Global Cross-Attention Transformer for Zero-Shot Metasurface Inverse Design

Li Zhu[1,2], Yichen Wang[1,2], Guangming He[1,2], Ouling Wu[1,2], Hongsheng Chen[1,2*], and Chao Qian[1,2*]

[1] *ZJU-UIUC Institute, Interdisciplinary Center for Quantum Information, State Key Laboratory of Extreme Photonics and Instrumentation, Zhejiang University, Hangzhou 310027, China*
[2] *ZJU-Hangzhou Global Science and Technology Innovation Center, Zhejiang Key Laboratory of Intelligent Electromagnetic Control and Advanced Electronic Integration, Zhejiang University, Hangzhou 310027, China*

(Received 27 January 2026; accepted manuscript online 4 March 2026)

The integration of artificial intelligence with electromagnetic metasurfaces has inaugurated a new era of intelligent metasurfaces, enabling self-adaptive ability for various user demands and in complex environments. However, inverse design, as the core of intelligent metasurfaces, is typically trained based on an assumption of ideal input, thus failing to maintain robustness against complex real-world signal distortions. Here, we present a zero-shot inverse design approach based on the global cross-attention transformer (GCAT), which is immune to complex environmental interferences despite being trained exclusively on pristine, interference-free data. Distinct from convolutional or recurrent architectures, GCAT treats inverse design as a query-based feature retrieval process, employing a set of learnable physical queries to interact with global far-field features. Through a cascaded cross-attention mechanism, the model learns to actively attend to dominant structural invariants while suppressing irrelevant interference components via soft-thresholding. Consequently, GCAT demonstrates exceptional zero-shot robustness, maintaining high reconstruction fidelity with Pearson correlation coefficient larger than 0.97 across diverse unseen interference scenarios whereas baseline models collapse. Our work establishes a resilient algorithmic foundation for the development of intelligent metasurfaces, paving the way for reliable inverse design in dynamic real-world environments.

DOI: 10.1088/0256-307X/43/4/040401 CSTR: 32039.14.0256-307X.43.4.040401

1. Introduction. With the rapid development of modern electromagnetic technologies, metasurfaces, two-dimensional (2D) artificial electromagnetic materials, have achieved precise control over the amplitude, phase, and polarization of electromagnetic waves through sub-wavelength structural design, and thus hold increasingly broad application prospects in electromagnetic manipulation and beyond.[1–6] Artificial intelligence (AI) technology has developed rapidly and has gradually emerged as a pivotal driver for advances in scientific research across various disciplines. Ranging from mainstream applications such as computer vision and machine translation to cutting-edge explorations in fields like physics and materials science, it has brought about revolutionary transformations.[7,8] The integration of AI with electromagnetic metasurfaces has inaugurated a new era of intelligent adaptive metasurfaces, capable of automatic sensing, decision-making, and real-time execution without human intervention.[9–15] This paradigm shift has catalyzed breakthroughs across diverse fields, including electromagnetic cloaking,[16–21] intelligent sensing,[22–26] wireless communications,[27–34] and optical computing.[35–40] Furthermore, recent explorations of novel physical mechanisms (e.g., human-eye-inspired focusing, topological protection, and strong coupling) have further expanded the electromagnetic spectrum control capabilities of metasurfaces, laying a crucial foundation for the cross-scenario applications of intelligent metasurfaces.[41–45]

However, a critical technical adaptation gap remains between current intelligent metasurface systems, developed under ideal environmental assumptions (e.g., no electromagnetic interference, perfect hardware compatibility), and real-world application scenarios characterized by complex factors such as electromagnetic interference and non-ideal hardware characteristics. As the core of intelligent metasurface systems, inverse design algorithms[46,47] inevitably encounter distorted data stemming from complex physical interferences in practical scenarios. These interferences include noise, channel fading, and information loss due to occlusion, as well as their cumulative superposition effects, all of which severely compromise the accuracy and stability of inverse design.[48–50] While classical numerical optimization methods provide high precision, they are often computationally prohibitive for real-time applications.[51,52] Conventional deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), although accelerating the design process, are constrained by local feature extraction or directional sequence bias, making them ill-equipped to capture long-range dependencies in electromagnetic responses.[53] Although advanced strategies such as physics-informed neural networks, generative diffusion models, and transfer learning have been proposed[54–60] to enhance adaptability and physical consistency,[61–67] ex-

*Corresponding authors. Email: hansomchen@zju.edu.cn; chaoq@intl.zju.edu.cn
© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

isting solutions still lack universal robustness and struggle to directly handle diverse, complex interferences without prior training on corrupted data.[68–71]

To bridge this gap, we propose a zero-shot inverse design framework based on a global cross-attention mechanism. This framework aims to endow inverse design algorithms with inherent immunity against far-field data distorted by complex environmental interferences (Fig. 1), thereby ensuring high-accuracy design of the required phase distribution. Distinct from conventional approaches relying on local feature convolution or sequential modeling, we introduce the global cross-attention transformer (GCAT) as the inverse design network (Fig. 2). Its core innovation lies in achieving intrinsic zero-shot robustness against unseen interferences while being trained exclusively on ideal, interference-free datasets. By discarding the conventional direct spatial-mapping paradigm, GCAT employs a two-layer cross-attention mechanism where learnable physical queries iteratively extract information from global feature memory. This compels the network to learn intrinsic electromagnetic topologies rather than overfitting to local regions, effectively filtering interference during retrieval. Extensive benchmarking demonstrates that GCAT significantly outperforms traditional baselines in reconstruction fidelity and zero-shot robustness across diverse interference scenarios (Figs. 3 and 4). This work provides a robust algorithmic foundation for metasurface applications in harsh electromagnetic environments and establishes a new paradigm for the "AI anti-interference" design of intelligent metasurface systems.

2. Results.

2.1. Far-Field Propagation and Distortion in Non-ideal Environments. The transition from idealized simulations to real-world deployment puts forward high requirements for metasurface inverse design. As schematically illustrated in Fig. 1, the signal received is not the pristine theoretical far-field directly radiated by the metasurface. Instead, the electromagnetic wavefront undergoes a non-ideal physical propagation chain where various environmental interferences inevitably accumulate,[72] leading to severe morphological distortions.

Scenario 1: Noise. The noise interference commonly used in research is typically modeled as additive white Gaussian noise (AWGN), which primarily originates from ambient electromagnetic radiation. Characterized by a Gaussian amplitude distribution and uniform spectral power density, this interference tends to submerge critical structural information within low-energy sidelobes, thereby hindering reliable information reception.

