G-WAVE: BIT-TO-WAVEFORM GENERATION FOR ADAPTIVE WIRELESS COMMUNICATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Wireless waveforms are traditionally hand-designed for general propagation conditions, rather than adapted to each specific environment. Existing neural transceivers seek such adaptation, but often rely on relatively simple architectures with limited long-range modeling capability. More expressive generative models such as Transformers are difficult to apply directly because wireless communication demands extremely accurate bit recovery, often with BER below 0.1%, while providing no ground-truth waveform supervision. To address these limitations, we introduce G-Wave, a first generative transceiver that directly synthesizes environment-adaptive physical time-domain waveforms from bits. G-Wave consists of a new Transformer-based transmitter and receiver models equipped with a multiscale architecture that combines long-range sequence modeling with fine-grained signal synthesis to meet the stringent precision required for communication. The two models are jointly optimized end-to-end using only bit-reconstruction error, allowing the waveform itself to be discovered without waveform labels. Across diverse 3GPP wireless environments, G-Wave achieves lower BER than the SOTA neural transceiver (DeepOFDM) and a strong conventional baseline (OFDM) under estimated channel information, reducing mean BER by 42.4% and 72.6%, respectively.

1 INTRODUCTION

Wireless communication is a cornerstone of modern society, enabling information exchange among vast numbers of interconnected devices (International Telecommunication Union, 2025). At the heart of this connectivity is *waveform design*, which determines how digital information is embedded in physical signals for transmission over wireless channels (Tse & Viswanath, 2005). This representation directly affects how reliably the information can be recovered after the transmitted signals undergo channel distortion and noise (Ait Aoudia & Hoydis, 2022b). Modern waveforms such as orthogonal frequency-division multiplexing (OFDM) are the result of decades of theory development and engineering refinement, combining modulation, orthogonal frequency-domain signaling, and Fourier transforms into a robust communication framework (Weinstein & Ebert, 1971; Tse & Viswanath, 2005). However, these waveforms are largely designed as general-purpose solutions across diverse propagation conditions, rather than being optimized for each specific wireless environment. This motivates a question: *can we train a model to automatically discover optimal waveforms tailored to a given environment and communication objective?*

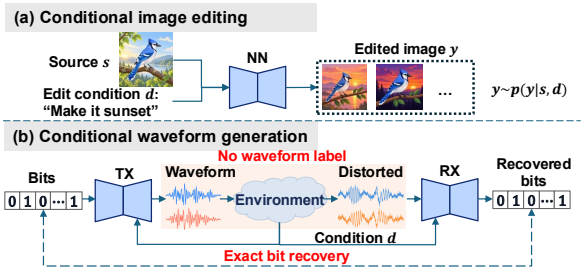


Figure 1: Comparison between conventional conditional generation and bit-to-waveform-to-bit wireless communication.

Recent advances in generative modeling offer a promising path toward such environment-adaptive waveform design through condition-guided generation and structural discovery (Rombach et al., 2022; Peebles & Xie, 2023; Watson et al., 2023). From our perspective, environment-adaptive waveform design can be viewed as a condition-guided sequence transformation problem, where the bit sequence specifies the information to be conveyed, while the wireless environment controls how that information is physically realized as a time-domain waveform. Despite this conceptual similarity,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

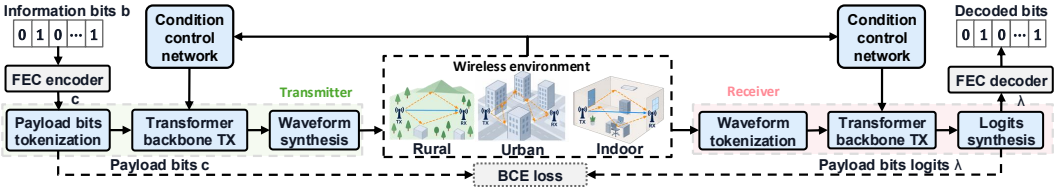


Figure 2: **Overview of G-Wave.** Environmental conditions modulate waveform synthesis and recovery; the illustrated propagation scenes denote alternative operating conditions. The transmitter and receiver are trained jointly through the bit-reconstruction loss.

existing conditional generative models for images, text, and audio cannot be directly transferred to communication waveform synthesis, which imposes substantially stricter requirements (Fig. 1). (1) Every payload bit must be recovered accurately (i.e., error rate $< 0.1\%$), whereas conventional generative tasks are typically evaluated at the semantic or perceptual level and can often tolerate stochastic variations in the generated output. (2) The communication chain traverses heterogeneous domains, from binary coded bits to complex-valued continuous waveforms and finally back to bits, whose statistical structures and functional roles differ substantially. (3) No ground-truth waveform specifies the desired physical realization for a given payload and propagation environment; the transmitter and receiver must therefore jointly discover the waveform representation using only the bit-reconstruction error backpropagated through the communication chain. We address these challenges with G-Wave, a wireless-environment-conditioned end-to-end generative transceiver (Fig. 2). The key contributions are:

- **End-to-end label-free bit-to-waveform generation.** We develop a fully differentiable generative transceiver that directly maps coded bits to physical waveforms and jointly optimizes the transmitter and receiver through the wireless channel. This end-to-end formulation removes the requirement for waveform labels, allowing the signaling waveform to be discovered only from bit-reconstruction supervision. Moreover, by generating waveforms directly from bits without predefined modulation formats, G-Wave exposes a broader waveform design space for discovering better environment-adaptive signaling structures and decreasing the bit error rate.
- **Communication-inspired multiscale feature representation and precise synthesis.** Inspired by the multi-subcarrier organization of classical communication waveforms and recent coarse-to-fine generative modeling (Tian et al., 2024; Feng et al., 2026), we develop a new residual multiscale architecture that captures coarse-to-fine temporal information modulation structures, facilitating precise waveform synthesis and reliable bit recovery. Projectors map coded bits, complex IQ waveforms, and soft bit logits into continuous feature spaces, allowing these heterogeneous sequences to be processed and conditioned within a unified neural framework. Unlike token-based generative models that rely on discrete codebooks or stochastic token sampling, our architecture operates directly on continuous features, enabling precise waveform synthesis and bit estimation while remaining fully differentiable for end-to-end optimization.
- **Strong empirical performance with interpretability and ablations.** Under a controlled single-link evaluation, G-Wave achieves lower mean information-bit error rate (BER) than DeepOFDM (Ankireddy et al., 2025), DeepOFW (Greidi & Cohen, 2026), and conventional OFDM under both perfect and estimated channel state information (CSI). On the mixed-channel test distribution with estimated CSI, it reduces mean information-bit BER across the evaluated E_b/N_0 points by 42.4% relative to DeepOFDM and 72.6% relative to conventional OFDM. Across the mixed-channel and scenario-specific evaluations, G-Wave maintains reliable bit recovery under diverse propagation conditions, including estimated CSI, while synthesizing waveforms that respond to the supplied channel condition.

2 Related Work

End-to-End and Environment-Conditioned Learned Communication. Early learned communication systems modeled the transmitter and receiver as an autoencoder and demonstrated end-to-end optimization mainly over relatively simple channel models (O'Shea & Hoydis, 2017; Dörner et al., 2018; Ait Aoudia & Hoydis, 2019; Cammerer et al., 2020). For more challenging frequency-selective channels, subsequent work commonly introduced neural networks into established communication pipelines, replacing or optimizing selected modules (Ait Aoudia & Hoydis, 2022b; Felix et al.,

2018; Ait Aoudia & Hoydis, 2022a; Honkala et al., 2021; Ankireddy et al., 2025; Wu et al., 2022). Feedback-based approaches instead use receiver-side observations or channel-related feedback to adapt subsequent transmitted representations (Kim et al., 2018; Song et al., 2020; Shao et al., 2023; Kurka & Gündüz, 2020; Wu et al., 2024). Another line of work uses hypernetworks to map CSI into the parameters of neural transceivers (Goutay et al., 2020; Zhang et al., 2022; Greidi & Cohen, 2026). Nevertheless, these approaches generally retain designer-specified physical-layer signal structures and rely on relatively simple mechanisms for incorporating channel conditions, limiting their flexibility in channel-adaptive waveform generation. In contrast, G-Wave learns the bit-to-waveform mapping directly in a latent feature space, where wireless environment conditions are injected to modulate the learned representation and guide the synthesis of channel-adaptive complex time-domain waveforms.

Generative Models for Wireless Communication. Generative models have been applied to wireless communication mainly for RF signal and data synthesis (Shi et al., 2019; Karunaratne et al., 2021; Chi et al., 2024; Kaasaragadda & Kokalj-Filipovic, 2025), wireless channel modeling and estimation (Balevi & Andrews, 2021; Arvinte & Tamir, 2023; Fesl et al., 2024; Sengupta et al., 2023; Zhou et al., 2025), and semantic-level communication (Qiao et al., 2024; Erdemir et al., 2023; Wang et al., 2025; Fu et al., 2026; Chen et al., 2026). These approaches primarily synthesize RF data for augmentation and downstream tasks, infer channel-related information, or reconstruct semantically equivalent content, rather than directly generating a physical waveform that encodes a designated bit sequence. In contrast, our goal is precise bit-to-waveform generation, which lies beyond the scope of these existing approaches.

