

6G Twin: Hybrid Gaussian Radio Fields for Channel Estimation and Non-Linear Precoder Design for Radio Access Networks

Muhammad Ahmed Mohsin, *Graduate Member, IEEE*, Muhammad Umer, Ahsan Bilal, Muhammad Ali Jamshed, *Senior Member, IEEE*, Dean F. Hougen, John M. Cioffi, *IEEE Fellow*

Abstract—This work introduces *6G Twin*, the *first end-to-end* artificial intelligence (AI)-native radio access network (RAN) design that unifies (i) neural Gaussian Radio Fields (GRF) for compressed channel state information (CSI) acquisition, (ii) continual channel prediction with handover persistence, and (iii) an energy-optimal nonlinear precoder (*minPMAC*). GRF replaces dense pilots with a sparse Gaussian field, cutting pilot overhead by $\sim 100\times$ while delivering ~ 1.1 ms inference and ≤ 2 min on-site training, thus enabling millisecond-scale closed-loop operation. A replay-driven continual learner sustains accuracy under mobility and cell transitions, improving channel normalized mean square error (NMSE) by >10 dB over frozen predictors and an additional 2-5 dB over uniform replay, thus stabilizing performance across UMi/UMa handovers. Finally, *minPMAC* solves a convex, order-free MAC precoder design that recovers the globally optimal order from Broadcast Channel (BC) duals and minimizes transmit energy subject to minimum-rate guarantees, achieving $4\text{--}10\times$ lower energy (scenario dependent) with *monotonically increasing* bits/J as SNR grows; this translates to up to $5\times$ higher data rate at comparable power or the same rates at substantially lower power. Together, these components form a practical, GPU-ready framework that attains real-time CSI, robust tracking in dynamic networks with efficient handovers, and state-of-the-art throughput-energy tradeoffs under 3GPP-style settings.

Index Terms—6G, digital twin, channel state information (CSI), Gaussian Radio Fields (GRF), continual learning, loss-aware reservoir sampling (LARS), handover, massive MIMO, nonlinear precoding, multiple access channel (MAC) polynomial, successive interference cancellation (SIC), energy efficiency, MA channel, Broadcast channel, 3GPP NR, CUDA

I. INTRODUCTION

Emerging 6G Networks aim for high spectral efficiency and low latency to support applications like virtual reality (VR), extended reality (XR), and internet of everything (IoE) with immersive connectivity (data rate ≥ 500 mbps). Achieving these goals hinges on massive MIMO systems that form highly-directional beams. These in turn require fast, precise beam forming via transmitter precoding¹. However, precoding critically depends on timely channel state information (CSI) at the transmitter, and the conventional pilot based CSI renders a feedback delay time of more than 8 ms causing a bottleneck to MIMO beam forming. AI for CSI prediction in [1] summarizes

CSI overhead and channel aging issues in modern massive MIMO (mMIMO). Accurate CSI enables the transmitters to adapt their waveforms, power levels, and spatial precoding to channel conditions to achieve required data rates and link reliability.

The mMIMO spatial multiplexing boosts data rates; with N antennas at each end, the theoretical maximum achievable rates scales by a factor of $\min(N_t, N_r, \rho_h)$ where N_t, N_r, ρ_h are the number of transmitter antennas, receiver antennas and channel rank respectively. Although this increases the data rate, the pilot overhead for CSI estimation in current systems can be as high as 20%- 40% (on some respective tones) of each coherence interval. For instance, with a coherence time of 1 ms and a pilot duration of 0.2 ms, the overhead is 20% and if coherence time in a dynamic environment drops to 0.5ms then the overhead increases substantially to 40%, on that respective tone.

[2] documents tens of Gb/s rate and mitigates time variation. 5G New Radio (NR) schedules resources in units called *physical resource blocks* (PRBs). Per 3GPP Release 18, a PRB occupies 12 contiguous sub-carriers across the full length of one slot, which, under normal cyclic prefix corresponds to 14 OFDM symbols. Hence, a single PRB contains $N_{\text{PRB}} = 12 \times 14 = 168$ resource elements in the downlink grid. With $\mu = 1$ numerology (30 kHz sub-carrier spacing), the pilot overheads are $18/168 \approx 11\%$ and $36/168 \approx 21\%$ of the transmission bandwidth.

Along with the CSI bottleneck, the current linear precoders do not perform well under rank-deficient scenarios, e.g., where the total number of device's receiver antennas exceeds the number of transmit antennas, as in Fig 1. Conventional orthogonal multiple-access (OMA) power-allocation schemes implemented in the latest Wi-Fi revisions—IEEE 802.11 b/g/n/ac/be—use linear reception strategies that treat the linearly processed interference from co-scheduled users as additive noise. [3] overviews linear MU-MIMO and OMA assumptions in practice. When the underlying MIMO channel is low rank, single-user decoding leaves the rich inter-user crosstalk unexploited, resulting in severe spectral-efficiency penalties. Moreover, meeting the gigabit-class uplink throughput targets of Wi-Fi 7 with these linear-receiver architectures necessitates transmit powers that are incompatible with the stringent energy budgets of battery-constrained hardware, e.g., augmented-reality (AR) glasses. Hence, next-generation WLANs require power-allocation and receiver designs that can jointly harness inter-user interference while respecting ultra-low-power operating limits.

State-of-the-art non-orthogonal multiple-access (NOMA)

Muhammad Ahmed Mohsin, Muhammad Umer, and John M. Cioffi are associated with Stanford University. Email: {muahmed, mumer, cioffi}@stanford.edu.

Ahsan Bilal and Dean F. Hougen are associated with University of Oklahoma. Email: {a.bilal-1, hougen}@ou.edu.

Muhammad Ali Jamshed is associated with University of Glasgow. Email: muhammadali.jamshed@glasgow.ac.uk.

¹It is note in this paper that we assume a duality between broadcast channel and multiple access channel and use *minPMAC* to address precoder designs.

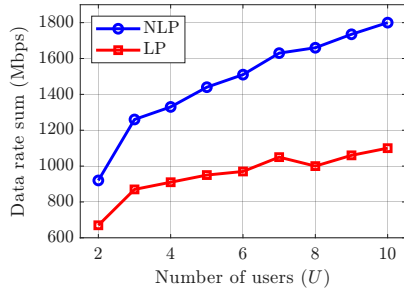


Fig. 1: Linear vs. Non-Linear Precoder under rank deficient scenarios for FWA

power-allocation algorithms deliver appreciable gains in spectral efficiency and energy consumption relative to their OMA counterparts [4], [5]. In the single-input single-output (SISO) regime, the optimal successive-interference-cancellation (SIC) schedule is trivially ordered by the users' channel magnitudes [6], [7]. This ordering principle, however, does not generalize to multi-user MIMO channels [8]–[11], where no inherent hierarchy emerges from the vector-valued channel realizations [12]. Consequently, most MIMO-NOMA studies resort to ad-hoc SIC heuristics, e.g., limiting the system to two users [13], forming disjoint user pairs and applying NOMA only within each pair [14], and similar simplifications. Such heuristic scheduling degrades both achievable rates and energy efficiency, falling short of contemporary WLAN requirements.

A further limitation of the prior art is its implicit assumption that all users share the entire system bandwidth, confining multiplexing to the power (or code) domain alone. Multi-carrier hybrid NOMA (MC-NOMA) [15] extends this framework by jointly optimizing power and sub-carrier allocation in the downlink [16], yet even here the SIC order is still chosen by a scalar channel-gain metric, an approach that breaks down when the access point (AP) or the users are equipped with multiple antennas. Robust, globally optimal decoding-order work for MIMO MC-NOMA appear in this problem, critical to realizing the full throughput and energy-saving potential of next-generation wireless LANs.

Contributions. To the best of the current knowledge, this study presents *first end-to-end* AI-native RAN design that jointly covers channel acquisition (GRF), continual channel prediction with handover persistence, and energy-optimal non-linear scheduling (minPMAC). (i) **GRF** (neural Gaussian Radio Fields) compress CSI acquisition, cutting pilot overhead by $\sim 100\times$ while delivering ~ 1.1 ms inference and state-of-the-art reconstruction accuracy across indoor/outdoor MIMO settings; this enables sub-millisecond closed-loop updates without sacrificing fidelity. (ii) **Continual learning for prediction & handovers** uses experience replay (LARS) to adapt online under mobility and cell transitions, improving channel NMSE by > 10 dB over static predictors and a further $2\text{--}5$ dB over uniform replay, all while maintaining accuracy across UMi/UMA handovers without per-scenario retraining. (iii) **minPMAC** is a convex, order-free MAC polymatroid design that recovers the globally optimal SIC sequence from dual variables and minimizes transmit energy for given minimum-rate targets;

empirically it reduces the required energy by $4\text{--}10\times$ (scenario-dependent) and yields monotonically increasing bits/J with SNR. Together, these components form a practical, GPU-ready pipeline that achieves real-time CSI, robust tracking through dynamic network conditions and handovers, and substantial energy savings under multiuser MIMO operation.

II. RELATED WORK

Channel Estimation. Neural radiance-field (NeRF) approaches such as NeRF² [17] and NeWRF [18] give much better channel-estimation accuracy than classic methods. Still, they need very dense measurements (about 178 samples per cubic foot), hours of GPU training, and 200–350 ms inference times, far too slow for real-time vehicle links. To speed inference time, researchers turned to 3D Gaussian Splatting (3DGS): WRF-GS [19] uses spherical-harmonic features on 3DGS to map RF power, but this ignores full MIMO CSI and depends on a Mercator projection that does not match electromagnetic physics; RadSplatter [20] improves RSSI prediction with a relaxed-mean features yet suffers unstable optimization. RFSPM [21] gains speed by adding tiny multi-layer perceptrons (MLPs) per Gaussian and custom CUDA kernels, making deployment harder. In short, both NeRF and 3DGS based models borrow optical-imaging assumptions that fail to capture true 3-D radio wave superposition and still demand too much data or compute for millisecond-level wireless use [22], [23].

Channel Prediction. Data-driven channel prediction has evolved from long-short term memory (LSTM) model baselines to transformers, large language models, and diffusion networks. Recent studies use transformers for the channel prediction [24]. LSTMs outperform autoregressive integrated moving average (ARIMA) but stumble on long memories [25], [26]; transformers ease that issue yet demand heavy, matched condition training and large computation, and their error jumps 15–30% when array spacing changes [24], [27], [28]. Generative pretrained transformer (GPT-2) adaptation enables near-zero-shot multiple input single output orthogonal frequency division multiplexing (MISO-OFDM) prediction after brief fine-tuning, but the language-to-CSI gap still caps accuracy [29]. Diffusion models can synthesize scarce MIMO data or handle joint source-channel estimation [30]–[33], yet they remain slow and lack online drift adaptation. These limitations motivate a continual-learning MIMO predictor.

