

MR2P: Diffusion Model-Driven Pixelated-Antenna Inverse Design Framework Generalizable Across Extensible Electromagnetic Figures of Merit

Minje Kim, *Graduate Student Member, IEEE*, Euiho Shin, *Graduate Student Member, IEEE*,
Hooyoung Kim, *Graduate Student Member, IEEE*, Taeyeong Yoon, *Graduate Student Member, IEEE*,
Semin Lee, *Graduate Student Member, IEEE*, and Jungsuek Oh, *Senior Member, IEEE*

Abstract—Electromagnetic (EM) inverse synthesis of pixelated antennas is a high-dimensional discrete problem that maps desired impedance and radiation responses to a fabrication-ready binary metallization layout. This paper introduces Multi-Response-to-Patch (MR2P), a conditional binary-diffusion framework that directly generates one-bit microstrip patch layouts from heterogeneous EM response specifications. Unlike full-wave or surrogate-assisted optimization pipelines that require a new search for each target, MR2P samples a binary layout through a single response-conditioned reverse process. The framework consists of a Multi-Response Encoder (MRE), pre-trained by a Multi-Response Masked AutoEncoder to embed broadband reflection coefficients $|S_{11}|$ and frequency-wise X - and Y -polarized beam patterns into a unified latent sequence, and a Binary Diffusion Model (BDM), which denoises Bernoulli noise into a 32×32 metal/void layout under cross-attention conditioning. Because the reverse process is defined directly on binary variables, no post-hoc thresholding is required. MR2P is trained on 24 k HFSS-labeled one-bit layouts and evaluated by closed-loop full-wave re-simulation of generated layouts. Ablation studies verify the importance of the MRE pretraining, while artificial wideband and triple-band targets, nearest-neighbor comparisons, and a fabricated multi-response example demonstrate threshold-free, fabrication-ready inverse synthesis from quantitative EM specifications.

Index Terms—Inverse Design, Antenna Inverse Design, Design Automation, Artificial Intelligence, Generative AI, Diffusion Models, Multi-Modal Model,

I. INTRODUCTION

THE rapid evolution toward sixth-generation (6G) wireless systems is driving antennas to support compact integration, multi-band operation, polarization diversity, beam control, and coexistence with densely packed RF front ends [1]–[14]. Under these requirements, the conventional forward-design workflow—selecting a topology, tuning geometric parameters, running full-wave electromagnetic (EM) simulations, and iterating until the specifications are satisfied—becomes increasingly costly. This motivates antenna inverse design, where the desired EM response is specified first and the corresponding physical structure is synthesized automatically.

This work was supported by Samsung Research Funding Incubation Center for Future Technology of Samsung Electronics under Project Number SRFC-IT2502-05. (Corresponding author: Jungsuek Oh.)

M. Kim, E. Shin, H. Kim, T. Yoon, S. Lee, and J. Oh are with the Department of Electrical and Computer Engineering (ECE), Institute of New Media and Communications (INMC), Seoul National University, Seoul 08826, Republic of Korea (e-mail: sbi06035@snu.ac.kr, jungsuek@snu.ac.kr)

Machine-learning-assisted antenna design has been widely explored to reduce repeated full-wave simulation and manual tuning. Surrogate-assisted optimization replaces the EM solver with a learned predictor and then searches the design space using evolutionary or heuristic algorithms [15]–[18]. Direct inverse models instead learn a response-to-geometry mapping and can output design variables without an outer optimization loop [19], [20]. A closely related recent work demonstrated surrogate-driven multi-objective synthesis of polygonal patch antennas by learning heterogeneous responses—reflection coefficient, input impedance, and 3-D radiation patterns—and searching for geometries that satisfy return-loss, gain, and radiation-pattern constraints [21]. While this shows the feasibility of multi-response antenna synthesis, its design representation remains a polygonal parameterization. To enlarge the accessible design space and improve generalization, pixelated and freeform antenna design research has been actively pursued. Pixelated layouts discretize the metallization into metal/void elements, allowing non-intuitive current paths and multi-resonant topologies that are difficult to express using a small number of template parameters. Tandem neural networks have been used to synthesize multiband pixelated antennas from desired impedance spectra [22], and machine-learning-assisted quasi-bisection has refined pixelated patches for bandwidth optimization [18]. Generative antenna design has also begun to emerge, including discriminator-generator schemes for arbitrary node-based geometries [17] and diffusion-generated antenna candidates selected by characteristic-mode-analysis-based evaluation [23]. These studies demonstrate the promise of broader topology representations, but they still leave open the problem of directly generating a fabrication-ready binary layout from quantitative, heterogeneous EM specifications.

This problem is technically challenging because pixelated antenna inverse design requires a response-conditioned mapping from mixed-format EM data to a high-dimensional discrete topology. The input response may include broadband reflection coefficients, frequency-dependent radiation patterns, polarization responses, gain, efficiency, and other FOMs, each with different dimensions and physical meanings. The output, however, must be a manufacturable metal/void pattern rather than a continuous image. Many neural inverse-design methods therefore rely on continuous intermediate variables followed by post-hoc thresholding, but this conversion can distort the

learned layout–response relation, particularly when the design space contains hundreds or thousands of binary pixels.

Conditional diffusion models provide a natural way to represent this one-to-many inverse mapping. Diffusion models generate samples by reversing a gradual noise-injection process and can be conditioned on external information to model distributions of the form $p(x|y)$ [24]–[28]. Text-to-image diffusion models are a representative example, where a text embedding conditions the denoising network to generate images aligned with the input prompt [29]–[31]. For antenna inverse design, the same principle can be transferred by replacing the text embedding with an EM-response embedding and replacing the natural image with a binary metallization layout. However, this transfer requires a response encoder that can jointly represent heterogeneous EM modalities and a generative process that natively preserves binary metal/void states.

In this paper, we propose the Multi-Response-to-Patch diffusion model (MR2P), a conditional binary diffusion framework that directly maps multiple EM responses—broadband reflection coefficients $|S_{11}|$, frequency-wise X - and Y -polarized beam patterns $\mathbf{BP}_X$ and $\mathbf{BP}_Y$ —to a one-bit microstrip patch layout. MR2P consists of a Multi-Response Encoder (MRE), which embeds heterogeneous response modalities into a unified latent representation, and a Binary Diffusion Model (BDM), which denoises Bernoulli noise into a fabrication-ready binary layout conditioned on the MRE latent representation [32]. Unlike continuous regression or image-generation approaches that require post-processing, MR2P operates directly on one-bit layouts and therefore preserves manufacturability throughout generation. Although this paper instantiates MR2P on pixelated microstrip patch antennas, the underlying response-conditioned discrete-generation view can be extended by augmenting the response-token set, consistent with recent inverse-design studies on arbitrary RF/sub-terahertz passives, metasurfaces, photonic devices, and mechanical metamaterials [33]–[39].

The main contributions of this work are summarized as follows.

- 1) We formulate antenna inverse synthesis as conditional binary diffusion, enabling direct generation of one-bit metal/void layouts from desired EM responses without post-hoc thresholding.
- 2) We introduce an MRE that jointly embeds heterogeneous response modalities, allowing impedance and dual-polarized radiation specifications to condition the generative process in a unified latent space.
- 3) We integrate the MRE with a BDM to generate fabrication-ready binary patch layouts in a single generative pass.
- 4) We validate MR2P using full-wave simulations, ablation studies, artificial-target examples, nearest-neighbor comparisons, and measurements, demonstrating both impedance-only and multi-response generation.

The remainder of this paper is organized as follows. Section II introduces the MR2P architecture, including the Multi-Response Masked AutoEncoder pre-training strategy and the

Binary Diffusion Model. Section III describes the dataset generation procedure and presents quantitative evaluations, ablation studies, distinctive artificial-target examples, and measurement results. Section IV discusses the scope, limitations, and future extensions of the framework, and Section V concludes the paper.

II. MULTI-RESPONSE-TO-PATCH DIFFUSION MODEL

A. Direct Binary EM Inverse-Synthesis Formulation

Let $x_0 \in \{0, 1\}^{32 \times 32}$ denote a one-bit patch layout, where each pixel represents either metal or void, and let $y = \{|S_{11}|, \mathbf{BP}_X, \mathbf{BP}_Y\}$ denote the target EM response set. Here, $|S_{11}|$ is the broadband reflection-coefficient magnitude, and $\mathbf{BP}_X$ and $\mathbf{BP}_Y$ are the frequency-wise beam-pattern responses for the two polarization channels. A conventional optimization-based inverse-synthesis formulation searches for a layout whose full-wave EM response $F_{\text{EM}}(x)$ matches the target response y :

$$x^* = \arg \min_{x \in \{0, 1\}^{32 \times 32}} \mathcal{L}(F_{\text{EM}}(x), y).$$

This formulation generally requires a new search for each target response. In contrast, MR2P formulates inverse synthesis as response-conditioned discrete generation:

$$x_0 \sim p_\theta(x_0|y), \quad x_0 \in \{0, 1\}^{32 \times 32}.$$

The objective is therefore not to optimize a single layout from scratch, but to learn a conditional distribution from heterogeneous EM responses to fabrication-ready binary layouts. Learning this distribution requires a conditional generative model capable of sampling layouts from $p_\theta(x_0|y)$. Conditional diffusion provides such a mechanism by representing generation as an iterative denoising process guided by the condition y .

