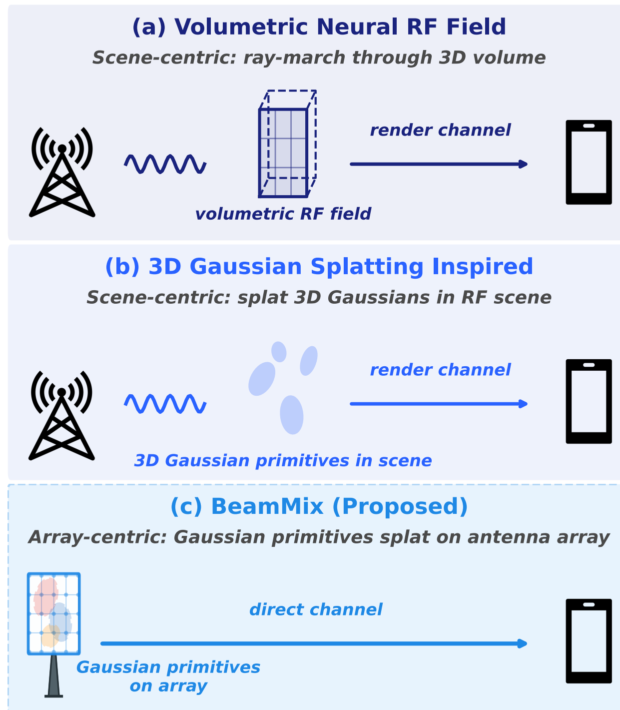


**Figure 1: Scene-centric vs. array-centric RF representations. Top (a): volumetric neural RF fields model the environment as a continuous 3D RF scene and render the channel via ray-marching. Middle (b): 3D Gaussian splatting replaces the volume with a cloud of Gaussian primitives placed in the physical scene. Bottom (c): BeamMix places a Gaussian mixture directly on the antenna array, bypassing the need for an intermediate scene representation.**

## 1 Introduction

Multiple-input, multiple-output (MIMO) enables high-rate cellular systems. A base station (BS) with many antennas boosts capacity by forming sharp beams and multiplexing users in space. To aim these beams correctly and construct multi-user MIMO (MU-MIMO) precoders, the BS must rapidly select beams and compute precoding weights from uplink (UL) measurements [2, 4, 9, 10, 15]. This task depends on accurate, real-time estimates of the downlink

(DL) channel at the level schedulers and precoders actually consume. Specifically, per-antenna, or element-space, channel state information (CSI). As arrays grow larger and bandwidths widen, the network must make many more fine-grained scheduling decisions per interval. The resolution of these decisions scales with the number of elements [3].

The element-space channel is determined by the surrounding three-dimensional (3D) environment. Obstacles and reflectors create multipath. Signals reflect, diffract, and scatter, so that each BS element sees a complex-valued superposition of a small number of multipath components (MPCs) [1, 12, 14]. Classical geometry-based stochastic channel models (GSCMs) capture this by decomposing the channel into a sum over clusters and rays [1, 12, 16]. Each cluster has a mean direction and delay, and the element-space channel is a coherent sum of steering vectors weighted by complex gains. In practice, only a handful of clusters carry most of the power. This clustered structure, together with recent research in the computer graphics domain for neural radiance fields (NeRFs) and 3D Gaussian Splatting (3DGS), has inspired a new line of wireless radiance/radiation field models.

NeRFs represent a static scene as a continuous 5D function that maps a 3D location and viewing direction to volume density and color, and render images by numerically integrating this radiance field along camera rays [11]. This achieves high-fidelity novel-view synthesis, but requires many multilayer perceptron (MLP) evaluations per ray. 3DGS replaces per-sample MLP queries with an explicit set of anisotropic 3D Gaussian primitives whose projected splats are composited in screen space. This enables real-time view synthesis while preserving a volumetric interpretation of the scene [8]. The underlying image formation is closely related to classical elliptical weighted average (EWA) splatting, which uses Gaussian reconstruction kernels combined with Gaussian low-pass filters to obtain anti-aliased footprints when projecting anisotropic kernels [24].

In the wireless domain, NeRF$^2$ and subsequent RF radiance field work transfer these graphics ideas to wireless by representing the environment as a volumetric RF field $F(\mathbf{x}, \mathbf{v}, \mathbf{t})$ or a cloud of 3D Gaussian virtual transmitters. Ray marching or electromagnetic splatting is used to integrate attenuation and phase along rays before finally aggregating into CSI or angular spectra [17, 18, 21, 23]. These scene-centric models treat the wireless radiation field as the primary object. Given a transmitter (TX) position and a fixed receiver (RX) location, they can synthesize spatial spectra, render CSI at new sites, or support RF digital twins. While physically elegant, this formulation is indirect for scheduling. What the BS actually needs is DL CSI in element space (the channel seen at each antenna element) and concrete decisions such as "which beam from this codebook?" or "what MU-MIMO gain does this precoder provide?". Those quantities only become available after a separate rendering step converts the volumetric RF field into element-space channels.

In machine learning, mixture-of-experts (MoE) models offer a complementary tool for modeling complex, heterogeneous functions. In an MoE, a trainable gating network routes each input to a small subset of expert subnetworks, and the outputs of these experts are combined to form the final prediction [6, 7, 13]. This sparse routing increases model capacity without proportional compute and encourages experts to specialize on different regions of the input space. For wireless channels, this aligns with the sparsity of real propagation, where dominant components carry most of the power. Different UL observations may activate different subsets of experts, and a gating network can learn to select which are relevant for a particular TX–RX configuration.

In this work we take an opposite approach to scene-centric RF radiance fields. Rather than learning a volumetric RF field and then rendering down to CSI, we adopt an array-centric view in which the element-space downlink channel itself is the primary object. We parameterize the channel on the array with a set of 3D Gaussian primitives. An illustrative comparison between the array-centric approach and alternative scene-centric approaches is shown in Fig. 1. Then an MoE gate, conditioned on UL CSI, is used to select only the few primitives that matter for each link. This produces a representation that is directly aligned with GSCM intuition and with the quantities consumed by schedulers and precoders.

We propose BeamMix, a map-free model that bypasses scene rendering and operates natively in CSI space. BeamMix consists of a sparse MoE built from lightweight, physics-aware 3D Gaussian primitives. Each primitive has a learnable 3D center, anisotropic covariance and rotation, preferred direction, delay, and amplitude. These parameters define a spatial footprint and phase pattern across the array. An EWA kernel projects each primitive onto the BS antenna coordinates, as in classical anti-aliased splatting [8, 24]. A UL-conditioned MoE gate selects the active primitives for each observation. The active primitives are then blended, along with a shallow UL$\rightarrow$DL MLP and a global complex scalar head, to produce the predicted DL CSI. In this way, BeamMix focuses modeling capacity directly on the decision quantity the scheduler uses rather than on an intermediate volumetric RF field.

Beyond its native element-space task, we also evaluate BeamMix on angular spatial spectrum synthesis as a diagnostic task. In the RF radiance field literature, this task plays the same role that novel-view synthesis plays in graphics. It tests whether a model has correctly internalized the scene's angular multipath structure and provides a standardized point of comparison to scene-centric baselines. BeamMix uses the same Gaussian primitive representation, but queries it on a hemispherical grid around the RX rather than on the array manifold. A predicted spectrum image is rendered that can be compared against NeRF and 3DGS inspired methods for wireless radiance fields.

As a whole, BeamMix is an array-space Gaussian MoE that models the channel directly while keeping outputs immediately actionable for scheduling. It avoids extra spectrum-rendering projections, preserves full complex information for MU-MIMO design, and decomposes each channel prediction into a set of spatially localized components with explicit footprints and gate weights. Our experiments show that BeamMix can both outperform scene-centric RF radiance fields on site-level UL$\rightarrow$DL CSI and spectrum synthesis, and match a strong per-antenna MLP baseline on full-array UL$\rightarrow$DL CSI while retaining this component-wise transparency.

## Contributions

We publicly release the BeamMix code. Our main contributions are summarized as follows:

- **Array-space 3D Gaussian MoE for CSI.** We propose BeamMix, a map-free, per-antenna formulation that models the DL channel directly on the array as a sparse mixture of 3D Gaussian primitives. The model produces decision-ready element-space CSI for beam selection and MU-MIMO precoding without relying on a volumetric RF field or ray marching through a scene representation.
- **UL-conditioned sparse routing.** BeamMix uses a UL encoder to produce a MoE gate over Gaussian primitives. For each UL observation, the gate activates a top-$k$ set of primitives whose parameters align with the observed UL CSI. This sparse routing is inspired by the clustered multipath structure assumed in GSCMs and produces interpretable Gaussian footprints on the array.
- **Unified primitive representation across tasks.** The same Gaussian primitives support both site-level and element-space UL→DL CSI prediction and angle-domain spatial-spectrum synthesis. This provides a coherent representation that can be queried on the array manifold for scheduling or on a hemispherical grid for spectrum evaluation.
- **Empirical gains and practicality.**
  - Section 4: On a site-level MIMO CSI benchmark with five methods (NeRF$^2$, GSRF, WRF-GS, WRF-GS+, and BeamMix) evaluated under the same protocol, BeamMix achieves the highest median DL SNR (25.1 dB) while training in 10.1 minutes and running inference in 0.08 ms/frame.
  - Section 5: Using line-of-sight (LOS), non-line-of-sight (NLOS), and continuous mobility scenarios from the Argos Channel Survey [14], BeamMix attains median UL→DL SNRs of $\approx$ 27 dB at 2.4 GHz and $\approx$ 18.8 dB at 5 GHz. BeamMix achieves near-perfect Discrete Fourier Transform (DFT) beam selection and 2-user MU-MIMO performance. It matches a strong per-antenna MLP baseline while exposing inspectable, spatially localized components.
  - Section 6: On an RF spectrum synthesis dataset, BeamMix achieves the best spectrum fidelity among NeRF and 3DGS inspired methods for wireless radiance fields. Specifically, BeamMix outperforms NeRF$^2$, WRF-GS, WRF-GS+, and GSRF, with median SSIM of 0.93, PSNR of 25.9 dB, and NMSE of $-19.6$ dB.