To break this accuracy–complexity log-jam, each link is bootstrapped with a few high-fidelity snapshots from a NeRF-based estimator, then a lightweight continual-learning predictor updates online as the channel evolves. This single pipeline keeps the instant CSI precision of NeRF during the initial iterations while avoiding its data hunger and heavy retraining thereafter, yielding a unified framework for scalable MIMO estimation and long-horizon prediction under changing propagation conditions.

NOMA Methods Power-domain NOMA is widely recognized to outperform OMA at a given power budget—especially in SISO settings where SIC follows the users' channel strengths—and extensive surveys establish these fundamentals

TABLE I: Key notation used in the paper.

Symbol	Description
U	Number of users; index set $\mathcal{U} = \{1, \dots, U\}$.
N	Number of OFDM subcarriers (tones); index set $\mathcal{N} = \{1, \dots, N\}$.
$L_{x,u}, L_y$	Transmit antennas of user u ; receive antennas at BS.
$\mathbf{H}_{u,n} \in \mathbb{C}^{L_y \times L_{x,u}}$	MIMO channel matrix of user u on tone n .
$\hat{\mathbf{H}}_t, \hat{\mathbf{H}}_{t+\Delta}$	GRF channel estimate at time t ; Δ -step forecast from the predictor.
$\mathbf{R}_{xx}(u, n) \succeq \mathbf{0}$	Transmit covariance of user u on tone n .
$\mathbf{R}_{nn} = \sigma^2 \mathbf{I}_{L_y}$	Receiver noise covariance.
$b_{u,n}, \mathbf{b}_{\min}$	Achievable rate of user u on tone n (bits/s/Hz); per-user minimum-rate targets.
$w_u \geq 0$	Energy weight for user u in the precoder objective.
$\mathcal{T} \subseteq \mathcal{U}$	User subset used in MAC polymatroid constraints.
\mathbf{X}_t	Sliding CSI window of length T used by the predictor.
$\mathcal{M}, N_{\text{buf}}$	Replay buffer and its size for continual learning.
$\lambda \in [0, 1]$	Mixture weight for current vs. replay loss in training.
NMSE	Normalized MSE: $\ \mathbf{H} - \hat{\mathbf{H}}\ _F^2 / \ \mathbf{H}\ _F^2$.
SNR _{dB}	$10 \log_{10} (\ \mathbf{H}_{\text{gt}}\ _F^2 / \ \mathbf{H}_{\text{pred}} - \mathbf{H}_{\text{gt}}\ _F^2)$.
GRF	Gaussian Radio Fields (compressed CSI acquisition).
LARS	Loss-Aware Reservoir Sampling (replay buffer management).
minPMAC	Energy-optimal nonlinear precoder on the MAC polymatroid.
Boldface upper-case (lower-case) letters denote matrices (vectors); $\text{tr}(\cdot)$ is the trace, $(\cdot)^H$ the Hermitian transpose, and $\ \cdot\ _F$ the Frobenius norm. The set $\{1, \dots, N\}$ indexes OFDM sub-carriers, and time is slotted at pilot instants $t \in \mathbb{N}$. For any quantity A , the notation A_t (resp. $A_{t+\Delta}$) refers to the present (resp. predicted) slot.	

and their 5G/6G relevance [34], [35]. However, once channels are vector-valued (MIMO), there is no natural scalar ordering; most MIMO–NOMA designs therefore rely on *heuristics* such as two-user pairing per beam/subcarrier, fixed or gain-based SIC schedules, or small disjoint clusters—choices that simplify scheduling but break global optimality and often induce non-convex formulations [34]. MC-NOMA improves matters by co-optimizing power and subcarriers, yet mainstream formulations still impose two-users-per-subchannel and gain-based SIC rules, so the core limitations persist [36]. In contrast, *minPMAC* treats the capacity region explicitly as a polymatroid, optimizes power–subcarrier *covariances* without embedding the decoding order (thus preserving convexity), and then *recovers* the globally optimal SIC sequence from the dual structure—thereby avoiding combinatorial search; the resulting MAC solution maps cleanly to the downlink via classical uplink–downlink duality [37]–[43].

III. PROBLEM SCOPE AND SYSTEM ARCHITECTURE

A single-cell multi-carrier MIMO uplink model has U users, each equipped with $L_{x,u}$ antennas. These transmit to a single radio twin base station (BS) equipped with L_y receive antennas. The BS hosts the proposed *AI-driven digital twin* composed of three sequential modules. First, *Gaussian Radio Fields* (GRF), performs compressed CSI acquisition by rendering the instantaneous MIMO channel from a learned set of 3-D Gaussian primitives. Second, a continual-learning (CL) channel predictor exploits temporal correlations to forecast future CSI and thereby compensates for channel aging as well as mobility-induced hand-over drift [44]. Third, the *minPMAC* nonlinear precoder consumes the predicted channel and outputs the globally optimal power–sub-carrier allocation together with the SIC decoding order. The outputs include optional time-sharing coefficients when multiple orders are required [45]. All three stages execute on the BS’s GPU cluster,

enabling a sub-millisecond closed-loop adaptation cycle. The next sections discuss theoretical and experimental details of each module separately.

Channel Estimation Formulation. Let $\mathbf{E}(\mathbf{r}, \omega) \in \mathbb{C}^3$ denote the steady-state electric field at spatial location $\mathbf{r} \in \mathbb{R}^3$ and angular frequency ω . In a linear, source-driven medium, the field obeys the vector Helmholtz equation

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}, \omega) - k^2(\omega) \mathbf{E}(\mathbf{r}, \omega) = -j\omega\mu_0 \mathbf{J}(\mathbf{r}, \omega), \quad (1)$$

where $k(\omega) = \omega\sqrt{\mu\epsilon}$ is the wave number and \mathbf{J} the impressed current density. The *channel state information* (CSI) between a transmit array with phase centers $\{\mathbf{r}_m^{\text{tx}}\}_{m=1}^{L_x}$ and a reception array $\{\mathbf{r}_\ell^{\text{rx}}\}_{\ell=1}^{L_y}$ is then captured by the frequency-dependent MIMO transfer matrix

$$[\mathbf{H}(\omega)]_{\ell,m} = \mathbf{g}^\top \mathbf{E}(\mathbf{r}_\ell^{\text{rx}}, \omega; \mathbf{r}_m^{\text{tx}}), \quad (2)$$

with \mathbf{g} the receiver’s effective dipole vector. Obtaining (2) from sparse pilot measurements is an ill-posed *inverse electromagnetic problem* that must infer the volumetric Green’s function satisfying (1) under highly heterogeneous, partly unknown, boundary conditions. A direct numerical solution—e.g., via finite-difference time-domain (FDTD) or finite-element methods—samples the domain at sub-wavelength resolution, leading to linear systems of order $\mathcal{O}((\ell/\lambda)^3)$ for an environment of characteristic size ℓ . Such solvers incur terabyte-scale memory footprints and runtimes far beyond the milliseconds coherence interval of 5G/6G links. Geometric optics (ray-tracing) offers an $\mathcal{O}(N_{\text{ray}} N_s N_b^R)$, where N_{ray} is the number of rays, N_s is the number of surfaces and N_b^R is the number of depth rays, surrogate by taking the $\lambda \rightarrow 0$ limit, but these neglect diffraction and scattering, and still become prohibitive once the number of surfaces N_s and bounce depth R grow in realistic indoor or urban scenes. Consequently, state-of-the-art CSI estimation adopts data-driven approximations that trade physical fidelity

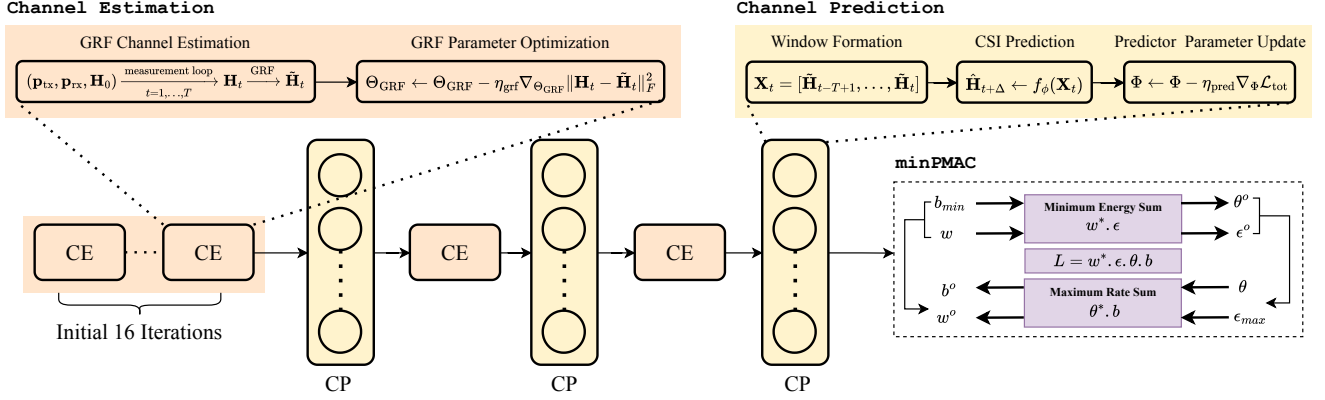


Fig. 2: Flow diagram of the proposed hybrid RF-channel pipeline. During the first 16 iterations, a NeRF-based Gaussian Radiance Field (GRF) module performs high-fidelity channel estimation (CE). Once initial parameters converge, the system transitions to a lightweight continual-learning predictor (CP) that: (i) forms sliding CSI windows, (ii) updates model weights online, and (iii) delivers long-horizon forecasts alongside link-quality metrics—minimum data-rate (minPMAC), optimal decoding order, receiver matrix, and minimum energy-rate.

for tractability in compressive pilot recovery, learned NeRF-style implicit fields, or, as proposed here, the explicit *Gaussian Radio Field* representation. The latter replaces the volumetric mesh by a sparse set of adaptive anisotropic Gaussian primitives, yielding closed-form superposition in (2) and enabling sub-millisecond channel synthesis suitable for real-time digital-twin operation.