Scenario 2: Channel fading. This represents a prevalent interference in electromagnetic propagation environments and primarily stems from multipath propagation. In urban environments with minimal structural shadowing, signals reflect, scatter, and diffract off obstacles to create multiple paths that superimpose at the receiver. These multipath components superimpose at the detector to cause constructive or destructive interference, thereby distorting the amplitude and phase of the received signal. This phenomenon is typically modeled as Rician fading, which utilizes the K-factor to quantify the ratio between the dominant line-of-sight (LoS) path and the scattered multipath components, thereby characterizing the resulting amplitude and phase distortions.

Scenario 3: Information loss. This distorted state is caused by physical blockage and material absorption from obstacles such as buildings or trees. Such obstructions lead to flattening of partial signal envelopes and the resulting unknown or missing status of the corresponding information bitstreams. Ultimately, these distorted far-fields,

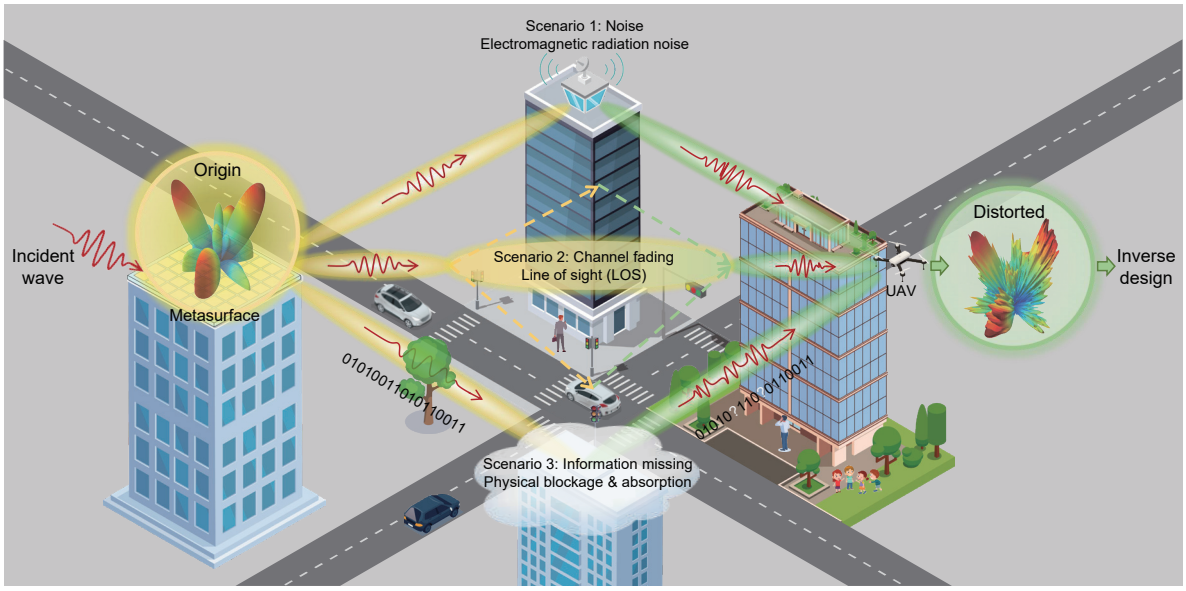

Fig. 1. Far-field propagation and distortion in complex environments. The conceptual propagation chain illustrates the transition of an ideal far-field radiation pattern, generated by a metasurface, into a distorted far-field through three distinct physical interference scenarios: noise, channel fading, and information loss. These distorted far-fields are ultimately collected by a UAV receiver, yielding the severely distorted far-field pattern highlighted in the glowing green circle.

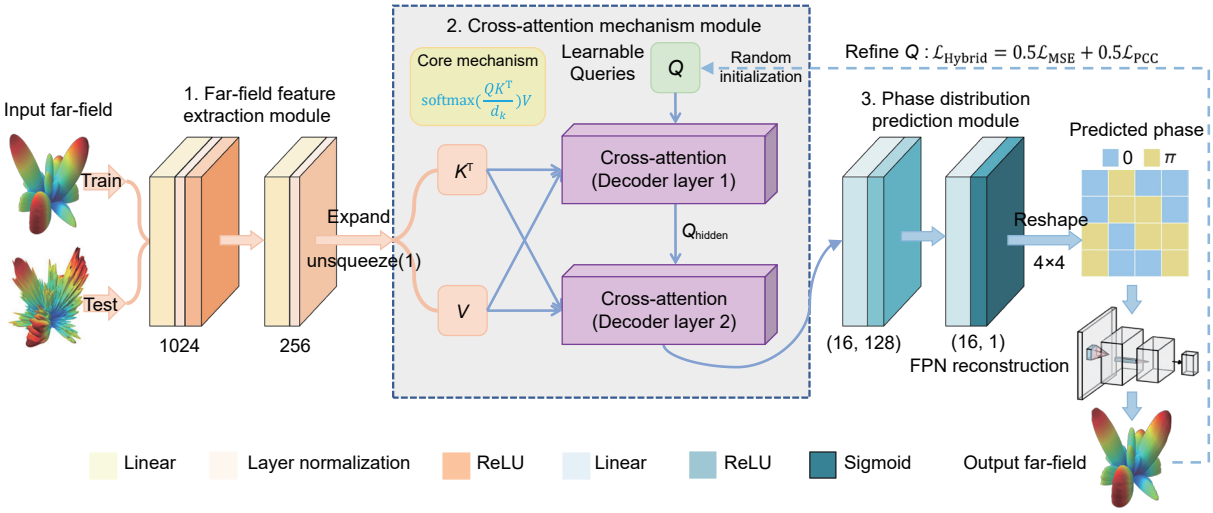

Fig. 2. Architecture of the GCAT and tandem framework training mechanism. The tandem framework consists of GCAT and a pre-trained FPN. The core mechanism of GCAT involves three key modules: (1) The far-field feature extraction module flattens the input far-field pattern into a 1D global vector and maps it into a high-dimensional latent space using linear projection and layer normalization to form the key (K) and value (V) memory bank; (2) The cross-attention mechanism module serves as the core engine, where a set of randomly initialized learnable physical queries (Q) iteratively interact with the far-field features during the training process. Through cascaded decoder layer 1 and decoder layer 2, the queries are refined via the cross-attention mechanism $\mathrm{softmax}[(QK^{\mathrm{T}})/\sqrt{d_k}]V$ to capture global electromagnetic correlations. By computing correlation scores, this module retrieves valid structural information while suppressing irrelevant interference; (3) The phase distribution prediction module maps the attention output to a 16-dimensional phase vector, which is then reshaped into a 4×4 matrix. Finally, the pre-trained FPN serves as a physical constraint, calculating the reconstruction loss of the predicted far-field pattern to ensure physical consistency of the inverse design and refine Q.

collected by receivers such as a single unmanned aerial vehicle, constitute the distorted inputs that the inverse design algorithm must confront. This "reality gap" between idealized training data and distorted operational inputs necessitates a zero-shot inverse network. Such a model acts as an "AI-based filte" to actively denoise and repair far-field signals based on underlying physical laws, achieving high-accuracy retrieval of the required phase distributions without any prior exposure to interference-distorted data.