3 PRELIMINARIES: MATH MODEL FOR WIRELESS COMMUNICATION

A typical wireless communication system consists of a transmitter and a receiver connected through a wireless propagation channel. At the transmitter, bits information is first encoded into a complex-valued time-domain waveform represented by its in-phase and quadrature (IQ) samples. These digital IQ samples are converted by the RF front end into a real-valued waveform and transmitted over the air. During propagation, environmental effects, including multipath, attenuation, and noise, distort the transmitted waveform and produce a received IQ sequence. The receiver processes this sequence to characterize the propagation effects and recover the transmitted bits. This communication process can be described mathematically through the following sequence of mappings. Before waveform generation, a forward error correction (FEC) encoder maps the information bits $\boldsymbol{b} \in \{0,1\}^K$ to a coded sequence with structured redundancy,

$$\boldsymbol{c} = f_{\mathrm{enc}}(\boldsymbol{b}) \in \{0,1\}^C, \tag{1}$$

where K and C denote the information-bit and coded-bit lengths, respectively, with coding rate $K/C < 1$. The redundancy introduced by FEC enables the receiver to correct transmission errors during decoding. Then, the transmitter maps $\boldsymbol{c}$ to an unnormalized time-domain complex-baseband waveform $\widetilde{\boldsymbol{x}}$ and rescales it to satisfy the average transmit-power constraint:

$$\widetilde{\boldsymbol{x}} = f_{\eta}^{\mathrm{TX}}(\boldsymbol{c}) \in \mathbb{C}^N, \qquad \boldsymbol{x} = \rho\widetilde{\boldsymbol{x}}. \tag{2}$$

where f_{η}^{TX} denotes the transmitter waveform mapping parameterized by η, and N is the length of the transmitted waveform sequence $\boldsymbol{x}$. The normalization factor is $\rho = \sqrt{NP/\|\widetilde{\boldsymbol{x}}\|_2^2}$, which ensures that the transmitted waveform has average transmit power P. During over-the-air propagation, multipath propagation and additive complex Gaussian noise $\boldsymbol{n}$ distort the transmitted waveform, yielding the received baseband sequence

$$\boldsymbol{y} = \boldsymbol{h} \circledast \boldsymbol{x} + \boldsymbol{n}, \qquad \boldsymbol{n} \sim \mathcal{CN}(0, N_0\boldsymbol{I}), \tag{3}$$

where $\boldsymbol{h} \in \mathbb{C}^{L_h}$ denotes the complex channel impulse response (CIR), $\circledast$ denotes linear convolution, and N_0 is the noise variance. The receiver processes the received sequence $\boldsymbol{y}$ and produces a soft estimate for each coded bit:

$$\boldsymbol{\lambda} = f_{\xi}^{\mathrm{RX}}(\boldsymbol{y}) \in \mathbb{R}^C, \qquad \widehat{\boldsymbol{b}} = f_{\mathrm{dec}}(\boldsymbol{\lambda}), \tag{4}$$

where f_{ξ}^{RX} denotes the receiver processing pipeline parameterized by ξ, encompassing receiver-side operations such as channel estimation, equalization, and soft bit inference, and $\boldsymbol{\lambda}$ contains the resulting coded-bit logits. The function f_{dec} denotes the FEC decoder paired with the encoder f_{enc}, which uses these soft estimates to recover the information bits $\widehat{\boldsymbol{b}}$. The reliability of the complete communication pipeline is measured by the information-bit error rate (BER), defined as $\mathrm{BER} = d_{\mathrm{H}}(\boldsymbol{b}, \widehat{\boldsymbol{b}})/K$, where $d_{\mathrm{H}}(\cdot,\cdot)$ denotes the Hamming distance.

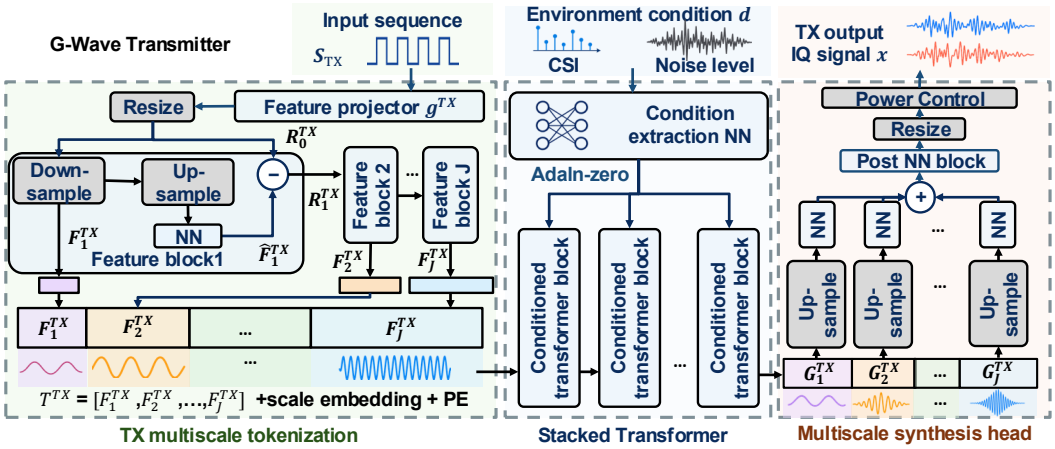


Figure 3: **G-Wave Multiscale neural transmitter.** Residual multiscale representation extracts coarse-to-fine features, which are concatenated with scale and positional embeddings and processed by an environment-conditioned bidirectional Transformer. The multiscale synthesis head then reconstructs and fuses the transformed features across scales before mapping them to the time-domain waveform.

4 G-WAVE: INTEGER-ALIGNED CONDITIONAL GENERATIVE TRANSCEIVER

4.1 SYSTEM MODEL AND END-TO-END PIPELINE

G-Wave parameterizes the transmitter and receiver as environment-conditioned neural networks:

$$\widetilde{\boldsymbol{x}} = f_\theta^{\text{NN-TX}}(\boldsymbol{c}, \boldsymbol{d}), \qquad \boldsymbol{\lambda} = f_\phi^{\text{NN-RX}}(\boldsymbol{y}, \boldsymbol{d}), \tag{5}$$

where $\boldsymbol{d}$ denotes the wireless environment descriptor constructed from the channel state information (CSI) available to the transceiver and the noise-variance information. θ and ϕ are the learnable parameters of the neural transmitter and receiver, respectively, and are jointly optimized end-to-end. Figure 2 provides an overview of the G-Wave architecture. The neural transmitter and receiver are independently parameterized but share a similar high-level design, consisting of three main components: a residual multiscale representation and synthesis interface, a bidirectional Transformer, and an environment-conditioning mechanism.

4.2 RESIDUAL MULTISCALE REPRESENTATION AND SYNTHESIS INTERFACE

To enable precise bit-preserving waveform synthesis, we design a fully differentiable multiscale representation-and-synthesis interface that can be optimized jointly with the transceiver. The design draws inspiration from the multicarrier structure of OFDM, where information-bearing symbols are distributed across orthogonal subcarriers and combined through an inverse Fourier transform to form a time-domain waveform. Frequency components at lower temporal frequencies contribute more slowly varying structures, whereas higher-frequency components introduce faster temporal variations. Accordingly, we seek a latent representation that captures temporal structure at multiple resolutions, from slowly varying global patterns to increasingly local variations. Combining this communication-inspired organization with recent coarse-to-fine generative modeling (Tian et al., 2024; Feng et al., 2026), we develop a residual multiscale representation-and-synthesis interface that progressively decomposes the latent sequence into coarse-to-fine features and reconstructs the transformed features in parallel for waveform synthesis.

Transmitter multiscale representation. To facilitate subsequent neural processing, the transmitter first converts the binary coded-bit sequence into a zero-centered numerical representation. Specifically, we center the coded bits as $\boldsymbol{s}_{\text{TX}} = \boldsymbol{c} - \frac{1}{2}\mathbf{1}_C$, yielding a one-channel sequence in $\{-\frac{1}{2}, \frac{1}{2}\}^{C\times 1}$. We next project the centered bit sequence into a latent feature space and align its sequence length with the transmitter latent grid used for waveform synthesis. This is implemented by a 1D convolutional feature projector followed by sequence-length resizing:

$$\boldsymbol{R}_0^{\text{TX}} = \mathcal{I}_{C\to L^{\text{TX}}}\left(g_{\text{TX}}(\boldsymbol{s}_{\text{TX}})\right) \in \mathbb{R}^{L^{\text{TX}}\times D}, \tag{6}$$

where g_{TX} lifts the scalar coded-bit sequence into a D-dimensional latent feature space, and $\mathcal{I}_{C\to L^{\text{TX}}}$ resizes the sequence length from C to L^{TX}. The resulting length-L^{TX} latent grid is used for the

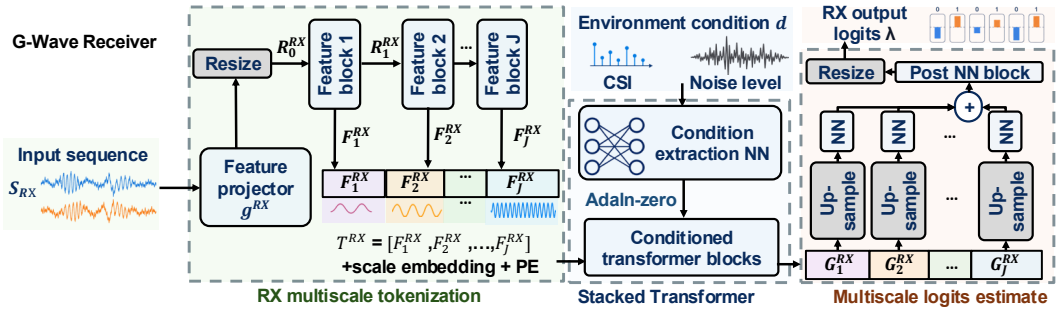


Figure 4: **Receiver architecture of G-Wave.** The received IQ sequence is projected into receiver-specific multiscale tokens, conditioned on CSI and noise variance through AdaLN-Zero Transformer blocks, and fused by the multiscale output head to produce coded-bit logits.