minPMAC Precoder Formulation. Let $\mathcal{U} = \{1, \dots, U\}$ index the users and $\mathcal{N} = \{1, \dots, N\}$ the OFDM sub-carriers. For user $u \in \mathcal{U}$ on tone $n \in \mathcal{N}$, $\mathbf{R}_{xx}(u, n) \in \mathbb{C}^{L_{x,u} \times L_{x,u}}$ is the transmit covariance (equivalently, power-allocation) matrix and $b_{u,n}$ is the achievable rate under a fixed but *unspecified* SIC decoding order. The overall design objective is to meet a vector rate requirement $\mathbf{b}_{min} = [b_{min,1}, \dots, b_{min,U}]^T$ while minimizing a weighted sum of user energies. Formally, the task is cast as the following *energy-sum minimization* problem:

$$\begin{aligned} \min_{\{\mathbf{R}_{xx}(u, n)\}} \quad & \sum_{u=1}^U \sum_{n=1}^N w_u \cdot \text{trace}\{\mathbf{R}_{xx}(u, n)\} \\ \text{C1: } \mathbf{b} = \sum_{n=1}^N [b_{1,n}, b_{2,n}, \dots, b_{U,n}]^T \succeq \mathbf{b}_{min} \succeq \mathbf{0} \\ \text{C2: } \sum_{u \in T} \sum_{n=1}^N b_{u,n} \leq \log_2 \left| \mathbf{R}_{nn} + \sum_{u \in T} H_{u,n} \mathbf{R}_{xx}(u, n) H_{u,n}^* \right| \\ & - \log_2 |\mathbf{R}_{nn}|, \quad \forall T \subseteq \{1, 2, \dots, U\} \\ \text{C3: } \mathbf{R}_{xx}(u, n) \succeq \mathbf{0}, \quad \forall u \in \{1, 2, \dots, U\}, \forall n \in \{1, 2, \dots, N\} \end{aligned} \quad (3)$$

where $\mathbf{H}_{u,n} \in \mathbb{C}^{L_y \times L_{x,u}}$ is the MIMO channel for user u on tone n , $\mathbf{R}_{nn} = \sigma^2 \mathbf{I}_{L_y}$ ² is the noise covariance matrix, and $w_u \geq 0$ assigns relative importance to user u 's transmit energy. Constraint **C1** enforces the per-user minimum-rate targets,

²Note here that $H_{u,n} = \mathbf{R}_{nn}^{-1/2} H$ which allows use of an equivalent white noise model.

while **C2** guarantees that *every* subset of users \mathcal{T} operates inside the MAC capacity polymatroid. Finally, **C3** ensures that each covariance matrix is Hermitian positive semi-definite. Problem (3) is a convex semidefinite program: the objective is linear in the decision variables $\mathbf{R}_{xx}(u, n)$, **C1** is affine, **C2** is jointly concave in $\{\mathbf{R}_{xx}(u, n)\}$ by MIMO capacity theory, and **C3** defines a convex cone. Crucially, the formulation is *decoupled* from the SIC decoding order π ; therefore no non-convex ‘ordering variables’ are introduced, in contrast to heuristic approaches that embed the order into the optimization and render the problem non-convex [4], [12], [46], [47]. Associating dual variables $\{\theta_u\}$ with the minimum-rate constraints **C1** yields the Lagrange dual *weighted sum-rate maximization under an energy budget*, from which the globally optimal decoding order can be inferred (§VI-B). Thus, the overall design procedure proceeds in two convex steps: (i) solve (3) to obtain the power-sub-carrier allocation, and (ii) determine the SIC sequence by sorting the optimal dual multipliers. This separation preserves convexity and ensures an efficient, globally optimal solution pipeline for nonlinear precoding. The resultant $\mathbf{R}_{xx}(u, n)$, along with learned $H_{u,n}$, then must accordingly be used in a GDFE that the minPMAC optimization assumes.

IV. HYBRID GAUSSIAN OBSERVATION FOR CHANNEL ESTIMATION

The proposed GRF model [48] consists mainly of 3 sub-modules, which are explained step-by-step as:

3D Gaussian Primitive Initialization. The proposed *Gaussian Radio Field* (GRF) models the propagation environment by a sparse set $\mathcal{G} = \{\mathcal{G}_i\}_{i=1}^N$ of anisotropic three-dimensional Gaussians. Primitive i is characterized by its centre $\boldsymbol{\mu}_i \in \mathbb{R}^3$ and positive-definite covariance $\boldsymbol{\Sigma}_i \in \mathbb{S}_+^3$; additional electromagnetic attributes are supplied by the neural field (see Fig. 3). Owing to the linearity of Maxwell’s equations in passive, non-magnetic media, the total field superposes of the fields

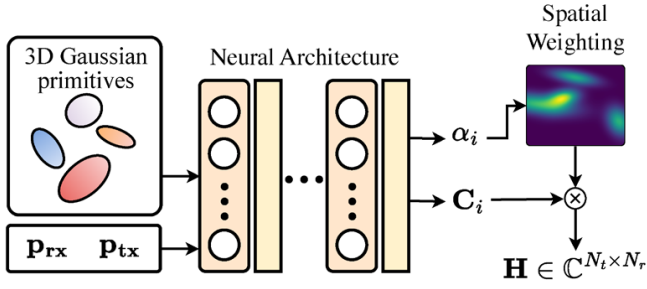


Fig. 3: Schematic of the differentiable MIMO channel model. Transmitter and receiver positions \mathbf{p}_{tx} and \mathbf{p}_{rx} are fused with M learnable 3D Gaussian primitives (means $\boldsymbol{\mu}_i$, covariances $\boldsymbol{\Sigma}_i$) and passed through a neural network to produce, for each primitive i , a base activation α_i and a complex channel contribution \mathbf{C}_i . Each contribution is gated by the spatial weight $w_i(\mathbf{p}_{\text{rx}}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ and summed to yield the predicted channel matrix $\mathbf{H} = \sum_{i=1}^M w_i \alpha_i \mathbf{C}_i \in \mathbb{C}^{N_t \times N_r}$.

generated by these primitives. Each Gaussian, therefore, acts as an adaptive basis function, chosen to capture fine spatial channel variations.

For a query location $\mathbf{x} \in \mathbb{R}^3$ the unnormalized kernel weight of primitive i is

$$w_i(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right), \quad (4)$$

which omits the constant Gaussian prefactor because only relative Weights are required.

Factoring $\boldsymbol{\Sigma}_i = \mathbf{R}_i \mathbf{S}_i^2 \mathbf{R}_i^\top$, guarantees $\boldsymbol{\Sigma}_i \succ 0$ while keeping the parameters unconstrained. Here $\mathbf{R}_i \in \text{SO}(3)$ is unitary, and $\mathbf{S}_i = \text{diag}(e^{s_{i,1}}, e^{s_{i,2}}, e^{s_{i,3}})$ ensures strictly positive eigenvalues through the exponential map. This compact parameterization facilitates the GRF's end-to-end differentiable optimization.

Attribute Network. The electromagnetic response of each Gaussian primitive is generated in real-time by two lightweight neural modules (Fig. 3). The *attribute network* $\mathcal{F}_{\text{attr}} : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^d \times \mathbb{R}$ receives the primitive center $\boldsymbol{\mu}_i$ and the transmitter location \mathbf{p}_{tx} , and returns a latent descriptor $\mathbf{z}_i \in \mathbb{R}^d$ together with a scalar activation α_i :

$$(\mathbf{z}_i, \alpha_i) = \mathcal{F}_{\text{attr}}(\gamma_L(\boldsymbol{\mu}_i), \gamma_L(\mathbf{p}_{\text{tx}}); \Theta_{\text{attr}}).$$

where $\gamma_L(\cdot)$ is the multi-resolution positional encoding [49], [50] and \mathbf{z}_i is the extracted features in latent space later utilized in the channel contribution \mathbf{C}_i ,

$$\gamma_L(\mathbf{x}) = [\mathbf{x}, \sin(2^0 \pi \mathbf{x}), \cos(2^0 \pi \mathbf{x}), \dots, \sin(2^{L-1} \pi \mathbf{x}), \cos(2^{L-1} \pi \mathbf{x})]. \quad (5)$$

Choosing L such that $2^{L-1} \geq \ell/\lambda$ (with ℓ the scene diameter and λ the carrier wavelength) guarantees Nyquist coverage of spatial harmonics. Throughout, $L=16$ is sufficient for $\ell \approx 50$ m and $\lambda \approx 5$ cm.

The latent vector \mathbf{z}_i captures local scattering traits, while α_i modulates their overall strength. A separate *decoder network* $\mathcal{F}_{\text{dec}} : \mathbb{R}^d \rightarrow \mathbb{R}^{N_t \times N_r} \times \mathbb{R}^{N_t \times N_r}$ maps \mathbf{z}_i to the real and imaginary parts of the complex contribution, $\mathbf{C}_i = \mathbf{C}_i^{\text{re}} +$

$j \mathbf{C}_i^{\text{im}} = \mathcal{F}_{\text{dec}}(\mathbf{z}_i; \Theta_{\text{dec}})$, thereby endowing each Gaussian with an effective $N_t \times N_r$ MIMO transfer matrix.

Channel Rendering. GRF evaluates the MIMO channel by a single weighted sum over N Gaussian primitives, thus bypassing ray-marching or volumetric integration. For a transmitter at \mathbf{p}_{tx} and a receiver at \mathbf{p}_{rx} the $N_t \times N_r$ channel matrix is

$$\mathbf{H}(\mathbf{p}_{\text{rx}}, \mathbf{p}_{\text{tx}}) = \sum_{i=1}^N w_i(\mathbf{p}_{\text{rx}}) \mathbf{C}_i, \quad (6)$$

where \mathbf{C}_i is the complex contribution of primitive i and the spatial weight. Figure 4 illustrates this spatial weighting mechanism.

$$w_i(\mathbf{p}_{\text{rx}}) = \alpha_i \exp\left(-\frac{1}{2}(\mathbf{p}_{\text{rx}} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{p}_{\text{rx}} - \boldsymbol{\mu}_i)\right) \quad (7)$$

decays with the Mahalanobis distance between the receiver and the Gaussian centre. The scalar α_i modulates the overall strength of the primitive i .

Equation (6) requires $\mathcal{O}(N N_t N_r)$ operations, exploiting the intrinsic sparsity of radio propagation; traditional NeRF/3DGS pipelines incur $\mathcal{O}(N_{\text{ray}} N_{\text{sample}} |\Theta|)$ evaluations of a large MLP and are therefore two-to-three orders of magnitude slower. Because GRF directly computes the field at the true 3-D receiver location, it avoids the projection artifacts that arise when 3-D Gaussian are rasterized onto 2-D planes. Fig. 3 demonstrates the attribute network and the spatial weighting pipeline of our GRF model.