Diffusion models [24]–[26] learn data distributions by defining a forward process that gradually corrupts a sample and a reverse process that reconstructs the original sample from noise. In the conditional setting, the goal is to model $p(x|y)$, where y is an external condition. For continuous data, the forward process is commonly written as

$$q(x_t|x_{t-1}, y) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t \mathbf{I}), \quad (1)$$

where $\beta_t \in (0, 1)$ is a predefined noise variance. Recursively applying this process gives

$$q(x_t|x_0, y) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)\mathbf{I}), \quad (2)$$

with

$$\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s). \quad (3)$$

The noisy sample can be reparameterized as

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}). \quad (4)$$

The reverse process learns to remove the injected noise. When the network predicts the noise component $\epsilon_\theta(x_t, t, y)$, the

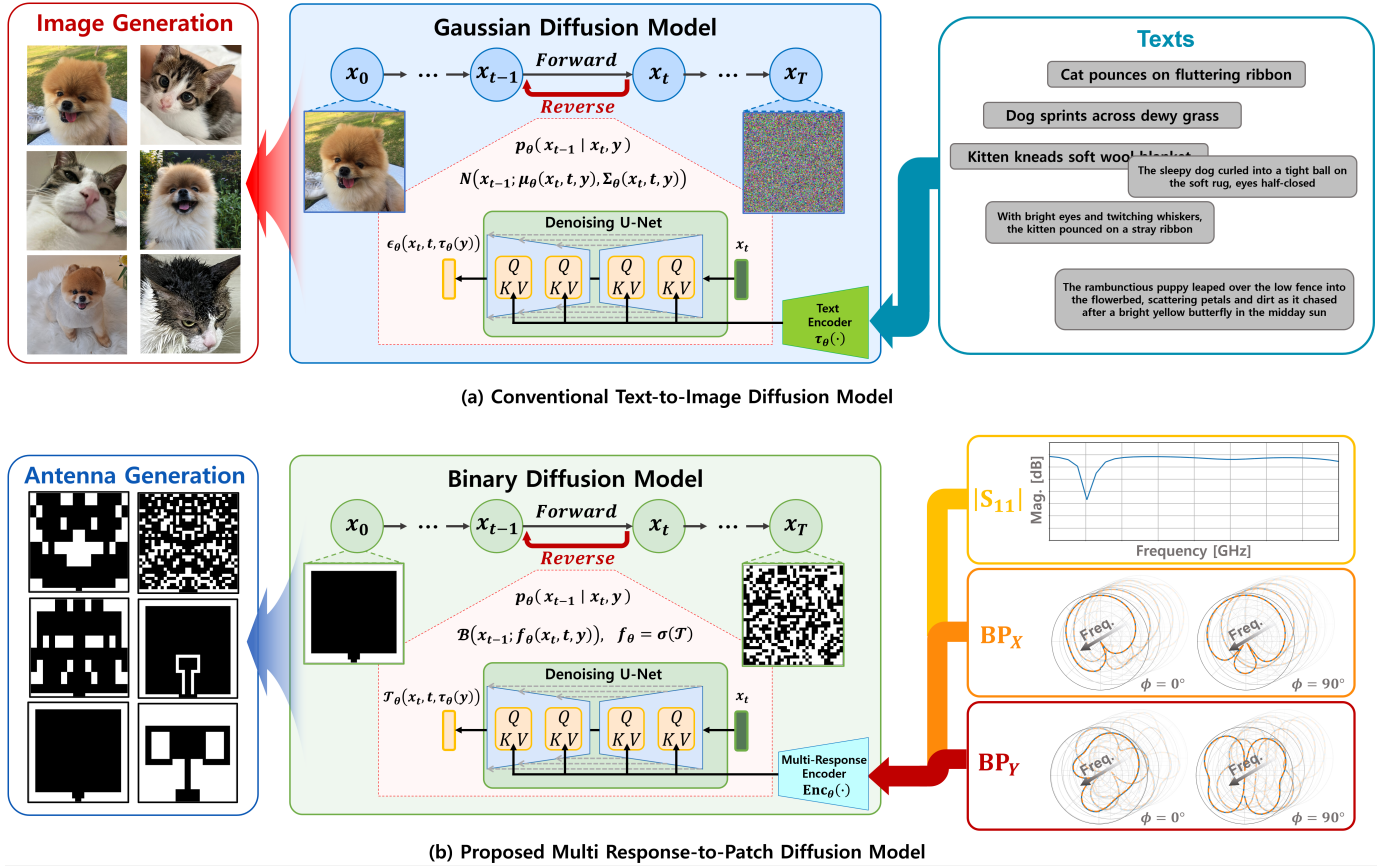


Fig. 1. Comparison between Text-to-Image diffusion model and Multi-Response-to-Patch diffusion model. (a) Conventional Text-to-Image diffusion model receives text prompts as input and generates images that are semantically aligned with the input text. (b) The proposed Multi-Response-to-Patch diffusion model takes multi-modal EM responses as input and generates a 1-bit patch antenna layout that is electromagnetically aligned with the input responses.

reverse mean can be written as

$$\mu_{\theta}(x_t, t, y) = \frac{1}{\sqrt{1 - \beta_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_{\theta}(x_t, t, y) \right), \quad (5)$$

and the reverse transition is modeled as

$$p_{\theta}(x_{t-1}|x_t, y) = \mathcal{N}(x_{t-1}; \mu_{\theta}(x_t, t, y), \Sigma_{\theta}(x_t, t, y)). \quad (6)$$

The simplified training objective is

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t, x_0, \epsilon, y} \left[\|\epsilon - \epsilon_{\theta}(x_t, t, y)\|^2 \right]. \quad (7)$$

Text-to-image diffusion models provide a representative example of this conditional-generation principle. In these models, a text encoder $\tau_{\theta}(\cdot)$ extracts a conditioning representation from the text prompt y , and this representation is injected into the denoising network, typically through cross-attention [29]–[31]. The corresponding denoising function is written as $\epsilon_{\theta}(x_t, t, \tau_{\theta}(y))$, and the training objective becomes

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t, x_0, \epsilon, y} \left[\|\epsilon - \epsilon_{\theta}(x_t, t, \tau_{\theta}(y))\|^2 \right]. \quad (8)$$

Fig. 1 illustrates how this conditional diffusion paradigm is transferred from text-to-image generation to EM inverse synthesis. In conventional text-to-image diffusion, the text encoder $\tau_{\theta}(\cdot)$ conditions the denoising network to generate a natural image aligned with a text prompt. In MR2P, the text encoder is replaced by an MRE $\text{Enc}_{\theta}(\cdot)$, the text prompt is replaced by

heterogeneous EM responses, and the generated natural image is replaced by a one-bit antenna layout. Moreover, because the target output is a printed metal/void pattern, MR2P replaces the continuous Gaussian image prior with a BDM so that the sampling trajectory remains discrete.

This specialization introduces two technical components. First, the MRE must convert heterogeneous EM responses into a unified conditioning sequence. Second, the BDM must generate a one-bit layout without relaxing the output into continuous values. These components are described in Sections II-B and II-C, respectively.

B. Multi-Response EM Representation Learning

The conditioning signal in MR2P is not a single scalar objective but a heterogeneous set of EM responses. To represent these responses in a unified form, we use a Multi-Response Masked AutoEncoder (MRMAE) to pre-train an MRE. The role of the MRE is to embed mixed-format EM responses into a shared latent sequence that can condition the BDM.

Masked AutoEncoders learn representations by masking a subset of input tokens and reconstructing the missing content from the visible tokens [40]. This idea has also been extended to multi-modal representation learning, where different modalities are projected into a common token space and reconstructed through modality-specific decoders [41], [42].