subsequent multiscale representation and is chosen to support integer-aligned waveform synthesis. Starting from $\boldsymbol{R}_0^{\mathrm{TX}}$, we decompose the latent representation into a coarse-to-fine hierarchy of J scales with lengths $\ell_1^{\mathrm{TX}} < \cdots < \ell_J^{\mathrm{TX}} = L^{\mathrm{TX}}$, where ℓ_1^{TX} denotes the coarsest scale and ℓ_J^{TX} corresponds to the full latent resolution. At each scale $j = 1, \ldots, J-1$, the current residual $\boldsymbol{R}_{j-1}^{\mathrm{TX}} \in \mathbb{R}^{L^{\mathrm{TX}} \times D}$ is first resized to length ℓ_j^{TX} to obtain the scale-specific feature $\boldsymbol{F}_j^{\mathrm{TX}} \in \mathbb{R}^{\ell_j^{\mathrm{TX}} \times D}$. The extracted feature is then projected back to the full latent resolution to form a reconstructed feature $\widehat{\boldsymbol{F}}_j^{\mathrm{TX}} \in \mathbb{R}^{L^{\mathrm{TX}} \times D}$, which is subtracted from the current residual:

$$\begin{aligned} \boldsymbol{F}_j^{\mathrm{TX}} &= \mathcal{I}_{L^{\mathrm{TX}} \to \ell_j^{\mathrm{TX}}}(\boldsymbol{R}_{j-1}^{\mathrm{TX}}), \quad \widehat{\boldsymbol{F}}_j^{\mathrm{TX}} = \phi_j^{\mathrm{TX}}\left(\mathcal{I}_{\ell_j^{\mathrm{TX}} \to L^{\mathrm{TX}}}(\boldsymbol{F}_j^{\mathrm{TX}})\right), \\ \boldsymbol{R}_j^{\mathrm{TX}} &= \boldsymbol{R}_{j-1}^{\mathrm{TX}} - \widehat{\boldsymbol{F}}_j^{\mathrm{TX}}, \quad j = 1, \ldots, J-1. \end{aligned} \tag{7}$$

After residual extraction, the multiscale features are concatenated from coarse to fine to form $Q^{\mathrm{TX}} = \sum_{j=1}^{J} \ell_j^{\mathrm{TX}}$ tokens. We add a learned scale embedding shared by tokens from the same scale, together with positional encoding (PE), before feeding the sequence into the bidirectional Transformer. The scale embedding identifies each token's temporal resolution, while PE provides its positional information. The resulting multiscale tokens are then jointly processed by the bidirectional Transformer, with the wireless environment condition injected through AdaLN-Zero (Peebles & Xie, 2023) to coordinate information across temporal positions and scales. As illustrated in Figure 3, this stage comprises the transmitter-side multiscale representation and environment-conditioned feature transformation, producing coordinated multiscale tokens for subsequent waveform synthesis.

Transmitter waveform-synthesis head. To reconstruct the physical waveform from the Transformer output, we follow the same multiresolution principle used in the input representation. The coordinated features at different scales encode complementary information, with coarse-scale features capturing global waveform structure and finer-scale features refining local temporal variations. Rather than decoding these scales serially, we reconstruct them in parallel on a common latent grid and combine their contributions to synthesize the final waveform. Specifically, the Transformer output $\boldsymbol{G}^{\mathrm{TX}} \in \mathbb{R}^{Q^{\mathrm{TX}} \times D}$ is partitioned according to the original scale boundaries as $\boldsymbol{G}^{\mathrm{TX}} = [\boldsymbol{G}^{\mathrm{TX},(1)}, \ldots, \boldsymbol{G}^{\mathrm{TX},(J)}]$, where $\boldsymbol{G}^{\mathrm{TX},(m)} \in \mathbb{R}^{\ell_m^{\mathrm{TX}} \times D}$ denotes the feature block associated with the m-th scale. Each block is resized to the full latent length L^{TX} and processed by a scale-specific branch ψ_m^{TX}, after which the reconstructed features are summed and jointly refined by a post-fusion block:

$$\boldsymbol{U}^{\mathrm{TX}} = \psi_{\mathrm{post}}^{\mathrm{TX}}\left(\sum_{m=1}^{J} \psi_m^{\mathrm{TX}}\left(\mathcal{I}_{\ell_m^{\mathrm{TX}} \to L^{\mathrm{TX}}}(\boldsymbol{G}^{\mathrm{TX},(m)})\right)\right) \in \mathbb{R}^{L^{\mathrm{TX}} \times D}. \tag{8}$$

The fused representation $\boldsymbol{U}^{\mathrm{TX}}$ is finally mapped to the physical waveform through an integer-aligned readout. Let $p_{\mathrm{TX}} = N/L^{\mathrm{TX}}$ denote the integer output ratio, such that each latent token directly generates p_{TX} complex IQ samples. The readout produces the corresponding real and imaginary components, which are paired into complex samples and concatenated along the temporal dimension to form the unnormalized waveform $\widetilde{\boldsymbol{x}} \in \mathbb{C}^N$.

Receiver-side adaptation. As illustrated in Figure 4, the receiver mirrors the high-level architecture of the transmitter while adapting its input and output interfaces to waveform-to-bit inference. The

received complex waveform is first represented by stacking its real and imaginary components along the channel dimension, $\boldsymbol{s}_{\mathrm{RX}} = [\mathrm{Re}(\boldsymbol{y}), \mathrm{Im}(\boldsymbol{y})] \in \mathbb{R}^{N_{\mathrm{RX}} \times 2}$, and projected into the receiver latent space. The resulting sequence is decomposed into coarse-to-fine multiscale features, concatenated into receiver tokens, and processed by a receiver-specific bidirectional Transformer, with the wireless environment condition injected through AdaLN-Zero in the same manner as at the transmitter. At the output, the transformed tokens are split by scale and lifted to a common latent resolution through the multiscale synthesis head; the scale-wise features are then fused, refined by a post-fusion neural block, and mapped to coded-bit logits through an integer-aligned readout. Let $p_{\mathrm{RX}} = C/L^{\mathrm{RX}}$, such that each latent position produces p_{RX} logits and the complete receiver output is $\boldsymbol{\lambda} \in \mathbb{R}^C$. These logits provide soft estimates of the transmitted coded bits and are subsequently used by the fixed channel decoder for information-bit recovery. Detailed pseudocode for the shared multiscale representation and output head is provided in Algorithms 1 and 2 in Appendix A.

Environment conditioning. To adapt the synthesized waveform and receiver inference to the current wireless environment, we condition both sides of the transceiver on CSI and noise-variance information. We represent CSI by the CIR taps $\boldsymbol{h}$ and convert these quantities into a real-valued environment descriptor:

$$\boldsymbol{d} = \left[\nu(N_0), \mathrm{Re}(\boldsymbol{h})^{\mathsf{T}}, \mathrm{Im}(\boldsymbol{h})^{\mathsf{T}}\right]^{\mathsf{T}} \in \mathbb{R}^{1+2L_h}, \tag{9}$$

where $\nu(\cdot)$ denotes a normalized logarithmic transformation of the noise variance. Under estimated CSI, $\widehat{\boldsymbol{h}}$ and $\widehat{N}_0$ replace $\boldsymbol{h}$ and N_0 only in the descriptor; physical propagation still uses $\boldsymbol{h}$. To effectively inject the environmental information into the latent representation, each side $a \in \{\mathrm{TX}, \mathrm{RX}\}$ employs a small condition-extraction MLP q^a to map the environment descriptor $\boldsymbol{d}$ to a condition embedding $\boldsymbol{e}^a = q^a(\boldsymbol{d})$. Here $\boldsymbol{e}^a$ is the resulting condition embedding used to modulate the corresponding encoder-only bidirectional Transformer through AdaLN-Zero. PE provides positional information for the multiscale sequence, while bidirectional self-attention captures dependencies across temporal positions and scales. The transmitter and receiver use independent conditioning networks and Transformer parameters.

End-to-end training and practical deployment. Since no target waveform is available for either the transmitter output or the received signal, the complete neural transceiver is trained directly through a differentiable wireless channel. At each training iteration, we sample a coded-bit sequence, a propagation channel, a noise variance N_0, and a noise realization $\boldsymbol{n}$. The transmitter synthesizes a waveform without waveform-level supervision, which is propagated through the sampled channel and processed by the receiver to produce coded-bit logits $\boldsymbol{\lambda}$. The transmitter and receiver are then jointly optimized using only the coded-bit reconstruction loss $\mathcal{L} = \mathcal{L}_{\mathrm{BCE}}(\boldsymbol{c}, \boldsymbol{\lambda})$. The channel coding scheme is fixed rather than learned: the encoder provides the coded sequence $\boldsymbol{c}$, while the channel decoder is used only for information-bit evaluation. Consequently, gradients propagate directly from the coded-bit loss through the receiver, differentiable channel, and synthesized waveform to the transmitter, without requiring intermediate waveform labels or separately pretrained components. At inference, the trained transmitter and receiver share an environment descriptor to synthesize the waveform and recover coded-bit logits; practical deployment details are given in Appendix B.

5 Experiments

5.1 Experimental setup

Link configuration. We evaluate a single-user, single-antenna frequency-selective link. Each frame contains 384 information bits encoded by a rate-$1/2$ 5G LDPC code into 768 coded bits. The transmitter emits 490 complex time-domain samples. Linear convolution with a 19-tap channel produces 508 received samples, which are zero-padded to 512 samples before neural reception. The sampling rate is 480 kHz. Propagation follows the clustered delay line (CDL) model in 3GPP TR 38.901 (3rd Generation Partnership Project, 2024), with channel normalization disabled.

