WDNN: Weighted Diffractive Neural Network for Physical-layer RF Signal Processing

Yezhou Wang*‡
Shanghai Jiao Tong University
yezhouwang@sjtu.edu.cn

Yongjian Fu*‡
Central South University
fuyongjian@csu.edu.cn

Hao Pan†
Microsoft Research Asia, Shanghai Jiao Tong University
panhao@microsoft.com

Qinyun Hu
Shanghai Jiao Tong University
hqy0126@sjtu.edu.cn

Lili Qiu
Microsoft Research Asia, UT Austin
liliqiu@microsoft.com

Yi-Chao Chen
Shanghai Jiao Tong University
yichao@sjtu.edu.cn

Guangtao Xue‡
Shanghai Jiao Tong University
gt_xue@sjtu.edu.cn

Ju Ren
Tsinghua University
renju@tsinghua.edu.cn

Abstract

Diffractive neural networks (NNs) have garnered attention for directly implementing wireless signal processing at the physical layer. However, they are limited by a constrained weight learning space and activation functions, which restricts their data processing capabilities. To address this, we propose an RF circuit-based weighted diffraction NN (WDNN) that rivals digital NNs in processing ability. We design a weighted asymmetric RF coupler unit that, when stacked into a network, enables diffractive propagation with arbitrary connection weights. Additionally, an activation module is introduced that utilizes RF amplifiers operating in their nonlinear regions. We validate the effectiveness of the proposed WDNN through three tasks: 32-level amplitude modulated (AM) signal decoding, 31-class angle of arrival (AoA) estimation, and 2-class Wi-Fi based fall detection. After training, WDNN achieves the accuracy of 98.5%, 93.7%, and 90.8% in the AM decoding, AoA estimation, and fall detection tasks, respectively; while the diffractive NN SOTA achieves only 21.6%, 16.9%, and 63.3%. We also implement the prototypes of WDNN and SOTA, and real-world experimental results demonstrate that our method achieves an average accuracy improvement of up to 76.85% across various tasks compared to SOTA.

CCS Concepts

• **Hardware** → **Digital signal processing**; **Wireless devices**;
• **Computing methodologies** → *Machine learning*.

Keywords

Diffractive neural network; Signal processing; Radio-frequency hardware

ACM Reference Format:
Yezhou Wang, Yongjian Fu, Hao Pan, Qinyun Hu, Lili Qiu, Yi-Chao Chen, Guangtao Xue, and Ju Ren. 2025. WDNN: Weighted Diffractive Neural Network for Physical-layer RF Signal Processing. In *The 31st Annual International Conference on Mobile Computing and Networking (ACM MOBICOM '25), November 4–8, 2025, Hong Kong, China.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3680207.3765240

1 Introduction

In wireless systems, direct signal processing in the analog domain attains widespread attention due to its inherent advantages. Recent studies have enhanced feature extraction in wireless signals at the physical layer. Minimalist vision [22] introduces a 3D-printed cover as a physical feature extraction layer for computer vision, enabling lightweight tasks such as head detection and traffic monitoring with an 8-pixel CMOS sensor. In acoustics, specially designed 3D-printed structures improve imaging[1, 10] and communication [49]. In the radio frequency (RF) domain, optimized metasurfaces enhance communication coverage, data rate [9, 24, 25, 30, 35], and indoor localization [42]. These approaches extract or enhance

*Yezhou Wang and Yongjian Fu are the co-primary authors.
†Hao Pan is the corresponding author.
‡Also with Shanghai Key Laboratory of Trusted Data Circulation and Governance in Web3.

ACM MOBICOM '25, Hong Kong, China

ACM ISBN 979-8-4007-1129-9/25/11
https://doi.org/10.1145/3680207.3765240

features at the physical layer, reducing the burden of digital data processing and improving wireless system performance.