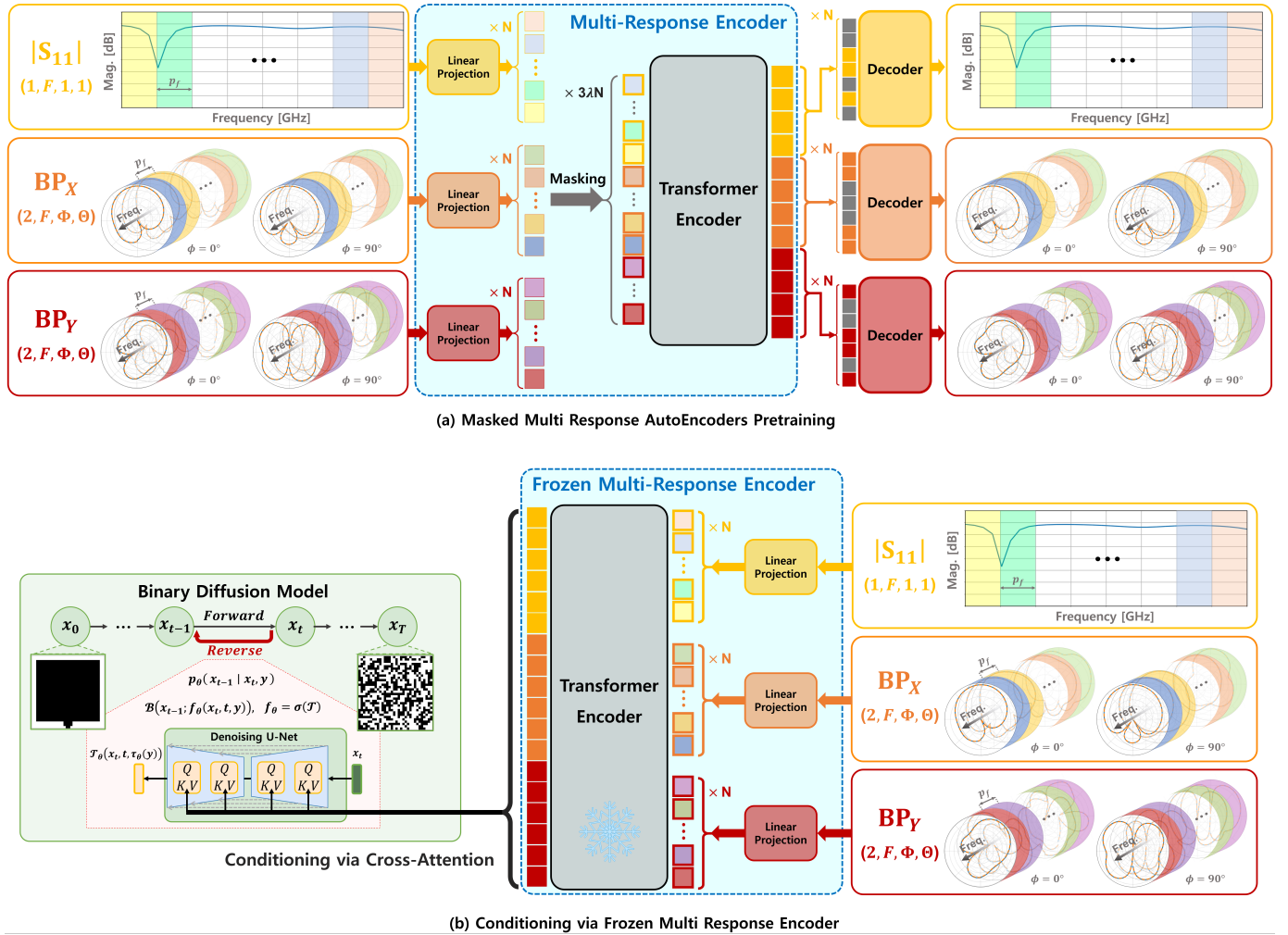


Fig. 2. Multi-Response Masked AutoEncoder (MRMAE) pipeline. (a) Pre-training. Tokenized antenna responses are partially masked. The encoder processes only unmasked tokens, and three modality-specific decoders reconstruct the complete $|S_{11}|$, BP_X , and BP_Y sequences. (b) Conditioning. After pre-training, the encoder and its linear projections are frozen. All tokens are encoded without masking, yielding a latent sequence c that conditions a BDM through cross-attention, ultimately generating a 1-bit 32×32 patch layout.

MRMAE follows this principle, but its modalities are EM response tensors rather than natural images or language tokens.

We regard $|S_{11}|$, BP_X , and BP_Y as three distinct response modalities. Specifically, they are represented as

$$|S_{11}| \in \mathbb{R}^{1 \times F \times 1 \times 1}, \quad BP_X, BP_Y \in \mathbb{R}^{2 \times F \times \Phi \times \Theta}, \quad (9)$$

where F is the number of frequency samples, Φ is the number of ϕ -planes, and Θ is the number of θ -angle samples. In this work, $F = 32$, $\Phi = 2$, and $\Theta = 90$. Expressing all EM responses in this tensorized format allows the MRE to process heterogeneous response modalities using a common token-based interface.

As shown in Fig. 2(a), MRMAE tokenizes the three antenna response modalities $\{|S_{11}|, BP_X, BP_Y\}$ along the frequency axis. With a frequency patch length of $p_f = 4$, each modality is divided into $N = \lceil F/p_f \rceil = 8$ tokens, yielding $N_{\text{tot}} = 24$ total tokens. Each token is flattened, passed through a modality-specific linear projection, and combined with a sinusoidal position embedding and a learnable modality tag. During pre-training, a fraction λ of the tokens is randomly

masked, and only the unmasked tokens are processed by the transformer encoder. The masked positions are then filled with learnable mask tokens, and three lightweight modality-specific decoders reconstruct the complete response sequences.

Let $[y_i^{(m)}]_{i=1}^N$ denote the sequence of N patches for modality $m \in \{|S_{11}|, BP_X, BP_Y\}$, and let $\hat{y}_i^{(m)}$ denote its reconstructed counterpart. We introduce a binary indicator $\gamma_i^{(m)} \in \{0, 1\}$, which equals 1 when the i -th patch of modality m is included in the reconstruction loss and 0 otherwise. When the loss is accumulated only over masked patches, $\gamma_i^{(m)}$ is set according to the mask; when the loss is accumulated over all patches, $\gamma_i^{(m)} = 1$ for every i . The MRMAE pre-training loss is

$$\mathcal{L}_{\text{MRMAE}} = \sum_{m \in \{|S_{11}|, BP_X, BP_Y\}} \xi_m \sum_{i=1}^N \frac{\gamma_i^{(m)}}{P_m} \left\| \hat{y}_i^{(m)} - y_i^{(m)} \right\|_2^2, \quad (10)$$

where ξ_m balances the contribution of each modality and P_m normalizes the patch-wise reconstruction error by the number of scalar entries in each patch. In this work, ξ_m is fixed to 1

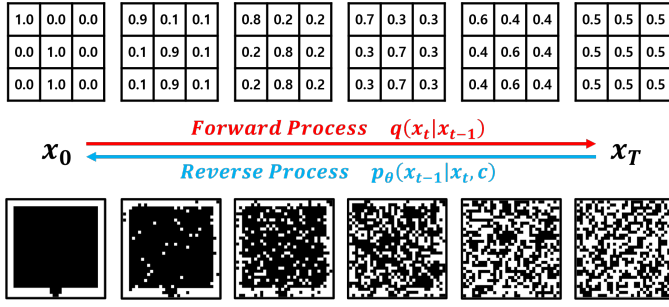


Fig. 3. Illustration of the BDM's forward and reverse processes. A binary representation is progressively diffused towards a random Bernoulli distribution with 0.5 everywhere.

for all modalities.

After pre-training, the modality-specific decoders are discarded, and the MRE Enc_θ is frozen. Given a new target response set $y = \{|S_{11}|, \mathbf{BP}_X, \mathbf{BP}_Y\}$, the MRE encodes all response tokens without masking to produce the conditioning sequence $c = \text{Enc}_\theta(y) \in \mathbb{R}^{N_{\text{tot}} \times D}$. As shown in Fig. 2(b), this latent sequence conditions the BDM through cross-attention, capturing intra-modal spectral trends and cross-modal correlations across impedance and beam-pattern responses.

C. Conditional Binary Diffusion for 1-Bit Layout Generation

The target structure in MR2P is a printed patch layout whose pixels must remain binary. Therefore, instead of applying Gaussian diffusion to continuous image values, MR2P adopts a BDM [32], where noise injection is implemented by stochastic bit flipping. Consequently, every intermediate state in the diffusion trajectory remains a one-bit layout.

1) *Forward process*: Let $x_0 \in \{0, 1\}^{32 \times 32}$ be a clean binary layout and let $\{\beta_t\}_{t=1}^T$ be a non-decreasing flip schedule. At each time step, each pixel is preserved with probability $1 - \beta_t$ and redrawn from a fair Bernoulli distribution with probability β_t :

$$q(x_t | x_{t-1}) = B\left(x_t; (1 - \beta_t)x_{t-1} + \frac{1}{2}\beta_t\right). \quad (11)$$

Iterating the forward process gives the closed-form marginal

$$q(x_t | x_0) = B(x_t; k_t x_0 + b_t), \quad \text{with} \quad (12)$$

$$k_t = \prod_{i=1}^t (1 - \beta_i),$$

$$b_t = (1 - \beta_t)b_{t-1} + \frac{1}{2}\beta_t, \quad b_1 = \frac{1}{2}\beta_1.$$

As t increases, k_t decreases and the Bernoulli parameter approaches 0.5, so the layout is gradually diffused toward random binary noise, as illustrated in Fig. 3.