Model and training. The G-Wave transmitter and receiver use multiscale latent lengths $(1, 7, 49, 245)$ and $(4, 16, 64, 256)$, respectively. Both sides contain a six-layer bidirectional Transformer with feature width 256, four attention heads, scale embeddings, PE, and AdaLN-Zero conditioning. We optimize it with Adam, batch size 16, learning rate 2×10^{-4}, and global gradient clipping at 1.0. At each training step, information bits are sampled i.i.d. from $\mathrm{Bernoulli}(1/2)$. For each batch, the CDL

profile is drawn uniformly from A–E and its RMS delay spread is drawn log-uniformly from 0.1 to $0.6\ \mu$s; terminal speed is fixed at 0 m/s. Training E_b/N_0 is sampled uniformly from $[0, 15]$ dB.

Baselines. We compare G-Wave against three representative baselines: conventional OFDM, DeepOFDM (Ankireddy et al., 2025), and DeepOFW (Greidi & Cohen, 2026). Conventional OFDM uses QPSK modulation and LMMSE equalization with either perfect CSI or pilot-based LS channel estimates. DeepOFDM jointly learns a convolutional modulator and neural receiver while retaining a predefined OFDM resource grid; for channel-aware comparison, we concatenate CSI with its transmitter grid features. DeepOFW instead uses a CSI-conditioned parameter generator to produce the waveform matrix and detector coefficients of a structured transceiver. To ensure a fair comparison, all methods use identical communication interfaces. All methods also share the same frame-energy budget and LDPC decoder for information-bit BER evaluation.

Evaluation protocol. Unless stated otherwise, all experiments follow the same paired evaluation protocol. Each BER point is evaluated using at least 16,000 frames. For configurations operating in the low-BER regime, the evaluation set is extended to 48,000 frames to improve statistical reliability. BER refers to information-bit BER after 20 LDPC decoding iterations and is reported over $E_b/N_0 \in \{0, 2, \ldots, 14, 15\}$ dB. For estimated CSI, all learned methods use the same 19-tap CIR estimate $\widehat{\boldsymbol{h}}$ obtained from a shared pilot observation given by OFDM. Tabulated mean BERs are arithmetic means over the evaluated E_b/N_0 points. In the tables, boldface marks the lowest observed BER in each BER column, including ties. A dash indicates that no post-decoding bit errors were observed over the evaluated frames, not that the underlying error probability is zero.

5.2 End-to-end BER performance

We first conduct an aggregate evaluation to evaluate the overall communication reliability of G-Wave across diverse propagation conditions and determine whether its gains persist beyond any particular channel profile or delay-spread regime. G-Wave and all baselines are evaluated over a mixed test distribution with randomly sampled CDL-A–E profiles and delay spreads across the considered E_b/N_0 range. We report results under both perfect and estimated CSI, with the latter examining whether the performance gains remain when channel-state information must be obtained through estimation. Figure 5 summarizes the resulting BER performance.

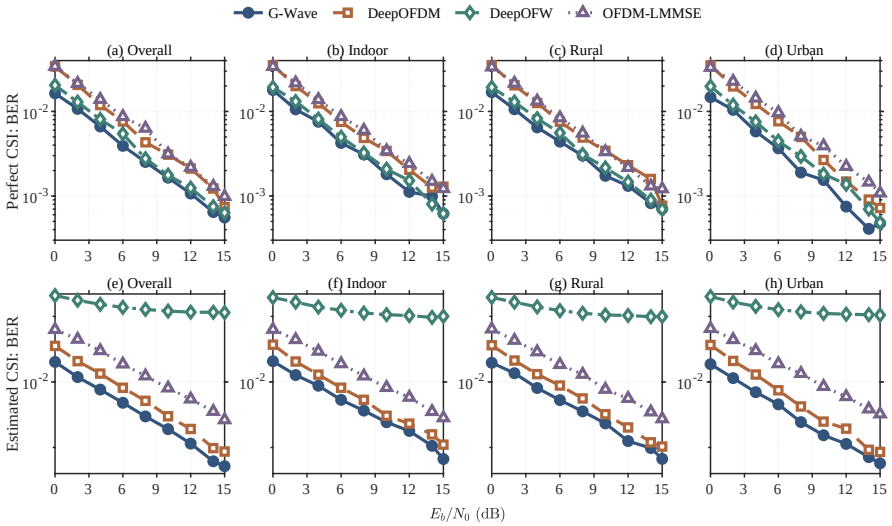


Figure 5: End-to-end information-bit BER under perfect CSI (a)–(d) and estimated CSI (e)–(h). Columns show the overall CDL setting, indoor office, RMa, and UMa, respectively.

G-Wave achieves the lowest mean information-bit BER on the mixed CDL-A–E test distribution under both CSI settings. Averaged over the evaluated E_b/N_0 points, its BER is 4.90×10^{-3} with perfect CSI and 5.74×10^{-3} with estimated CSI; the corresponding values for DeepOFDM are 9.52×10^{-3} and 9.96×10^{-3}. G-Wave thus retains low BER and its performance advantage with estimated CSI, indicating robustness to errors in the environmental descriptor. DeepOFW, by contrast, has mean BER 6.01×10^{-3} with perfect CSI but 1.41×10^{-1} with estimated CSI.

Scenario-specific generalization. We next evaluate whether G-Wave generalizes across three representative 3GPP deployment scenarios: indoor office, rural macro (RMa), and urban macro (UMa), which exhibit substantially different propagation characteristics. Within the CDL framework, these scenarios are instantiated using their characteristic RMS delay-spread ranges specified in 3GPP TR 38.901 (3rd Generation Partnership Project, 2024): 20–60 ns for indoor office, 32–153 ns for RMa, and 93–1148 ns for UMa. For each frame, we sample CDL-A–E uniformly and draw the RMS delay spread log-uniformly within the corresponding scenario range while keeping the trained transceivers fixed. As shown in Figure 5, G-Wave achieves the lowest mean BER in each scenario under both CSI settings. Averaging over the evaluated E_b/N_0 points and weighting the three scenarios equally, its mean BER is 4.95×10^{-3} under perfect CSI, compared with 5.90×10^{-3} for DeepOFW. With estimated CSI, the corresponding mean is 5.94×10^{-3} for G-Wave, versus 1.01×10^{-2} for DeepOFDM, 2.11×10^{-2} for conventional OFDM, and 1.28×10^{-1} for DeepOFW. Together, these results indicate that G-Wave provides a more effective and robust transceiver design across the evaluated propagation conditions, particularly when only estimated CSI is available.

Per-CDL-profile performance. To evaluate the consistency of G-Wave across a diverse set of standardized channel profiles, we separately test CDL-A–E and report the information-bit BER in Table 1. G-Wave achieves the lowest mean BER on CDL-A–C under both perfect and estimated CSI. On CDL-D/E, several methods have no observed post-decoding bit errors over the evaluated frames. These results suggest that G-Wave maintains reliable BER under each of the evaluated CDL profiles.

Table 1: Mean information-bit BER over the evaluated E_b/N_0 points for each CDL profile.

	Information-bit BER ($\times 10^{-3}$) ↓									
	CDL-A		CDL-B		CDL-C		CDL-D		CDL-E	
Method	(Perfect/Estimated)		(Perfect/Estimated)		(Perfect/Estimated)		(Perfect/Estimated)		(Perfect/Estimated)	
G-Wave (Ours)	**6.93**	**8.39**	**7.85**	**9.65**	**9.30**	**11.0**	—	—	—	—
DeepOFDM	13.6	14.6	18.1	19.3	17.8	18.6	2.35×10^{-4}	1.18×10^{-3}	—	—
DeepOFW	8.35	123	10.6	125	11.2	124	—	132	—	141
OFDM-LMMSE	14.3	29.4	19.9	40.8	19.2	37.7	1.03×10^{-3}	0.246	—	0.210

5.3 Channel-conditioned waveform adaptation

Shared spectral structure. To evaluate channel-specific adaptation, we first characterize the spectral structure shared across generated waveforms. Let $X(f)$ denote the frequency-domain representation of the transmitted waveform $\boldsymbol{x}$, where f denotes normalized frequency. We estimate a universal transmit-power template as $T_{\text{raw}}(f) = \mathbb{E}|X(f)|^2$ using 256 randomly sampled channels and 64 payloads per channel. As shown in Figure 6(a), the waveforms exhibit a common power pattern, while the coherent shared component $|\mathbb{E}X(f)|^2$ accounts for only 0.074% of the total power. This suggests that the generated waveforms are not dominated by a fixed preference for particular frequency regions; instead, most of the transmit-spectrum variation is shaped by the specific channel condition.

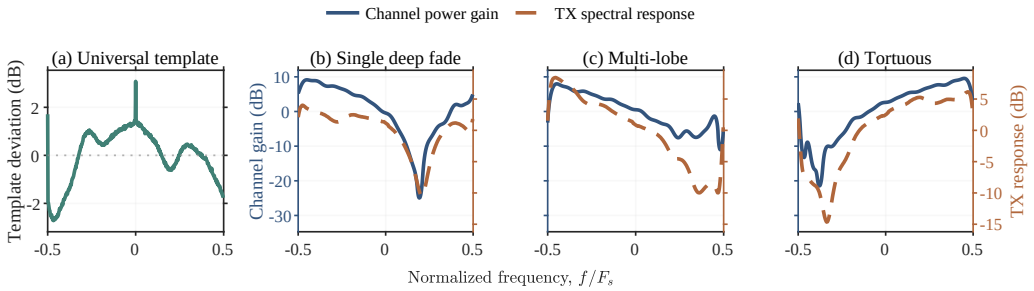


Figure 6: Channel-conditioned spectral adaptation learned by G-Wave. (a) Universal transmit-power template shared across channel realizations. Template-normalized TX spectral responses for three selected NLOS realizations: (b) CDL-B at $0.54\,\mu$s with a single deep fade; (c) CDL-A at $0.36\,\mu$s with a multi-lobe gain profile; and (d) CDL-C at $0.35\,\mu$s with a tortuous gain profile.