Moving beyond traditional physical-layer feature extraction techniques, recent research has introduced diffractive deep neural networks (D2NN [27]), which achieve neural network computations within the physical domain. D2NN leverages multi-layer metasurfaces to diffract signals, with each meta-atom functioning analogously to a neuron in a digital neural network. This approach fundamentally shifts signal manipulation from digital computation towards analog processing, showcasing a pathway to ultra-fast and energy-efficient inference. However, D2NN requires sophisticated and precise inter-layer alignment, along with complex calibration procedures, which renders practical implementation challenging. In response to these limitations, planar diffractive neural networks (pla-NN [15]) have emerged, utilizing analog circuits to perform signal processing. By shifting the diffraction operation from wireless channels to wired circuits, pla-NN aims to simplify the overall architecture and improve system stability and scalability.

However, existing state-of-the-art (SOTA) diffraction-based hardware NN models are limited to relatively simple data processing tasks. For example, [15] can only achieve 4-label classification of angle of arrival (AoA) and objects in the 10 GHz band. Meanwhile, [12, 27] demonstrate digit recognition in the optical and terahertz bands. The limited computational capabilities of SOTA diffractive NNs are mainly due to two factors: (1) fixed or minimally adjustable propagation weights, which hinder effective weight learning through backpropagation; and (2) the absence of nonlinear activation modules. As a result, these methods are restricted to basic linear function fitting, in contrast to the nonlinear capabilities of digital NNs.

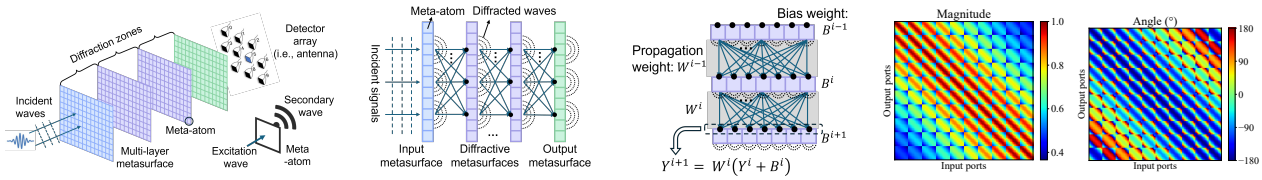
In this study, we propose a weighted diffractive NN (named WDNN) for processing RF signals at the physical layer, aiming to approximate the powerful nonlinear function fitting capabilities of digital NNs. To achieve this, we employ an analog circuit-based approach to design WDNN, addressing the SOTA's limitations as follows: (i) we develop a novel diffractive propagation module with fully adjustable weights, enhancing the learnability of propagation weights; (ii) we integrate nonlinear modules to facilitate activation functions.

To achieve the arbitrarily adjustable diffractive propagation weights, we conduct an in-depth analysis of RF coupler designs and the overall propagation characteristics of networks composed of them. Based on these insights, we design a novel RF coupler that supports asymmetric propagation with flexible and learnable weights. By carefully arranging multiple RF couplers as a propagation network, we can realize an efficient propagation module for connecting two layers with arbitrary weights. Additionally, we use RF attenuators and phase delay lines at each output port of the RF coupler network to represent bias vectors in NNs. To implement nonlinear activation functions, we utilize RF amplifiers as nonlinear modules. By carefully adjusting supply voltages beyond typical ranges, we ensure they operate in the cutoff regions to achieve the desired nonlinear characteristics. In Sec. 3, we provide a detailed description of the design specifics of WDNN. These designs enable WDNN to achieve nonlinear fitting capabilities comparable to digital NNs, including arbitrary propagation weight matrices, bias vectors, and nonlinear activation functions.