2.2. Global Cross-Attention Transformer Framework and Training Mechanism. To implement zero-shot inverse design, we propose the GCAT, an architecture that fundamentally shifts the inverse design paradigm from conventional "image-to-image translation" to "query-based feature retrieval". Distinct from traditional approaches, GCAT employs a set of learnable physical queries to interact with global far-field features via a two-layer cross-attention decoder, enabling high-accuracy retrieval of required phase distributions without any prior exposure to interference-distorted data. To facilitate the training of GCAT, we constructed a zero-shot tandem framework comprising the core trainable GCAT and a pre-trained forward prediction network (FPN). The FPN acts as a high-fidelity physical simulator, Pearson correlation coefficient (PCC) is greater than 0.99, and provides essential gradient guidance to ensure the predicted phase distributions adhere to electromagnetic principles. A critical distinction of our experimental setup is the "clean-training, noisy-testing" protocol (Fig. 2), where all models are trained on ideal, interference-free data. This rigorous setting evaluates the intrinsic zero-shot robustness of GCAT and its ability to generalize to unseen, distorted environments—not merely to fit specific noise patterns.

As illustrated in Fig. 2, the superior zero-shot robustness of GCAT is underpinned by its query-driven global cross-attention architecture, which consists of three functional modules. First, the far-field feature extraction module flattens input 2D far-field patterns into one-dimensional (1D) global context vectors. This captures long-range dependencies instantly, circumventing the locality bias of CNNs and the sequential information loss of RNNs. As a result, this linear projection preserves global far-field features, forming a shared memory bank (keys K and values V) representing the features of the input far-field. Second, the cross-attention mechanism module employs a set of learnable physical queries (Q) to interact with this memory bank. The physical queries are initialized using a standard Gaussian distribution with a mean of 0 and a standard deviation of 1 to ensure a diverse representation of the initial search space. Unlike CNNs, which process local pixels indiscriminately, the cross-attention operation calculates a similarity score between the queries and the global far-field features. For the "one-to-many" far-field-to-phase mapping, cross-attention outperforms self-attention in efficiency and feature discrimination, reducing computational complexity and enabling more effective learning of the far-field's global features. This mechanism acts as a natural "soft filter": since the learnable queries are optimized to recognize physical signal patterns (e.g., main lobes), they generate high attention weights for signal components while assigning near-zero weights to irrelevant noise or interference.

Furthermore, we employ a two-layer cross-attention decoder architecture. In the first decoder layer, initialized queries perform an active "coarse exploratory learning" of the global context via the following operation:

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^{\text{T}}}{\sqrt{d_k}}\right)V\,, \tag{1}$$

where d_k denotes the scaling factor derived from the embedding dimension of the hidden features to prevent gradient vanishing during the softmax operation. Since queries search for structural invariants (e.g., lobe orientation) rather than artifacts, they act as matched filters orthogonal to random noise. In the second layer, queries perform "fine-grained feature refinement", locking onto the intrinsic topological skeleton. This iterative refinement decouples stable physical signals from high-dimensional interference, ensuring that the final attention output contains clean physical semantics. Finally, the attention output is mapped to the discrete phase distribution via the phase prediction module. By leveraging this query-driven cross-attention mechanism, GCAT actively extracts structural invariants from corrupted fields rather than passively memorizing data patterns, thereby achieving intrinsic zero-shot robustness.

Unlike conventional design paradigms—where the inverse network learns to map the far-field input directly to the ground truth phase distribution, potentially leading to violations of physical laws or failure to resolve the non-uniqueness problem—our tandem strategy leverages an unsupervised reconstruction loss to facilitate the indirect training of GCAT, ensuring that the network learns to prioritize physical consistency. Technically, the GCAT operates by taking the flattened global vector of the original far-field pattern Y_{orig} as input, from which it subsequently predicts the required phase distribution P_{pred}. This P_{pred} is then fed into the frozen FPN to generate a reconstructed far-field pattern Y_{reco}. The FPN provides a physics-consistency constraint for GCAT's gradient backpropagation. This compels the outcomes of GCAT's inverse design to align with real-world physical laws and overcome the non-uniqueness problem. The indirect training objective is to minimize the discrepancy between Y_{reco} and the original Y_{orig} using a hybrid loss function $\mathcal{L}_{\text{Hybrid}}$ combining mean squared error (MSE) and PCC. The dual-metric hybrid loss function leverages the complementary strengths of MSE and PCC, which prevents the model from falling into local optima that favor numerical accuracy over physical structure, given by

$$\begin{aligned}\mathcal{L}_{\text{Hybrid}} &= w_{\text{MSE}}\cdot\mathcal{L}_{\text{MSE}}(Y_{\text{reco}},Y_{\text{orig}})\\ &\quad + w_{\text{PCC}}\cdot\mathcal{L}_{\text{PCC}}(Y_{\text{reco}},Y_{\text{orig}})\,,\end{aligned} \tag{2}$$

where $w_{\text{MSE}}=0.5$ and $w_{\text{PCC}}=0.5$ are the weighting factors. The PCC is used to evaluate the linear correlation and structural similarity between the patterns, given by

$$\begin{aligned}&\text{PCC}\\ &=\frac{\sum_{i=1}^{n}(Y_{\text{reco},i}-\mu_{Y_{\text{reco}}})(\hat{Y}_{\text{reco},i}-\mu_{Y_{\text{reco}}})}{\sqrt{\sum_{i=1}^{n}(Y_{\text{reco},i}-\mu_{Y_{\text{reco}}})^2\sum_{i=1}^{n}(\hat{Y}_{\text{reco},i}-\mu_{Y_{\text{reco}}})^2}}\,,\end{aligned} \tag{3}$$

where $\mu_{Y_{\text{reco}}}$ and $\mu_{Y_{\text{orig}}}$ represent the average values of the predicted and target far-field intensities across all n sampling points, respectively.

This "prediction-reconstruction" loop forces GCAT to learn the intrinsic, robust mapping rules that are inherently transferable to external interference. Consequently, the proposed framework eliminates the need to retrain the network with specific types of distorted data beforehand, as it can proceed directly to inference and robustness testing. Subsequent interference experiments will demonstrate its superior performance and zero-shot robustness compared to classical baseline network architectures, such as CNNs (e.g., U-Net) and RNNs [e.g., bidirectional gated recurrent unit (BiGRU), bidirectional long short-term memory (BiLSTM), and long short-term memory (LSTM)].