Channel-specific spectral adaptation. To better understand the BER gains of G-Wave over conventional OFDM, we examine how the transmit spectrum changes with channel quality. Using the shared spectral template as a reference, for each test channel we compute the payload-averaged spectrum

$P_{\boldsymbol{h}}(f) = \mathbb{E}_{\boldsymbol{c}}|X(f)|^2$ and its relative response $R_{\mathrm{dB}}(f) = 10\log_{10}(P_{\boldsymbol{h}}(f)/T_{\mathrm{raw}}(f))$. Figure 6(b)–(d) compares this response with the channel gain for three representative frequency-selective realizations. The gain response correlations are 0.93, 0.89, and 0.93 respectively. Across all cases, G-Wave allocates more transmit power to frequency regions with stronger channel gain while suppressing deeply faded regions, thereby directing the limited transmit energy toward more favorable parts of the channel. Such adaptive power allocation helps reduce severe signal degradation in deeply faded frequencies and facilitates more reliable delivery to the receiver. Further spectral-response results across CDL profiles and delay spreads are provided in Appendix D. Per-scale contributions and waveform-encoding/CSI-perturbation analyses are reported in Appendices E and F, respectively.

5.4 Ablation study

The full G-Wave model combines the residual multiscale, AdaLN-Zero conditioning, PE, and direct time-domain waveform generation. Each ablation changes one design dimension relative to this reference. *Single-scale* removes the serial coarse-to-fine residual decomposition and its associated tokenizer branches while retaining a single-scale neural interface. *No scale embedding* deletes only the additive scale-label vectors from the TX and RX tokenizers. *Concatenative conditioning* replaces layer-wise AdaLN-Zero modulation with broadcast condition features followed by dense reprojection, whereas *No conditioning* zeros the environment descriptor at both endpoints. *Frequency-domain I/O* preserves the neural backbone but inserts a unitary FFT/IFFT interface so that the network reads and writes frequency-domain coefficients. *OFDM resource-grid I/O* further constrains waveform generation to a prescribed OFDM grid with IFFT and cyclic-prefix framing. All variants use the same channel distribution, frame-energy constraint, and paired perfect-CSI evaluation protocol. Table 2 summarizes the results.

The ablation results collectively support the main design choices in G-Wave. Replacing the multiscale representation with a single-scale variant or removing scale embeddings consistently degrades BER, highlighting the value of explicit multiscale modeling. A similar trend appears in environment conditioning, where weaker or absent condition injection leads to poorer performance. Finally, constraining the transceiver with frequency-domain or OFDM-grid interfaces further reduces reliability, suggesting that direct bit-to-time-domain synthesis benefits from greater waveform-design freedom.

Table 2: Ablation under perfect CSI. BER entries are means across the evaluated E_b/N_0 points.

	Coded-bit BER ↓		Information-bit BER ↓	
Variant	Mean	× Full	Mean	× Full
G-Wave	$\mathbf{7.04\times10^{-3}}$	1.00	$\mathbf{4.57\times10^{-3}}$	1.00
Single-scale	7.79×10^{-3}	1.11	5.07×10^{-3}	1.11
No scale embedding	8.10×10^{-3}	1.15	5.21×10^{-3}	1.14
Concatenative conditioning	8.09×10^{-3}	1.15	5.24×10^{-3}	1.15
No conditioning	1.26×10^{-2}	1.79	9.22×10^{-3}	2.02
Frequency-domain I/O	7.78×10^{-3}	1.11	5.04×10^{-3}	1.10
OFDM resource-grid I/O	2.06×10^{-2}	2.93	9.69×10^{-3}	2.12

6 Conclusion

We introduced G-Wave, a new end-to-end generative transceiver that adapts waveform generation to diverse wireless channel conditions and enables highly reliable bit transmission with very low BER. Its key technical contributions are a Transformer-based multiscale bit-to-waveform architecture for precise long-range signal modeling and an end-to-end bit-supervised training framework that jointly learns waveform generation and recovery without waveform labels. Beyond wireless bit transmission, we envision that this framework could inspire generative learning systems for other applications that require both adaptive signal synthesis and extremely high output accuracy, such as secure communications, cybersecurity, and other precision-critical learning tasks.

AI USE STATEMENT

Generative AI tools assisted with manuscript drafting and editing, literature search, feedback on the methodology and experimental design, preparation of illustrative figure drafts and plotting code, and discussion of experimental results. The authors reviewed and revised AI-assisted materials and take responsibility for the final text, code, figures, analyses, and claims.

ETHICS STATEMENT

This work evaluates learned waveform design in simulated, single-user wireless links using standardized channel models and randomly generated bits. It does not involve human participants or personal data. The experiments do not assess over-the-air interference or coexistence with other spectrum users; practical deployment would require such evaluation and compliance with applicable spectrum regulations.

REPRODUCIBILITY STATEMENT

Section 5.1 specifies the link configuration, channel sampling, model and training settings, baselines, and evaluation protocol. Appendix A provides pseudocode for the multiscale components, while the appendix also documents comparison-model details and the inference procedure. The experiments use randomly generated bits and simulated channels rather than a private dataset. We will publicly release the implementation and evaluation code after the paper is accepted.

REFERENCES

3rd Generation Partnership Project. Study on channel model for frequencies from 0.5 to 100 GHz. Technical Report TR 38.901, 3rd Generation Partnership Project (3GPP), January 2024. URL https://www.etsi.org/deliver/etsi_tr/138900_138999/138901/17.01.00_60/tr_138901v170100p.pdf. Version 17.1.0, Release 17.

Fayçal Ait Aoudia and Jakob Hoydis. Model-free training of end-to-end communication systems. *IEEE Journal on Selected Areas in Communications*, 37(11):2503–2516, 2019. doi: 10.1109/JSAC.2019.2933891.

Fayçal Ait Aoudia and Jakob Hoydis. End-to-end learning for OFDM: From neural receivers to pilotless communication. *IEEE Transactions on Wireless Communications*, 21(2):1049–1063, 2022a. doi: 10.1109/TWC.2021.3101364.

Fayçal Ait Aoudia and Jakob Hoydis. Waveform learning for next-generation wireless communication systems. *IEEE Transactions on Communications*, 70(6):3804–3817, 2022b. doi: 10.1109/TCOMM.2022.3164060.

Sravan Kumar Ankireddy, S Ashwin Hebbar, Pramod Viswanath, and Hyeji Kim. Deep-ofdm: Neural modulation for high mobility. *arXiv e-prints*, pp. arXiv–2506, 2025.

Marius Arvinte and Jonathan I. Tamir. MIMO channel estimation using score-based generative models. *IEEE Transactions on Wireless Communications*, 22(6):3698–3713, 2023. doi: 10.1109/TWC.2022.3220784.

Eren Balevi and Jeffrey G. Andrews. Wideband channel estimation with a generative adversarial network. *IEEE Transactions on Wireless Communications*, 20(5):3049–3060, 2021. doi: 10.1109/TWC.2020.3047100.

Sebastian Cammerer, Fayçal Ait Aoudia, Sebastian Dörner, Maximilian Stark, Jakob Hoydis, and Stephan ten Brink. Trainable communication systems: Concepts and prototype. *IEEE Transactions on Communications*, 68(9):5489–5503, 2020. doi: 10.1109/TCOMM.2020.3002915.

Xingyu Chen, Zihao Feng, Wuqiong Zhao, Jianrong Ding, Ke Sun, and Xinyu Zhang. From bits to tokens: Knowledge-driven generative communication of multimodal data. In *23rd USENIX Symposium on Networked Systems Design and Implementation (NSDI 26)*, pp. 191–207. USENIX Association, 2026. URL https://www.usenix.org/conference/nsdi26/presentation/chen-xingyu.

Guoxuan Chi, Zheng Yang, Chenshu Wu, Jingao Xu, Yuchong Gao, Yunhao Liu, and Tony Xiao Han. RF-Diffusion: Radio signal generation via time-frequency diffusion. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pp. 77–92, 2024. doi: 10.1145/3636534.3649348.

Sebastian Dörner, Sebastian Cammerer, Jakob Hoydis, and Stephan ten Brink. Deep learning based communication over the air. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):132–143, 2018.

Ecenaz Erdemir, Tze-Yang Tung, Pier Luigi Dragotti, and Deniz Gündüz. Generative joint source-channel coding for semantic image transmission. *IEEE Journal on Selected Areas in Communications*, 41(8):2645–2657, 2023. doi: 10.1109/JSAC.2023.3288243.

Alexander Felix, Sebastian Cammerer, Sebastian Dörner, Jakob Hoydis, and Stephan ten Brink. OFDM-autoencoder for end-to-end learning of communications systems. In *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2018. doi: 10.1109/SPAWC.2018.8445920.

Shibo Feng, Zhicheng Chen, Xi Xiao, Zhong Zhang, Qing Li, Xingyu Gao, and Peilin Zhao. MSDformer: Multi-scale discrete transformer for time series generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(8):10129–10142, 2026. doi: 10.1109/TPAMI.2026.3681981.