To validate the advantages of WDNN in NN computation, we design three wireless tasks: 32-level amplitude modulation (AM) communication at 5 GHz, 31-class angle-of-arrival (AoA) estimation with 4° resolution over the range of -60° to 60° at 5 GHz, and 2-class fall detection based on Wi-Fi signals. For each task, WDNN, pla-NN (SOTA), and digital NN are individually trained on simulated datasets, and their performance is subsequently compared. Simulation results show that WDNN outperforms pla-NN by 76.1% and is only 0.2% behind the digital NN in the AM modulation task; it exceeds pla-NN by 75.9% and is only 3.8% lower than digital NN in the AoA estimation task; and in the WiFi-based pose recognition task, WDNN outperforms pla-NN by 27.5% and is just 1.9% less accurate than the digital NN. For the AM and AoA tasks, we further implement hardware prototypes for WDNN and pla-NN and conduct the real-world experiments. Design parameters for the RF components in each methods are obtained through reverse engineering and the trained network weights, and both are fabricated using printed circuit board (PCB) technology. Experimental results show that: in the AM decoding task, WDNN achieves a raw data decoding accuracy of 98.5%, significantly higher than pla-NN's 21.6%; in the AoA estimation task, WDNN achieves a classification accuracy of 93.7% with 4° resolution, which is also much higher than pla-NN's 16.9%. These results demonstrate that WDNN achieves outstanding performance in RF signal feature extraction and processing at the physical layer.

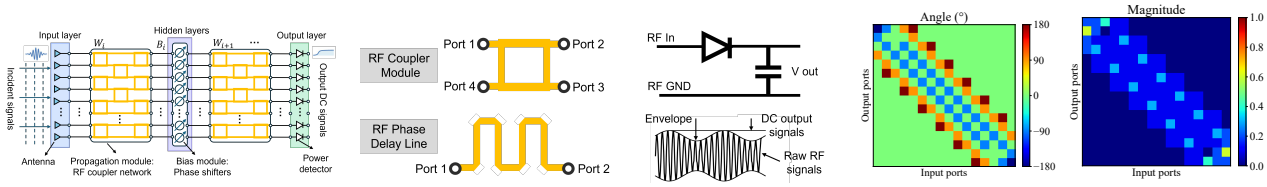
Our contributions are summarized as follows:

- We propose a weighted diffractive neural network, called WDNN, as the first physical-layer NN capable of RF signal processing, featuring arbitrary diffraction propagation weights and nonlinear activation functions.
- We design an innovative RF coupler that achieves arbitrary and asymmetric coupling weights. Leveraging this coupler, we construct an RF coupler network that serves as a propagation module, efficiently establishes connections between layers with arbitrary weights.
- We utilize the nonlinear region of RF amplifiers to implement the nonlinear activation functions.
- We conduct extensive experiments to validate the capabilities of WDNN in RF data processing at the physical layer, showing that it significantly outperforms SOTA and achieves performance close to digital NNs.



(a) Multi-layer metasufaces-based D2NN (3D view and side view) (b) Diffractive propagation (c) Weight matrix of the diffraction zone

Figure 1: An exemplar multi-layer metasurface-based D2NN [27], where each metasurface consists of 10×10 meta-atoms, with each atom acting as a neuron



(a) Analog circuit-based pla-NN structure (b) RF coupler & phase shifter (c) RF power detector (d) Weight matrix of the RF coupler network

Figure 2: An exemplar analog circuit-based pla-NN [15], where each layer is equipped with 16 RF ports, with each port functioning as a neuron

2 Background and Motivation

Neural networks (NNs) are powerful for signal processing. In a NN, neurons are connected across layers through weighted links, with each neuron having an associated bias and activation function. In each hidden layer, the output is computed by $\mathbf{z} = \mathbf{W}\mathbf{x} + \mathbf{b}$, then passed through an activation function τ to yield $\mathbf{a} = \tau(\mathbf{z})$, which serves as the layer's output. The weights $\mathbf{W}$ and biases $\mathbf{b}$ are learned via backpropagation, enabling strong nonlinear fitting. This section examines the existing analog-domain NN computation researches, outlines SOTA methods, and analyzes their limitations.