2.3. Visual and Statistical Validation under Specific Noise Interference. To validate the superior performance and robustness of GCAT, we conducted benchmarks where sidelobe regions of far-field patterns were subjected to 10 dB signal-to-noise ratio (SNR) AWGN interference. The main lobe remained intact to isolate the network's ability to distinguish signals in low-energy regions.

Figure 3 presents a visual comparison of the reconstruction results, where all models were trained on clean data but tested on noisy inputs. Figure 3(a) displays the ground truth: the first plot shows the target three-dimensional (3D) far-field radiation pattern corresponding to a randomly selected metasurface phase distribution example; the second plot shows the input fed to these models—a distorted pattern where the sidelobes are corrupted by 10 dB noise. The coordinate system indicates a standard right-handed Cartesian system used for plotting. The subsequent plots in Fig. 3(a) illustrate the reconstruction outputs from three representative networks: our proposed GCAT, the BiGRU (representing RNNs), and the U-Net (representing CNNs). Visually, the GCAT reconstruction is nearly indistinguishable from the ground truth, effectively filtering artifacts while restoring smooth topology and deep nulls. This confirms that the cross-attention mechanism accurately retrieves signal components from noise. In contrast, BiGRU exhibits reduced fidelity in null regions, while U-Net suffers from severe morphological distortion. U-Net's degradation is largely attributable to its local inductive bias: convolutional kernels prioritize local features, leading the network to misinterpret stochastic noise as valid structural patterns.

To further quantify these visual observations, figures 3(b) and 3(c) provide a detailed comparative analysis of the reconstruction fidelity and statistical stability across the test set. The 2D Cartesian cuts at zenith angle $\theta=90^\circ$ [Fig. 3(b)] reveal that the normalized amplitude curves reconstructed by GCAT achieve a near-perfect overlap with the ground-truth target, maintaining a PCC of 0.9991. Notably, GCAT accurately recovers the sharp transitions and the precise depths of the nulls between radiation lobes, whereas the baseline models exhibit significant deviations. For instance, the BiGRU-reconstructed

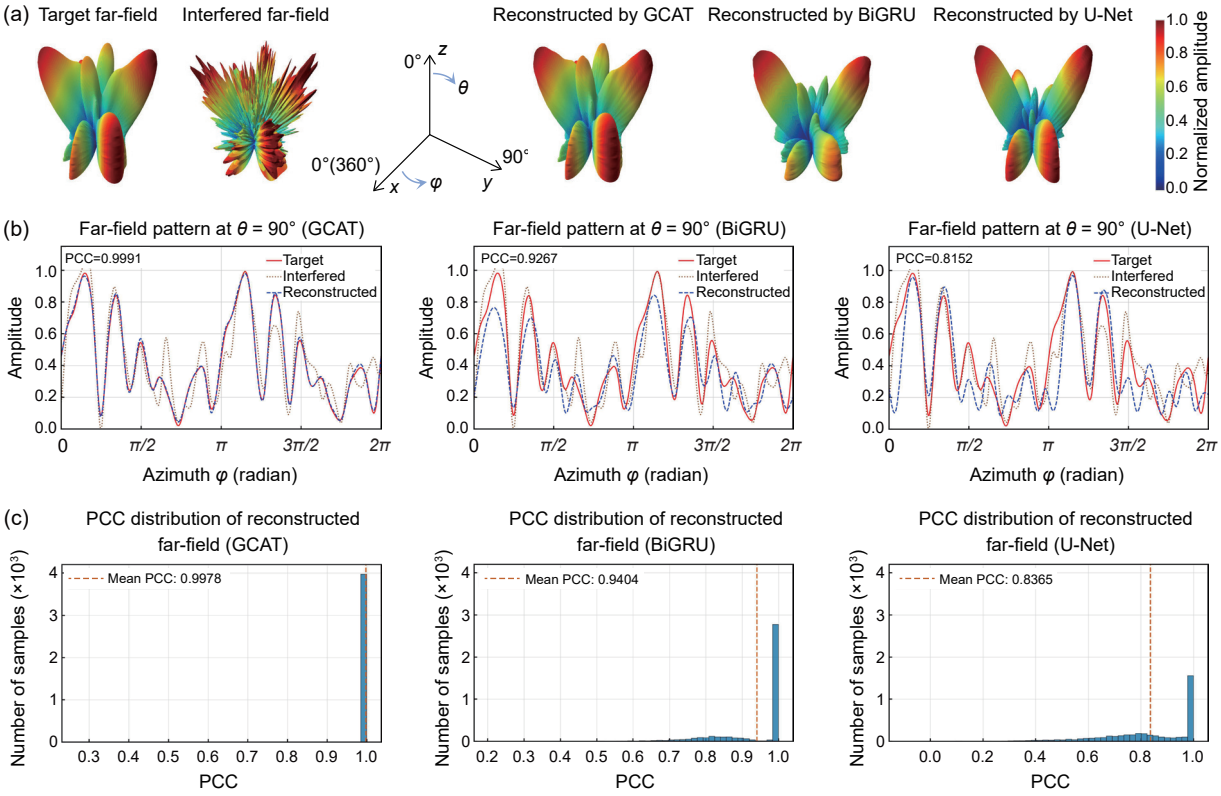

Fig. 3. Visual and quantitative comparison of inverse design performance under AWGN interference. (a) 3D visualization of far-field patterns for a random sample. The sequence from left to right presents the ground truth target far-field, the interfered input far-field at an SNR of 10 dB, and the reconstructed patterns generated by the proposed GCAT, BiGRU, and U-Net, respectively. GCAT successfully recovers both main lobe and sidelobe details, whereas other models exhibit significant beam distortion. (b) 2D visual and quantitative comparison for the same random sample. The panels display the 2D Cartesian cuts at $\theta = 90°$ and the resulting PCC between the reconstructed and target far-fields. The reconstructed normalized amplitude curve of GCAT shows near-perfect overlap with the target, achieving a reconstruction accuracy of 99.91%, whereas a noticeable gap remains for the other baseline models. (c) Statistical distribution of the PCC for reconstructed far-fields across the entire test set. To eliminate individual sampling bias, the evaluation is extended to 4000 test samples. The histograms illustrate that the proposed GCAT achieves a mean PCC of 0.9978 with a narrow distribution, demonstrating superior stability and precision compared to the broader, lower-accuracy distributions of BiGRU (mean PCC = 0.9404) and U-Net (mean PCC = 0.8365).

curve shows noticeable amplitude fluctuations in the valley regions, and the U-Net output fails to track the target morphology, resulting in a substantially lower PCC of 0.8152 for this specific sample. To eliminate individual sampling bias and evaluate the universal robustness of the framework, we conducted a large-scale statistical evaluation across 4000 independent test samples [Fig. 3(c)]. The statistical results demonstrate that GCAT achieves a mean PCC of 0.9978, with the vast majority of samples clustered in an exceptionally narrow, high-accuracy distribution. In sharp contrast, the PCC distributions for BiGRU (mean PCC = 0.9404) and U-Net (mean PCC = 0.8365) are significantly broader and shifted toward lower fidelity, indicating a lack of stability when confronting stochastic interference. These quantitative results prove that GCAT leverages global cross-attention to extract structural invariants and suppress local distortions, ensuring superior robustness and statistical reliability.