Benedikt Fesl, Michael Baur, Florian Strasser, Michael Joham, and Wolfgang Utschick. Diffusion-based generative prior for low-complexity MIMO channel estimation. *IEEE Wireless Communications Letters*, 13(12):3493–3497, 2024. doi: 10.1109/LWC.2024.3474570.

Jingwen Fu, Ming Xiao, Mikael Skoglund, and Dong In Kim. Land-then-transport: A flow matching-based generative decoder for wireless image transmission, 2026. URL https://arxiv.org/abs/2601.07512.

Mathieu Goutay, Fayçal Ait Aoudia, and Jakob Hoydis. Deep hypernetwork-based MIMO detection. In *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2020. doi: 10.1109/SPAWC48557.2020.9154283.

Ran Greidi and Kobi Cohen. DeepOFW: Deep learning-driven OFDM-flexible waveform modulation for peak-to-average power ratio reduction, 2026.

Mikko Honkala, Dani Korpi, and Janne MJ Huttunen. DeepRx: Fully convolutional deep learning receiver. *IEEE Transactions on Wireless Communications*, 20(6):3925–3940, 2021.

International Telecommunication Union. Measuring digital development: Facts and figures 2025. Technical report, International Telecommunication Union, 2025. URL https://www.itu.int/hub/publication/d-ind-ict_mdd-2025-3/.

Yagna Kaasaragadda and Silvija Kokalj-Filipovic. ReFormer: Generating radio fakes for data augmentation, 2025. URL https://arxiv.org/abs/2501.00282.

Samurdhi Karunaratne, Samer S. Hanna, and Danijela Cabric. Open set RF fingerprinting using generative outlier augmentation. In *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, 2021. doi: 10.1109/GLOBECOM46510.2021.9685335.

Hyeji Kim, Yihan Jiang, Sreeram Kannan, Sewoong Oh, and Pramod Viswanath. Deepcode: Feedback codes via deep learning. In *Advances in Neural Information Processing Systems*, volume 31, pp. 9458–9468, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/31f81674a348511b990af268ca3a8391-Abstract.html.

David Burth Kurka and Deniz Gündüz. DeepJSCC-f: Deep joint source-channel coding of images with feedback. *IEEE Journal on Selected Areas in Information Theory*, 1(1):178–193, 2020. doi: 10.1109/JSAIT.2020.2987203.

Timothy O'Shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):563–575, 2017.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

Li Qiao, Mahdi Boloursaz Mashhadi, Zhen Gao, Chuan Heng Foh, Pei Xiao, and Mehdi Bennis. Latency-aware generative semantic communications with pre-trained diffusion models. *IEEE Wireless Communications Letters*, 13(10):2652–2656, 2024. doi: 10.1109/LWC.2024.3429295.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

Ushnish Sengupta, Chinkuo Jao, Alberto Bernacchia, Sattar Vakili, and Da-shan Shiu. Generative diffusion models for radio wireless channel modelling and sampling. In *GLOBECOM 2023-2023 IEEE Global Communications Conference*, pp. 4779–4784. IEEE, 2023.

Yulin Shao, Emre Ozfatura, Alberto Perotti, Branislav M. Popovic, and Deniz Gündüz. AttentionCode: Ultra-reliable feedback codes for short-packet communications. *IEEE Transactions on Communications*, 71(8):4437–4452, 2023. doi: 10.1109/TCOMM.2023.3280563.

Yi Shi, Kemal Davaslioglu, and Yalin E. Sagduyu. Generative adversarial network for wireless signal spoofing. In *Proceedings of the ACM Workshop on Wireless Security and Machine Learning*, pp. 55–60, 2019. doi: 10.1145/3324921.3329695.

Jinxiang Song, Bile Peng, Christian Häger, Henk Wymeersch, and Anant Sahai. Learning physical-layer communication with quantized feedback. *IEEE Transactions on Communications*, 68(1): 645–653, 2020. doi: 10.1109/TCOMM.2019.2951563.

Keyu Tian, Yi Jiang, Zehuan Yuan, Bingyue Peng, and Liwei Wang. Visual autoregressive modeling: Scalable image generation via next-scale prediction. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005. doi: 10.1017/CBO9780511807213.

Sixian Wang, Jincheng Dai, Kailin Tan, Xiaoqi Qin, Kai Niu, and Ping Zhang. DiffCom: Channel received signal is a natural condition to guide diffusion posterior sampling. *IEEE Journal on Selected Areas in Communications*, 43(7):2651–2666, 2025. doi: 10.1109/JSAC.2025.3559158.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620:1089–1100, 2023. doi: 10.1038/s41586-023-06415-8.

S. B. Weinstein and Paul M. Ebert. Data transmission by frequency-division multiplexing using the discrete fourier transform. *IEEE Transactions on Communication Technology*, 19(5):628–634, 1971. doi: 10.1109/TCOM.1971.1090705.

Haotian Wu, Yulin Shao, Krystian Mikolajczyk, and Deniz Gündüz. Channel-adaptive wireless image transmission with OFDM. *IEEE Wireless Communications Letters*, 11(11):2400–2404, 2022. doi: 10.1109/LWC.2022.3204837.

Haotian Wu, Yulin Shao, Emre Ozfatura, Krystian Mikolajczyk, and Deniz Gündüz. Transformer-aided wireless image transmission with channel feedback. *IEEE Transactions on Wireless Communications*, 23(9):11904–11919, 2024. doi: 10.1109/TWC.2024.3386052.

Jing Zhang, Chao-Kai Wen, and Shi Jin. Adaptive MIMO detector based on hypernetwork: Design, simulation, and experimental test. *IEEE Journal on Selected Areas in Communications*, 40(1): 65–81, 2022. doi: 10.1109/JSAC.2021.3126064.

Xingyu Zhou, Le Liang, Jing Zhang, Peiwen Jiang, Yong Li, and Shi Jin. Generative diffusion models for high dimensional channel estimation. *IEEE Transactions on Wireless Communications*, 24(7): 5840–5854, 2025. doi: 10.1109/TWC.2025.3549592.

A MULTISCALE TOKENIZER AND SYNTHESIS ALGORITHMS

Algorithms 1 and 2 give the shared TX/RX steps for the multiscale tokenizer and output head, respectively. The side index a selects the corresponding input, latent lengths, and learned parameters.

Algorithm 1: Shared multiscale tokenizer

Require: side $a \in \{\mathrm{TX}, \mathrm{RX}\}$; input $\boldsymbol{s}_a$ of length L^a_{in}
Require: latent length L^a; scales $\ell^a_1 < \cdots < \ell^a_J = L^a$
$\boldsymbol{R}^a_0 \leftarrow \mathcal{I}_{L^a_{\mathrm{in}} \to L^a}(g^a(\boldsymbol{s}_a))$
for $j = 1, \ldots, J-1$ **do**
  $\boldsymbol{F}^a_j \leftarrow \mathcal{I}_{L^a \to \ell^a_j}(\boldsymbol{R}^a_{j-1})$
  $\widehat{\boldsymbol{F}}^a_j \leftarrow \phi^a_j(\mathcal{I}_{\ell^a_j \to L^a}(\boldsymbol{F}^a_j))$
  $\boldsymbol{R}^a_j \leftarrow \boldsymbol{R}^a_{j-1} - \widehat{\boldsymbol{F}}^a_j$
end for
$\boldsymbol{F}^a_J \leftarrow \boldsymbol{R}^a_{J-1}$
$\boldsymbol{T}^a \leftarrow [\boldsymbol{F}^a_1 \| \cdots \| \boldsymbol{F}^a_J] + \boldsymbol{E}^a_{\mathrm{scale}} + \boldsymbol{E}^a_{\mathrm{pos}}$
return $\boldsymbol{T}^a$ and $\{\ell^a_j\}_{j=1}^J$

Algorithm 2: Shared multiscale output head

Require: side $a \in \{\mathrm{TX}, \mathrm{RX}\}$; coordinated tokens $\boldsymbol{G}^a$
Require: scales $\{\ell^a_m\}_{m=1}^J$; latent length L^a
Require: $p_{\mathrm{TX}} = N/L^{\mathrm{TX}}$; $p_{\mathrm{RX}} = C/L^{\mathrm{RX}}$
$\{\boldsymbol{G}^{a,(m)}\}_{m=1}^J \leftarrow \textsc{SplitByScale}(\boldsymbol{G}^a, \{\ell^a_m\})$
$\boldsymbol{U}^a \leftarrow \psi^a_{\mathrm{post}}\Big(\sum_{m=1}^J \psi^a_m\Big(\mathcal{I}_{\ell^a_m \to L^a}(\boldsymbol{G}^{a,(m)})\Big)\Big)$
if $a = \mathrm{TX}$ **then**
  $\boldsymbol{o}^a \leftarrow \mathrm{ComplexReshape}(\mathrm{Linear}_{2p_{\mathrm{TX}}}(\boldsymbol{U}^a))$
else
  $\boldsymbol{o}^a \leftarrow \mathrm{Reshape}(\mathrm{Linear}_{p_{\mathrm{RX}}}(\boldsymbol{U}^a))$
end if
return $\boldsymbol{o}^a$, where $\boldsymbol{o}^{\mathrm{TX}} = \widetilde{\boldsymbol{x}}$ and $\boldsymbol{o}^{\mathrm{RX}} = \boldsymbol{\lambda}$

B INFERENCE AND PRACTICAL DEPLOYMENT

This appendix details the inference and deployment procedure for G-Wave. Deployment does not require waveform labels, backpropagation, or online parameter updates: the trained neural transmitter and receiver are fixed, and environmental information enters only through their condition embeddings.