## 2 Related Work

Classical wireless channel models span probabilistic, deterministic, and data–driven approaches [1]. Geometry–based stochastic and ray–tracing methods explicitly parameterize paths via angles, delays, and material properties. However, these approaches are limited when the environment is only coarsely known and when many–antenna, wideband effects are significant.

A line of systems work has studied UL→DL prediction in FDD systems by exploiting the shared physical environment. R2-F2 introduces a signal–processing pipeline that estimates a small set of frequency–invariant path tuples and transports them across bands before re–synthesizing the channel on the target frequency band [15]. FIRE instead uses a variational autoencoder to learn a low–dimensional latent representation of the propagation environment from uplink CSI and decodes it to downlink CSI. It achieves accurate cross–band MIMO operation without explicit path recovery [9]. Both systems work at the site level. They model one effective channel per user (or per small antenna set) and do not expose full element–space structure.

BeamMix differs in two ways. First, it learns a UL→DL mapping in full element space, enabling beam selection and MU–MIMO precoding on top of the predicted channels. Second, it parameterizes this mapping with sparse Gaussian primitives on the array manifold, rather than a purely latent path representation. This enables BeamMix to natively support element-space CSI prediction and related tasks, as well as RF spectrum synthesis.

Recent work has imported neural radiance fields (NeRFs) from graphics into wireless. NeRF$^2$ defines a neural RF radiance field that maps voxel position, view direction, and transmitter position to attenuation and phase, and uses volumetric ray–marching [23]. It demonstrates that a continuous MLP–based volumetric field can capture multipath structure well enough to support tasks like RF spectrum synthesis, indoor localization and frequency division duplexing CSI prediction. However, NeRF$^2$ operates primarily at a site-level, with one RF "view" per receiver. It uses a relatively resource-heavy MLP with stochastic sampling, and does not predict full element–space channels.

Subsequent work replaces the NeRF MLP with 3D Gaussian primitives to improve efficiency. WRF-GS and WRF-GS+ reconstruct a wireless radiation field using 3DGS techniques. They represent the scene with Gaussian virtual transmitters whose learned attributes are rendered into spatial spectra and used for other tasks like RF spectrum synthesis and CSI prediction [17, 18]. GSRF further adopts complex–valued 3D Gaussians with Fourier–Legendre radiance, orthographic splatting, and complex–valued ray tracing to synthesize spatial spectra efficiently, and reports gains in data efficiency over NeRF$^2$ and WRF–GS+ [21].

BeamMix is complementary to these wireless radiance field methods. Similar to WRF–GS, WRF–GS+, and GSRF, BeamMix uses Gaussian primitives and splatting ideas. Unlike them, it parameterizes the channel directly on the BS array and is conditioned on uplink CSI rather than on 3D TX/RX positions and explicit scene geometry. Our representation is thus array–centric and task–centric. The same Gaussian mixture supports both RF spectrum synthesis and UL→DL element–space CSI prediction without requiring map data of the environment.

Our primitive design is inspired by classical splatting and its modern 3D Gaussian approaches in graphics. EWA splatting formulates volume and surface rendering as a resampling problem and uses elliptical Gaussian reconstruction kernels combined with an anti–aliasing low–pass filter. This produces the EWA footprint with analytic projection to screen space [24]. 3DGS extends this idea to real–time radiance fields: scenes are represented by anisotropic 3D Gaussians with learned opacities and spherical-harmonic colors, and a visibility–aware, tile–based rasterizer splats them into images at 1080p in real time [8].

BeamMix borrows both the elliptical footprint idea and the few anisotropic primitives philosophy, but applies them in a novel domain. Instead of splatting onto a camera image plane, we splat onto the array manifold to produce complex channel coefficients across antennas and tones. Our EWA–style footprint controls spatial anti–aliasing on the array, and the mixture weights are modulated by a learned UL–conditioned gate and frequency–dependent amplitudes. In summary, prior work has explored (i) end–to–end UL→DL site-level prediction for FDD (R2-F2, FIRE), (ii) NeRF–style volumetric RF radiance fields and their 3DGS adaptations for spectrum and CSI synthesis (NeRF$^2$, WRF-GS, WRF-GS+, GSRF), and (iii) classical EWA/3DGS splatting in the graphics domain. BeamMix combines and extends beyond these approaches. We introduce a physics–guided mixture–of–Gaussian experts that lives directly in element space. It is conditioned on uplink CSI, and supports both RF spectrum synthesis and UL→DL CSI prediction. Our evaluations show that this representation outperforms wireless radiance field methods on site-level CSI prediction and RF spectrum synthesis tasks. Simultaneously, BeamMix performs near-perfect beam selection accuracy and MU-MIMO on real-world many-antenna measurements. We now describe the BeamMix architecture in detail.

## 3 Methodology

This section discusses the BeamMix approach. It models the DL channel as a set of Gaussian primitives that are splatted directly onto the antenna array. Each primitive induces a smooth spatial footprint across the array elements, so nearby antennas see similar channel responses. A sparse MoE gate selects the few primitives relevant to each observation, producing element-space CSI that schedulers and precoders can use immediately. Our method is a shift from recent wireless radiance field models, which first learn a 3D RF field and then perform a rendering step to recover CSI or spatial spectra.

### 3.1 Channel Modeling Background

The DL MIMO channel between a BS and a user is the coherent superposition of MPCs. Classical GSCMs represent this channel as a sum over clusters and rays, where each ray carries a complex-valued gain, a delay, and an angle of departure/arrival [1, 12]. For a BS array with $M$ antennas and a narrowband carrier, the DL element-space channel vector takes the form

$$\mathbf{h}_{\mathrm{dl}}(f) \;=\; \sum_{c=1}^{C}\sum_{r=1}^{R_c} \alpha_{c,r}(f)\, \mathbf{a}\big(\theta_{c,r},\phi_{c,r}\big)\, \mathrm{e}^{-j2\pi f \tau_{c,r}}, \tag{1}$$

where $C$ is the number of clusters, $R_c$ is the number of rays in cluster $c$, $\alpha_{c,r}(f)$ is a complex gain (including path-loss and phase), $\mathbf{a}(\theta,\phi) \in \mathbb{C}^M$ is the array steering vector at azimuth/elevation $(\theta,\phi)$, and $\tau_{c,r}$ is the propagation delay. In practice, only a handful of MPCs carry most of the power, corresponding to strong LOS and a few dominant reflections. The estimated element-space UL CSI is fed into a small MoE gate, which produces primitive weights indicating how strongly each Gaussian primitive contributes to the downlink channel.

Eq. (1) suggests a natural parametrization. The channel can be decomposed into spatially coherent primitives, each associated with paths at similar angles and delays. Prior wireless radiance field methods such as NeRF$^2$ and WRF-GS+ embed this idea into a continuous RF field defined over all points and directions in 3D space, and reconstruct CSI by integrating through that field. In contrast, BeamMix collapses the representation down to the array manifold. Each Gaussian primitive serves as a spatially localized mixture component, and its contribution is splatted directly onto the antenna positions via Eq. (3), producing a decision-centric model of $\mathbf{h}_{\mathrm{dl}}$. Fig. 2 visualizes this decomposition for three example primitives. Each primitive has a smooth footprint across the array, and the full DL channel is obtained by summing such primitive responses. In orthogonal frequency-division multiplexing (OFDM) systems we observe $\mathbf{h}_{\mathrm{dl}}(f_k)$ on a discrete grid of subcarriers $f_k$, yielding a complex-valued tensor of size (tones × antennas). Prior wireless radiance field methods such as NeRF$^2$ [23] and WRF-GS+ [18] represent the environment as a volumetric RF field and reconstruct CSI by ray-marching through that field. This scene-centric formulation is indirect for scheduling, since the element-space channel $\mathbf{h}_{\mathrm{dl}}$ only appears after a separate rendering step. We denote the DL CSI across tones and antennas by $\mathbf{H}_{\mathrm{dl}} \in \mathbb{C}^{K_{\mathrm{dl}}\times M}$, where $[\mathbf{H}_{\mathrm{dl}}]_{k,m}$ is the DL channel from the BS antenna $m$ on subcarrier $f_k$.

BeamMix takes the opposite stance. Instead of first reconstructing the scene and then rendering the channel, we directly model the channel as a sparse, physics-guided mixture of Gaussian primitives. The primitives are inspired by clustered MPCs in GSCMs, and their contributions are splatted onto the array manifold in closed form.

### 3.2 Physics-Guided Gaussian Primitives

We parameterize the channel on the array as a finite set of $G$ Gaussian primitives, indexed by $g \in \{1,\ldots,G\}$. Each primitive is a learnable spatial component parameterized by:

- a 3D center $\boldsymbol{\mu}_g \in \mathbb{R}^3$;
- an anisotropic covariance encoded as a diagonal scale vector $\boldsymbol{\sigma}_g \in \mathbb{R}^3_+$ and a rotation matrix $\mathbf{R}_g \in \mathrm{SO}(3)$;
- a preferred departure direction $\mathbf{u}_g \in \mathbb{S}^2$ (unit sphere);
- a propagation delay parameter $\tau_g$ and phase offset $\phi_g$;
- an amplitude parameter $a_g$.

The covariance and rotation define an anisotropic Gaussian density in a local coordinate frame,

$$p_g(\mathbf{x}) \;\propto\; \exp\!\Big(-\tfrac{1}{2}\big\|\Sigma_g^{-1/2}\mathbf{R}_g^{\top}(\mathbf{x}-\boldsymbol{\mu}_g)\big\|_2^2\Big), \tag{2}$$

where $\Sigma_g = \mathrm{diag}(\boldsymbol{\sigma}_g^2)$. This defines a learned spatial component. The Gaussian density concentrates the primitive's influence near $\boldsymbol{\mu}_g$, and the direction $\mathbf{u}_g$ controls the phase pattern it induces across the array.