2) *Reverse process*: The reverse process models $p_\theta(x_{t-1}|x_t, c)$, where $c = \text{Enc}_\theta(y)$ is the MRE conditioning sequence. Following the predict- x_0 parameterization, a U-Net predicts a clean layout estimate

$$\hat{x}_0 = f_\theta(x_t, t, c).$$

In practice, f_θ is implemented as $\sigma(T_\theta(x_t, t, c))$, where T_θ produces unnormalized logits and $\sigma(\cdot)$ converts them into

Algorithm 1 MR2P inference for direct binary layout synthesis

Require: responses $y = \{|S_{11}|, \mathbf{BP}_X, \mathbf{BP}_Y\}$, MRE Enc_θ , U-Net f_θ , flip schedule $\{\beta_t\}$

- 1: $c \leftarrow \text{Enc}_\theta(y)$
- 2: $x_T \sim B(0.5)$ ▷ pure noise
- 3: **for** $t = T, T-1, \dots, 1$ **do**
- 4: $\hat{x}_0 \leftarrow f_\theta(x_t, t, c)$
- 5: compute $\phi_\theta(x_t, t, \hat{x}_0)$ by (14)
- 6: $x_{t-1} \sim B(\phi_\theta)$
- 7: **end for**

Ensure: x_0 ▷ fabrication-ready layout

Bernoulli probabilities. Bayes' rule for Bernoulli variables gives

$$p_\theta(x_{t-1}|x_t, c) = B(x_{t-1}; \phi_\theta(x_t, t, \hat{x}_0)), \quad (13)$$

where

$$\phi_\theta = \frac{[(1 - \beta_t)x_t + \frac{1}{2}\beta_t] \odot [k_t \hat{x}_0 + \frac{1}{2}b_t]}{Z(x_t, t, \hat{x}_0)}, \quad (14)$$

and $Z(\cdot)$ normalizes the Bernoulli probability for each pixel.

3) *Training objective*: Because x_0 is binary, training the reverse model reduces to pixel-wise binary classification. The BDM is trained using binary cross-entropy:

$$\mathcal{L}_{\text{CE}} = \mathbb{E}_{t, x_0, y} [\text{BCE}(f_\theta(x_t, t, c), x_0)], \quad (15)$$

where $x_t \sim q(x_t|x_0)$ is sampled from the forward process and $c = \text{Enc}_\theta(y)$.

4) *Generation*: At inference, MR2P starts from pure Bernoulli noise $x_T \sim B(0.5)$ and applies the learned reverse chain conditioned on the target response set. The complete response-conditioned binary sampling procedure is summarized in Algorithm 1.

During sampling, the target EM responses are encoded by the frozen MRE and used as cross-attention conditioning for the BDM. The BDM then transforms uniform Bernoulli noise into a one-bit 32×32 patch layout. Since the reverse process is defined directly on binary variables, the generated output is already a metal/void pattern and requires no post-hoc thresholding.

III. EXPERIMENTAL RESULTS

A. Dataset and Full-Wave Response Generation

To train and validate the proposed response-conditioned binary generator, we constructed a full-wave-labeled EM dataset of 24 k one-bit 32×32 patch layouts. Each layout is annotated with EM responses at 32 uniformly spaced frequencies from 6.00 GHz to 13.75 GHz with a 0.25 GHz step. Two complementary strategies—rule-based random generation and human sketching—were adopted to balance structural diversity with practical relevance.

Rule-based random layouts (20 k). To generate synthetic yet fabricable patterns, we devised the four principles illustrated in Fig. 4(a). (i) *Subset structuring*. The designable area is tessellated into rectangular sub-sets of size $S_1 \times S_2$ ($1 \times 1, 2 \times 1, 2 \times 2, 4 \times 2$). A sub-set behaves like a single

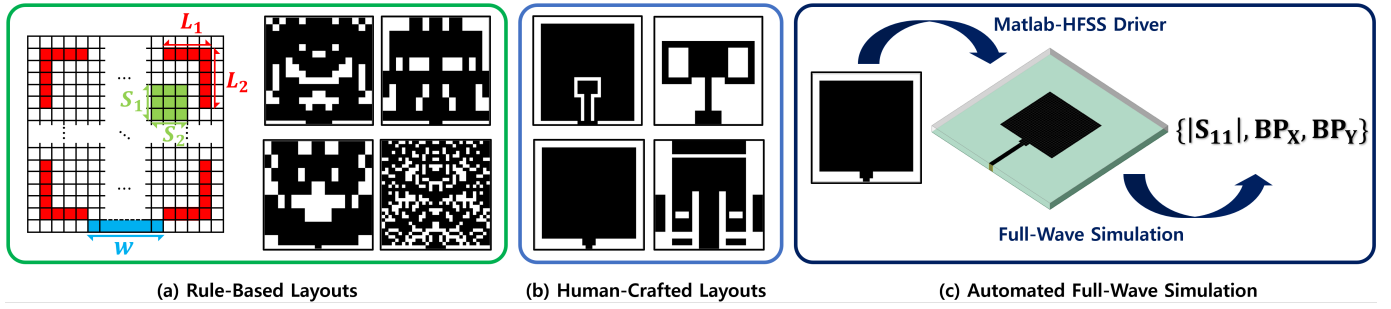


Fig. 4. (a) Rule-based layouts. (b) Human-crafted layouts. (c) Automated full-wave simulation.

metal pixel, discouraging isolated one-cell islands that would violate minimum etch-gap rules and complicate meshing. (ii) *Probability-driven binary assignment*. Inside each subset, metal (1) or void (0) is drawn from a Bernoulli law with controllable occupancy ($p_{\text{metal}} = 0.7$ or 0.8), allowing the overall fill ratio to be adjusted while retaining stochastic variability. (iii) *Fixed boundary condition*. The outer frame is forced to metal over lengths L_1 and L_2 and mirrored about the horizontal axis; this frame acts as a current return path and isolates the learnable interior from edge-diffraction artifacts. (iv) *Feed-width sweep*. The final row encodes a microstrip feed of width $w \in \{2, 4, 6, 8\}$ pixels, enabling the dataset to sample different input-coupling regimes without altering the interior rules. By combining these principles, we obtain well-formed patches that span a wide spectrum of sparsity, symmetry, and resonance characteristics without human intervention.

Human-crafted layouts (4 k). Experienced antenna engineers supplemented the dataset with hand-drawn canonical shapes: classical rectangular patches, inset-feed variants, dual-slot and multi-resonant structures, and other motifs commonly found in literature and industrial catalogues, as shown in Fig. 4(b). These samples inject real-world priors that purely stochastic rules may miss and provide an anchor for qualitative comparisons.

Automated full-wave simulation. As depicted in Fig. 4(c), a MATLAB-HFSS driver transforms each patch layout into an HFSS 3-D model. Both the ground and patch are treated as perfect electric conductors on a Taconic TLY-5 substrate with a thickness of 1.52 mm. The antenna footprint is confined to $\frac{1}{2}\lambda_0$ at 10 GHz, giving a pixel pitch of 0.469 mm. This single-stack-up configuration, consisting of PEC metals and a fixed low-loss substrate, is used for all layouts in the dataset. Therefore, conductor and dielectric losses are negligible, and the simulated gain is approximately equal to the directivity. Accordingly, in the present work, the BP_X and BP_Y channels should be interpreted as dual-polarized directivity-pattern responses rather than as explicit control of radiation efficiency. For every layout, HFSS full-wave simulations automatically extract the magnitude of S_{11} , and E-/H-plane gain/directivity responses for both polarizations, forming the multi-modal EM responses used throughout this paper.

Train-test partition. After data collection, the dataset is randomly partitioned into 95% training and 5% test sets, guaranteeing that no layout appears in both splits. The 5%

TABLE I
MAIN ARCHITECTURAL HYPERPARAMETERS OF MRMAE AND BDM.

Parameter	Value	Parameter	Value
MRMAE / MRE			
Encoder dimension	256	Encoder depth	4
Encoder heads	4	Decoder dimension	64
Decoder depth	1	Decoder heads	2
BDM			
Base dimension	64	Maximum dimension	512
Downsampling stages	3	Blocks per stage	3
Attention resolutions	16, 8	Attention head dimension	64
Dropout	0.1	Diffusion timesteps	100

test split corresponds to 1,200 held-out target responses. Unlike forward surrogate modeling, evaluating a direct inverse-design model requires generating a layout for each target response and re-simulating the generated layout using HFSS. Therefore, even the 5% test split requires a large number of closed-loop full-wave validations. We did not introduce an additional validation split for hyperparameter tuning because no validation-driven architecture search was performed; the main network configuration is fixed across experiments, and only the explicitly reported ablation factors are varied.