Algorithm 1 GRF Channel Rendering ($\mathcal{O}(N N_t N_r)$)

Require: primitive set $\{(\boldsymbol{\mu}_i, \mathbf{q}_i, \mathbf{s}_i, \alpha_i, \mathbf{z}_i)\}_{i=1}^N$,

1: transmitter position \mathbf{p}_{tx} , receiver position \mathbf{p}_{rx}

Ensure: channel matrix $\mathbf{H}(\mathbf{p}_{\text{rx}}, \mathbf{p}_{\text{tx}}) \in \mathbb{C}^{N_t \times N_r}$

2: $\mathbf{H} \leftarrow \mathbf{0}$

▷ initialize accumulator

3: **for** $i = 1$ to N **do**

4: $\mathbf{R}_i \leftarrow \text{QUATTOROT}(\mathbf{q}_i)$

5: $\boldsymbol{\Lambda}_i^{-1} \leftarrow \mathbf{R}_i \text{diag}(e^{-2s_i}) \mathbf{R}_i^\top$

6: $\Delta \mathbf{p} \leftarrow \mathbf{p}_{\text{rx}} - \boldsymbol{\mu}_i$

7: $w_i \leftarrow \alpha_i \exp(-\frac{1}{2} \Delta \mathbf{p}^\top \boldsymbol{\Lambda}_i^{-1} \Delta \mathbf{p})$

8: $\mathbf{C}_i \leftarrow \text{DECODE}(\mathbf{z}_i)$

9: $\mathbf{H} \leftarrow \mathbf{H} + w_i \mathbf{C}_i$

10: **end for**

11: **return** \mathbf{H}

V. CONTINUAL CHANNEL TRACKING AND PREDICTION

Between two successive GRF estimates, the propagation medium may evolve appreciably (mobility, blockage, antenna reconfiguration, cell handoff), rendering a static predictor obsolete. This work, therefore, develops a continual-learning (CL) channel-prediction algorithm that adapts online across heterogeneous network configurations and handoff situations without requiring per-scenario retraining. Figure 5 shows why this is needed: offline models trained once and frozen fail to track shifting channel statistics as the UE moves from UMi-compact \rightarrow UMi-dense \rightarrow UMi-standard (and as SNR

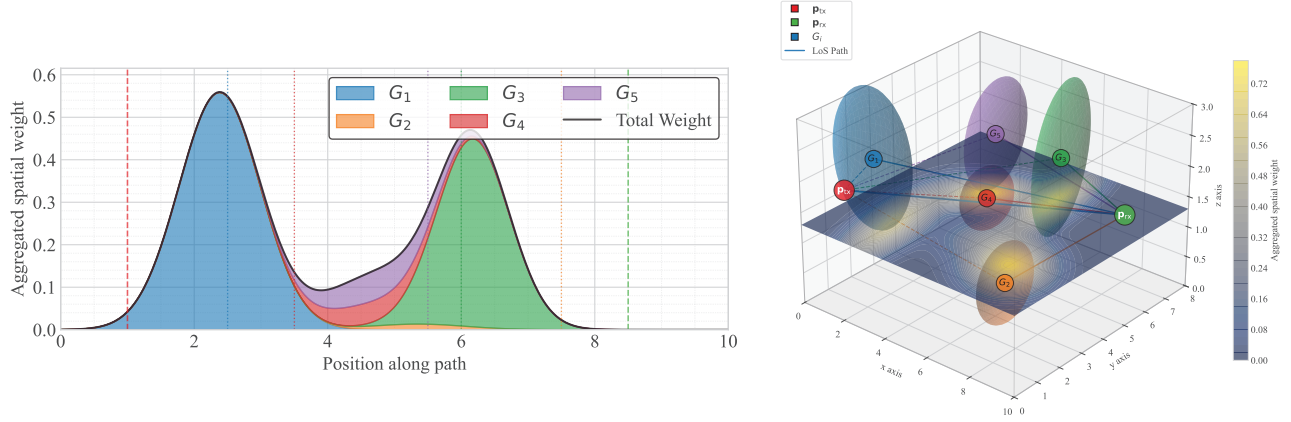


Fig. 4: **Visualization of the spatial weighting in GRF.** (Right:) Anisotropic 3D Gaussians (G_1 - G_5) are shown as ellipsoids with their shapes determined by covariance Σ_i , along with the transmitter (\mathbf{p}_{tx}), receiver (\mathbf{p}_{rx}), and line-of-sight path. The heat map represents the spatial weight distribution at receiver height. (Left:) Weight profile along the \mathbf{p}_{tx} - \mathbf{p}_{rx} path showing individual Gaussian contributions and their cumulative effect (black outline).

improves). Dense variants eventually reach -30 dB NMSE, but Standard and Compact topologies saturate in the -22 to -25 dB range. This difference between the NMSE for Dense and other scenarios (Standard and Compact) gives evidence of distribution mismatch and weight staleness. The proposed CL approach interleaves each new mini-batch with a bounded replay buffer of historical samples, regularizing updates and mitigating catastrophic forgetting as conditions drift across cells. At every GRF refresh, prediction errors are logged; high-loss (“hard”) fades are preferentially retained via experience replay (ER)s, ensuring that rare yet performance-critical states remain represented in the buffer. As shown in Fig. 6, this ER loop yields a step-change in accuracy: all three topology families collapse to a low-error band near -38 dB NMSE by ~ 10 dB SNR and continue improving to ≈ -42 dB at 30 dB SNR (best: Compact-LARS), more than 10 dB better than the non-CL Standard/Compact baselines and a clear gain even over Dense.

A. Experience Replay for GRF-aware Prediction

Let $\{\mathcal{D}^{(m)}\}_{m \geq 1}$ be the sequence of mini-batches collected in successive cells. At iteration k , the predictor f_{Φ} trains on the current batch $\mathcal{D}^{(k)}$ while regularizing against forgetting using experience replay drawn from $\bigcup_{m < k} \mathcal{D}^{(m)}$. After each GRF rendering, the latest estimate $\hat{\mathbf{H}}_t \in \mathbb{C}^{N_t \times N_r}$ is appended to a sliding window $\mathbf{X}_t = [\hat{\mathbf{H}}_{t-T+1}, \dots, \hat{\mathbf{H}}_t]$, which the predictor maps to a Δ -step forecast $\hat{\mathbf{H}}_{t+\Delta} = f_{\Phi}(\mathbf{X}_t)$. The forecast is consumed by the scheduler immediately, while the pair $(\mathbf{X}_t, \hat{\mathbf{H}}_{t+\Delta})$ is pushed to the replay buffer \mathcal{M} . When the next GRF estimate $\hat{\mathbf{H}}_{t+\Delta}$ arrives, the prediction error is logged as a replay score, ensuring that difficult samples are revisited. Thus, GRF and the predictor form a closed loop: GRF supplies ground-truth snapshots; the predictor fills the gaps; and its continually refined outputs feed back into the digital-twin pipeline, promoting ergodic tracking of the evolving channel distribution.

Replay Buffer. A fixed-size buffer $\mathcal{M} = \{e_1, \dots, e_{N_{\text{buf}}}\}$ stores

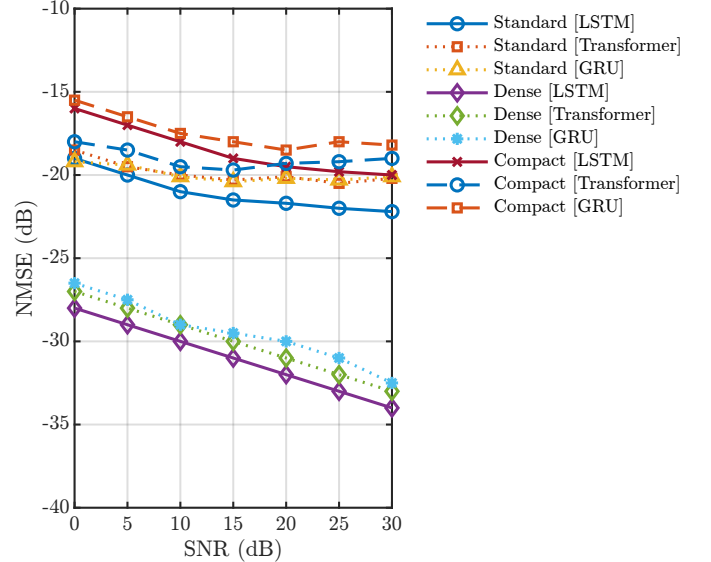


Fig. 5: Baseline SNR trained on UMi dense and tested on all scenarios under different architectures.

past channel samples, where an element $e = (\hat{\mathbf{H}}_t, \mathbf{p}_{tx}, \mathbf{p}_{rx})$ comprises the latest GRF estimate and the corresponding link geometry. Upon observing a new sample, e_t , it is inserted into \mathcal{M} and, if full, evicts the oldest entry (FIFO policy). A FIFO reservoir $\mathcal{M} = \{e^{(n)}\}_{n=1}^{N_{\text{buf}}}$ stores triples

$$e^{(n)} = (\mathbf{X}^{(n)}, \mathbf{H}^{(n)}, \phi^{(n)}),$$

where $\mathbf{H}^{(n)}$ is the measured channel at $t^{(n)} + \Delta$ and $\phi^{(n)}$ may encode scenario ID, SNR, etc. Each new sample $(\mathbf{X}_t, \hat{\mathbf{H}}_{t+\Delta})$ is added to \mathcal{M} and, if full, the oldest entry is discarded.

Interleaved Optimization. For every incoming batch, \mathcal{D}_k , the composite training set $\mathcal{B} = \mathcal{D}_k \cup \text{SAMPLE}(\mathcal{M}, B_{\text{replay}})$, where B_{replay} replay items are drawn uniformly at random.

The predictor is updated by minimizing,

$$\mathcal{L}_{\text{ER}} = \frac{1}{|\mathcal{B}|} \sum_{(\hat{\mathbf{H}}, \mathbf{p}_{\text{tx}}, \mathbf{p}_{\text{rx}}) \in \mathcal{B}} \|f_{\Phi}(\hat{\mathbf{H}}) - \mathbf{H}_{\text{gt}}\|_F^2, \quad (8)$$

thereby balancing plasticity to new fading statistics against retention of earlier conditions. In effect, ER allows the same predictor to track continuous channel evolution, e.g. UMi-compact \rightarrow UMi-dense \rightarrow UMi-standard, without catastrophic forgetting, and to feed refined CSI back to the GRF estimator for the next cycle.