2.4. Benchmarking Zero-Shot Robustness across Diverse Interference Scenarios. To further validate the systemic superiority and reliability of the proposed GCAT framework, we expanded our evaluation to include five distinct network architectures: GCAT, BiGRU, U-Net, BiLSTM, and LSTM. While the previous evaluation focused on static noise and single-factor analysis, this section evaluates the model's ability to maintain zero-shot robustness across distinct dynamic electromagnetic interference scenarios. As illustrated in Fig. 4(a), the evaluation is conducted within a zero-shot tandem framework, where the inverse network module serves as the primary focus of the performance comparison. Crucially, all models are tested on distorted data after being trained exclusively on pristine samples, while the pre-trained FPN remains fixed to provide a stable physical constraint. Furthermore, the training trajectories in Fig. 4(b) show rapid convergence and tight alignment between training and validation losses. This ensures that GCAT learns generalized, transferable physical mapping rules rather than overfitting, enabling robust deployment in unseen environments.

To systematically quantify this resilience, we conducted extensive benchmark tests across the following four dynamic interference scenarios.

Scenario 1: AWGN. We varied the intensity of AWGN applied to the sidelobes from a severe 0 dB to a mild 20 dB

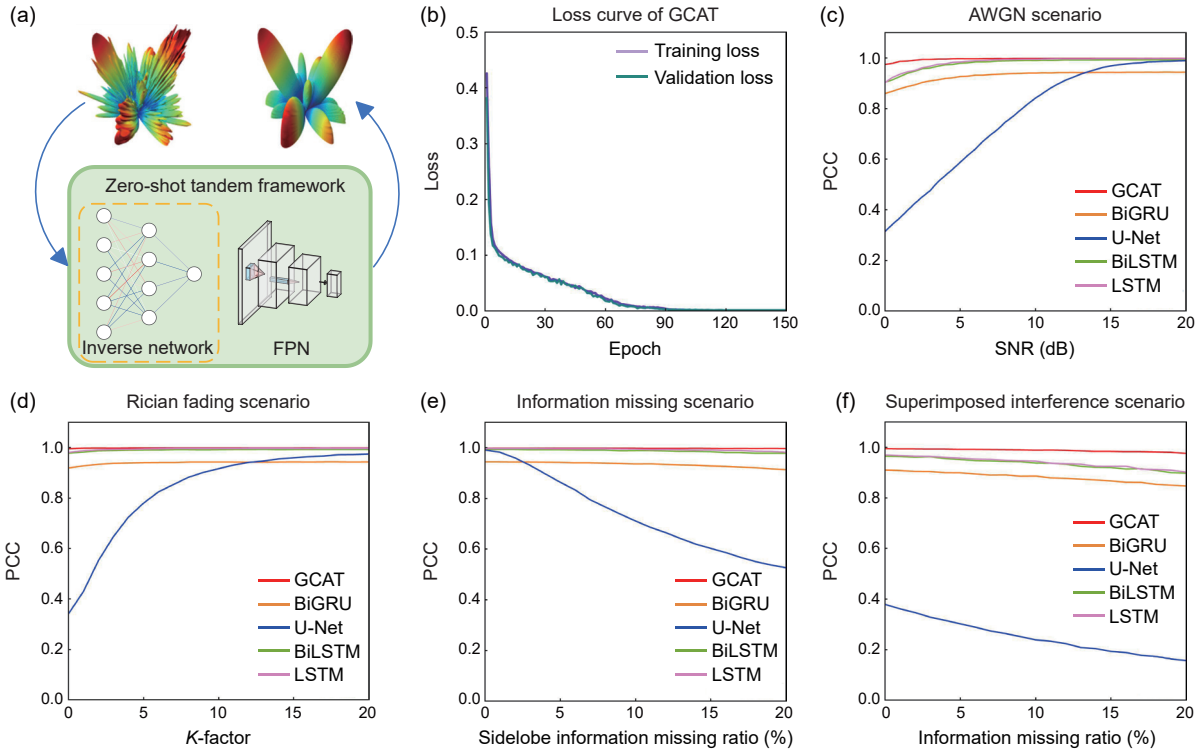

Fig. 4. Comprehensive zero-shot robustness evaluation across multiple electromagnetic interference scenarios. (a) Simplified schematic of the zero-shot tandem framework under multiple interference scenarios test. The inverse network module serves as the primary focus of this evaluation, encompassing the proposed GCAT and various baseline models for performance comparison, while the FPN remains fixed via pre-training. (b) The training and validation loss curves of GCAT, indicating rapid and stable convergence on pristine data without overfitting. (c) AWGN scenario: mean PCC versus sidelobe SNR. GCAT maintains high fidelity (PCC > 0.97), while RNN-based models exhibit diminished fidelity and CNN-based models undergo linear degradation. (d) Rician fading scenario: mean PCC versus the Rician K-factor. As fading becomes severe ($K \to 0$), GCAT remains robust, whereas U-Net collapses. (e) Information loss scenario: mean PCC versus the sidelobe information loss ratio. GCAT exhibits a flat curve, proving its ability to infer global structure from incomplete data. (f) Superposition interference scenario: a stress test combining fading ($K = 5$), noise (SNR $= 5$ dB), and information loss ratio (0–20%). Even under this triple-interference superposition, GCAT maintains a PCC > 0.97, vastly superior to U-Net (PCC < 0.4) and other baseline models, validating its applicability in harsh environments.

(interval: 1 dB), while all models were trained exclusively on clean data. The performance trajectories are plotted in Fig. 4(c). As expected, all models improve as SNR increases. However, the GCAT framework exhibits superior noise immunity compared to conventional architectures. While the performance of the CNN-based U-Net degrades linearly with decreasing SNR, GCAT maintains high fidelity (PCC $>$ 0.97) even at an extreme noise level of 0 dB. This exceptional zero-shot robustness stems from the global cross-attention mechanism. Mathematically, the AWGN introduces uncorrelated random fluctuations to the far-field features (K and V). However, since the cross-attention score is derived from the dot product between the physical queries (Q) and the keys (K), the stochastic noise components—orthogonal to the structured physical queries—yield near-zero attention weights. Consequently, the softmax operation effectively suppresses these noise artifacts via soft thresholding, allowing the network to selectively aggregate valid electromagnetic information while bypassing the noise floor. This mechanism ensures that GCAT recovers the intrinsic topological structure of the far-field even when the signal is buried in noise, a capability that local convolution operations fundamentally lack.