The BS antenna positions $\{\mathbf{r}_m\}_{m=1}^{M}$ are known. For a given primitive $g$ and antenna $m$, we define an EWA weight

$$w_{g,m} \;=\; \exp\!\Big(-\tfrac{1}{2}\big\|\Sigma_g^{-1/2}\mathbf{R}_g^{\top}(\mathbf{r}_m-\boldsymbol{\mu}_g)\big\|_2^2\Big). \tag{3}$$

This weight is highest when $\mathbf{r}_m$ lies near the Gaussian's footprint and decays smoothly away. In computer graphics, EWA splatting projects Gaussian kernels onto pixels for anti-aliased rendering [19, 20, 24]. In BeamMix, the same mechanism becomes a differentiable spatial aggregation over the array manifold: $w_{g,m}$ encodes how strongly primitive $g$ couples into antenna $m$. Each primitive acts as a spatially localized mixture component on the array.

In addition to spatial footprint, each primitive has a directional and frequency structure: $\phi_{g,m,k} = 2\pi f_k \tau_g \;+\; 2\pi\kappa\,(\mathbf{r}_m\cdot\mathbf{u}_g) \;+\; \phi_g$ and $A_{g,k} = \sigma(a_g)\,\sigma_f(f_k)$. $\phi_{g,m,k}$ is the total phase of primitive $g$ at antenna $m$ and subcarrier $k$, $f_k$ is the (normalized) frequency index, $\kappa$ is a learnable global phase-per-meter parameter (initialized to

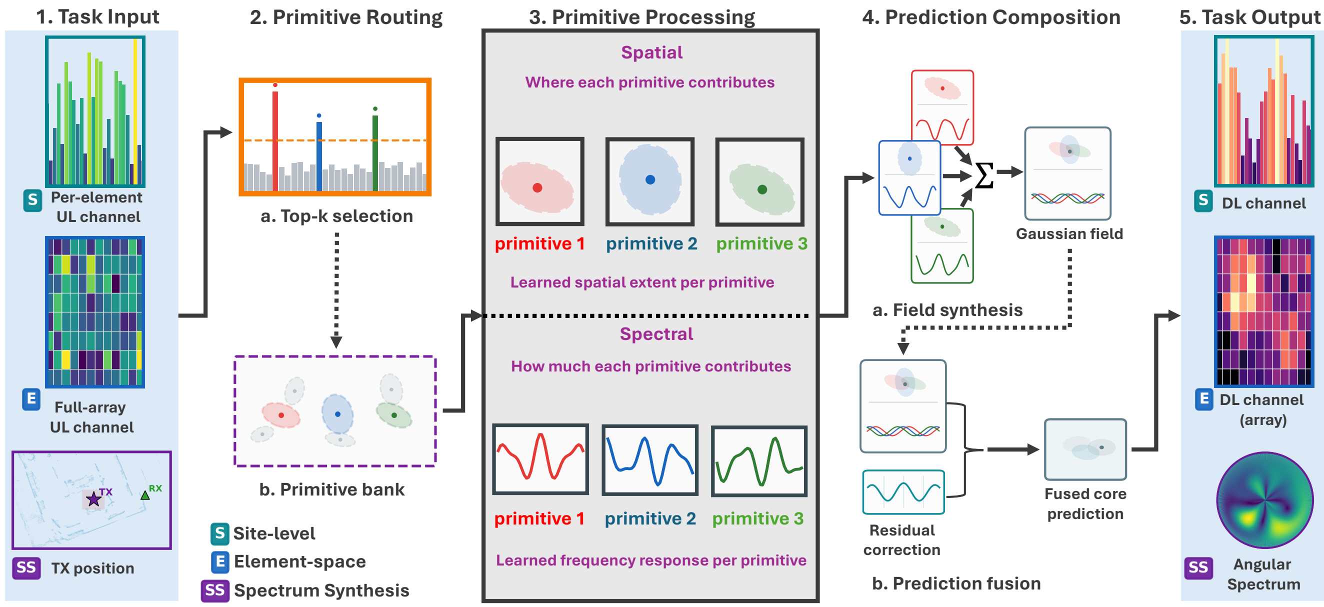


**Figure 2: Unified BeamMix architecture. BeamMix uses the same primitive-based core across site-level uplink-to-downlink CSI prediction (S), element-space uplink-to-downlink CSI prediction (E), and RF spectrum synthesis (SS). Task-specific inputs drive primitive routing, where top-$k$ gate weights select a sparse subset of a shared Gaussian primitive bank. Each selected primitive produces a task-domain spatial response and a frequency-dependent response; field synthesis aggregates these responses into the Gaussian field, which is fused with a lightweight residual-correction path to form the final prediction. The output is task-specific: site-level downlink CSI, element-space downlink CSI, or a synthesized spectrum image. Spatial responses are shown abstractly, with primitives evaluated on the array manifold for S/E and on room/query coordinates for SS.**

zero and optimized via backpropagation), $\sigma(\cdot)$ denotes a softplus or sigmoid nonlinearity, and $\sigma_f(\cdot)$ is a small MLP that models the smooth frequency response of the primitive. Together, these terms give each primitive a consistent direction, a delay, and a smooth frequency envelope. In graphics, Gaussian splats approximate surfaces and emitted radiance. In BeamMix, they serve as spatially localized mixture components on the array, encoding the directional and frequency structure of the channel rather than scene geometry and appearance. The raw contribution of the Gaussian field to the DL channel is then

$$\left[\mathbf{H}_{\mathrm{prop}}\right]_{k,m} = \sum_{g=1}^{G} w_{g,m}\, A_{g,k}\, e^{j\phi_{g,m,k}}. \tag{4}$$

This is the core of our physics-guided field. A finite mixture of anisotropic Gaussian primitives with direction, delay, and frequency structure, splatted onto the array.

In a many-antenna array, different signal components naturally couple to different subsets of elements. Because each BS element sits at a different location $\mathbf{r}_m$ on the array, a component with direction $\mathbf{u}_g$ and delay $\tau_g$ induces a distinct complex gain at each antenna, $h_g[m,k] \propto \exp(j2\pi f_k \tau_g)\, \exp(j2\pi\kappa\, (\mathbf{r}_m \cdot \mathbf{u}_g))$. So, even a single such component produces a characteristic phase progression and amplitude pattern across the array. In BeamMix, this spatial coupling is captured by the EWA weights $w_{g,m}$ and direction vectors $\mathbf{u}_g$. Each Gaussian primitive defines where on the array its contribution is large and how its phase varies across antennas. The observed DL CSI $\mathbf{H}_{\mathrm{dl}}$ is then modeled as the coherent sum of such primitive responses, which is why different learned primitives give rise to different lit-up patterns in both Fig. 2 (concept illustration) and Fig. 4 (real-world interpretability).

## 3.3 Sparse Mixture-of-Experts Routing

While Eq. (4) defines a global field, only a fraction of primitives meaningfully contribute for a given user and frame. This matches RF intuition, where a few dominant components carry most of the channel energy. Therefore, Eq. (4) is embedded into a MoE architecture.

For each sample, we use a lightweight encoder on the available UL CSI to produce gates over primitives. In all cases the gate is a single shared MLP, not $G$ separate networks. At the site level, where we predict a single CSI vector per BS element, the encoder takes the UL CSI $\mathbf{h}_{\mathrm{ul}} \in \mathbb{C}^{K_{\mathrm{ul}}}$, flattens its real and imaginary parts, and a single MLP outputs $G$ logits followed by a softmax distribution over $G$ experts,

$$\mathbf{g} = \mathrm{softmax}\big(\mathrm{MLP}_{\mathrm{site}}(\mathrm{ReIm}(\mathbf{h}_{\mathrm{ul}}))\big), \tag{5}$$

$$g_g \equiv \mathbf{g}[g] \in [0,1], \quad \sum_g g_g = 1. \tag{6}$$

At the element-space level, a single shared MLP processes each antenna's UL spectrum independently and outputs a per-antenna scalar gate in $[0,1]$ via sigmoid. This gate modulates the EWA weights: $w_{g,m} \leftarrow w_{g,m} \cdot g_m$, so that each primitive can be selectively activated on different parts of the array.

The gating weights modulate the primitive amplitudes,

$$\tilde{A}_{g,k} = g_g \, A_{g,k}, \tag{7}$$

and optionally sparsify via top-$k$ truncation, keeping only the largest gates per sample. Plugging $\tilde{A}_{g,k}$ into Eq. (4) produces a sparse sum in which a set of Gaussian experts are active.

### 3.4 Local UL→DL Mapper & Global Scalar Head

While the Gaussian field captures spatially structured multipath, it cannot represent the full per-antenna UL-to-DL mapping on its own. We therefore attach two data-driven heads that capture per-antenna effects that do not decompose into the Gaussian field's smooth spatial and frequency-coherent form. At both site level and element-space, we add a small MLP that maps the UL spectrum to the DL spectrum per antenna or per BS element:

$$\widehat{\mathbf{h}}_{\text{local}} = f_{\text{local}}(\mathbf{h}_{\text{ul}}), \tag{8}$$

implemented as a multi-layer real-valued MLP on concatenated real/imag parts. This head acts as a flexible local interpolator, taking in residual non-geometric effects. For notational simplicity we write $\mathbf{h}_{\text{ul}}$ as the UL feature vector provided to each head. At the site level this is the UL CSI vector for a single BS element. At the element-space level we apply $f_{\text{local}}$ per antenna using that antenna's UL spectrum, and we feed a flattened version of $\mathbf{H}_{\text{ul}}$ into the scalar head. A large fraction of the UL→DL mismatch can be attributed to a global complex gain per link: frequency-flat phase rotation and amplitude rescaling across all DL tones. We introduce a global scalar head $s(\mathbf{h}_{\text{ul}}) \in \mathbb{C}$ that predicts a single complex scalar per sample:

$$s(\mathbf{h}_{\text{ul}}) = s_{\text{real}} + j s_{\text{imag}}, \tag{9}$$

$$[s_{\text{real}}, s_{\text{imag}}] = \text{MLP}_{\text{scalar}}(\text{ReIm}(\mathbf{h}_{\text{ul}})). \tag{10}$$

This scalar is applied multiplicatively to the physics-guided output:

$$\widehat{\mathbf{h}}_{\text{core}} = (1 - \alpha - \beta)\, \mathbf{H}_{\text{prop}} + \beta\, \widehat{\mathbf{h}}_{\text{local}} + \alpha\, \mathbf{h}_{\text{ul,ref}}, \tag{11}$$

$$\widehat{\mathbf{h}}_{\text{dl}} = s(\mathbf{h}_{\text{ul}})\, \widehat{\mathbf{h}}_{\text{core}}, \tag{12}$$

where $\alpha, \beta$ are learnable blending coefficients and $\mathbf{h}_{\text{ul,ref}} = \mathbf{h}_{\text{ul}}[K_{\text{ul}}] \otimes \mathbf{1}_{K_{\text{dl}}}$ is the last UL subcarrier replicated across all DL subcarriers as a passthrough seed. This head is the learnable analog of the optimal complex alignment scalar commonly used to evaluate NMSE. It is predicted solely from UL and is available at test time.