Mini-batch Composition. At global step k the user equipment (UE) resides in cell $c(k)$ with local dataset $\mathcal{D}_{c(k)}$. $\mathcal{B}_{\text{curr}} \subset \mathcal{D}_{c(k)}$ and $\mathcal{B}_{\text{rep}} \subset \mathcal{M}$ is drawn to form $\mathcal{B} = \mathcal{B}_{\text{curr}} \cup \mathcal{B}_{\text{rep}}$. Let θ denote the current network parameters ($\theta \equiv \Phi$). The overall loss is a convex mixture

$$\mathcal{L}_{\text{tot}}(\theta) = \lambda \mathcal{L}_{\text{curr}}(\theta, \mathcal{B}_{\text{curr}}) + (1 - \lambda) \mathcal{L}_{\text{rep}}(\theta, \mathcal{B}_{\text{rep}}), \quad (9)$$

with mixing coefficient $\lambda \in [0, 1]$. Both terms in (9) use the normalized mean-square error (NMSE):

$$\mathcal{L}_{\text{curr}} = \frac{1}{|\mathcal{B}_{\text{curr}}|} \sum_{(\mathbf{X}, \mathbf{H}) \in \mathcal{B}_{\text{curr}}} \frac{\|\mathbf{H} - f_{\theta}(\mathbf{X})\|_F^2}{\|\mathbf{H}\|_F^2}, \quad (10)$$

$$\mathcal{L}_{\text{rep}} = \frac{1}{|\mathcal{B}_{\text{rep}}|} \sum_{(\mathbf{X}, \mathbf{H}) \in \mathcal{B}_{\text{rep}}} \frac{\|\mathbf{H} - f_{\theta}(\mathbf{X})\|_F^2}{\|\mathbf{H}\|_F^2}. \quad (11)$$

6G Twin Channel Estimator. During each prediction horizon, $(t, t + \Delta)$ the ER-enhanced predictor delivers the look-ahead CSI $\hat{\mathbf{H}}_{t+\Delta}$ to the resource scheduler, while its instantaneous error with the subsequent GRF update is back-propagated to refine both the predictor and the Gaussian field. This closed GRF-ER feedback loop (i) preserves high-fidelity CSI in the gaps between consecutive GRF renderings and (ii) tracks non-stationary propagation changes, e.g. UMi-compact \rightarrow UMi-dense, without catastrophic forgetting, thereby sustaining 6G-level accuracy throughout mobility and hand-over events.

Reservoir Sampling. As the UE traverses successive UMi- $\{\text{compact}, \text{dense}, \text{standard}\}$ cells it generates a time-ordered stream $e_t = (\mathbf{X}_t, \mathbf{H}_{t+\Delta})$, where $\mathbf{X}_t \in \mathbb{C}^{2 \times T \times N_t \times N_r}$ stacks the *past* T GRF estimates and $\mathbf{H}_{t+\Delta} \in \mathbb{C}^{2 \times N_t \times N_r}$ is the *next-slot* ground-truth channel. A replay reservoir \mathcal{M} of fixed size N_{buf} (small enough to be cached at the serving gNB stores a uniform sample of all observations seen so far. Classical reservoir sampling guarantees

$$\Pr[e_i \in \mathcal{M}_t] = \frac{N_{\text{buf}}}{t}, \quad i = 1, \dots, t,$$

So a channel recorded in the first cell is as likely to be rehearsed as one collected moments before the current GRF update.

1) *Data acquisition.* Upon arrival of e_t the algorithm inserts it into \mathcal{M} with probability $\frac{N_{\text{buf}}}{t}$; otherwise the sample is discarded.

2) *Mini-batch assembly.* Each stochastic gradient descent (SGD) step forms $\mathcal{B}_k = \mathcal{B}_{\text{curr}} \cup \mathcal{B}_{\text{rep}}$, where $\mathcal{B}_{\text{curr}} \subset \mathcal{D}^{(k)}$ is drawn from the live cell trace and $\mathcal{B}_{\text{rep}} \subset \mathcal{M}$ from the reservoir.

Algorithm 2 6G-Twin Channel Estimation & Prediction (GRF + ER/LARS)

Require: buffer size N_{buf} , mixing weight λ , sampling mode $s \in \{\text{UNIFORM}, \text{LARS}\}$,

- 1: learning rates $\eta_{\text{pred}}, \eta_{\text{grf}}$, GRF refresh period Δ_{GRF}
- 2: Initialize replay buffer $\mathcal{M} \leftarrow \emptyset$, counter $t \leftarrow 0$
- 3: Initialize GRF parameters Θ_{GRF} , predictor parameters Φ
- 4: **function** INSERT($\mathbf{X}, \mathbf{H}, \ell$)
- 5: $t \leftarrow t + 1$
- 6: **if** $|\mathcal{M}| < N_{\text{buf}}$ **then**
- 7: $\mathcal{M} \leftarrow \mathcal{M} \cup \{(\mathbf{X}, \mathbf{H}, \ell)\}$
- 8: **else if** $\text{rand}() < \frac{N_{\text{buf}}}{t}$ **then** ▷ reservoir test
- 9: **if** $s = \text{LARS}$ **then**
- 10: pick victim v with prob. $\Pr[v = i] \propto (\ell_i + \varepsilon)^{-1}$
- 11: **else**
- 12: $v \leftarrow \text{randint}(1, |\mathcal{M}|)$
- 13: **end if**
- 14: $\mathcal{M}[v] \leftarrow (\mathbf{X}, \mathbf{H}, \ell)$
- 15: **end if**
- 16: **end function**
- 17: **for** each visited cell k **do**
- 18: **for** each measurement $(\mathbf{p}_{\text{tx}}, \mathbf{p}_{\text{rx}}, \mathbf{H}_t)$ **do**
- 19: **if** $t \bmod \Delta_{\text{GRF}} = 0$ **then** ▷ full GRF update
- 20: $\hat{\mathbf{H}}_t \leftarrow \text{RENDERGRF}(\Theta_{\text{GRF}}, \mathbf{p}_{\text{rx}}, \mathbf{p}_{\text{tx}})$
- 21: $\Theta_{\text{GRF}} \leftarrow \Theta_{\text{GRF}} - \eta_{\text{grf}} \nabla_{\Theta_{\text{GRF}}} \|\hat{\mathbf{H}}_t - \mathbf{H}_t\|_F^2$
- 22: **end if**
- 23: Form window $\mathbf{X}_t = [\hat{\mathbf{H}}_{t-T+1}, \dots, \hat{\mathbf{H}}_t]$
- 24: $\hat{\mathbf{H}}_{t+\Delta} \leftarrow f_{\Phi}(\mathbf{X}_t)$ ▷ look-ahead CSI
- 25: $\ell_t \leftarrow \|\mathbf{H}_{t+\Delta} - \hat{\mathbf{H}}_{t+\Delta}\|_F^2 / \|\mathbf{H}_{t+\Delta}\|_F^2$ ▷ NMSE
- 26: INSERT($\mathbf{X}_t, \mathbf{H}_{t+\Delta}, \ell_t$)
- 27: **if** ready to update predictor **then**
- 28: Sample $\mathcal{B}_{\text{curr}} \subset \mathcal{D}^{(k)}$, $\mathcal{B}_{\text{rep}} \subset \mathcal{M}$
- 29: Compute mixed loss \mathcal{L}_{tot} via Eq. (8)
- 30: $\Phi \leftarrow \Phi - \eta_{\text{pred}} \nabla_{\Phi} \mathcal{L}_{\text{tot}}$
- 31: **end if**
- 32: **end for**
- 33: **end for**

The predictor parameters Φ are updated by minimizing the mixed NMSE of Eq. (8).

3) *Hand-over persistence.* When the UE hands off from cell k to $k+1$, the same buffer \mathcal{M} travels with it, ensuring that previously visited UMi statistics continue to influence training even though they are no longer observed. Coupled with the GRF estimator, this reservoir-based replay maintains a history-balanced memory, allowing the 6G-Twin channel estimator to track distribution shifts without sacrificing legacy performance.

Loss Aware Reservoir Sampling (LARS). Uniform reservoir sampling treats every past observation equally; however, deep fades or cell-edge links precisely those that limit throughput may then be under-represented. Therefore, uniform eviction is replaced with LARS [51], [52]. Immediately after the forward pass on mini-batch, \mathcal{B}_k per-sample NMSE is recorded $\ell_i = \|\mathbf{H}_i - f_{\Phi}(\mathbf{X}_i)\|_F^2 / \|\mathbf{H}_i\|_F^2$ for each element $e_i \in \mathcal{B}_k$. If the buffer is full ($|\mathcal{M}| = N_{\text{buf}}$) and a new sample

e_t is *selected* for insertion with reservoir probability N_{buf}/t , LARS evicts an existing entry v drawn according to

$$\Pr[v = i] = (\ell_i + \varepsilon)^{-1} / \sum_{j=1}^{N_{\text{buf}}} (\ell_j + \varepsilon)^{-1}, \quad (12)$$

with $\varepsilon > 0$ preventing division by zero. Hence, items whose prediction loss is already small are more likely to be removed, while hard-to-predict channels (large loss value ℓ_i) persist longer. This bias keeps rare yet performance-critical propagation states deep fades, rich scattering, extreme cell-edge SNRs, actively rehearsed, enhancing the robustness of the GRF-aware continual predictor.

VI. NON-LINEAR PRECODER

At scheduler time t , the GRF estimator delivers the most recent CSI snapshot $\hat{\mathbf{H}}_{u,n,t} \in \mathbb{C}^{L_y \times L_{x,u}}$ and the CL predictor provides a Δ -step look-ahead $\hat{\mathbf{H}}_{u,n,t+\Delta}$ together with an error summary (e.g., innovation covariance) $\Sigma_{u,n,t+\Delta}$. The true channel is modeled as

$$\mathbf{H}_{u,n,t+\Delta} = \hat{\mathbf{H}}_{u,n,t+\Delta} + \mathbf{E}_{u,n,t+\Delta} \quad (13)$$

$$\mathbb{E}[\text{vec}(\mathbf{E}) \text{vec}(\mathbf{E})^H] = \Sigma_{u,n,t+\Delta}, \quad (14)$$

and use either (i) a worst-case ellipsoidal set $\{\mathbf{E} : \|\Sigma^{-1/2} \text{vec}(\mathbf{E})\|_2 \leq \rho\}$, or (ii) a chance-constraint with violation level ε .