Scenario 2: Channel Fading. Following the physical model of channel fading caused by multipath propagation—introduced previously and illustrated in Fig. 1—we evaluated GCAT under Rician fading to simulate urban environments. We varied the K-factor from 0 (Rayleigh fading; worst case: no dominant LoS, intense multipath interference) to 20 (mild fading; LoS-dominant, minimal distortion) to cover a sufficient range of channel fading severity levels. The quantitative evaluation in Fig. 4(d) demonstrates that as the K-factor approaches 0, different architectures exhibit a clear divergence in statistical stability. Even under intense multiplicative distortion, where the LoS path is significantly weakened, GCAT leads the pack by maintaining a consistently high PCC. By processing the far-field as a holistic entity via the global cross-attention mechanism, GCAT learns to recognize the invariant topological "skeleton" of the radiation pattern. In contrast, U-Net fails completely at low K values. Since fading acts as a multiplicative global distortion, the local convolution kernels of a CNN capture only small-scale details and cannot perceive the overall topological deformation, rendering them unable to distinguish valid signal features from interference-induced distortion.

Scenario 3: Information Missing. In this scenario, we randomly masked 0 to 20% of the sidelobe regions to zero, simulating data information loss resulting from physical blockage and material absorption by obstacles. The result in Fig. 4(e) is striking. GCAT maintains a flat performance curve, sustaining a PCC > 0.99 even with 20% information loss. Conversely, U-Net shows a nearly linear decay in performance proportional to the missing ratio. This validates the global nature of our GCAT approach: since the queries retrieve information from the global context, they can reconstruct the full phase distribution using the remaining unmasked regions—a capability fundamentally absent in locally constrained CNNs, which fail immediately when local features are masked.

Scenario 4: Superimposed Interference. Real-world environments are rarely governed by a single interference type. To simulate a "worst-case" battlefield or urban communication scenario, we constructed a superposition model: the signal is simultaneously subjected to multipath fading ($K = 5$), AWGN (SNR $= 5$ dB), and data information loss (0–20%). Crucially, unlike the previous single-factor tests, this superposition affects the entire far-field, including the main lobe. As shown in Fig. 4(f), under this crushing interference load, the stratification of model performance becomes absolute. The CNN-based model (U-Net) suffers catastrophic failure. Its PCC values hover below 0.4, rendering it effectively useless. GCAT demonstrates exceptional resilience, maintaining a PCC > 0.97. This conclusively proves that the global cross-attention mechanism provides intrinsic zero-shot robustness, enabling the model to generalize to unseen and complex electromagnetic interference environments. While the RNN models (BiGRU, BiLSTM, and LSTM) maintain basic functionality, they exhibit noticeable performance degradation compared to GCAT. This disparity is primarily attributed to the fundamental difference in feature integration: unlike RNNs that rely on sequential recurrence and can accumulate errors under heavy distortion, GCAT leverages a query-based global attention mechanism. By interacting learnable physical queries with the entire global context, GCAT effectively filters out superimposed noise and multipath artifacts through its soft-thresholding capability, whereas RNNs lack the requisite core mechanism to distinguish structural invariants from complex, non-linear interference.

3. Conclusion. In summary, we have addressed the critical challenge of zero-shot inverse design for intelligent metasurfaces. We proposed the GCAT, an inverse design framework that achieves zero-shot robustness against complex electromagnetic interference. Unlike traditional approaches that require noise-augmented training, GCAT is trained solely on pristine data yet generalizes to heavily corrupted environments. This is achieved through learnable physical queries combined with a two-layer cross-attention decoder, which mimics an active exploratory learning and feature refinement process: the queries iteratively filter out unseen interference (noise, channel fading, and information loss) by locking onto the structural invariants of the far-field via the global cross-attention mechanism. Extensive benchmarks confirm that GCAT significantly outperforms conventional CNN and RNN architectures (trained under identical conditions), maintaining high fidelity (PCC > 0.97) where others fail. This research establishes a resilient algorithmic foundation for robustness in the AI-assisted electromagnetic design field and demonstrates that robust AI models can be designed to actively "filter out" real-world interference, promoting the application of metasurfaces in intelligent communications and radar sensing in complex "out-of-lab" settings.

4. Methods.

4.1. Data Generation. We constructed a synthetic dataset by executing MATLAB simulation code based on antenna array theory for a 4×4 unit cell metasurface array operating at 5 GHz. The unit cell periodicity was set to 42 mm. The phase response of each unit cell was binary-coded (0 or π). We generated 40000 random phase configuration samples and calculated their corresponding far-field radiation patterns using the array factor summation formula. Then we sampled the far-field on the upper hemisphere, where the zenith angle θ ranges over linspace $(0, \pi/2, 50)$ and the azimuth angle φ ranges over linspace $(0, 2\pi, 50)$. This forms a 50×50 sampling grid. All far-field radiation patterns were normalized to $[0, 1]$. We split the dataset into training (32000), validation (4000), and test (4000) sets in an $8 : 1 : 1$ ratio.

4.2. Training Details. Training employed an end-to-end physics-constrained approach. GCAT takes the original far-field as input, and its output is fed into the frozen FPN to compute the reconstruction loss against the target far-field. GCAT and FPN were implemented in PyTorch. The proposed GCAT model takes a flattened 50×50 far-field radiation pattern (input dimension of 2500) as input and outputs a 4×4 phase matrix (16 neurons). The feature extraction backbone consists of a 4-layer multilayer perceptron with hidden dimensions reducing from 2500 to 1024 and finally to 256. This is followed by a 2-layer Transformer decoder module, which utilizes 8 attention heads and an embedding dimension of 256. Each layer incorporates layer normalization and ReLU activation functions to ensure numerical stability. The learnable physical queries were initialized as a (1, 16, 256) parameter tensor. Crucially, the training dataset consisted entirely of ideal, interference-free samples generated by the analytical forward model. No noise, fading, or masking was applied during the training phase. This setup was designed to enforce the model to learn the underlying physics rather than memorize interference patterns. Training employed the advanced hybrid loss combining MSE and PCC. We used the Adam optimizer with a learning rate of 0.0005, a batch size of 128, and trained for 150 epochs with a cosine annealing learning rate scheduler on an NVIDIA GeForce RTX 5060 Ti GPU. Our GCAT model has approximately 5 million (5M) trainable parameters; with batch size 128, each epoch takes approximately 19 seconds, and total training time is about 1 hour on a standard GPU, showing high practical feasibility. All baseline models (U-Net, BiGRU, etc.)

were trained using the exact same clean dataset to ensure a fair comparison. In the actual inverse design process, once the interfered data and model are loaded, the model takes only approximately 2 seconds to perform inference on the entire test set of 4000 samples. This translates to a single-sample inference latency of approximately 0.5 ms. This millisecond-level response speed is several orders of magnitude faster than traditional numerical optimization methods, which often require minutes to converge for a single configuration. Such low latency ensures that our proposed GCAT framework meets the demands for rapid execution of inverse design in dynamic, real-world application environments.