The main architectural hyperparameters are summarized in Table I. The MRMAE follows a MultiMAE-style encoder-decoder configuration [41], while the BDM follows the U-Net configuration used in latent binary diffusion models [32]. These hyperparameters were fixed across the main experiments; only the explicitly reported ablation factors were varied.

B. Closed-Loop Full-Wave Validation Protocol

The proposed MR2P is evaluated through a closed-loop full-wave validation protocol rather than by training loss or by a learned surrogate. Given an input target response y_{tar} , MR2P generates a binary layout. This generated layout is then converted back into an HFSS model and re-simulated to obtain the reproduced response y_{rep} . This protocol directly measures whether the generated binary layout realizes the prescribed target response y_{tar} after full-wave verification.

The alignment between the input target response y_{tar} and the reproduced full-wave response y_{rep} is quantified using a

normalized dB-domain cosine dissimilarity:

$$E_{\cos} = \left(1 - \frac{y_{\text{tar,dB}}^T y_{\text{rep,dB}}}{\|y_{\text{tar,dB}}\|_2 \|y_{\text{rep,dB}}\|_2 + \epsilon} \right) \times 100\%, \quad (16)$$

where $y_{\text{tar,dB}}$ and $y_{\text{rep,dB}}$ denote the vectorized dB-scale forms of y_{tar} and y_{rep} , respectively, and ϵ is a small constant for numerical stability.

Because MR2P is a conditional diffusion model, multiple candidate layouts can be sampled from the same target response y_{tar} . Directly re-simulating all stochastic samples in HFSS would be prohibitively time-consuming, especially for the full test set and ablation studies. Therefore, for the baseline evaluation and Ablations 1–4, we generate 20 candidate layouts for each y_{tar} and select one representative layout as the Hamming-distance medoid, i.e., the sample with the smallest average normalized Hamming distance to the other generated layouts. This medoid selection uses only layout-space distances and does not use EM-response errors, so it is not a best-response or cherry-picked sample. The medoid layout is then re-simulated by HFSS and used for the closed-loop quantitative results.

C. Quantitative Test-Set Performance and Ablation Studies

The proposed MR2P is composed of several principal design choices: (i) how the MRMAE is self-supervised trained, (ii) how the MRE is coupled to the BDM, (iii) how multi-modal EM responses are grouped into tokens, and (iv) which response modalities are supplied as conditioning information. The ablation studies vary one factor at a time while holding the remaining choices fixed at the baseline. This analysis identifies which ingredients govern response-layout alignment accuracy and which choices have comparatively minor impact.

Baseline configuration. Unless otherwise specified, the MRMAE is pre-trained with masking ratio $\lambda = 0.50$, and the reconstruction loss is accumulated only on masked tokens ($L_{\text{mask}} = \text{True}$). The MRE is frozen during BDM training, and all three modalities $\{|S_{11}|, \text{BP}_X, \text{BP}_Y\}$ are supplied. Table II summarizes the resulting closed-loop full-wave alignment.

The baseline model reproduces broadband impedance responses with an average $|S_{11}| E_{\cos}$ of 3.10%, while the BP_X and BP_Y responses are reproduced with average $E_{\cos}$ values of 5.20% and 8.40%, respectively. The larger BP_Y error is consistent with the fact that this channel is often dominated by low-level cross-polarized or null-region responses, where small field variations can cause larger normalized response-shape deviations.

Ablation 1: MRMAE pre-training. Table III sweeps the mask-loss switch $L_{\text{mask}} \in \{\text{True}, \text{False}\}$ and masking ratio $\lambda \in \{0.25, 0.50, 0.75\}$, while the encoder remains frozen. The $|S_{11}|$ and BP_X $E_{\cos}$ values remain relatively stable across the tested pre-training recipes, whereas the BP_Y $E_{\cos}$ value shows moderate variation. Overall, the differences caused by the masking ratio and loss-accumulation option are smaller than the changes caused by the encoder-coupling strategy in Table IV, indicating that downstream performance is governed primarily by the quality and stability of the pre-trained

TABLE II
BASELINE CLOSED-LOOP COSINE DISSIMILARITY RESULTS. AVG. = MEAN OVER ALL TEST SAMPLES; WORST 10% = MEAN OVER WORST 10%. EACH RESULT IS OBTAINED FROM THE HAMMING-DISTANCE MEDOID AMONG 20 STOCHASTIC SAMPLES PER TARGET.

	Avg. $E_{\cos}$ (%)			Worst 10% $E_{\cos}$ (%)		
	$ S_{11} $	BP_X	BP_Y	$ S_{11} $	BP_X	BP_Y
Baseline	3.10	5.20	8.40	4.22	6.72	10.74

TABLE III
IMPACT OF MRMAE PRE-TRAINING.

λ	L_{mask}	Avg. $E_{\cos}$ (%)			Worst 10% $E_{\cos}$ (%)		
		$ S_{11} $	BP_X	BP_Y	$ S_{11} $	BP_X	BP_Y
0.75	True	3.24	5.42	8.78	4.36	6.94	11.16
0.50	True	3.10	5.20	8.40	4.22	6.72	10.74
0.25	True	3.30	5.55	8.95	4.40	7.04	11.32
0.75	False	3.55	5.66	8.70	4.70	7.32	11.10
0.50	False	3.18	5.35	8.50	4.30	6.91	10.95
0.25	False	3.14	5.30	8.72	4.25	6.84	11.24

TABLE IV
IMPACT OF ENCODER COUPLING STRATEGIES.

Strategy	Avg. $E_{\cos}$ (%)			Worst 10% $E_{\cos}$ (%)		
	$ S_{11} $	BP_X	BP_Y	$ S_{11} $	BP_X	BP_Y
Frozen (baseline)	3.10	5.20	8.40	4.22	6.72	10.74
Fine-tuned	3.62	5.86	9.05	4.85	7.53	11.58
From scratch	6.15	8.05	11.20	8.25	10.22	14.12

response representation rather than by the exact masking configuration.

Ablation 2: MRE coupling strategy. Three coupling strategies are compared in Table IV: keeping the encoder frozen, fine-tuning it jointly with the BDM, and training it from scratch. Freezing the pre-trained MRE gives the lowest $E_{\cos}$ values. Fine-tuning slightly degrades all three response modalities, while training the encoder from scratch produces substantially larger errors, especially for $|S_{11}|$ and BP_X . These results indicate that stable EM-response conditioning is critical in the small-data regime and that MRMAE pre-training is most effective when the learned MRE is preserved during BDM training.

Ablation 3: Modality grouping. To test whether the encoder benefits from explicit semantic separation, we collapse the four beam-pattern channels into a single tensor BP and feed the pair $\{|S_{11}|, \text{BP}\}$ to the model. Table V shows that retaining the original three-modality input consistently improves both mean and worst-case $E_{\cos}$. This suggests that keeping X- and Y-polarized information as separate modalities aids latent response modeling.

Ablation 4: $|S_{11}|$ -only versus joint response conditioning. To examine the effect of conditioning scope, we compare

TABLE V
IMPACT OF MODALITY GROUPING.

Setup	Avg. $E_{\cos}$ (%)		Worst 10% $E_{\cos}$ (%)	
	$ S_{11} $	$\mathbf{BP}_{\text{avg}}$	$ S_{11} $	$\mathbf{BP}_{\text{avg}}$
$\mathbf{BP}_X\text{-}\mathbf{BP}_Y$ (baseline)	3.10	6.80	4.22	8.73
$\mathbf{BP}$	3.46	7.28	4.52	9.10

TABLE VI
EFFECT OF CONDITIONING SCOPE ON $|S_{11}|$ REPRODUCTION.

Condition	Avg. $ S_{11} $ $E_{\cos}$ (%)	Worst 10% $E_{\cos}$ (%)
$ S_{11} $ -only	1.80	2.58
$ S_{11} + \mathbf{BP}_X + \mathbf{BP}_Y$	3.10	4.22

an impedance-only condition with the full joint condition. The $|S_{11}|$ -only variant uses only the $|S_{11}|$ response encoder and conditions the BDM only on impedance response tokens. The joint condition corresponds to the baseline MR2P setting, where $|S_{11}|$, $\mathbf{BP}_X$, and $\mathbf{BP}_Y$ are all supplied.

Table VI shows that the $|S_{11}|$ -only condition achieves lower impedance-response error than the joint condition. This is expected because impedance-only synthesis is less constrained: the generated layout only needs to reproduce the reflection response. In contrast, the joint condition must simultaneously satisfy impedance and dual-polarized beam-pattern responses, so part of the $|S_{11}|$ -only accuracy is traded for stronger multi-response conditioning.