A. Energy-Optimal Nonlinear Precoding on the MAC Polymatroid

For each user u and tone n let $\mathbf{R}_{xx}(u, n) \succeq \mathbf{0}$ denote the (possibly multi-stream) transmit covariance. Denote noise covariance by $\mathbf{R}_{nn} = \sigma^2 \mathbf{I}_{L_y}$. The achievable *per-tone* rate tuple under successive interference cancellation (SIC) must lie in the MAC capacity region, which is a polymatroid: for every subset $\mathcal{T} \subseteq \mathcal{U}$,

$$\sum_{u \in \mathcal{T}} b_{u,n} \leq \log_2 \left| \mathbf{R}_{nn} + \sum_{u \in \mathcal{T}} \hat{\mathbf{H}}_{u,n} \mathbf{R}_{xx}(u, n) \hat{\mathbf{H}}_{u,n}^H \right| - \log_2 |\mathbf{R}_{nn}|. \quad (15)$$

Summing over tones yields the standard MAC polymatroid for the block. This description is exact for Gaussian codebooks and SIC and is independent of the decoding order. Weighted sum of user energies is minimized while satisfying minimum rate targets $\mathbf{b}_{\min} \succeq \mathbf{0}$:

$$\begin{aligned} \min_{\{\mathbf{R}_{xx}(u,n) \succeq \mathbf{0}\}} \quad & \sum_{u=1}^U \sum_{n=1}^N w_u \text{tr}(\mathbf{R}_{xx}(u, n)) \\ \text{s.t.} \quad & \mathbf{b} = \sum_{n=1}^N [b_{1,n}, \dots, b_{U,n}]^T \succeq \mathbf{b}_{\min}, \\ & (15) \text{ holds for all } \mathcal{T} \subseteq \mathcal{U}, \forall n \in \mathcal{N}. \end{aligned} \quad (16)$$

Problem (16) is a convex semidefinite program (SDP): the objective is linear in $\{\mathbf{R}_{xx}\}$, constraints are concave in the covariances by MIMO-MAC capacity, and PSD constraints

are conic. The formulation is order-free; SIC ordering is recovered from the dual variables.

Dual structure and SIC order. Associate multipliers $\{\theta_u \geq 0\}$ to the per-user minimum-rate constraints and $\{\lambda_{\mathcal{T},n} \geq 0\}$ to (15). At optimality, only a *chain* of subset constraints $\emptyset \subset \mathcal{T}_1 \subset \dots \subset \mathcal{T}_U = \mathcal{U}$ is active on each tone, yielding a corner point of the polymatroid that corresponds to a specific SIC order π ; the active chain determines π and vice-versa. Practically, the globally optimal π can be inferred by sorting the optimal duals $\{\theta_u^*\}$ (ties broken by the active-set structure). This is the standard order-duality relationship for (weighted) MAC boundary points and underpins efficient order recovery without combinatorial search.

B. Recovering Decoding Vectors $\theta_{u,n}$ and Implementing Non-linear Precoding

While (16) determines optimal covariances $\{\mathbf{R}_{xx}(u, n)\}$ and the SIC order π , practical receivers need *decoding vectors* for each stream. Suppose user u sends $d_{u,n}$ streams on tone n via the rank- $d_{u,n}$ factorization

$$\mathbf{R}_{xx}(u, n) = \sum_{s=1}^{d_{u,n}} p_{u,n,s} \mathbf{f}_{u,n,s} \mathbf{f}_{u,n,s}^H, \quad \|\mathbf{f}_{u,n,s}\|_2 = 1, p_{u,n,s} \geq 0. \quad (17)$$

Under the optimal SIC order π , let $\mathcal{I}_n^{(k)}$ be the set of users not yet decoded before stage k (i.e., residual interferers when decoding user π_k). The linear MMSE-SIC decoder for stream $(u = \pi_k, s)$ is

$$\theta_{u,n,s} \propto \left(\mathbf{R}_{nn} + \sum_{i \in \mathcal{I}_n^{(k)}} \hat{\mathbf{H}}_{i,n} \mathbf{R}_{xx}(i, n) \hat{\mathbf{H}}_{i,n}^H \right)^{-1} \hat{\mathbf{H}}_{u,n} \mathbf{f}_{u,n,s}. \quad (18)$$

Equations (16)–(18) fully specify the transmit covariances, the optimal SIC sequence, and the per-stream decoding vectors. When the downlink (BC) is ultimately to be served, the MAC solution is mapped to the BC via uplink-downlink duality and realize the non-linear transmitter with *dirty-paper coding* (DPC) or its practical approximations—Tomlinson-Harashima precoding (THP) and vector perturbation (VP). DPC achieves the BC capacity region; THP/VP are near-DPC at substantially lower complexity [42], [43].

Worst-case (ellipsoidal) robust Semi-Definite Programming (SDP). Replace $\hat{\mathbf{H}}_{u,n}$ in (15) by $\hat{\mathbf{H}}_{u,n} + \mathbf{E}_{u,n}$ and require the inequality to hold for all $\mathbf{E}_{u,n}$ in the ellipsoid $\|\Sigma^{-1/2} \text{vec}(\mathbf{E}_{u,n})\|_2 \leq \rho$. Using the S-procedure, each set-membership log-det constraint admits a safe LMI relaxation, producing a tractable SDP; for linear-beamforming special cases, semidefinite relaxation (SDR) is often tight.

Chance-constrained design. Enforce, for each \mathcal{T}, n ,

$$\Pr \left\{ \sum_{u \in \mathcal{T}} b_{u,n} \leq \log_2 \left| \mathbf{R}_{nn} + \sum_{u \in \mathcal{T}} (\hat{\mathbf{H}}_{u,n} + \mathbf{E}_{u,n}) \mathbf{R}_{xx}(u, n) (\hat{\mathbf{H}}_{u,n} + \mathbf{E}_{u,n})^H \right| - \log_2 |\mathbf{R}_{nn}| \right\} \geq 1 - \varepsilon. \quad (19)$$

Directly optimizing (19) is intractable; *safe convex approximations* are adopted based on measure-concentration inequalities

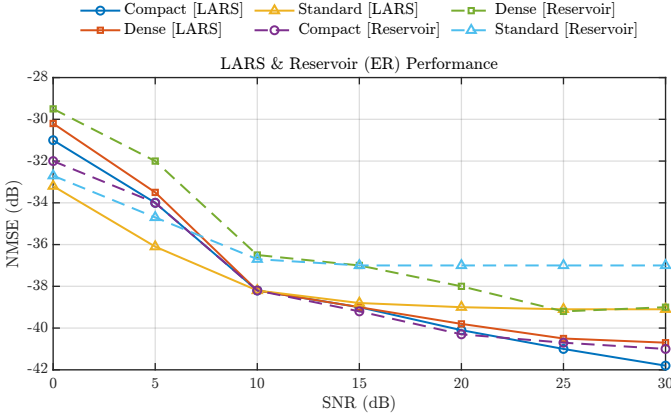


Fig. 6: SNR [dB] vs. NMSE [dB] for Experience Replay.

or scenario/sample approximations, which lead to explicit LMIs or conic constraints and preserve feasibility with probability at least $1 - \varepsilon$. The resulting program remains convex in $\{\mathbf{R}_{xx}\}$.

VII. COMPLEXITY ANALYSIS

GRF Channel Estimation. Let N_G be the number of Gaussian primitives and (N_t, N_r) the Tx/Rx antenna counts used in GRF rendering (Eq. (6)). The rendering cost is

$$\mathcal{C}_{\text{GRF-render}} = \Theta(N_G N_t N_r). \quad (20)$$

With GRF refresh period Δ_{GRF} (Alg. 2), the amortized per-slot cost is

$$\mathcal{C}_{\text{GRF,slot}} = \Theta\left(\frac{1}{\Delta_{\text{GRF}}} N_G N_t N_r\right). \quad (21)$$

Continual Channel Prediction. Let $B = |\mathcal{B}|$ be the mixed mini-batch size, T the window length, and N_{buf} the replay-buffer size (Sec. *Reservoir Sampling*). Denote the model-dependent forward/backward cost by $\mathcal{C}_{\text{pred}}(B; T, N_t, N_r)$. One CL update step incurs

$$\mathcal{C}_{\text{CL-step}} = \mathcal{C}_{\text{pred}}(B; T, N_t, N_r) + \Theta(B) + \Theta(1) + \Theta(N_{\text{buf}}), \quad (22)$$

where $\Theta(B)$ is NMSE evaluation over \mathcal{B} , $\Theta(1)$ is the reservoir test, and $\Theta(N_{\text{buf}})$ is the LARS eviction (Eq. (12)).

Non-Linear Precoder (energy-sum minimization on MAC polymatroid). For each tone n , constraint (15) requires a log-det of an $L_y \times L_y$ matrix. Naively checking all subset constraints $(2^U - 1)$ per tone gives

$$\mathcal{C}_{\text{poly-check}} = \Theta(N (2^U - 1) L_y^3). \quad (23)$$

The SIC order recovery from duals (sorting $\{\theta_u^*\}$) costs

$$\mathcal{C}_{\text{order}} = \Theta(U \log U). \quad (24)$$

MMSE-SIC decoding vectors (Eq. (18)) require one $L_y \times L_y$ solve per decoded user per tone:

$$\mathcal{C}_{\text{MMSE-SIC}} = \Theta(N U L_y^3). \quad (25)$$

One closed-loop adaptation cycle. With GRF amortized by Δ_{GRF} , a full slot that includes prediction, optimization, order recovery, and decoders has

$$\begin{aligned} \mathcal{C}_{\text{cycle}} = & \Theta\left(\frac{1}{\Delta_{\text{GRF}}} N_G N_t N_r\right) + \mathcal{C}_{\text{pred}}(B; T, N_t, N_r) + \\ & \Theta(B + N_{\text{buf}}) + \Theta(N (2^U - 1) L_y^3) + \Theta(N U L_y^3 + U \log U). \end{aligned} \quad (26)$$

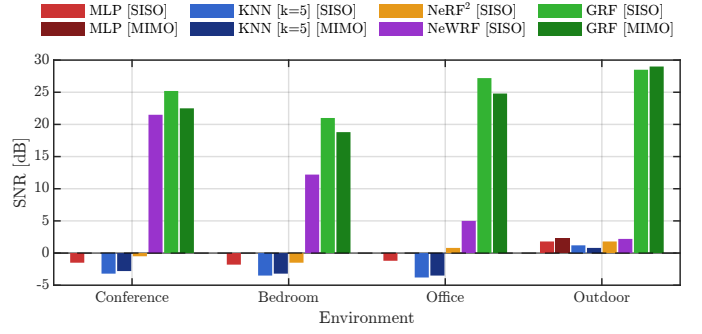


Fig. 7: Comparison of SNR [dB] across environments and methods. Our proposed **GRF** achieves consistently higher SNR in both SISO and MIMO modes, outperforming all prior models across diverse environments.