2.1 Diffractive Neural Network

2.1.1 Metasurface-based solution. The diffractive deep neural network (D2NN), first introduced in [27], is constructed using multi-layer metasurfaces, as shown in Fig. 1(a). Each metasurface layer corresponds to an NN layer: the first is the input layer, the middle serves as hidden layers, and the last is the output layer. Each meta-atom functions as a neuron with a complex transmission coefficient. The term "diffractive" derives from the Huygens-Fresnel principle [5], where each meta-atom receives signals from the previous layer, generates near-field diffraction, and transmits secondary waves to the next layer, as shown in Fig. 1(b). The bias of each meta-atom can be flexibly adjusted via transmission attenuation and phase delay. As signals traverse the metasurface layers, they are transformed into energy-based features, which are detected by a power detector array in the D2NN.

Limitation: The propagation weights between meta-atoms in adjacent layers are determined by the wireless channel, reflecting attenuation and phase delay. As shown in Fig. 1(c), the propagation matrix $\boldsymbol{W}$ between two 10×10 metasurfaces exhibits high amplitude and phase periodicity. This is caused by the periodic arrangement of meta-atoms in the metasurface. Hence, metasurface-based solutions cannot flexibly adjust the propagation matrix. Moreover, in RF scenarios, unlike in the light frequency band [47, 50], meta-atoms struggle with nonlinearity, resulting in the absence of activation functions.

2.1.2 Analog circuit-based solution. The planar diffractive NN (pla-NN), proposed by [15], replaces wireless channels with wired ones to implement diffraction using RF analog circuits. As shown in Fig. 2(a), it consists of the input layer, propagation modules, hidden layer (bias modules), and output layer. The input layer includes multiple antennas to receive RF signals. The propagation module uses an RF coupler network for energy transfer to achieve inter-layer diffraction, the RF coupler unit design [18] is shown in Fig. 2(b). These couplers split energy from one input port (*e.g.*, port 1) to two output ports (*e.g.*, ports 2 & 3), and similarly for another input port (*e.g.*, port 4). The hidden layer applies neuron biases through phase delay lines. In the output layer, RF power detectors, as shown in Fig. 2(c), are used to convert RF signal features into voltage signals, which can then be sampled by ADC chips.

Limitations: The pla-NN shares two key limitations with D2NN. First, its propagation network relies on -3dB RF couplers that split energy evenly with equal phase delays, resulting in fixed, non-learnable weights $\mathbf{W}$ and limited amplitude and phase transfer, as shown in Fig. 2(d) and examined in Sec.3.1. Second, the absence of nonlinear RF modules precludes activation functions. Additionally, pla-NN uses only phase delay lines for bias modulation, which are inadequate for complex signals and provide a smaller learning space for bias vectors than D2NN.

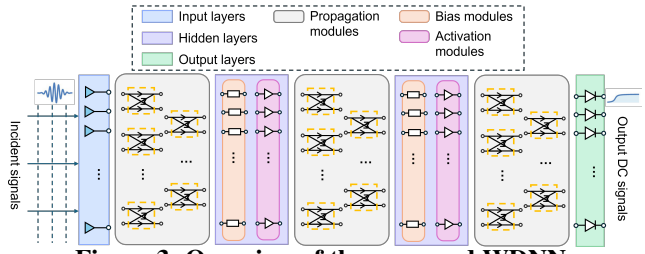


Figure 3: Overview of the proposed WDNN

2.2 Our Tasks

Our objective is to address the limitations of current SOTA diffractive NNs. We seek to develop an innovative solution that demonstrates strong nonlinear fitting capabilities, comparable to those of digital NNs. To achieve this, we need to implement: (1) propagation connections between two layers that allows maximal arbitrary weight adjustments, and (2) support for nonlinear activation functions. Given these requirements, a metasurface-based diffractive NN design is nearly impossible, as it relies entirely on wireless channels to determine propagation weights. Furthermore, designing nonlinear meta-atoms at RF frequencies is extremely challenging. Therefore, in this paper, we adopt an analog circuit design approach to develop a hardware-layer NN, which we name the Weighted Diffractive Neural Network (WDNN).