Data Availability. The data supporting the findings of this study are available from the authors upon reasonable request.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Grant Nos. 62550050, 62422514, and 62471432 for C. Q.; U25A20520 and 62475228 for H.C.), and the Natural Science Foundation of Zhejiang Province, China (Grant No. LZ26F010004 for C. Q.).

References

[1] Fan Z X, Qian C, Jia Y T, Feng Y M, Qian H L, Li E P, Fleury R, and Chen H S 2024 *Nat. Commun.* **15** 9416
[2] Fan Z X, Qian C, Jia Y T, Wang Z D, Ding Y Z, Wang D P, Tian L W, Li E P, Cai T, Zheng B, Kaminer I, and Chen H S 2022 *Sci. Adv.* **8** eabn7905
[3] Wen E D, Yang X Z, and Sievenpiper D F 2023 *Nat. Commun.* **14** 7736
[4] Huang M, Zheng B, Cai T, Li X F, Liu J, Qian C, and Chen H S 2022 *Nanophotonics* **11** 2001
[5] Wang X Y, Xi R, Hou X, Qiu H R, Liu Z H, Xia D X, Kang X K, Ma S Y, Zhang Y H, Li L, Lan L, and Liao G S 2025 *Prog. Electromagn. Res.* **183** 67
[6] Saifullah Y, Wang H P, Qian C, and Chen H S 2025 *Adv. Funct. Mater.* e21623
[7] Krizhevsky A, Sutskever I, and Hinton G E 2017 *Commun. ACM* **60** 840
[8] Feng N X, Wang H, Wang X, Zhang Y X, Qian C, Huang Z X, and Chen H S 2024 *Prog. Electromagn. Res.* **180** 1
[9] Qian C, Kaminer I, and Chen H S 2025 *Nat. Commun.* **16** 1154
[10] Feng F, Huo D W, Zhang Z Y, Lou Y J, Wang S Y, Gu Z J, Liu D S, Duan X H, Wang D Q, Liu X W, Qi J, Yu S L, Du Q Y, Chen G Y, Lu C C, Yu Y, Ren X F, and Yuan X C 2025 *Adv. Photonics* **7** 024001
[11] Wu N F, Sun Y X, Hu J T, Yang C, Bai Z C, Wang F L, Cui X Z, He S J, Li Y J, Zhang C, Xu K, Guan J, Xiao S M, and Song Q H 2025 *eLight* **5** 5
[12] Qian C and Chen H S 2025 *J. Opt.* **27** 012501
[13] Ma Q, Bai G D, Jing H B, Yang C, Li L, and Cui T J 2019 *Light Sci. Appl.* **8** 98
[14] Qian C, Tian L W, and Chen H S 2025 *Light Sci. Appl.* **14** 93
[15] Qian C, Zheng B, Shen Y C, Jing L, Li E P, Shen L, and Chen H S 2020 *Nat. Photonics* **14** 383
[16] Qian C, Jia Y T, Wang Z D, Chen J T, Lin P J, Zhu X Y, Li E P, and Chen H S 2024 *Adv. Photonics* **6** 016001
[17] Xu Z H, Xu S, Qian C, Xu W Y, Ren H, Su W M, Chen Q D, Chen H S, and Sun H B 2024 *Proc. Natl. Acad. Sci. U.S.A.* **121** e2309096120
[18] Wang Z D, Qian C, Lin P J, Zheng B, Kim G, Noh J, Li E P, Rho J, and Chen H S 2024 *Adv. Mater.* **36** 2400797
[19] Saifullah Y, He Y, Boag A, Yang G, and Xu F 2022 *Adv. Sci.* **9** 2203747
[20] Hua Y F, Qian C, Chen H S, and Wang H P 2022 *Mater. Today Phys.* **27** 100754
[21] Qian C and Chen H S 2021 *Appl. Phys. Lett.* **118** 180501
[22] Liu C, Huang Y S, Zhang M Y, Ma Y M, Yu W M, Wu J J, Li L L, Ma Q, and Cui T J 2025 *Adv. Opt. Mater.* **13** e01899
[23] Li W H, Ma Q, Liu C, Zhang Y F, Wu X N, Wang J W, Gao S Z, Qiu T S, Liu T H, Xiao Q, Wei J X, Gu T T, Zhou Z Z, Li F S, Cheng Q, Li L L, Tang W X, and Cui T J 2023 *Nat. Commun.* **14** 989
[24] Weng J K, Ding Y J, Hu C B, Zhu X F, Liang B, Yang J, and Cheng J C 2020 *Nat. Commun.* **11** 6309
[25] Qian C, Wang Z D, Qian H L, Cai T, Zheng B, Lin X, Shen Y C, Kaminer I, Li E P, and Chen H S 2022 *Nat. Commun.* **13** 2694
[26] Wang Z D, Qian C, Cai T, Tian L W, Fan Z X, Liu J, Shen Y C, Jing L, Jin J M, Li E P, Zheng B, and Chen H S 2021 *Adv. Intell. Syst.* **3** 2100066
[27] Basar E 2020 *IEEE Trans. Commun.* **68** 3187
[28] Wu Q Q, Zheng B X, You C S, Zhu L P, Shen K M, Shao X D, Mei W D, Di B Y, Zhang H L, Basar E, Song L Y, Di Renzo M, Luo Z Q, and Zhang R 2024 *Proc. IEEE* **112** 724
[29] Chen Z N, Qing X M, Liu W, Sheng H W, Li T, and Zhang B 2024 *IEEE Commun. Mag.* **62** 94
[30] Yuan S, Ding Z Z, Zhang X G, Wang D J, Sun J K, and Jiang W X 2025 *Prog. Electromagn. Res.* **184** 24
[31] Liao D S, Wang C, Zhu X K, Jing L Q, Li M, and Wang Z J 2025 *Prog. Electromagn. Res.* **183** 81
[32] Djoudi M, Tounsi M L, Sarrazin J, and Casaletti M 2025 *Prog. Electromagn. Res.* **183** 9
[33] Qian C Y, Qi Z J, Wang Z X, Xu H, Li X J, Shao R W, Zhang J W, Fan L G, Chen Y J, Luo Z J, Dai J Y, Wu J W, Cheng Q, and Cui T J 2025 *Prog. Electromagn. Res.* **184** 64
[34] Zhao Y F, Lv Q H, Guan Y L, Ismail A M, Feng Y, Wang Z Y, Lin D Y, Niu H, Chen R, and Yuen C 2025 *Electromagn. Sci.* **3** 0110022
[35] Rong C G, Wu L, Tao J, Cheng Y Z, Wang K Y, Chen L, Luo H, Chen F, and Li X C 2025 *J. Light. Technol.* **43** 8538
[36] Qian C, Lin X, Lin X B, Xu J, Sun Y, Li E P, Zhang B L, and Chen H S 2020 *Light Sci. Appl.* **9** 59
[37] Huang C X, Zhang J J, Cheng Q, and Cui T J 2021 *Adv. Funct. Mater.* **31** 2103379
[38] Liu J, Wu Q H, Sui X B, Chen Q, Gu G H, Wang L P, and Li S C 2021 *PhotoniX* **2** 5
[39] Tan Q Z, Qian C, Cai T, Zheng B, and Chen H S 2022 *Prog. Electromagn. Res.* **175** 13947
[40] Yu Z C, Gu Z, Chen L, Su J L, Qin S L, Cai Z X, Yu X Y, Ma Q, You J W, and Cui T J 2025 *Adv. Opt. Mater.* **13** e01180
[41] Qian C, Yang Y, Hua Y F, Wang C, Lin X, Cai T, Ye D X, Li E P, Kaminer I, and Chen H S 2022 *Nat. Commun.* **13** 4383
[42] Lu H, Zhao J W, Zheng B, Qian C, Cai T, Li E P, and Chen H S 2023 *Nat. Commun.* **14** 3301
[43] Qin S L, Chen L, Cai Z X, Gu Z, Yu X Y, Yu Z C, Su J L, Xiao Q, Li X Y, Ma Q, You J W, and Cui T J 2025 *Adv. Opt. Mater.* **13** 2500044
[44] He G M, Qian C, Jia Y T, Fan Z X, Wang H P, and Chen H S 2025 *Adv. Opt. Mater.* **13** 2401933
[45] Allayarov I, Aita V, Roth D J, Van Casteren B, Bykov A Yu, Evlyukhin A B, Zayats A V, and Calà Lesina A 2025 *Light Sci. Appl.* **14** 387
[46] Wei H R, He X Y, and Cao W H 2025 *Optica* **12** 505