Key observations. (i) Freezing the pre-trained MRE is the most influential design choice. Compared with training the encoder from scratch, the frozen MRE reduces the average $E_{\cos}$ by 3.05 percentage points for $|S_{11}|$, 2.85 percentage points for $\mathbf{BP}_X$, and 2.80 percentage points for $\mathbf{BP}_Y$. (ii) The self-supervised pre-training recipe is robust relative to the encoder-coupling strategy. Changing the masking ratio or switching between masked-only and all-token reconstruction changes the average errors only moderately. (iii) Keeping X - and Y -polarized beam patterns as separate modalities yields a small but repeatable benefit, reducing the average beam-pattern $E_{\cos}$ from 7.28% to 6.80%. (iv) The $|S_{11}|$ -only condition achieves lower impedance-response error than the joint condition, because it solves a less constrained inverse problem. The joint condition sacrifices part of the $|S_{11}|$ -only accuracy to simultaneously condition the generated layout on $\mathbf{BP}_X$ and $\mathbf{BP}_Y$.

D. Response-Conditioned Synthesis Examples

We first use an $|S_{11}|$ -only MR2P variant to isolate impedance-conditioned generation, and then present the full multi-response case to demonstrate the complete MR2P capability. The fabricated examples are characterized using the anechoic-chamber setup shown in Fig. 6. The AUT is mounted on a low-permittivity foam pedestal on the turntable, a standard-gain horn antenna is used as the receive horn, and the VNA is placed outside the quiet zone and covered with absorbers to reduce unwanted scattering.

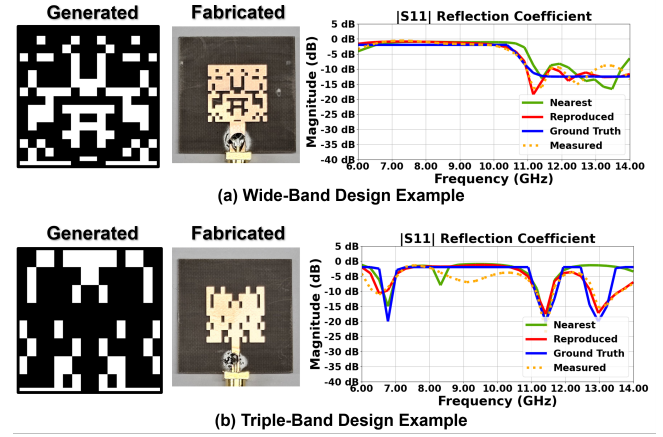


Fig. 5. **Distinctive examples.** Each row shows the generated patch layout, fabricated prototype, and $|S_{11}|$ curves: nearest-neighbor from the training set, target, reproduced, and measured result. (a) wideband matching target. (b) Triple-band matching target.

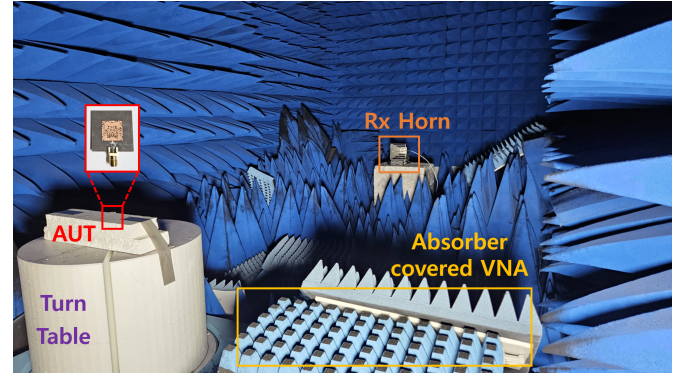


Fig. 6. Measurement setup in the anechoic chamber. The antenna under test (AUT) is mounted on a low-permittivity foam pedestal on the turntable, a standard-gain horn antenna is used as the receive (Rx) horn, and the vector network analyzer (VNA) is located outside the quiet zone and covered with absorbers to reduce unwanted scattering.

Distinctive $|S_{11}|$ -only generation from artificial targets. To demonstrate the extendability of MR2P across EM figures of merit, we train an $|S_{11}|$ -only MR2P variant consisting of a pre-trained $|S_{11}|$ response encoder and its conditioned binary diffusion model. We then provide two artificial $|S_{11}|$ targets—a wideband matching target and a triple-band matching target—and generate the corresponding binary patch layouts. For each target, we additionally plot the nearest-neighbor (NN) $|S_{11}|$ curve from the training set, where the NN is selected by the dB-domain cosine dissimilarity $E_{\cos}$ between $|S_{11}|$ curves. This comparison verifies that the generated design is not a near-duplicate retrieval but a new layout better aligned to the prescribed target.

Fig. 5 summarizes both cases. In the wideband case, the reproduced response of the generated layout achieves a contiguous -10 dB matching band over approximately 11–14 GHz, corresponding to a fractional bandwidth of approximately 24% about 12.5 GHz. In the triple-band case, the reproduced response exhibits three separated -10 dB bands centered near 6.5 GHz, 11.4 GHz, and 13.0 GHz, with

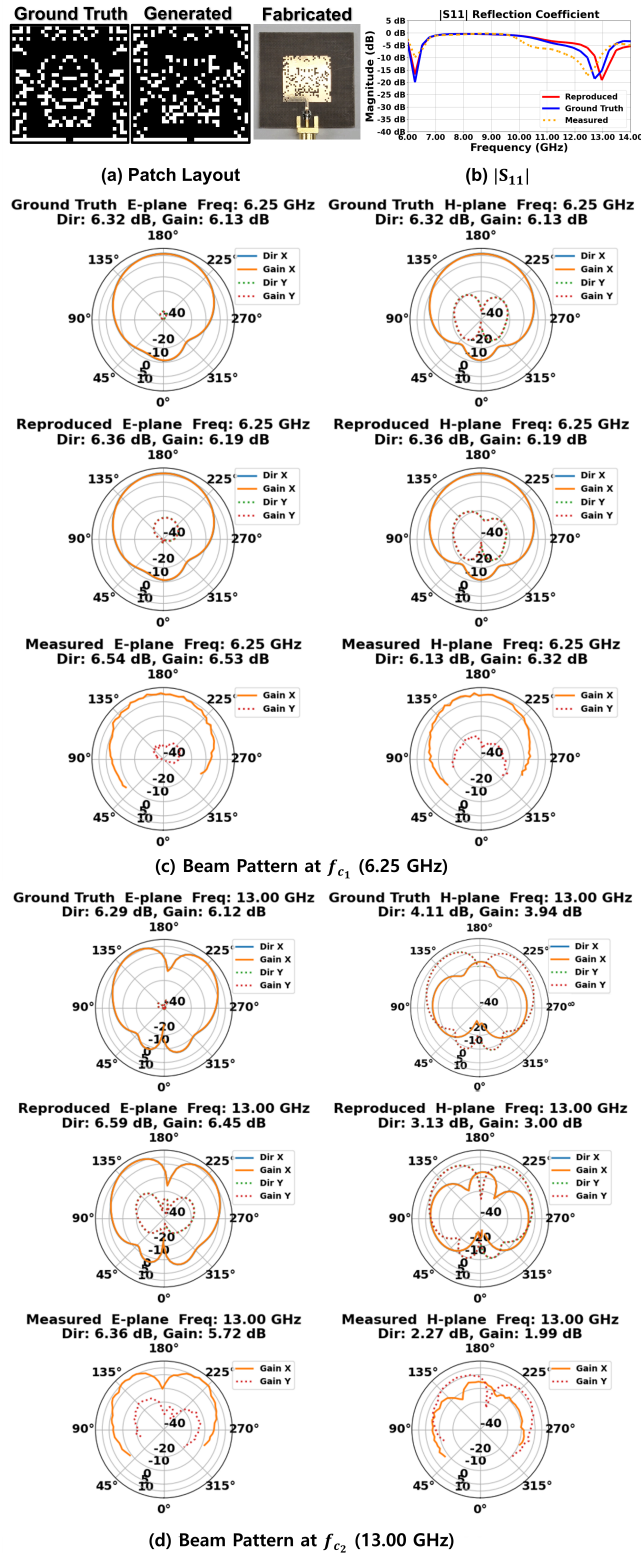


Fig. 7. **Multi-response example.** (a) Ground-truth, generated, and fabricated layouts. (b) $|S_{11}|$ curves: reproduced, ground truth, and measured. (c) E-/H-plane patterns at $f_{c1} = 6.25$ GHz. (d) E-/H-plane patterns at $f_{c2} = 13.00$ GHz.

approximate bandwidths of 0.3 GHz, 0.5 GHz, and 0.9 GHz, respectively. For both targets, the measured $|S_{11}|$ trends follow the reproduced results. These examples validate that MR2P

can be specialized to impedance-only generation and can realize nontrivial bandwidth shapes beyond the closest training examples.