3 WDNN Design

The architecture of WDNN is shown in Fig. 3. The input layer uses antenna arrays. Our propagation module allows flexible weighted diffractive propagation. The hidden layer includes a bias module for adjustable weights and a nonlinear activation module for nonlinear processing. At the output layer, power detectors convert RF signals to voltage signals for ADC sampling and analysis. Next, we detail the implementations of the propagation, bias, and nonlinear activation modules, which are the core innovations of WDNN.

3.1 Propagation Module

Connecting multiple RF ports with arbitrary propagation weights is challenging. Although RF power dividers and combiners using PCB vias can achieve these connections [7], the via capacitance degrades signal quality, and handling a large number of ports becomes inefficient [6]. The SOTA approach, pla-NN [15], employs a stacked RF coupler network to connect RF ports between layers. We start by analyzing this network and modeling its propagation matrix. We then identify it's limitations and propose an efficient RF coupler design for constructing adjustable propagation matrices.

3.1.1 RF coupler network modeling. A simplified diagram of a standard RF coupler is shown in Fig. 4(b). In a scattering parameter matrix (**S**) of the RF coupler, each element S_{ij} is a complex number, representing the ratio of the output voltage wave at the ith port to the input voltage wave at the jth port. For a well-designed RF coupler, the scattering matrix should exhibit almost zero reflection, implying that $S_{41}, S_{14}, S_{23}, S_{32}, S_{11}, S_{22}, S_{33}, S_{44}$ are almost zero. As shown in Fig. 4(a), we focus on the propagation relationships among the input and output vectors. To simplify the matrix expression, we omit the zeros and define the propagation matrix $\boldsymbol{p}$ for an RF coupler as:

$$\begin{bmatrix} out_1 \\ out_2 \end{bmatrix} = \boldsymbol{p} \begin{bmatrix} in_1 \\ in_2 \end{bmatrix}, \quad \boldsymbol{p} = \begin{bmatrix} S_{21} & S_{31} \\ S_{24} & S_{34} \end{bmatrix} \tag{1}$$

Consider a network consisting of $M \times N$ of RF couplers, arranged as shown in Fig. 4(b). Odd-numbered columns contain M couplers, while even-numbered columns have $M-1$ couplers. Thus, each layer consists of $2M$ "neurons", making the propagation matrix $\mathbf{W}$ a $2M \times 2M$ matrix. Since RF couplers propagate signals column by column, we analyze the network by columns. Starting with the odd-numbered kth column, we define the propagation matrices of each RF coupler in the odd-numbered column are $\mathbf{p}_{k,1}$, $\mathbf{p}_{k,2}$, ..., $\mathbf{p}_{k,2M}$, and each matrix is of size 2×2, then, we can derive the matrix expression for $\mathbf{P}_k$ as follows:

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{p}_{k,1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \cdots & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times2} & \mathbf{p}_{k,2} & \mathbf{0}_{2\times2} & \cdots & \mathbf{0}_{2\times2} \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \mathbf{p}_{k,3} & \cdots & \mathbf{0}_{2\times2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{2\times2} & \mathbf{0}_{2\times2} & \cdots & \mathbf{0}_{2\times2} & \mathbf{p}_{k,2M} \end{bmatrix} \tag{2}$$