Condition acquisition. At the beginning of a condition-update interval, the receiver estimates the channel impulse response $\widehat{\boldsymbol{h}}$ and noise variance $\widehat{N}_0$. A conventional estimator can obtain these quantities from a known preamble or OFDM pilot symbols. The receiver then constructs $\widehat{\boldsymbol{d}} = [\nu(\widehat{N}_0), \mathrm{Re}(\widehat{\boldsymbol{h}})^{\mathsf{T}}, \mathrm{Im}(\widehat{\boldsymbol{h}})^{\mathsf{T}}]^{\mathsf{T}}$ using the same ordering and normalization as during training, and G-Wave uses this fixed-length descriptor as its condition. Perfect-CSI evaluation supplies the true descriptor, whereas estimated-CSI evaluation constructs it from the shared pilot observation described in Section 5.1.

Descriptor exchange and buffering. When the channel varies slowly relative to the payload transmission rate, the receiver can send its estimated descriptor $\widehat{\boldsymbol{d}}$ to the transmitter over a control link. The transmitter stores the latest shared descriptor in a condition buffer, and both endpoints reuse it for subsequent payload frames. After one or more frames, the receiver can refresh its estimate and send an updated descriptor; both endpoints switch to the new value at an agreed update boundary. In this way, condition feedback can be less frequent than payload transmission while the transmitter uses the buffered descriptor to generate each waveform. The update policy, feedback quantization, and delay are protocol choices outside the learned transceiver.

Payload transmission. For each payload, the fixed channel encoder produces $\boldsymbol{c}$, and the neural transmitter generates the complete complex waveform in one conditioned forward pass followed by frame-energy normalization. After propagation, the neural receiver directly stacks the real and imaginary parts of the received sequence, applies its conditioned forward pass, and outputs the coded-bit logits $\boldsymbol{\lambda}$. The fixed channel decoder then recovers the information bits. Thus, condition acquisition and descriptor exchange occur on the control path, whereas bit-to-waveform synthesis and bit recovery remain on the payload path.

C COMPARISON-MODEL DETAILS

All methods use the same information and coded-bit lengths, channel distribution, transmit-energy normalization, and 490-sample frame budget. DeepOFDM uses its released two-dimensional residual-

grid transmitter and receiver. DeepOFW is reconstructed from the released channel-conditioned waveform generator and uses five blocks of 80 waveform samples with an 18-sample cyclic prefix per block, ensuring that its guard interval covers the complete channel memory. The conventional OFDM baseline uses 32 subcarriers, 14 symbols, a three-sample cyclic prefix, QPSK, and LMMSE equalization; we report both perfect-CSI and pilot-estimated variants. Coded BER is measured directly from receiver logits, and information BER is measured after the same 20-iteration LDPC decoder for every method.

D SPECTRAL RESPONSE ACROSS CDL PROFILES AND DELAY SPREADS

To examine whether the channel-dependent transmit-spectrum response in Figure 6 extends beyond the selected high-selectivity examples, we evaluate one random channel realization for each combination of CDL-A–E and RMS delay spreads of 0.10, 0.25, and $0.60\,\mu$s. We use the same universal power template $T_{\mathrm{raw}}(f)$ as in Section 5.3, estimated from 256 channels and 64 random-bit frames per channel. For each test channel at $10\,\mathrm{dB}$ E_b/N_0, we average the transmit power over 2048 random-bit frames on the native 490-bin whole-frame Fourier grid and compute $R_{\mathrm{dB}}(f) = 10\log_{10}(P_{\boldsymbol{h}}(f)/T_{\mathrm{raw}}(f))$. We compare it with the channel power gain $G_{\mathrm{dB}}(f) = 10\log_{10}|H(f)|^2$. After excluding the bins around DC and Nyquist, we mean-center both in-band vectors and calculate their Pearson correlation r, the regression slope of R_{dB} on G_{dB}, and the gain span $\max G_{\mathrm{dB}} - \min G_{\mathrm{dB}}$. The metrics use unsmoothed data; only the displayed curves use 15-bin circular smoothing and fixed axes across panels.

Figure 7 shows that the response depends on the particular channel realization, not just its CDL profile or nominal delay spread. Five of the seven panels with a gain span of at least $2\,\mathrm{dB}$ have $r \geq 0.6$, including CDL-B at $0.25\,\mu$s ($r = 0.77$, span $12.1\,\mathrm{dB}$) and CDL-D at $0.25\,\mu$s ($r = 0.82$, span $2.7\,\mathrm{dB}$). The relationship is not universal: CDL-A at $0.60\,\mu$s has $r = -0.07$ despite a $2.1\,\mathrm{dB}$ span, and CDL-E at $0.60\,\mu$s reaches only $r = 0.36$ with a $3.8\,\mathrm{dB}$ span. When the gain span is below $1\,\mathrm{dB}$, correlations and slopes are especially sensitive to small variations. Thus, the stronger correlations near 0.9 in Figure 6(b)–(d) characterize selected deep-fade realizations.

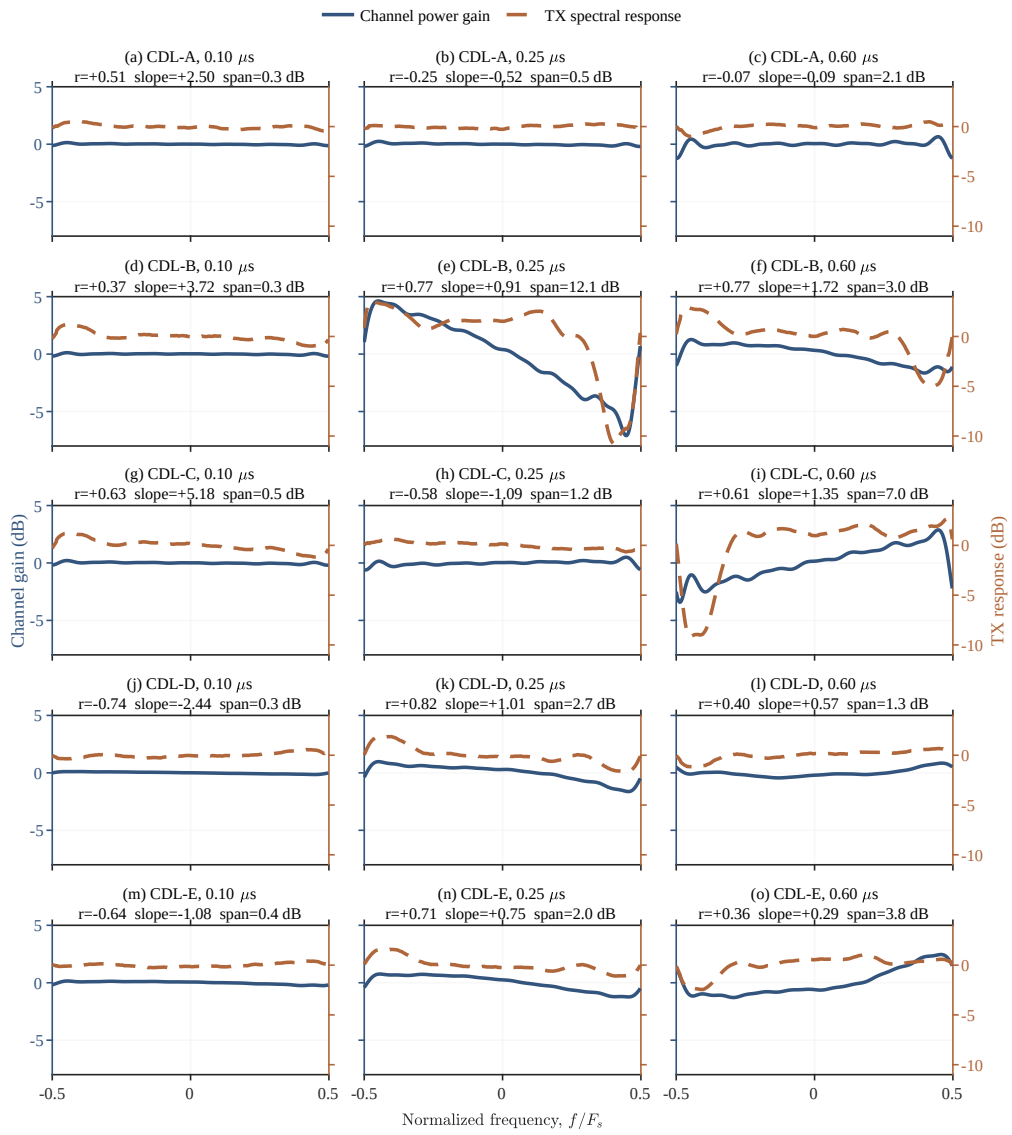


Figure 7: Channel gain and template-normalized transmit-spectrum response across CDL-A–E (rows) and three RMS delay spreads (columns). Each panel shows one channel realization. Titles report the in-band Pearson correlation r, regression slope, and gain span computed before smoothing; the plotted curves are smoothed for readability and share fixed axis limits.

E PER-SCALE CONTRIBUTION OF THE MULTISCALE SYNTHESIS HEADS

To examine how the trained model uses different temporal scales, we test the four branches of its synthesis head without changing their weights. Normally, each branch converts one scale's tokens to a common latent length, and their outputs are combined before the final readout. We first activate one branch at a time and measure the resulting TX waveform energy or RX logit magnitude. These measurements describe what each branch produces alone, not its fraction of the complete output. We then remove one branch at a time and measure the change in BER to assess how much bit recovery depends on that branch. When all four branches are active, this test reproduces the original model output exactly.