[47] Han X Q, Wang X D, Xu M Y, Feng Z, Yao B W, Guo P J, Gao Z F, and Lu Z Y 2025 *Chin. Phys. Lett.* **42** 027403
[48] Chen S Y, Feng Q, Ran L Y, Ding Z P, Xi F, Li H T, Tian S R, and Liu Z 2025 *IEEE Trans. Veh. Technol.* **74** 6222
[49] Pradhan A, Abd-Elmagid M A, Dhillon H S, and Molisch A F 2024 *IEEE Trans. Wirel. Commun.* **23** 1783
[50] Han H M, Shen T J, Chen Y X, Lu W D, Zheng S L, Wang X B, and Yang X N 2024 *IEEE Trans. Veh. Technol.* **73** 7351
[51] Elsawy M M R, Lanteri S, Duvigneau R, Fan J A, and Genevet P 2020 *Laser Photonics Rev.* **14** 1900445
[52] Jung I, Peng Z, and Rahmat-Samii Y 2024 *Electromagn. Sci.* **2** 1
[53] Saifullah Y, Wu N X, Wang H P, Zheng B, Qian C, and Chen H S 2025 *Adv. Photon.* **7** 034005
[54] Wu B, Qian C, Wang Z D, Lin P J, Li E P, and Chen H S 2025 *Light Sci. Appl.* **14** 211
[55] Wu N X, Qian C, Wang Z D, Zhu X Y, Xiao C, Li E P, and Chen H S 2025 *Adv. Funct. Mater.* **35** 2502678
[56] Liu Z C, Zhu D Y, Rodrigues S P, Lee K T, and Cai W S 2018 *Nano Lett.* **18** 6570
[57] Xu X, Feng H Z, Gu M N, Li C X, and Tian Y 2025 *J. Electron. Mater.* **54** 11661
[58] Shan H Y and Jiang T 2025 *IEEE Trans. Antennas Propag.* **73** 8852
[59] Fang Z W and Zhan J 2020 *IEEE Access* **8** 24506
[60] Xiao C, Wu N X, Hong Z Y, Wu X B, Chen H S, and Qian C 2025 *Adv. Theory Simul.* **8** e01063
[61] Fan Z X, Qian C, Jia Y T, Chen M, Zhang J, Cui X S, Li E P, Zheng B, Cai T, and Chen H S 2022 *Phys. Rev. Appl.* **18** 024022
[62] Zhu E Z, Zong Z, Li E J, Lu Y, Zhang J W, Xie H, Li Y, Yin W Y, and Wei Z 2025 *Nat. Commun.* **16** 2251
[63] Zhang J, Qian C, Fan Z X, Chen J T, Li E P, Jin J M, and Chen H S 2022 *Adv. Opt. Mater.* **10** 2200748
[64] Jia Y T, Qian C, Fan Z X, Cai T, Li E P, and Chen H S 2023 *Light Sci. Appl.* **12** 82
[65] Lin P J, Qian C, Jia Y T, Wang Z D, Wu B, Huangfu J T, and Chen H S 2024 *IEEE Trans. Antennas Propag.* **72** 5252
[66] You G F, Qian C, Tan S R, Li E P, and Chen H S 2024 *Phys. Rev. Appl.* **22** L041002
[67] Huang C, Jia Y T, Li M, Fang Y, Wu A W, Chen H S, and Qian C 2025 *Laser Photonics Rev.* e02298
[68] Wu N X, Jia Y T, Qian C, and Chen H S 2023 *Adv. Opt. Mater.* **11** 2202130
[69] Chen J T, Qian C, Zhang J, Jia Y T, and Chen H S 2023 *Nat. Commun.* **14** 4872
[70] Fu Q S, Wang D W, Hu Y, Zhou W Y, Wang X, Liu J, and Zhao W S 2025 *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 1
[71] Wang Z D, Qian C, Saifullah Y, Fang Y, Li E P, and Chen H S 2025 *IEEE Trans. Antennas Propag.* **73** 432
[72] Lin P J, Qian C, Zhang J, Chen J T, Zhu X Y, Wang Z D, Huangfu J T, and Chen H S 2023 *Prog. Electromagn. Res.* **178** 83