### 3.5 Training Details

All BeamMix components are trained jointly in a single backward pass using an unaligned complex NMSE loss:

$$\mathcal{L}_{\text{NMSE}} = \frac{1}{B} \sum_{i=1}^{B} \frac{\|\widehat{\mathbf{h}}_{\text{dl}}^{(i)} - \mathbf{h}_{\text{dl}}^{(i)}\|_2^2}{\|\mathbf{h}_{\text{dl}}^{(i)}\|_2^2}. \tag{13}$$

We use AdamW (lr = 5e-4, weight decay = 5e-5) with gradient clipping (norm 1.0). Gaussian centers are initialized uniformly, the scalar head bias is initialized to $[1, 0]$ (unit gain, zero phase), and all other parameters follow default PyTorch initialization. Top-$k$ selection uses hard index selection; the retained gate values remain differentiable but the index selection itself is not. For spectrum synthesis, the MoE gate is conditioned on the TX 3D position with Fourier positional encoding rather than UL CSI. The UL encoder is a $2K_{\text{ul}}{\to}64{\to}1$ MLP (sigmoid), the frequency amplitude MLP is $1{\to}64{\to}64{\to}1$ (softplus), the local UL→DL head is $2K_{\text{ul}}{\to}128{\to}128{\to}2K_{\text{dl}}$ (linear), and the global scalar head is $2K_{\text{ul}}M{\to}256{\to}256{\to}2$ (linear). All hidden layers use ReLU. The MLP baseline uses the same global scalar head (Eq. (10)), ensuring a fair comparison.

Primitive scales are stored in log-space (exponentiated for positivity), rotations as normalized quaternions, amplitudes use softplus, and directions $\mathbf{u}_g$ are $\ell_2$-normalized each forward pass. Delays $\tau_g$ and phase offsets $\phi_g$ are unconstrained. Blend coefficients $\alpha, \beta$ are each clamped to $[0, 1]$.

## 4 MIMO CSI Evaluation: Site-Level

This section shows that the site-level variant of our model (BeamMix-Site) predicts DL CSI from UL CSI more accurately than NeRF$^2$, while also training and running inference much faster.

### 4.1 Motivation and Problem Setup

Neural wireless radiance and radiation field models have recently been proposed for MIMO CSI prediction. They represent the environment as a continuous 3D RF field and learn to map UL measurements to DL channels or spatial spectra [17, 18, 23]. These methods explore whether a model can exploit geometric regularity in the propagation environment to interpolate or extrapolate CSI across frequency.

We consider the standard MIMO UL→DL prediction setting used by NeRF$^2$ on the Argos MIMO dataset. For each time index ("frame") $n \in \{1, \dots, N\}$ and BS element (gateway) $s \in \{1, \dots, 8\}$, the measured complex-valued UL and DL CSI along the OFDM subcarriers are represented as vectors

$$\mathbf{h}_{\text{ul}}^{(n,s)} \in \mathbb{C}^{K_{\text{ul}}}, \qquad \mathbf{h}_{\text{dl}}^{(n,s)} \in \mathbb{C}^{K_{\text{dl}}},$$

with $K = K_{\text{ul}} + K_{\text{dl}} = 52$ subcarriers per frame. Following NeRF$^2$, the first $K_{\text{ul}} = 26$ tones are treated as UL and the remaining $K_{\text{dl}} = 26$ tones as DL, separated by guard bands in frequency.

The goal is to learn a function

$$\widehat{\mathbf{h}}_{\text{dl}}^{(n,s)} = f_\theta\big(\mathbf{h}_{\text{ul}}^{(n,s)}, \text{geometry}\big), \tag{14}$$

that maps the UL CSI (and any available geometric side information) for a given BS element and frame to its corresponding DL CSI vector. Performance is evaluated by comparing the predicted DL vector $\widehat{\mathbf{h}}_{\text{dl}}^{(n,s)}$ to the ground-truth $\mathbf{h}_{\text{dl}}^{(n,s)}$ using the frame-level SNR metric defined in Sec. 4.3.

### 4.2 Datasets and Protocol

We use the Argos channel study measurement dataset [14], following the MIMO CSI evaluation protocol of NeRF$^2$. The dataset provides a complex-valued tensor

$$\mathbf{H} \in \mathbb{C}^{N \times 8 \times K},$$

where $N$ is the number of time frames, the second dimension indexes the eight BS elements, and $K = 52$ is the number of OFDM

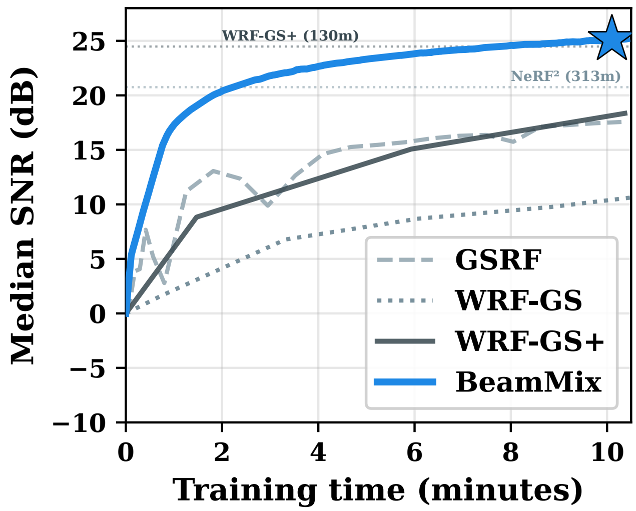


**Figure 3: Site-level training convergence on the NeRF² MIMO benchmark. Median SNR versus wall-clock training time. BeamMix reaches 25.1 dB in 10 minutes. Dashed lines mark the final converged SNR of WRF-GS+ (24.5 dB, 130 min) and NeRF² (20.7 dB, 313 min). Full results for all five methods are in Table 1.**

**Table 1: Site-level UL→DL CSI prediction on the NeRF² MIMO benchmark. All five methods share the same train/validation/test splits, global CSI amplitude normalization, and evaluation protocol. SNR is the per-frame metric in Eq. (18). NMSE is the normalized mean-squared error in linear scale. Train time is wall-clock minutes to the best validation checkpoint on one GPU. Latency is mean per-frame inference time on that GPU.**

| | SNR (dB) | | | | Train | Latency |
|---|---|---|---|---|---|---|
| Method | med | P10 | P90 | NMSE (lin.) | (min) | (ms/frame) |
| NeRF² [23] | 20.75 | 12.60 | 22.64 | 0.0084 | 312.7 | 15.50 |
| GSRF | 19.20 | 11.24 | 21.85 | 0.0120 | 14.2 | 0.22 |
| WRF-GS | 19.81 | 12.47 | 21.84 | 0.0104 | 80.6 | 4.40 |
| WRF-GS+ | 24.48 | 19.66 | 26.76 | 0.0036 | 130.2 | 6.78 |
| **BeamMix-Site (ours)** | **25.13** | **21.37** | **27.42** | **0.0031** | **10.1** | **0.08** |

subcarriers. As in NeRF², we apply a single global amplitude normalization across the full dataset:

$$\widetilde{\mathbf{H}}_{n,s,k} = \frac{\mathbf{H}_{n,s,k}}{\max_{n',s',k'} |\mathbf{H}_{n',s',k'}|}, \tag{15}$$

yielding a normalized CSI tensor $\widetilde{\mathbf{H}} \in \mathbb{C}^{N\times 8\times K}$.

We then partition the subcarriers into UL and DL halves as described above:

$$\mathbf{h}_{\mathrm{ul}}^{(n,s)} = \widetilde{\mathbf{H}}[n,s,1:26] \in \mathbb{C}^{K_{\mathrm{ul}}}, \tag{16}$$

$$\mathbf{h}_{\mathrm{dl}}^{(n,s)} = \widetilde{\mathbf{H}}[n,s,27:52] \in \mathbb{C}^{K_{\mathrm{dl}}}. \tag{17}$$

Each frame–antenna pair $(n,s)$ therefore yields one UL/DL sample. If $N_{\mathrm{train}}$ and $N_{\mathrm{test}}$ are the numbers of training and test frames, respectively, then the total number of training and test samples is $N_{\mathrm{train}} \times 8$ and $N_{\mathrm{test}} \times 8$. We follow the same frame-wise split as NeRF² and reserve 10% of the training frames as a validation set for checkpoint selection.

For the NeRF² baseline, we use the authors' publicly released MIMO CSI code [23]. It predicts DL CSI by ray-marching through a volumetric network conditioned on sampled 3D points, ray directions, and UL CSI. For the site-level comparison, we instantiate BeamMix on a single BS element at a time, matching NeRF²'s per-element formulation (Secs. 3.2–4.3). All primitive parameters and heads are shared across BS elements. Unlike NeRF², BeamMix-Site requires no explicit ray directions—it conditions only on UL CSI to route primitives and synthesize DL CSI.