Multi-response example: joint $|S_{11}|$, BP_X , and BP_Y generation. We next present a representative test-set example to highlight the core capability of MR2P: simultaneously conditioning on broadband impedance and frequency-wise dual-polarized beam patterns, and generating a fabrication-ready patch layout in one pass. Fig. 7(a) compares the ground-truth layout, the MR2P-generated layout, and the fabricated prototype. Fig. 7(b) shows that the reproduced $|S_{11}|$ closely follows the target curve, including the two operating bands near $f_{c1} = 6.25$ GHz and $f_{c2} = 13.00$ GHz. The reproduced -10 dB fractional bandwidths are 4.80% at f_{c1} and 5.80% at f_{c2} . Figs. 7(c)–(d) compare E-/H-plane patterns at the two center frequencies. The main-lobe directions and co-polar shapes are well preserved between the target and reproduced patterns, and the measured co-polar patterns follow the same overall trend. Larger discrepancies tend to appear in deep cross-polar null regions, where small perturbations can cause large dB-scale deviations while having limited impact on the main-lobe behavior.

IV. DISCUSSION

The results demonstrate that MR2P is best understood as a response-conditioned discrete EM inverse-synthesis framework rather than as a conventional antenna optimizer. Earlier inverse-design pipelines commonly either retain an outer optimization loop or attempt one-shot response-to-geometry regression. Many of these approaches are limited by at least one of the following constraints: a restricted geometric parameterization, a narrow response set, continuous intermediate layouts that require post-hoc thresholding, or the absence of a unified representation for heterogeneous EM responses. MR2P addresses these issues by combining two components: the MRE embeds multi-modal, frequency-dependent EM responses into a unified conditioning sequence, and the BDM samples a genuinely one-bit metal/void layout without thresholding.

This distinction is especially relevant when compared with recent surrogate-driven multi-objective antenna synthesis. Such methods can learn heterogeneous responses, including impedance and radiation-pattern information, and then search a parameterized geometry space using an optimizer. MR2P takes a different route: it directly learns a conditional distribution $p_\theta(x_0|y)$ from EM responses to binary layouts. Therefore, the generated structure is not obtained by repeatedly querying a forward surrogate during a new search, but by response-conditioned sampling from a learned discrete topology distribution.

The examples in Section III-D clarify the practical meaning of this capability beyond aggregate test-set metrics. First, the $|S_{11}|$ -only MR2P variant demonstrates FOM extendability: using the same training recipe with a reduced modality set, it generates layouts that realize artificial wideband and triple-band -10 dB matching targets. The nearest-neighbor comparison, selected by the dB-domain cosine dissimilarity $E_{\cos}$ in the training set, provides a concrete baseline showing

that the generated responses are not simple retrievals but are better aligned to the prescribed bandwidth shapes. Second, the multi-response example demonstrates joint conditioning: MR2P generates a binary patch layout that reproduces both impedance and co-polar beam-pattern trends at two operating frequencies.

Several limitations remain. First, only microstrip-line feeding is represented in the present dataset; coaxial probes, aperture coupling, and slot coupling are not yet modeled. Second, vias and plated through-holes are absent, which precludes multilayer, shorted, or cavity-backed designs. Third, the substrate permittivity, loss tangent, and thickness are fixed, so material-dependent impedance and radiation trends cannot yet be learned. Fourth, the present layout resolution is limited to 32×32 ; finer grids would require larger denoising networks, larger datasets, and careful tuning of the bit-flip schedule. Fifth, each pixel currently has only two states, metal or void, whereas practical RF structures may require multi-state descriptions involving vias, stacked metals, dielectric openings, lumped components, or feed transitions.

Future work will address these limitations along several directions. A multinomial or categorical diffusion process [43] can extend the BDM prior so that each pixel can represent multiple fabrication states, such as void, metal, via, feed, or stacked-metal features. Substrate parameters such as ϵ_r , loss tangent, and thickness can also be embedded as additional conditioning tokens and supplied through the same cross-attention pathway as the MRE outputs, enabling material- and profile-aware generation. For larger-aperture structures, such as arrays, metasurfaces, and board-level RF layouts, progressive or tile-based diffusion [44], [45] may provide a path toward higher effective resolution without prohibitive memory cost. Finally, the current training loss uses uniform weighting across response modalities. Re-weighting cross-polarized or null-region beam-pattern channels, or incorporating response-aware weighting, may further reduce the large dB-scale deviations observed in low-level BP_Y regions.

V. CONCLUSION

This paper presented MR2P, a conditional binary-diffusion framework for direct EM inverse synthesis of pixelated microstrip patch antennas. MR2P maps heterogeneous response specifications, including broadband $|S_{11}|$ and dual-polarized beam-pattern responses, into a one-bit metal/void layout. The MRE embeds multi-modal EM responses into a unified latent representation, while the BDM generates fabrication-ready binary layouts directly from Bernoulli noise, eliminating post-hoc thresholding.

Closed-loop full-wave validation shows that MR2P reproduces broadband $|S_{11}|$, BP_X , and BP_Y responses with average $E_{\cos}$ values of 3.10%, 5.20%, and 8.40%, respectively. Ablation studies confirm that freezing the pre-trained MRE is critical for stable response-conditioned generation. The $|S_{11}|$ -only conditioning experiment further shows that impedance-only synthesis achieves lower $|S_{11}|$ error, while the full MR2P setting trades part of this impedance-only accuracy for simultaneous conditioning on dual-polarized beam-pattern

responses. The $|S_{11}|$ -only and full multi-response examples further demonstrate that MR2P can generate fabrication-ready layouts for artificial matching targets and jointly condition on impedance and beam-pattern specifications.

Future work will extend MR2P beyond the present single-layer, microstrip-fed, fixed-substrate binary setting by incorporating multi-state diffusion for vias, feeds, and stacked metals; substrate and tolerance conditioning; and higher-resolution generation for larger-aperture EM structures.

REFERENCES

- [1] F. Vogelsang, J. Bott, D. Starke, M. Hamme, B. Sievert, H. Rücker, and N. Pohl, "A fully integrated 0.48 thz fmcw radar sensor in a sige technology," *IEEE Journal of Microwaves*, 2025.
- [2] L. Chen, M. T. Taba, Z. Khalifa, A. Cathelin, and E. Afshari, "A fast back-to-lock dpll-based 192–210-ghz chirp generator with+ 5.9-dbm peak output power for sub-thz imaging and sensing," *IEEE Journal of Solid-State Circuits*, vol. 59, no. 5, pp. 1461–1474, 2023.
- [3] N. Decarli, S. Bartoletti, A. Bazzi, R. A. Stirling-Gallacher, and B. M. Masini, "Performance characterization of joint communication and sensing with beyond 5 g nr-v2x sidelink," *IEEE Transactions on Vehicular Technology*, 2024.
- [4] K. Sehla, T. M. T. Nguyen, G. Pujolle, and P. B. Velloso, "Resource allocation modes in c-v2x: from lte-v2x to 5g-v2x," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8291–8314, 2022.
- [5] M. H. C. Garcia, A. Molina-Galan, M. Boban, J. Gozalvez, B. Coll-Perales, T. Şahin, and A. Kousaridas, "A tutorial on 5g nr v2x communications," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1972–2026, 2021.
- [6] S. Chen, S. Sun, and S. Kang, "System integration of terrestrial mobile communication and satellite communication—the trends, challenges and key technologies in b5g and 6g," *China communications*, vol. 17, no. 12, pp. 156–171, 2021.
- [7] D. Zhou, M. Sheng, J. Li, and Z. Han, "Aerospace integrated networks innovation for empowering 6g: A survey and future challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 975–1019, 2023.
- [8] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.
- [9] J. Jung, W. Lee, G. Lee, S. Hong, and J. Oh, "Ultra-thinned metasurface-embedded smartphone antenna-in-package for millimeter-wave 5g/6g coverage enhancement," *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 10, pp. 7766–7781, 2023.
- [10] B. Kim and J. Oh, "Dual-wideband low-profile three-notch patch antenna with indirect differential feeding for 5g millimeter-wave applications," *IEEE Antennas and Wireless Propagation Letters*, 2025.
- [11] J. Heo, B. Kim, M. Kang, H. Kim, Y. Lee, and J. Oh, "Tetra-plexing dual-broadband/dual-polarized antenna for 5g/6g millimeter-wave systems," *IEEE Transactions on Antennas and Propagation*, 2025.
- [12] H. Kim and J. Oh, "D-band scalable phased array antenna-in-package using stripline power divider-based sub-array," *IEEE Access*, 2024.
- [13] S. Oh and J. Oh, "140-ghz affordable miniaturized array antenna-on-package for sub-thz transceiver," *IEEE Access*, vol. 11, pp. 132 780–132 791, 2023.
- [14] D. Jang, T. H. Lim, S. Park, and H. Choo, "Broadband all-metal vivaldi array antenna with pyramidal-shaped wings for leo satellite applications," *Journal of Electromagnetic Engineering and Science*, vol. 23, no. 5, pp. 405–411, 2023.
- [15] N. Calik, M. A. Belen, and P. Mahouti, "Deep learning base modified mlp model for precise scattering parameter prediction of capacitive feed antenna," *International journal of numerical modelling: electronic networks, devices and fields*, vol. 33, no. 2, p. e2682, 2020.
- [16] Y. He, J. Huang, W. Li, L. Zhang, S.-W. Wong, and Z. N. Chen, "Hybrid method of artificial neural network and simulated annealing algorithm for optimizing wideband patch antennas," *IEEE Transactions on Antennas and Propagation*, vol. 72, no. 1, pp. 944–949, 2023.
- [17] Y. Zhong, P. Renner, W. Dou, G. Ye, J. Zhu, and Q. H. Liu, "A machine learning generative method for automating antenna design and optimization," *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, vol. 7, pp. 285–295, 2022.