VIII. QUANTITATIVE RESULT ANALYSIS

A. GRF Channel Estimation

All methods are evaluated using the signal-to-noise ratio (SNR), defined as

$$\text{SNR}_{\text{H,dB}} = 10 \log_{10} \frac{\|\mathbf{H}_{\text{gt}}\|_F^2}{\|\mathbf{H}_{\text{pred}} - \mathbf{H}_{\text{gt}}\|_F^2}, \quad (27)$$

where \mathbf{H}_{gt} and \mathbf{H}_{pred} are the ground-truth and predicted channel-gain tensors, respectively. A higher value indicates more accurate channel reconstruction. GRF is compared with four strong alternatives: NeWRF [18], NeRF² [17], a vanilla baseline MLP, and a k -nearest-neighbours (KNN) estimator. Each model is tested in both SISO and MIMO configurations. As summarised in Fig. 7, GRF delivers the highest SNR in every indoor setting. In SISO mode it attains **25.23 dB**, **21.14 dB**, and **26.53 dB** in the *conference room*, *bedroom*, and *office*, respectively—an average gain of 10.9 dB over the next best approach (NeWRF). Even in the more demanding MIMO case, GRF maintains a clear margin, achieving **22.73 dB**, **19.60 dB**, and **24.78 dB** in the same environments. Traditional learning-based baselines, particularly MLP and KNN, fall below 0 dB in most indoor scenes, confirming that simple interpolators lack the inductive bias required for fine-grained channel reconstruction.

The outdoor scenario further highlights GRF’s scalability. It achieves **28.32 dB** (SISO) and **27.92 dB** (MIMO), which is ~ 26 dB higher than NeWRF (2.03 dB) and ~ 27 dB above NeRF² (1.40 dB). These results indicate that GRF can effectively model both rich multipath indoor channels and the sparser, large-scale characteristics of outdoor propagation. Across all eight test cases, GRF is consistently the top performer, underscoring the value of integrating Gaussian radio fields with neural implicit representations for high-fidelity channel prediction. The substantial margin—often exceeding an order of magnitude in MSE—demonstrates that GRF generalizes better than prior NeRF-style methods and conventional learning-based baselines, particularly in challenging MIMO and large-scale settings.

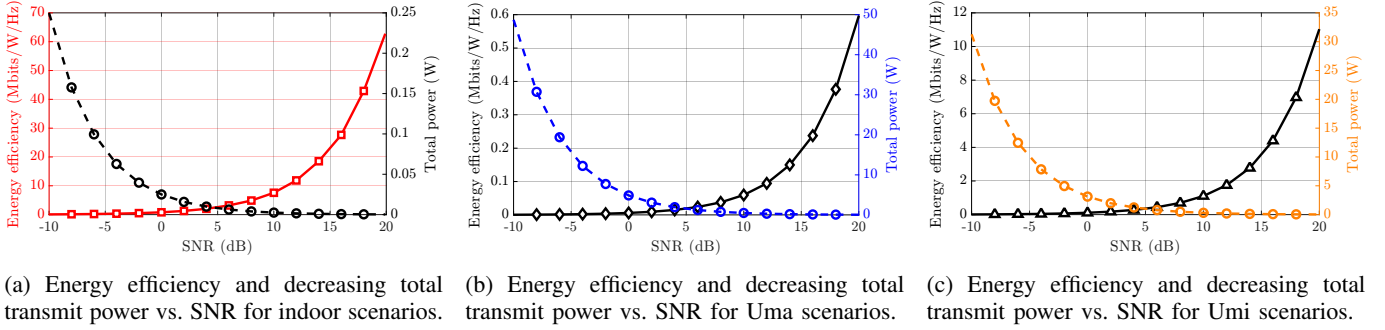


Fig. 8: Energy efficacy vs. SNR(db) and Total Transmit Power (W) vs. SNR for various scenarios for the minPMAC non linear precoder design.

B. Continual Adaptation with Experience Replay

Figure 6 reports NMSE versus SNR when the predictors are trained in a continual-learning loop that interleaves current samples with a bounded replay buffer. Two buffer management strategies are compared: LARS (loss-aware reservoir sampling; solid lines), which preferentially retains high-loss/hard channel states, and uniform Reservoir replay (dashed lines). Unlike the baseline models in Fig. 5, where Standard and Compact models saturated near -22 dB to -25 dB and even Dense variants stalled around -34 dB at high SNR, the online CL models rapidly adapt as conditions improve. By ~ 10 dB SNR all scenarios collapse to a low-error band around -38 dB, eliminating the baseline scenario-dependent spread.

In these experiments, uniform reservoir replay already improves over the static baseline: for example, the Standard model's NMSE drops from approximately -25 dB when frozen to around -37 dB at high SNR. However, it plateaus there because random sampling fails to emphasize the rare, challenging fades. This gap highlights the extra value of LARS's loss-aware sampling, which keeps the most challenging fades. By continuously retraining and focusing on these hard examples, the predictor adapts to changing channel conditions and maintains low NMSE across all scenarios, even under high-SNR conditions. LARS is adopted, which keeps the highest-loss examples in the buffer. With LARS, models drive NMSE even lower: at 30 dB SNR, Standard-LARS reaches about -39 dB, Dense-LARS about -41 dB, and Compact-LARS about -42 dB. These extra 2–5 dB improvements beyond uniform replay highlight how loss-aware sampling further adapts the predictor to clean channels and maintains low NMSE across all scenarios.

C. Non-Linear Precoder Design

Across all three environments (indoor, UMa, UMi), Fig 8's two-axis plots place *energy efficiency* (left y -axis, in Mbit/J/Hz) and *total transmit power* (right y -axis, in W) against the operating SNR (in dB). As SNR increases, the *minimum* power needed to satisfy the per-user rate vector \mathbf{b}_{\min} decreases monotonically, while the achieved energy efficiency increases monotonically over the same SNR range.

In the indoor panel, energy efficiency rises steeply across the SNR sweep, entering the $60\text{--}70$ Mb/J/Hz band by SNR \approx

20 dB; over the same range, the total transmit power required to meet the target minimum rates collapses into the $20\text{--}50$ mW band. The curve transitions from near-zero efficiency at deep negative SNRs to roughly 65 ± 5 Mb/J/Hz at 20 dB, while the total power falls from the $0.20\text{--}0.25$ W bracket at -10 dB to about $0.02\text{--}0.05$ W at 20 dB. This corresponds to reduction in required power over the plotted operating window, consistent with minPMAC's ability to re-allocate energy and pick the globally optimal SIC order as the feasible MAC region expands with SNR.

In the urban-macro (UMa) plot, absolute scales are harsher, yet the qualitative behavior remains: energy efficiency climbs from near zero to approximately $0.5\text{--}0.6$ Mb/J/Hz by 20 dB, while the total power to satisfy the same rate targets drops from tens of Watts at low SNR to the 10 ± 2 W band at the high-SNR end.

For UMi, the energy-efficiency trace increases into the $10\text{--}12$ Mb/J/Hz range by 20 dB, while the total-power curve descends from the 30 ± 5 W region at -10 dB to roughly $5\text{--}8$ W at 20 dB. These readings imply an $>4\times$ reduction in required power along with a near $10\times$ rise in bits/J/Hz across the plotted SNRs. Compared with UMa, UMi achieves noticeably higher energy efficiency at the top end (owing to shorter ranges and improved spatial conditioning), while still delivering multi-Watt savings in total power when operated in the moderate-to-high SNR regime.

Across indoor, UMa, and UMi, the two-axis figures consistently show *monotonically decreasing* total power and *monotonically increasing* energy efficiency as SNR grows. Quantitatively, indoor achieves the largest absolute efficiency (peaking near ~ 65 Mb/J/Hz and reducing to a few tens of milliwatts), UMi lands in the $\sim 10\text{--}12$ Mb/J/Hz band with $\sim 5\text{--}8$ W at high SNR, and UMa reaches $\sim 0.5\text{--}0.6$ Mb/J/Hz with $\sim 8\text{--}12$ W. These gains are attributable to minPMAC's order-free convex allocation (energy vectors + subcarrier assignment) followed by dual-driven SIC ordering, which together convert the SNR expansion of the MAC region into lower Joules per delivered bit and, consequently, lower required transmit power at the same service targets.

IX. CONCLUSION AND FUTURE WORK

Our experiments yield three core lessons that go beyond architecture minutiae: (i) GRF succeeds not because it is

larger, but because a sparse, additive field of anisotropic Gaussians respects superposition, collapsing CSI synthesis to a single weighted sum and enabling sub-ms operation; (ii) continual prediction improves most when memory prioritizes hard states (deep fades, handovers), with LARS outperforming uniform replay by keeping failure modes in the buffer; and (iii) minPMAC decouples SIC order from covariance allocation and recovers the order from duals, delivering consistent energy savings and monotonic bits/J with SNR without combinatorial search. The most important empirical finding is multiplicative: GRF reduces pilot burden and stabilizes predictor targets; the predictor supplies look-ahead CSI that the scheduler can trust; and the scheduler's energy savings feed back into cheaper, more frequent re-estimation, yielding better than additive gains. Operationally, three knobs consistently drove the quality-latency trade: 1) the GRF refresh cadence Δ_{GRF} (balancing drift versus overhead), 2) the replay budget and its loss-aware policy (small buffers work if they retain hard samples), 3) and the minimum-rate vector \mathbf{b}_{\min} (which selects the active face of the MAC polymatroid and thus power scaling). Problems arise when engineers ignore how frequency affects signals or overlook near-field effects. Systems also struggle when individual signal processing becomes too complex for large networks. The key insights include that 6G designers must understand physics, adapt to changing conditions, and use stable mathematical structures. Our combined GRF, CL(LARS), and minPMAC approach puts this principle into practice.

Future work will (a) extend GRF to multi-cell cooperative twins that share compact summaries to improve edge throughput and handover stability, (b) implement uncertainty-aware, cross-time/frequency pilot placement to further cut overhead and speed re-acquisition after abrupt changes, (c) add wideband and near-field support via frequency-dependent primitives and spherical-wave rendering, (d) integrate robust and chance-constrained scheduling using predictor innovation covariances for reliability guarantees, and (e) co-design CUDA/ASIC kernels for GRF, replay, and log-det blocks to ensure sub-ms closed-loop operation on realistic gNB budgets.