## 4.3 Metrics

We evaluate both NeRF² and BeamMix-Site using NeRF²'s official CSI SNR metric. For a single frame $n$, let $\widehat{\mathbf{H}}_n \in \mathbb{C}^{8\times K_{\mathrm{dl}}}$, $\mathbf{H}_n \in \mathbb{C}^{8\times K_{\mathrm{dl}}}$ denote the predicted and ground-truth DL CSI matrices for all eight BS elements and the $K_{\mathrm{dl}}$ downlink subcarriers, respectively (both extracted from $\widetilde{\mathbf{H}}$ as described above). The per-frame SNR is then

$$\mathrm{SNR}_n = -10\log_{10}\frac{\|\widehat{\mathbf{H}}_n - \mathbf{H}_n\|_F^2}{\|\mathbf{H}_n\|_F^2}, \tag{18}$$

where $\|\cdot\|_F$ denotes the Frobenius norm across the $8\times K_{\mathrm{dl}}$ entries of the DL CSI matrix for frame $n$. We report aggregate SNR statistics over all test frames in terms of the median, 10th percentile, and 90th percentile.

We also report wall-clock training time to the best validation checkpoint, time-to-quality (TTQ) at target SNR thresholds, and per-frame inference latency. All methods are trained and tested on a single NVIDIA RTX 3090 (24 GB VRAM).

## 4.4 Site-Level Results on the NeRF² Benchmark

Fig. 3 and Table 1 summarize the site-level UL→DL CSI prediction performance on the NeRF² MIMO (Argos) benchmark. All methods use the same frame-based train/validation/test splits and global CSI normalization described in Sec. 4.2, and all SNR values are computed using the per-frame metric in (18).

On the test split, Table 1 compares all five methods under the same protocol. NeRF² attains a median per-frame SNR of 20.75 dB, while GSRF and WRF-GS achieve 19.20 dB and 19.81 dB, respectively. WRF-GS+, which augments WRF-GS with additional supervision, is the strongest baseline at 24.48 dB median SNR and an NMSE of 0.0036. BeamMix-Site outperforms all baselines, achieving a median SNR of 25.13 dB (NMSE 0.0031), improving the lower tail to 21.37 dB at the 10th percentile and the upper tail to 27.42 dB at the 90th percentile. Compared to WRF-GS+, BeamMix-Site gains 0.65 dB in median SNR while training 12.9× faster (10.1 vs. 130.2 min) and running inference 84.8× faster (0.08 vs. 6.78 ms). The worst 10% of frames improve by nearly 9 dB relative to NeRF².

At inference time, BeamMix-Site processes the $N_{\mathrm{test}} = 800$-frame test set in 0.123 s, corresponding to an average latency of 0.15 ms per frame (about $6.5 \times 10^3$ frames/s). Under the same hardware and implementation settings, NeRF² requires 12.4 s to process the same test set, or 15.5 ms per frame. Overall, BeamMix-Site achieves

$\sim 24\times$ faster training and $\sim 100\times$ faster inference than NeRF$^2$, while simultaneously delivering higher SNR across the test distribution.

*Finding: At the site level, we can predict downlink CSI by modeling a set of array-anchored Gaussian primitives selected using the uplink CSI. This makes reconstructing a full 3D radio scene unnecessary.*

## 5 MIMO CSI Evaluation: Element Space

This section shows that BeamMix provides a decision-ready, interpretable channel model that is accurate enough to sit inside scheduling and beam management loops.

### 5.1 Motivation and Problem Setup

Existing wireless radiance field models for MIMO CSI prediction are primarily site-level. They predict a single complex channel per BS element/gateway given UL CSI and geometry [17, 18, 23]. This is sufficient for link-level studies on scalar SNR or spectral efficiency, but practical massive MIMO systems operate in element space. Schedulers and precoders consume per-antenna CSI over a full array to perform beam selection, interference nulling, and MU-MIMO user multiplexing. A site-level scalar channel cannot answer questions such as which beam in a codebook should be chosen or how much MU-MIMO gain is obtained with a specific precoder.

This element-space regime is the native operating point of BeamMix. To evaluate, we conduct a complementary experiment to the site-level benchmark. Per-antenna CSI is used from the Argos channel study [14], but now in its full-array form. Each sample consists of UL CSI $\mathbf{H}_{\text{ul}} \in \mathbb{C}^{K_{\text{ul}} \times M}$ and DL CSI $\mathbf{H}_{\text{dl}} \in \mathbb{C}^{K_{\text{dl}} \times M}$ for $M$ BS antennas, where $M = 96$ in a $12 \times 8$ uniform rectangular array (URA) and $K_{\text{ul}} = K_{\text{dl}} = 26$ tones. Frames are drawn from continuous mobility scenarios. We use both LOS and NLOS measurement sets, as well as 2.4 GHz and 5 GHz setups.

We follow Sec. 3.1 and split UL/DL either by interleaving (even tones as UL, odd as DL) or by halves. For each capture, we create a train/val/test split of frames (70%/10%/20%) and treat each (frame, user) pair as a sample. UL and DL are normalized by the average UL power per sample, matching our UL→DL NMSE objective.

### 5.2 Element-Space Models and Metric

We instantiate the element-space BeamMix model described in Sec. 3.2–3.4. A shared set of Gaussian primitives is splatted over the 3D array coordinates and phases, producing a complex-valued DL tensor $\widehat{\mathbf{H}}_{\text{prop}} \in \mathbb{C}^{K_{\text{dl}} \times M}$. As in the site-level case, we condition on UL CSI via a MoE gate and add a shared per-antenna UL→DL head. We also include the global scalar head $s(\mathbf{H}_{\text{ul}}) \in \mathbb{C}$, so that the final prediction is

$$\widehat{\mathbf{H}}_{\text{dl}} = s(\mathbf{H}_{\text{ul}})\,\widehat{\mathbf{H}}_{\text{core}}, \quad \widehat{\mathbf{H}}_{\text{core}} \in \mathbb{C}^{K_{\text{dl}} \times M}, \tag{19}$$

where $\widehat{\mathbf{H}}_{\text{core}}$ is the blend of Gaussian field, local UL→DL mapper, and a UL passthrough (cf. Sec. 3.4).

As a strong black-box baseline, we train a per-antenna UL→DL MLP with a global scalar head and no Gaussian primitives or geometry:

$$\widehat{\mathbf{H}}_{\text{dl}}^{\text{MLP}} = s_{\text{MLP}}(\mathbf{H}_{\text{ul}})\, f_{\text{local}}(\mathbf{H}_{\text{ul}}), \tag{20}$$

**Table 2: Element-space UL→DL CSI prediction and beam-selection accuracy on Argos captures with a DFT codebook. SNR$_{\text{el}}$ is Frobenius SNR over antennas and DL tones, reported as the median over test samples. Top-$k$ is the fraction of test samples where predicted and true CSI select the same beam.**

| Scenario | Method | SNR$_{\text{el}}$ (dB) | NMSE (lin.) | Top-1 | Top-3 |
|---|---|---|---|---|---|
| 2.4 GHz LOS | MLP | 26.8 | $2.09 \times 10^{-3}$ | 0.963 | 1.000 |
| | **BeamMix** | **26.9** | $\mathbf{2.04 \times 10^{-3}}$ | 0.960 | 1.000 |
| 2.4 GHz NLOS | MLP | 26.6 | $2.19 \times 10^{-3}$ | 0.954 | 1.000 |
| | **BeamMix** | **27.0** | $\mathbf{2.00 \times 10^{-3}}$ | 0.956 | 1.000 |
| 5 GHz LOS | MLP | 18.7 | $1.35 \times 10^{-2}$ | 0.932 | 1.000 |
| | **BeamMix** | **18.8** | $\mathbf{1.32 \times 10^{-2}}$ | 0.931 | 1.000 |

where $f_{\text{local}}$ is the shared per-antenna UL→DL head from Sec. 3.4 and $s_{\text{MLP}}$ is the same global scalar head used in our method, BeamMix. This MLP uses the exact same input normalization, train/val/test splits, and unaligned NMSE loss as BeamMix. It provides a fair comparison against a geometry-agnostic but high-capacity baseline.

Element-space NMSE (dB) is defined for a single (frame, user) as

$$\text{NMSE}_{\text{el}} = 10 \log_{10} \frac{\|\widehat{\mathbf{H}}_{\text{dl}} - \mathbf{H}_{\text{dl}}\|_F^2}{\|\mathbf{H}_{\text{dl}}\|_F^2}, \tag{21}$$

where $\|\cdot\|_F$ is the Frobenius norm over tones and antennas. We also report the corresponding SNR,

$$\text{SNR}_{\text{el}} = -\text{NMSE}_{\text{el}}, \tag{22}$$

and summarize the distribution over the test set by the median and 10th/90th percentiles. Both BeamMix and the MLP baseline are trained on unaligned NMSE.

### 5.3 Element-Space UL→DL Accuracy

Table 2 reports element-space SNR on the Argos captures. In all cases, BeamMix and the MLP baseline achieve strong UL→DL prediction quality, with median NMSE around −27 dB (SNR ≈ 27 dB) at 2.4 GHz and around −19 dB at 5 GHz. BeamMix consistently outperforms the MLP baseline by a small but repeatable margin in both LOS and NLOS scenarios.

At 2.4 GHz LOS, BeamMix attains a median SNR of 26.9 dB versus 26.8 dB for the MLP baseline, with slightly improved upper tails. The gap is more pronounced in the 2.4 GHz NLOS capture, where BeamMix reaches 27.0 dB median SNR and improves both the 10th and 90th percentiles by approximately 0.4–0.6 dB. At 5 GHz, both methods are challenged by the increased frequency selectivity and shorter wavelength. BeamMix still modestly improves the median SNR by 0.06 dB and the tails by 0.05–0.1 dB.

In absolute terms, these SNR levels are high: an element-space NMSE of −27 dB corresponds to an average prediction error that is only $\sim 0.2\%$ of the channel energy. This indicates that the UL→DL mapping is almost deterministic on these mobility traces, and both a geometry-aware model (BeamMix) and a strong per-antenna MLP can learn it accurately.