- [18] Q. Wang, Z. Pang, D. Gao, P. Liu, X. Pang, and X. Yin, "Machine learning-assisted quasi-bisection method for pixelated patch antenna bandwidth optimization," *IEEE Antennas and Wireless Propagation Letters*, 2024.
- [19] L.-Y. Xiao, W. Shao, F.-L. Jin, B.-Z. Wang, and Q. H. Liu, "Inverse artificial neural network for multiobjective antenna design," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 10, pp. 6651–6659, 2021.
- [20] Y. Su, Y. Yin, H. Chen, S. Li, H. Zhao, Z. Su, and X. Yin, "Time-domain scattering parameters based neural network inverse model for antenna designs," *IEEE Antennas and Wireless Propagation Letters*, 2024.
- [21] P. Singh, S. S. Panda, J. C. Dash, B. Riscob, S. K. Pathak, and R. S. Hegde, "Rapid multi-objective antenna synthesis via deep neural network surrogate-driven evolutionary optimization," *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, 2025.
- [22] A. Gupta, E. A. Karahan, C. Bhat, K. Sengupta, and U. K. Khankhoje, "Tandem neural network based design of multiband antennas," *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 8, pp. 6308–6317, 2023.
- [23] Q. Pang, J. Ouyang, F. Yang, S. Yang, and J. Hu, "Inverse design method of antenna based on generative artificial intelligence," *IEEE Antennas and Wireless Propagation Letters*, 2025.
- [24] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [25] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International conference on machine learning*. PMLR, 2021, pp. 8162–8171.
- [26] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.
- [27] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.
- [28] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.
- [29] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," *arXiv preprint arXiv:2112.10741*, 2021.
- [30] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans *et al.*, "Photorealistic text-to-image diffusion models with deep language understanding," *Advances in neural information processing systems*, vol. 35, pp. 36 479–36 494, 2022.
- [31] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.
- [32] Z. Wang, J. Wang, Z. Liu, and Q. Qiu, "Binary latent diffusion," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 22 576–22 585.
- [33] E. A. Karahan, Z. Liu, A. Gupta, Z. Shao, J. Zhou, U. Khankhoje, and K. Sengupta, "Deep-learning enabled generalized inverse design of multi-port radio-frequency and sub-terahertz passives and integrated circuits," *Nature Communications*, vol. 15, no. 1, p. 10734, 2024.
- [34] C. Niu, M. Phaneuf, and P. Mojabi, "A diffusion model for multi-layered metasurface unit cell synthesis," *IEEE Open Journal of Antennas and Propagation*, vol. 4, pp. 654–666, 2023.
- [35] Z. Zhang, C. Yang, Y. Qin, H. Feng, J. Feng, and H. Li, "Diffusion probabilistic model based accurate and high-degree-of-freedom metasurface inverse design," *Nanophotonics*, vol. 12, no. 20, pp. 3871–3881, 2023.
- [36] J.-H. Bastek and D. M. Kochmann, "Inverse design of nonlinear mechanical metamaterials via video denoising diffusion models," *Nature Machine Intelligence*, vol. 5, no. 12, pp. 1466–1475, 2023.
- [37] F. Mazé and F. Ahmed, "Diffusion models beat gans on topology optimization," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 8, 2023, pp. 9108–9116.
- [38] G. Giannone, A. Srivastava, O. Winther, and F. Ahmed, "Aligning optimization trajectories with diffusion models for constrained design generation," *Advances in Neural Information Processing Systems*, vol. 36, pp. 51 830–51 861, 2023.
- [39] L. Zhu, W. Hua, C. Lv, and Y. Liu, "Rapid inverse design of high degree of freedom meta-atoms based on the image-parameter diffusion model," *Journal of Lightwave Technology*, 2024.
- [40] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16 000–16 009.
- [41] R. Bachmann, D. Mizrahi, A. Atanov, and A. Zamir, "Multimae: Multi-modal multi-task masked autoencoders," in *European Conference on Computer Vision*. Springer, 2022, pp. 348–367.
- [42] X. Geng, H. Liu, L. Lee, D. Schuurmans, S. Levine, and P. Abbeel, "Multimodal masked autoencoders learn transferable representations," *arXiv preprint arXiv:2205.14204*, 2022.
- [43] E. Hooeboom, D. Nielsen, P. Jaini, P. Forré, and M. Welling, "Argmax flows and multinomial diffusion: Learning categorical distributions," *Advances in neural information processing systems*, vol. 34, pp. 12 454–12 465, 2021.
- [44] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, "Cascaded diffusion models for high fidelity image generation," *Journal of Machine Learning Research*, vol. 23, no. 47, pp. 1–33, 2022.
- [45] S. Yellapragada, A. Graikos, K. Triaridis, P. Prasanna, R. Gupta, J. Saltz, and D. Samaras, "Zoomldm: Latent diffusion model for multi-scale image generation," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 23 453–23 463.



Minje Kim (Graduate Student Member, IEEE) received the B.S. degree from the Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea, in 2023, where he is currently pursuing the integrated master's and Ph.D. degree. His current research interests include the generative AI for antenna inverse design and EMI/EMC prediction utilizing Text-to-Image diffusion model, and large-array wireless power transfer system.



Euiho Shin (Graduate Student Member, IEEE) received a B.S. degree in Electrical and Computer Engineering and Industrial Engineering from Seoul National University, Korea, in 2020, where he is currently pursuing a Ph.D. degree. His current research interests include radar systems, synthetic aperture radar (SAR), artificial intelligence (AI), and antenna arrays.



Hooyoung Kim (Graduate Student Member, IEEE) received the B.S. degree in electrical and computer engineering from Seoul National University, Seoul, Korea, in 2024. He is currently pursuing an integrated master's and Ph.D. degree in the Department of Electrical and Computer Engineering at Seoul National University, Seoul, Korea. His current research interests include ray-tracing EM analysis technique for 6G communication, wireless propagation channel measurement and AI-enabled EM analysis.



Taeyeong Yoon (Graduate Student Member, IEEE) received the B.S. degree (summa cum laude) in electronic engineering from Korea Aerospace University, Goyang, South Korea, in 2021, and the M.S. degree from the Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea, in 2023, where he is currently pursuing the Ph.D. degree.

His current research interests include suppressing dual-polarized multi-path coupling, dual-polarized antennas for sub-6 GHz applications, high-efficiency power amplifiers for millimeter-wave (mm-wave) applications, and RF/mm-wave/microwave integrated circuits.



Semin Lee (Graduate Student Member, IEEE) received a B.S. degree in electrical and computer engineering and civil environment engineering from Seoul National University, Korea, in 2025. He is currently pursuing an integrated master's and Ph.D. degree in the Department of Electrical and Computer Engineering at Seoul National University, Seoul, South Korea. His current research interests include ray-tracing EM analysis technique, EM Theory, AI-enabled ray-tracing for 6G channel modeling.



Jungsuek Oh (Senior Member, IEEE) Jungsuek Oh (Senior Member, IEEE) received his B.S. and M.S. degrees from Seoul National University, Korea, in 2002 and 2007, respectively, and a Ph.D. degree from the University of Michigan at Ann Arbor in 2012. From 2007 to 2008, he was with Korea Telecom as a hardware research engineer, working on the development of flexible RF devices. In 2012, he was a postdoctoral research fellow in the Radiation Laboratory at the University of Michigan. From 2013 to 2014, he was

a staff RF engineer with Samsung Research America, Dallas, working as a project leader for the 5G/millimeter-wave antenna system. From 2015 to 2018, he was a faculty member in the Department of Electronic Engineering at Inha University in South Korea. He is currently a full Professor with tenure in the School of Electrical and Computer Engineering, Seoul National University, South Korea. His research areas include mmWave beam focusing/shaping techniques, antenna miniaturization for integrated systems, radio propagation modeling for complex environments, and AI based RF/EM system simulators. He is the recipient of the 2011 Rackham Predoctoral Fellowship Award at the University of Michigan. His group has been recognized with numerous awards at leading academic and industrial conferences. He has published more than 200 technical journal and conference papers, and has served as a TPC member and as a session chair for the IEEE AP-S/USNC-URSI, ISAP and APMC. He has been a senior member of IEEE since 2017.