REFERENCES

- [1] C. Jiang, J. Guo, X. Li, S. Jin, and J. Zhang, "Ai for csi prediction in 5g-advanced and beyond," *arXiv preprint arXiv:2504.12571*, 2025.
- [2] A. Fayad, T. Cinkler, and J. Rak, "Toward 6g optical fronthaul: A survey on enabling technologies and research perspectives," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 1, pp. 629–666, 2024.
- [3] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–1, 2023.
- [4] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (noma) for cellular future radio access," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, 2013, pp. 1–5.
- [5] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge University Press, 2005.
- [6] S. Bhattacharya, K. Rajabalifardi, M. A. Mohsin, R. Pote, and J. M. Cioffi, "Optimal power allocation and time sharing in low rank multi-carrier wi-fi channels," 2024. [Online]. Available: <https://arxiv.org/abs/2410.21782>
- [7] M. A. Mohsin, S. Bhattacharya, K. Rajabalifardi, R. Pote, and J. M. Cioffi, "Optimum power allocation for low rank wi-fi channels: A comparison with deep rl framework," 2025. [Online]. Available: <https://arxiv.org/abs/2501.11266>
- [8] R. D. Wesel and J. M. Cioffi, "Achievable rates for tomlinson-harashima precoding," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 824–831, 2002.
- [9] S. M. Hamedoon, J. N. Chattha, and M. Bilal, "Towards intelligent user clustering techniques for non-orthogonal multiple access: a survey," *EURASIP Journal on Wireless Communications and Networking*, vol. 2024, no. 1, p. 7, 2024.
- [10] M. Halabouni, M. Roslee, S. Mitani, O. Abuajwa, A. Osman, F. Z. Binti Ali, and A. Waseem, "Noma-mimo in 5g network: a detailed survey on enhancing data rate," *PeerJ Computer Science*, vol. 11, p. e2388, 2025.
- [11] V. Fanibhare, N. I. Sarkar, and A. Al-Anbuky, "A study of downlink power-domain non-orthogonal multiple access performance in tactile internet employing sensors and actuators," *Sensors*, vol. 24, no. 22, p. 7220, 2024.
- [12] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "On the sum rate of mimo-noma and mimo-oma systems," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 534–537, 2017.
- [13] M. Umer, M. A. Mohsin, S. A. Hassan, H. Jung, and H. Pervaiz, "Performance analysis of star-ris enhanced comp-noma multi-cell networks," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 2000–2005.
- [14] Z. Ding, F. Adachi, and H. V. Poor, "The application of mimo to non-orthogonal multiple access," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 537–552, 2016.
- [15] C. Choy Chai and X.-P. Zhang, "Iterative water-filling power and subcarrier allocation for multicarrier noma downlink," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.
- [16] M. Ghasemlou, N. Neda, and J. S. Harsini, "A novel resource allocation scheme for full-duplex noma systems with power splitting-based swipt," *Physical communication*, vol. 63, p. 102272, 2024.
- [17] X. Zhao, Z. An, Q. Pan, and L. Yang, "Nerf2: Neural radio-frequency radiance fields," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, ser. ACM MobiCom '23. ACM, Oct. 2023, p. 1–15. [Online]. Available: <http://dx.doi.org/10.1145/3570361.3592527>
- [18] H. Lu, C. Vatteuer, B. Mirzasoileman, and O. Abari, "Newrf: A deep learning framework for wireless radiation field reconstruction and channel prediction," 2024. [Online]. Available: <https://arxiv.org/abs/2403.03241>
- [19] C. Wen, J. Tong, Y. Hu, Z. Lin, and J. Zhang, "Neural representation for wireless radiation field reconstruction: A 3d gaussian splatting approach," 2025. [Online]. Available: <https://arxiv.org/abs/2412.04832>
- [20] M. Niemeyer, F. Manhardt, M.-J. Rakotosaona, M. Oechsle, D. Duckworth, R. Gosula, K. Tateno, J. Bates, D. Kaeser, and F. Tombari, "Radsplat: Radiance field-informed gaussian splatting for robust real-time rendering with 900+ fps," 2025. [Online]. Available: <https://arxiv.org/abs/2403.13806>
- [21] K. Yang, G. Dong, S. Ji, W. Du, and M. Srivastava, "Scalable 3d gaussian splatting-based rf signal spatial propagation modeling," 2025. [Online]. Available: <https://arxiv.org/abs/2502.01826>
- [22] Y. Yang, O. Alhussein, A. Arani, Z. Zhang, and M. Debbah, "Generative diffusion receivers: Achieving pilot-efficient mimo-ofdm communications," *arXiv preprint arXiv:2506.18419*, 2025.
- [23] J. M. Cioffi, C.-S. Hwang, and K. J. Kerpez, "Ergodic spectrum management," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1794–1821, 2020.
- [24] S. Zhang, S. Zhang, Y. Mao, L. K. Yeung, B. Clerckx, and T. Q. Quek, "Transformer-based channel prediction for rate-splitting multiple access-enabled vehicle-to-everything communication," *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 12 717–12 730, 2024.
- [25] G. Liu, Y. Xu, H. Zongjiang, Y. Rao, J. Xia, and L. Fan, "Deep learning-based channel prediction for edge computing networks toward intelligent connected vehicles," *IEEE Access*, vol. PP, pp. 1–1, 08 2019.
- [26] W. Jiang and H. D. Schotten, "Recurrent neural networks with long short-term memory for fading channel prediction," in *2020 IEEE 91st vehicular technology conference (VTC2020-Spring)*. IEEE, 2020, pp. 1–5.
- [27] H. Jiang, M. Cui, D. W. K. Ng, and L. Dai, "Accurate channel prediction based on transformer: Making mobility negligible," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2717–2732, 2022.
- [28] H. Kim, J. Choi, and D. J. Love, "Machine learning for future wireless communications: Channel prediction perspectives," *arXiv preprint arXiv:2502.18196*, 2025.
- [29] B. Liu, X. Liu, S. Gao, X. Cheng, and L. Yang, "Llm4cp: Adapting large language models for channel prediction," *Journal of Communications and Information Networks*, vol. 9, no. 2, pp. 113–125, 2024.

- [30] T. Lee, J. Park, H. Kim, and J. G. Andrews, "Generating high dimensional user-specific wireless channels using diffusion models," 2024. [Online]. Available: <https://arxiv.org/abs/2409.03924>
- [31] S. Bhattacharya, M. A. Mohsin, K. Rajabalifardi, and J. M. Cioffi, "Successive interference cancellation-aided diffusion models for joint channel estimation and data detection in low rank channel scenarios," 2025. [Online]. Available: <https://arxiv.org/abs/2501.11229>
- [32] N. Zilberstein, A. Swami, and S. Segarra, "Joint channel estimation and data detection in massive mimo systems based on diffusion models," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 13 291–13 295.
- [33] Z. Hu, Y. Li, and C. Han, "Transfer learning enabled transformer-based generative adversarial networks for modeling and generating terahertz channels," *Communications Engineering*, vol. 3, no. 1, p. 153, 2024.
- [34] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Non-orthogonal multiple access for 5g: Solutions, challenges, opportunities, and future research trends," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2332–2370, 2017.
- [35] S. M. R. Islam, M. Deng, Ö. A. Dobre, and K.-S. Kwak, "Non-orthogonal multiple access (noma): How it meets 5g and beyond," *Wiley 5G Ref*, 2019, also available as arXiv:1907.10001.
- [36] B. Di, L. Song, and Y. Li, "Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7686–7698, Nov. 2016.
- [37] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.
- [38] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels—part i: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.
- [39] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate in MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
- [40] H. Weingarten, Y. Steinberg, and S. S. (Shitz), "The capacity region of the gaussian MIMO broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
- [41] M. F. Khan, M. A. Mohsin, Z. Alam, M. Saad, and M. Waqar, "Machine learning-driven performance analysis of compressed communication in aerial-ris networks for future 6g networks," 2025. [Online]. Available: <https://arxiv.org/abs/2508.01911>
- [42] J. M. Cioffi, "Generalized decision feedback equalization, chapter 5," Technical Report / Lecture Notes, Stanford University, Department of Electrical Engineering, Tech. Rep., 2025, available: "Chapter 5" in *Generalized Decision Feedback Equalization* book, <https://cioffi-group.stanford.edu/doc/book/chap5.pdf>.
- [43] J. M. Cioffi and G. D. Forney, *Generalized Decision-Feedback Equalization for Packet Transmission with ISI and Gaussian Noise*. Boston, MA: Springer US, 1997, pp. 79–127. [Online]. Available: https://doi.org/10.1007/978-1-4615-6281-8_4
- [44] M. A. Mohsin, M. Umer, A. Bilal, M. A. Jamshed, and J. M. Cioffi, "Continual learning for wireless channel prediction," 2025. [Online]. Available: <https://arxiv.org/abs/2506.22471>
- [45] S. Bhattacharya, K. Rajabalifardi, M. A. Mohsin, and J. M. Cioffi, "Optimum power-subcarrier allocation and time-sharing in multicarrier noma uplink," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.
- [46] J. Zuo, Y. Liu, Z. Ding, L. Song, and H. V. Poor, "Joint design for simultaneously transmitting and reflecting (star) ris assisted noma systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 611–626, 2023.
- [47] Q. Sun, S. Han, C.-L. I, and Z. Pan, "On the ergodic capacity of mimo noma systems," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 405–408, 2015.
- [48] M. Umer, M. A. Mohsin, A. Bilal, and J. M. Cioffi, "Neural gaussian radio fields for channel estimation," 2025. [Online]. Available: <https://arxiv.org/abs/2508.11668>
- [49] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," 2020. [Online]. Available: <https://arxiv.org/abs/2003.08934>
- [50] D. Fan, R. Meng, X. Xu, Y. Liu, G. Nan, C. Feng, S. Han, S. Gao, B. Xu, D. Niyato *et al.*, "Generative diffusion models for wireless networks: Fundamental, architecture, and state-of-the-art," *arXiv preprint arXiv:2507.16733*, 2025.
- [51] U. Mall, B. Hariharan, and K. Bala, "Change-aware sampling and contrastive learning for satellite images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5261–5270.
- [52] L. Kumari, S. Wang, T. Zhou, and J. A. Bilmes, "Retrospective adversarial replay for continual learning," *Advances in neural information processing systems*, vol. 35, pp. 28 530–28 544, 2022.