BeamMix's SNR improvement over the MLP is modest (0.1–0.4 dB). Its value in element-space is the spatially and frequency-coherent structure from the Gaussian field (which carries 74% of

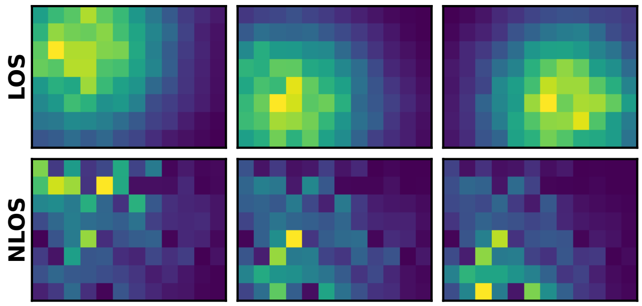


**Figure 4: Per-sample primitive footprints on the 12×8 URA at 2.4 GHz. Each heatmap shows the effective contribution of one primitive to a single test prediction, computed as the product of the per-antenna gate activation, the EWA spatial weight, and the primitive amplitude. Three primitives are shown for one LOS sample (top) and one NLOS sample (bottom). In LOS, primitives produce broad footprints covering approximately 40% of the array, consistent with a channel that varies smoothly across antennas. In NLOS, footprints are concentrated on 8–16% of the array, with sharp peaks and large suppressed regions, reflecting greater spatial variation when the direct path is blocked.**

output energy at 2.4 GHz), the decomposable mixture form, the unified architecture across tasks, and competitive downstream metrics confirming the predicted CSI is decision-ready.

## 5.4 Element-Space Evaluation: Beams and MU-MIMO

To move beyond aggregate NMSE and assess decision-relevant performance, we evaluate how the predicted element-space CSI affects (i) beam selection with a DFT codebook and (ii) two-user MU-MIMO precoding.

*5.4.1 Beam selection with a DFT codebook.* We adopt a separable DFT codebook $\mathbf{W} \in \mathbb{C}^{M\times B}$ for a $12 \times 8$ URA, where $B = M$ and each column corresponds to a phase-steered beam. For each (frame, user) test sample, we compute average beam gains over DL tones:

$$g_{\text{true}}[b] = \frac{1}{K_{\text{dl}}}\sum_{k=1}^{K_{\text{dl}}}\left|\mathbf{h}_{\text{dl}}[k]\mathbf{w}_b\right|^2, \quad g_{\text{pred}}[b] = \frac{1}{K_{\text{dl}}}\sum_{k=1}^{K_{\text{dl}}}\left|\widehat{\mathbf{h}}_{\text{dl}}[k]\mathbf{w}_b\right|^2, \tag{23}$$

and define the best beams

$$b^*_{\text{true}} = \arg\max_b g_{\text{true}}[b], \quad b^*_{\text{pred}} = \arg\max_b g_{\text{pred}}[b]. \tag{24}$$

We report top-1 and top-3 beam selection accuracy over all test samples. Table 2 shows results.

Across all three captures, both BeamMix and the MLP baseline achieve near-perfect beam selection accuracy. Top-1 accuracy ranges from 93%–96%, and top-3 accuracy is effectively 100%. In other words, almost every time, the beam that would be chosen from the codebook given BeamMix's predicted CSI is exactly the same beam (or within the top-3 candidates) that would be chosen from the true CSI.

**Table 3: Component ablation (2.4 GHz LOS). GF% is the Gaussian field's share of output energy.**

| Configuration | NMSE (dB) | Beam T1 | ZF (dB) | GF% |
|---|---|---|---|---|
| Default ($G$=6000, $k$=128) | −26.9 | 96.0% | 19.5 | 74 |
| No gate (uniform) | −27.0 | 96.0% | 19.5 | 56 |
| No EWA ($w$=1) | −27.2 | 96.1% | 19.5 | 35 |
| No local MLP | −0.6 | 0.4% | 8.5 | 42 |
| No scalar head | −27.4 | 96.0% | 19.5 | 6 |

*5.4.2 MU-MIMO precoding with predicted CSI.* We next study two-user MU-MIMO using standard ZF and MRT precoders designed either from true CSI or from predicted CSI (BeamMix or the MLP). For each test frame and DL tone we form a 2-user channel matrix and compute per-user SINR, summarizing by the median and 10th/90th percentiles.

Across all three captures, both BeamMix and the MLP track oracle (true-CSI) performance almost exactly. For example, on the 2.4 GHz LOS capture, ZF with true CSI yields median SINR $\approx$ 19.7 dB, while ZF with BeamMix- or MLP-predicted CSI differs by at most 0.01 dB in both median and tails. At 5 GHz, the median loss relative to oracle is at most 0.1–0.2 dB for both precoders. This indicates that, for the evaluated setups, predicted element-space CSI is effectively decision-perfect for MU-MIMO.

*Finding: By modeling the full complex signal at each antenna, BeamMix preserves the fine phase relationships needed to pick beams and separate users. This is reflected in its near-perfect beam selection accuracy and MU-MIMO performance.*

## 5.5 Ablation Study

Table 3 ablates key BeamMix components at 2.4 GHz LOS. The Gaussian field fraction GF% = $(1-\alpha-\beta) \times 100$ is the learned share of the output allocated to the Gaussian field after training, where $\alpha$ and $\beta$ are the UL-passthrough and local-MLP blend weights from Sec. 3.4. In the default configuration, the Gaussian field carries 74% of the output energy. Each primitive contributes jointly across antennas and subcarriers via its learned delay $\tau_g$, direction $\mathbf{u}_g$, and frequency response to encode spatially and frequency-coherent structure that a per-antenna MLP does not explicitly model. The local MLP then refines this base prediction with per-antenna corrections, and both paths are trained jointly through the shared loss. Removing the local MLP collapses NMSE to −0.6 dB and beam accuracy to 0.4%, because the Gaussian field's smooth parameterization cannot capture all per-antenna effects on its own. An MLP-only baseline (Table 2) matches accuracy but lacks explicit

**Table 4: RENEW FDD in-band results ($8 \times 8$ uniform planar array (UPA), 64 antennas). Both methods use identical training protocols. Only array coordinates differ from the Argos experiments.**

| Method | SNR med (dB) | NMSE (lin.) | Beam top-1 | ZF SINR (dB) |
|---|---|---|---|---|
| MLP | 29.17 | $1.21 \times 10^{-3}$ | 97.0% | 50.7 |
| BeamMix | 29.03 | $1.25 \times 10^{-3}$ | 96.4% | 49.4 |

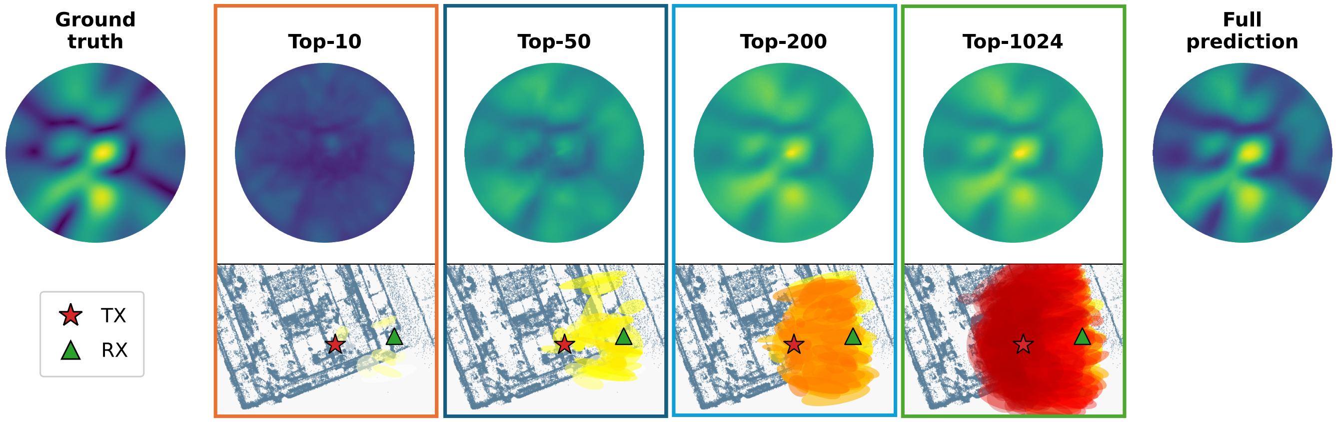


**Figure 5: Progressive primitive build-up and spatial locations for spectrum synthesis. Top: as more primitives are included, the predicted angular spectrum progressively recovers finer structure in the ground truth. The full prediction adds the local MLP head. Bottom: where the corresponding top-K primitives are located in the room, overlaid on the LiDAR point cloud.**

delay/direction structure and cannot decompose predictions. All other components can be individually removed without degrading accuracy, because the blend mechanism redistributes energy between the Gaussian field and local MLP. The GF% reveals this redistribution where removing the gate shifts GF from 74% to 56%, and removing EWA shifts it to 35%. At 5 GHz, GF% drops to 27% in the default configuration. The shorter wavelength increases the antenna spacing in wavelengths, producing finer spatial variation that the smooth Gaussian footprints capture less completely, so the optimizer allocates more weight to the local MLP. Reducing $G$ from 6000 to 128 or $k$ from 128 to 16 has negligible impact at either frequency.

## 5.6 BeamMix Interpretability

Beyond quantitative metrics, BeamMix produces a structured and interpretable decomposition of the channel. Its Gaussian primitives, EWA footprints, and UL-conditioned gates provide a channel representation that can be inspected component by component, which is not possible with a purely black-box MLP. The primitives are spatially localized mixture components on the array.

Fig. 4 visualizes the per-sample primitive footprints on the $12 \times 8$ URA. Each heatmap shows the effective contribution of one primitive, computed as the product of the per-antenna gate activation, the EWA spatial weight $w_{g,m}$ (Sec. 3.2), and the primitive amplitude. Three top-ranked primitives are shown for one LOS sample (top row) and one NLOS sample (bottom row).

In LOS (top row), primitives produce broad footprints covering approximately 40% of the array, consistent with a channel that varies smoothly across antennas. In NLOS (bottom row), footprints are concentrated on 8–16% of the array with sharp peaks and large suppressed regions, reflecting greater spatial variation when the direct path is blocked.