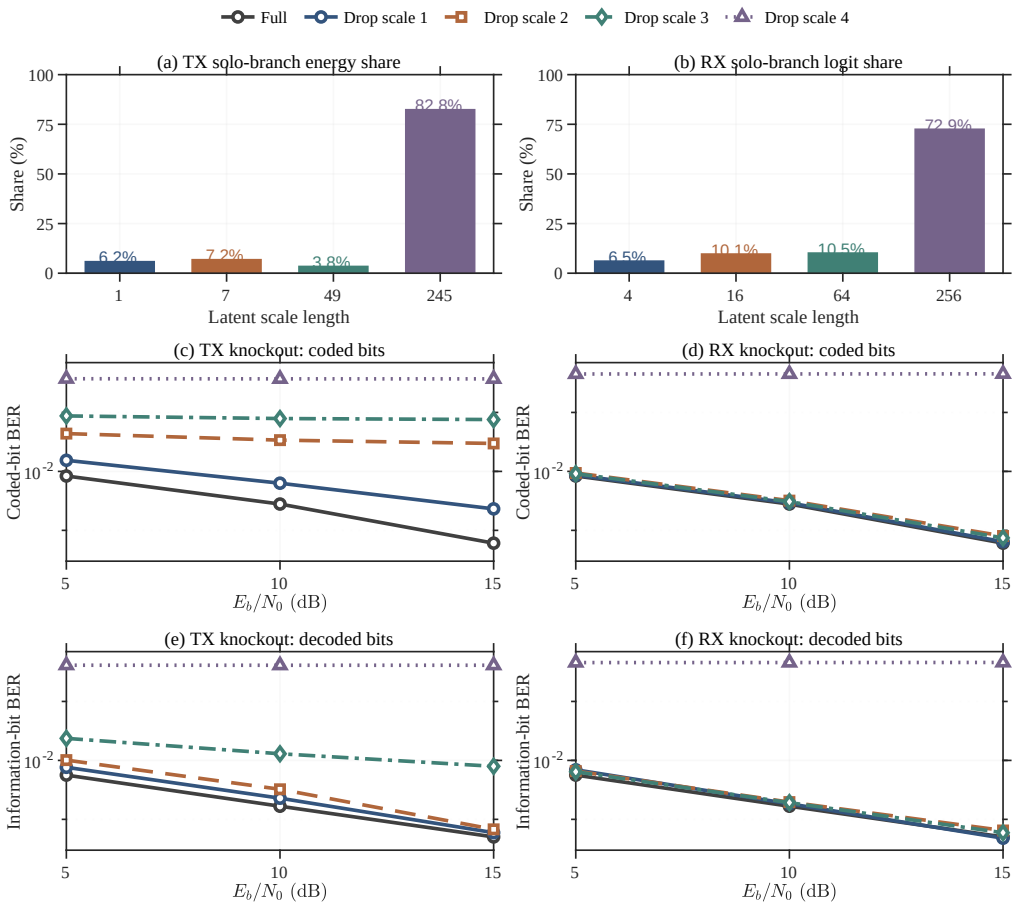


Figure 8: Contribution of individual multiscale-head branches. (a,b) Normalized solo-branch TX waveform-energy and RX mean-absolute-logit shares. (c,d) Coded-bit BER and (e,f) post-LDPC information-bit BER when one TX or RX branch is removed. Scale 1–4 correspond to latent lengths $(1, 7, 49, 245)$ at the TX and $(4, 16, 64, 256)$ at the RX. All knockout curves use the same trained weights and paired channel/noise pool.

The knockout evaluation uses a shared pool of 3200 frames per E_b/N_0 point at 5, 10, and 15 dB, with CDL-A–E sampled uniformly and RMS delay spread sampled log-uniformly from 0.1 to 0.6 μs. Figure 8(a)–(b) reports the single-branch measurements: when operated separately, the finest TX and RX branches produce 82.8% of the summed isolated TX waveform energy and 72.9% of the summed isolated RX absolute-logit magnitude, respectively. Removing a branch tests a different property. At 15 dB, disabling the TX branches with latent lengths 1, 7, 49, and 245 increases coded-bit BER by factors of 3.8, 49, 125, and 612 relative to the full model. Thus, even coarse TX branches with small isolated outputs are important for bit recovery. The post-LDPC information-bit BER is less affected when either of the first two TX branches is removed, consistent with error correction mitigating part of the coded-bit degradation. On the RX side, disabling the first three branches increases coded-bit BER by only 6%, 32%, and 23%, whereas removing the finest branch raises it to approximately 0.45. The trained RX readout is therefore much more sensitive to removal of the finest branch. These tests establish branch-level dependence.

F Influence and Perturbation Analysis

TX/RX influence. To characterize the encoding and decoding dependencies learned by G-Wave, we measure how much one output element changes when a single input element is perturbed, with the other inputs and environment descriptor held fixed. Let B_{TX} and B_{RX} denote the numbers of base coded-bit sequences used to probe the transmitter and received frames used to probe the receiver,

respectively. At the TX, each coded bit is flipped individually. We denote the change in its centered input value by Δ_b and the resulting waveform samples for base sequence i, with and without flipping bit j, by $x_t^{(i,j)}$ and $x_t^{(i)}$. The influence of bit j on complex transmit sample t, normalized by the input change, is

$$d_{\mathrm{TX}}(j,t) = \frac{1}{B_{\mathrm{TX}}\Delta_b} \sum_{i=1}^{B_{\mathrm{TX}}} \left| x_t^{(i,j)} - x_t^{(i)} \right|. \tag{10}$$

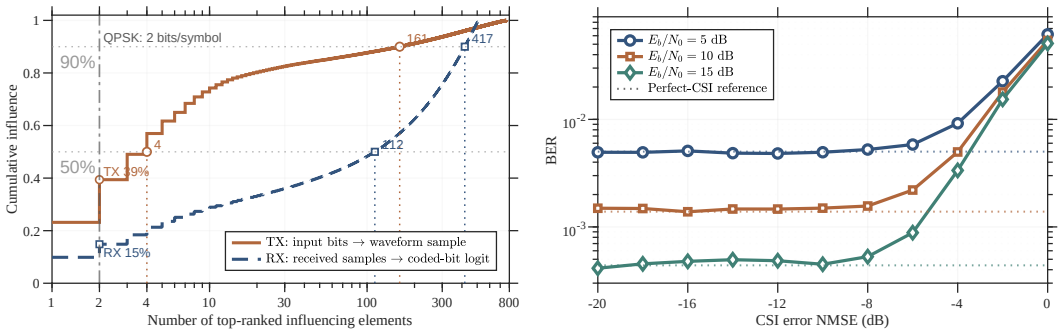
At the RX, we add a complex Gaussian perturbation δ_t to one received sample at a time, with RMS amplitude $\rho\ \mathrm{RMS}(\boldsymbol{y}^{(i)})$ for frame i. If $\lambda_j^{(i,t,\delta_t)}$ and $\lambda_j^{(i)}$ denote the coded-bit logits with and without this perturbation, the influence of received sample t on logit j is

$$d_{\mathrm{RX}}(j,t) = \frac{1}{B_{\mathrm{RX}}} \sum_{i=1}^{B_{\mathrm{RX}}} \mathbb{E}_{\delta_t} \left| \lambda_j^{(i,t,\delta_t)} - \lambda_j^{(i)} \right|, \tag{11}$$

where the expectation is estimated from the sampled perturbations. In this experiment, $B_{\mathrm{TX}} = 8$, $B_{\mathrm{RX}} = 32$, and $\rho = 0.05$. Since the centered TX input takes values $-\frac{1}{2}$ and $+\frac{1}{2}$, its bit-flip magnitude is $\Delta_b = 1$. We use one fixed CDL-D channel (seed 700, 0.21 μs delay spread) at 10 dB E_b/N_0, with $C = 768$ coded bits, $N = 490$ transmit samples, and $N_{\mathrm{RX}} = 508$ received samples. Flips and small additive perturbations probe the discrete TX input and continuous RX input, respectively; neither requires assumptions about the model internals.

For each of the N transmit samples, we rank its C bit influences in descending order, divide by their sum, and accumulate the top-k shares; the TX curve is the median of these cumulative curves across samples. Likewise, for each of the C logits, we rank and normalize the influences of the N_{RX} received samples, then take the median cumulative curve across logits. A common multiplicative normalization of either influence matrix cancels in these shares. Thus, influence measures finite-perturbation sensitivity, not a unique assignment of a bit to a waveform sample. Figure 9a shows that four and 161 bits are needed to reach 50% and 90% of the median TX cumulative influence, respectively; the corresponding RX counts are 112 and 417 samples. These results indicate a distributed bit-to-waveform representation and broad aggregation across the received waveform.

Sensitivity to CSI perturbations. We further perturb the injected CSI while keeping the physical propagation channel fixed. As shown in Figure 9b, G-Wave is largely insensitive to mild CSI errors and exhibits gradual performance degradation as the perturbation increases. The system remains operational even under severe CSI distortion approaching 0 dB NMSE, indicating strong tolerance to imperfect channel-condition information.



(a) Cumulative TX/RX influence.

(b) CSI-condition perturbation.

Figure 9: Complementary sensitivity analyses. (a) Coded bits are ranked by their effect on one transmitted sample, while received samples are ranked by their effect on one coded-bit logit; the rank-two marker denotes two bits per QPSK constellation symbol, not a time-domain OFDM baseline. (b) BER under controlled perturbations of the injected CSI condition, with the physical channel and transmitted waveform left unchanged.