The MLP baseline matches BeamMix numerically but cannot decompose its prediction into inspectable components. BeamMix exposes which primitives are active and where they sit on the array, enabling per-component reasoning that dense weight matrices do not provide.

*Finding: BeamMix predicts the channel and exposes which array regions and primitives are responsible, with explicit spatial footprints and gate weights for each active component.*

## 5.7 Generalization: RENEW FDD

To test generalization, we apply BeamMix to the RENEW FDD dataset (Rice University, $8 \times 8$ UPA, 64 antennas) using identical hyperparameters as Argos [5, 22]. The array coordinates are changed in this dataset. Table 4 reports in-band results.

Both methods achieve strong performance (SNR $\approx$ 29 dB, beam accuracy $\geq$ 96%). The MLP slightly leads by 0.14 dB SNR and 0.6% beam accuracy. BeamMix generalizes without retuning and retains its component-wise decomposition.

# 6 Spectrum Synthesis

This section provides a head-to-head comparison with existing wireless radiance field methods, showing BeamMix synthesizes spatial spectra more faithfully than competing methods.

## 6.1 RF Spectrum Synthesis

This section compares BeamMix to NeRF$^2$, WRF-GS, WRF-GS+, and GSRF on an RF spectrum dataset, following the original protocol. For each transmitter position $\mathbf{t}$ and fixed RX pose $\mathbf{G}$, the dataset provides a ground-truth spatial spectrum $S(\beta, \alpha) \in [0, 1]^{90\times 360}$ giving relative received power over elevation $\beta$ and azimuth $\alpha$, computed from real $4 \times 4$ array measurements as in NeRF$^2$.

BeamMix queries the same hemispherical grid using its Gaussian field, yielding a predicted spectrum $\widehat{S}$. We evaluate $\widehat{S}$ against $S$ using standard image-quality metrics (MAE, SSIM, PSNR, NMSE) and report per-spectrum render time.

## 6.2 Experimental Setup

In the RFID spectrum dataset, the RX array (gateway) is fixed in position, and the mobile TX locations are split into train/test sets.

| | Ground Truth | BeamMix (ours) | WRF-GS + | NeRF$^2$ | WRF-GS | GSRF |
|---|---|---|---|---|---|---|
| SSIM | N/A | 0.9267 | 0.8795 | 0.7872 | 0.7551 | 0.7194 |
| PSNR (dB) | N/A | 25.880 | 24.324 | 22.369 | 20.492 | 18.258 |

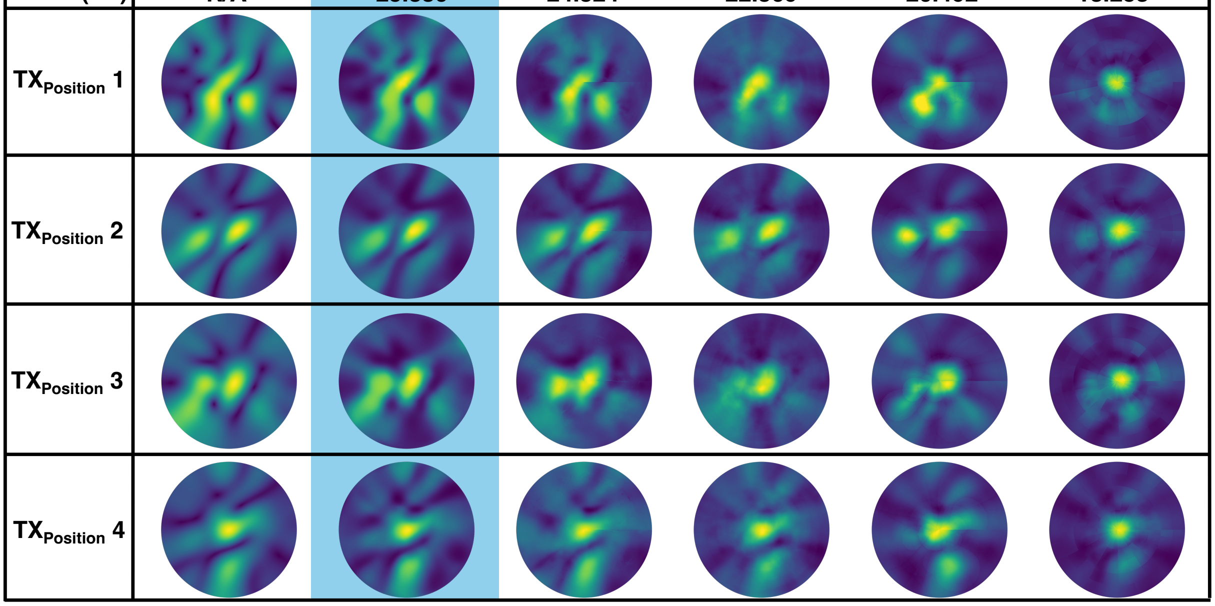


**Figure 6: Comparison of spatial spectrum synthesis images. The ground truth spatial spectrum shows the strength of the RF signal from a particular direction. This multipath profile is from the RX perspective ($4 \times 4$ antenna array) as the TX moves around the room. Selected TX positions from the held-out test set are shown [$\mathrm{TX_{Position}}1$, $\mathrm{TX_{Position}}4$].**

Each TX has a ground-truth $90 \times 360$ spatial spectrum computed from real array measurements [23]. All models are queried on the same hemispherical grid and evaluated against the corresponding ground truth.

We use the NeRF$^2$ RFID spectrum dataset and the authors' published code for spectrum synthesis [23]. In addition to the NeRF$^2$ comparison, BeamMix is evaluated against three 3DGS-inspired methods: WRF-GS, WRF-GS+, and GSRF [17, 18, 21]. We use each method's publicly released code for this experiment. Training and testing was done for each method on a single NVIDIA RTX 3090 with 24 GB of VRAM. All methods are trained and evaluated on the same set of training transmitter positions and tested on the same held-out TX positions. All produce spectra on the same $90 \times 360$ hemispherical grid.

NeRF$^2$ learns an implicit RF radiance field with a multilayer perceptron that takes (voxel position, view direction, TX position) features (with positional encodings) and predicts attenuation and phase terms [23]. A volumetric integration along each RX ray yields the predicted scalar spectrum value per angle bin. WRF-GS adopts an explicit field parameterization with anisotropic 3D Gaussian primitives for RF, rendering spectra via differentiable splatting/integration. Attenuation/phase are learned per primitive and composed along rays. WRF-GS+ refines the Gaussian-based field parameterization and rendering pipeline (e.g., improved primitive attributes and training) to further boost spectrum fidelity while keeping inference practical. GSRF extends 3D Gaussian Splatting to the RF domain by attaching complex-valued radiance and transmittance to each Gaussian and expanding the directional radiance in a Fourier–Legendre basis. An orthographic splatting step onto a spherical receive surface plus a complex-valued ray tracer aggregates amplitude and phase along each ray. This targets phase-aware RF synthesis in a MLP-free manner with high efficiency in training and inference [21]. BeamMix (ours) uses a physics-guided Gaussian field with an analytic quadratic form for per-ray distance metrics and tiled top-$k$ selection over primitives. Only the Gaussian primitives that align with a given ray are aggregated. A small local head modulates per-ray predictions to capture residual diffuse energy.

## 6.3 Results

Table 5 summarizes spectrum synthesis accuracy. BeamMix achieves the strongest median SSIM among all baselines, and also attains the best PSNR and the most negative NMSE. NeRF$^2$ integrates learned attenuation/phase along each ray without an explicit two–hop, TX-aware prior or enforced sparsity. It tends to fill in nearby directions and average over clusters, softening peaks and slightly lifting the noise floor. WRF-GS and WRF-GS+ both show energy bleed across neighboring angles. While the WRF-GS+ result is much stronger than WRF-GS, both render a bit blurrier and less structurally faithful compared to BeamMix. GSRF underperforms compared to NeRF$^2$ and the 3DGS inspired methods. We observe GSRF produces spectra with correct main lobe placement but noticeably softer sidelobes, especially compared to BeamMix and WRF-GS+.

**Table 5: Test-set metrics for spectrum synthesis evaluation. Arrows indicate whether higher (↑) or lower (↓) is better.**

| Method | $SSIM_{med}$ ↑ | $SSIM_{mean}$ ↑ | MAE ↓ | PSNR (dB) ↑ | NMSE (dB) ↓ |
|---|---|---|---|---|---|
| GSRF | 0.7194 | 0.6964 | 0.1029 | 18.258 | -11.33 |
| WRF-GS | 0.7551 | 0.7397 | 0.0737 | 20.492 | -13.68 |
| $NeRF^2$ | 0.7872 | 0.7724 | 0.0706 | 22.369 | -16.80 |
| WRF-GS+ | 0.8795 | 0.8437 | 0.0550 | 24.324 | -17.40 |
| **BeamMix (ours)** | **0.9267** | **0.8921** | **0.0523** | **25.880** | **-19.58** |

Fig. 5 shows how the predicted spectrum progressively recovers finer structure as more primitives are included. First, BeamMix represents the scene as a set of oriented primitives and, for each receive direction, mixes only the few primitives that actually align with that ray (top-$K$ selection). This sparse, direction-selective composition suppresses crosstalk between unrelated paths, which keeps peaks sharp and the sidelobe floor low. In practice, that is what pushes up SSIM (local structure preserved) and PSNR (large errors at peaks avoided), and results in more negative NMSE (total energy concentrated at the correct angles). This is consistent with Table 5 and the visuals in Fig. 6.

Second, BeamMix includes a two-hop attenuation prior (TX⟶primitive and primitive⟶RX) with learnable but bounded exponents. This acts like a soft path-loss constraint. Energy decays with distance in a physically reasonable way rather than being spread arbitrarily across the hemisphere. The result is fewer smeared lobes and a lower background floor, which directly improves MAE and NMSE without sacrificing sharpness.

Third, the model gates primitives using the transmitter position. As the TX moves, only the reflectors that are geometrically plausible from that vantage point are activated or strengthened. This TX-conditioned visibility echoes classical geometry-based channel models and stabilizes lobe positions and widths across the test trajectory. SSIM is raised by keeping beam locations consistent with the ground truth.

Lastly, a small local head accounts for genuinely diffuse scatter that no single lobe explains. This prevents overconfident needle-like artifacts and helps match the residual texture of the spectrum without blurring the dominant beams. Together, these components enable BeamMix to place beam power aligned with physics. These choices encourage spectra that respect the sparse, geometry-driven nature of multipath: strong, well-placed beams, low background, and sensible power roll-off.

*Finding: BeamMix mixes only the primitives that align with each direction in the spatial spectrum, rather than diffusing energy over many nearby directions as 3D field models tend to do. This concentrates power on the dominant multipath directions.*

## 7 Limitations & Future Work

BeamMix is evaluated in three settings: (i) site-level UL⟶DL MIMO CSI prediction, (ii) element-space UL⟶DL CSI, and (iii) spatial spectrum synthesis. These span both scene-centric (spectrum) and array-centric (CSI) tasks on real-world measurements. However, a limited range of deployment scenarios is covered. All experiments are based on sub-6 GHz measurements. How well the same Gaussian-primitive MoE scales to mmWave bands or deployment scenarios like multi-cell remains an open question.

BeamMix fixes the number of Gaussian primitives and the top-$k$ sparsity level in advance, which may not be optimal across environments with very different multipath richness. Future work could couple BeamMix with explicit scene information (e.g., LiDAR point clouds) to initialize primitives on physical surfaces, or adopt adaptive densification mechanisms similar to those found in 3DGS.

On the RENEW FDD dataset (Sec. 5.7), neither BeamMix nor the MLP baseline can extrapolate across the 72 MHz UL–DL frequency gap from data alone. Bridging this gap likely requires explicit frequency-domain priors or physics-based interpolation.

Our element-space evaluation uses standard DFT codebooks and simple linear precoders (ZF and MRT). Integrating BeamMix into a full scheduling and link adaptation stack is left to future work.

## 8 Conclusion

This work introduced BeamMix, a map-free MoE architecture built from physics-aware 3D Gaussian primitives for wireless channel modeling. Unlike wireless radiance field methods that first learn a volumetric RF field and then render down to the quantities needed by schedulers, BeamMix operates directly in CSI space. Gaussian primitives splat complex-valued footprints onto the array manifold, and a UL-conditioned gate activates only those primitives whose parameters align with the observed UL channel. A local UL⟶DL mapper and a global complex scalar head capture residual non-geometric effects, producing a decision-ready representation that is well suited for beam selection and MU-MIMO precoding.

Across four benchmarks, BeamMix demonstrates the effectiveness of the array-centric Gaussian mixture approach. When predicting site-level MIMO CSI against five baselines, BeamMix-Site attains the highest median DL-CSI SNR (25.1 dB) while training in 10.1 minutes and running inference in 0.08 ms/frame. On MIMO element-space traces, the same primitive design extended to the full array achieves median UL⟶DL SNRs of about 27.0 dB at 2.4 GHz and about 18.8 dB at 5 GHz. Produced beam-selection decisions are almost always identical to those obtained from true CSI, with small and consistent SINR gaps for two-user MU-MIMO precoding. On the spectrum synthesis task, BeamMix achieves the strongest median SSIM and PSNR among $NeRF^2$, WRF-GS, WRF-GS+, and GSRF. With a SSIM of 0.93, PSNR of 25.9 dB, and NMSE of −19.6 dB, BeamMix indicates it has internalized the scene's angular multipath structure more faithfully than prior RF radiance-field models. On a second real-world dataset (RENEW FDD, 64-element 8×8 UPA), BeamMix generalizes without dataset-specific tuning, matching the MLP baseline within 0.2 dB SNR and 0.6% beam accuracy.

BeamMix decomposes each prediction into spatially localized mixture components with explicit footprints and gate weights. Overall, mixtures of physics-guided 3D Gaussian primitives provide an effective middle ground between classical channel models and wireless radiance fields.

## Acknowledgments

This work is supported by the National Science Foundation (NSF) grant OAC-2512931.

# References


[1] 3rd Generation Partnership Project (3GPP). 2020. *Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 16.1.0 Release 16).* ETSI Technical Report TR 138 901 V16.1.0. ETSI, Sophia Antipolis, France. https://www.etsi.org/deliver/etsi_tr/138900_138999/138901/16.01.00_60/tr_138901v160100p.pdf ETSI TR 138 901 V16.1.0 (2020-11).

[2] Amir Mehdi Ahmadian Tehrani. 2020. *Precoding Algorithms for Massive Multiple-Input, Multiple-Output Wireless Systems.* Ph. D. Dissertation. Technische Universität München. https://mediatum.ub.tum.de/1524282

[3] Emil Björnson, Jakob Hoydis, and Luca Sanguinetti. 2017. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. *Foundations and Trends® in Signal Processing* 11, 3-4 (2017), 154–655. doi:10.1561/2000000093

[4] Emil Björnson, Luca Sanguinetti, Henk Wymeersch, Jakob Hoydis, and Thomas L. Marzetta. 2019. Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays. *Digital Signal Processing* 94 (2019), 3–20. doi:10.1016/j.dsp.2019.06.007 Special Issue on Source Localization in Massive MIMO.

[5] Xu Du and Ashutosh Sabharwal. 2021. Massive MIMO Channels with Inter-User Angle Correlation: Open-Access Dataset, Analysis and Measurement-Based Validation. *IEEE Transactions on Vehicular Technology* (2021), 1–1. doi:10.1109/TVT.2021.3131606

[6] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. 1991. Adaptive Mixtures of Local Experts. *Neural Computation* 3, 1 (03 1991), 79–87. arXiv:https://direct.mit.edu/neco/article-pdf/3/1/79/812104/neco.1991.3.1.79.pdf doi:10.1162/neco.1991.3.1.79

[7] Michael I. Jordan and Robert A. Jacobs. 1994. Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation* 6, 2 (03 1994), 181–214. arXiv:https://direct.mit.edu/neco/article-pdf/6/2/181/812708/neco.1994.6.2.181.pdf doi:10.1162/neco.1994.6.2.181

[8] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42, 4 (2023), 139–1.

[9] Zikun Liu, Gagandeep Singh, Chenren Xu, and Deepak Vasisht. 2021. FIRE: enabling reciprocity for FDD MIMO systems. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking* (New Orleans, Louisiana) *(MobiCom '21)*. Association for Computing Machinery, New York, NY, USA, 628–641. doi:10.1145/3447993.3483275

[10] Thomas L. Marzetta, Erik G. Larsson, Hong Yang, and Hien Quoc Ngo. 2016. *Fundamentals of Massive MIMO.* Cambridge University Press.

[11] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.

[12] A.F. Molisch. 2010. *Wireless Communications.* Wiley. https://books.google.com/books?id=5SoDDQAAQBAJ

[13] Noam Shazeer, *Azalia Mirhoseini, *Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. 2017. Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. In *International Conference on Learning Representations.* https://openreview.net/forum?id=B1ckMDqlg

[14] Clayton Shepard, Jian Ding, Ryan E. Guerra, and Lin Zhong. 2016. Understanding real many-antenna MU-MIMO channels. In *2016 50th Asilomar Conference on Signals, Systems and Computers.* 461–467. doi:10.1109/ACSSC.2016.7869082

[15] Deepak Vasisht, Swarun Kumar, Hariharan Rahul, and Dina Katabi. 2016. Eliminating Channel Feedback in Next-Generation Cellular Networks. In *Proceedings of the 2016 ACM SIGCOMM Conference* (Florianopolis, Brazil) *(SIGCOMM '16)*. Association for Computing Machinery, New York, NY, USA, 398–411. doi:10.1145/2934872.2934895

[16] Satyavrat Wagle, Akshay Malhotra, Shahab Hamidi-Rad, Aditya Sant, David J. Love, and Christopher G. Brinton. 2025. Physics-based generative models for geometrically consistent and interpretable wireless channel synthesis. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence* (Montreal, Canada) *(IJCAI '25)*. Article 1043, 9 pages. doi:10.24963/ijcai.2025/1043

[17] Chaozheng Wen, Jingwen Tong, Yingdong Hu, Zehong Lin, and Jun Zhang. 2025. Neural Representation for Wireless Radiation Field Reconstruction: A 3D Gaussian Splatting Approach. *IEEE Transactions on Wireless Communications* (2025), 1–1. doi:10.1109/TWC.2025.3631663

[18] Chaozheng Wen, Jingwen Tong, Yingdong Hu, Zehong Lin, and Jun Zhang. 2025. Wrf-gs: Wireless radiation field reconstruction with 3d gaussian splatting. In *IEEE INFOCOM 2025-IEEE Conference on Computer Communications.* IEEE, 1–10.

[19] Lee Westover. 1990. Footprint evaluation for volume rendering. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques* (Dallas, TX, USA) *(SIGGRAPH '90)*. Association for Computing Machinery, New York, NY, USA, 367–376. doi:10.1145/97879.97919

[20] Lee Westover. 1990. Footprint evaluation for volume rendering. *SIGGRAPH Comput. Graph.* 24, 4 (Sept. 1990), 367–376. doi:10.1145/97880.97919

[21] Kang Yang, Gaofeng Dong, Sijie Ji, Wan Du, and Mani Srivastava. 2025. GSRF: Complex-Valued 3D Gaussian Splatting for Efficient Radio-Frequency Data Synthesis. In *Proceedings of the 39th Conference on Neural Information Processing Systems (NeurIPS).*

[22] Xing Zhang, Lin Zhong, and Ashutosh Sabharwal. 2018. Directional Training for FDD Massive MIMO. *IEEE Transactions on Wireless Communications* 17, 8 (2018), 5183–5197. doi:10.1109/TWC.2018.2838600

[23] Xiaopeng Zhao, Zhenlin An, Qingrui Pan, and Lei Yang. 2023. Nerf2: Neural radio-frequency radiance fields. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking.* 1–15.

[24] Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. 2002. EWA Splatting. *IEEE Transactions on Visualization and Computer Graphics* 8, 3 (July 2002), 223–238. doi:10.1109/TVCG.2002.1021576