Similarly, when k is even, we can obtain the matrix:

$$\mathbf{P}_k = \begin{bmatrix} 1 & \mathbf{0}_{1\times2} & \mathbf{0}_{1\times2} & \cdots & \mathbf{0}_{1\times2} & 0 \\ \mathbf{0}_{2\times1} & \mathbf{p}_{k,1} & \mathbf{0}_{2\times2} & \cdots & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} \\ \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{p}_{k,2} & \cdots & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \cdots & \mathbf{0}_{2\times2} & \mathbf{p}_{k,2M-1} & \mathbf{0}_{2\times1} \\ 0 & \mathbf{0}_{1\times2} & \mathbf{0}_{1\times2} & \cdots & \mathbf{0}_{1\times2} & 1 \end{bmatrix} \tag{3}$$

We obtain the propagation matrix for a network comprised of the stacked RF couplers:

$$\boldsymbol{W} = \boldsymbol{P_1 P_2 ... P_N} \tag{4}$$

3.1.2 Weighted Asymmetric RF Coupler. Using unequal RF couplers is a straightforward approach to enable learnable weights. We explore this by employing an unequal RF coupler [44] with two degrees of freedom to design the propagation module. The detailed pattern is shown on the right part of Fig. 5, and its propagation matrix is as follows:

$$\boldsymbol{p_{cp}}(n_c, \theta_{c1}, \theta_{c2}) = \begin{bmatrix} n_c e^{j\theta_{c1}} & (1-n_c)e^{j\theta_{c2}} \\ (1-n_c)e^{j\theta_{c2}} & n_c e^{j\theta_{c1}} \end{bmatrix} \tag{5}$$

where n_c, θ_{c1}, and θ_{c2} are adjustable parameters and can be designed by the geometric parameters: $L_1, L_2, D_1, D_2, W_1, W_2$,

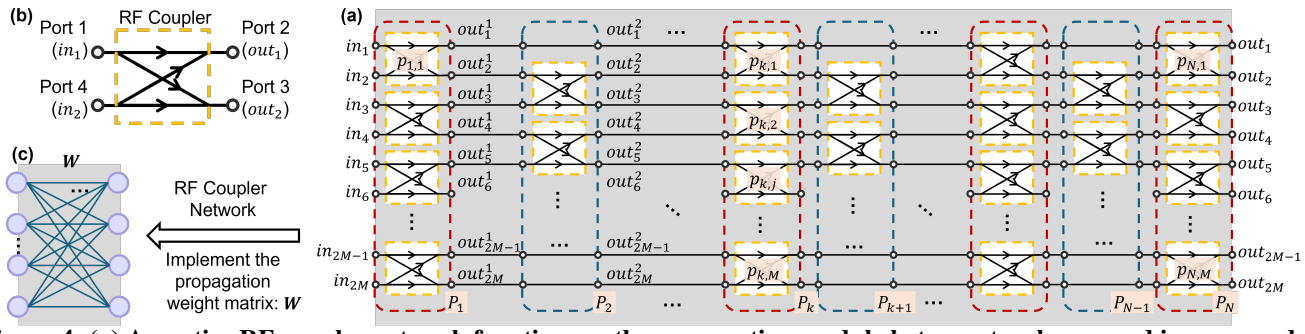


Figure 4: (a) An entire RF coupler network functions as the propagation module between two layers and is composed of numerous RF coupler units. (b) A simplified diagram of an RF coupler. (c) The propagation weight matrix in NNs

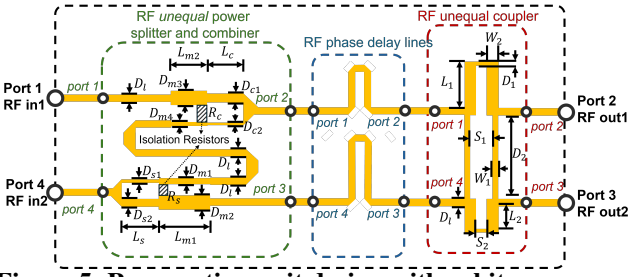


Figure 5: Propagation unit design with arbitrary propagation coefficients

S_1, S_2. Due to the symmetric nature and the principle of reciprocity [33], for the existing unequal RF coupler design, there always satisfy that $S_{21} = S_{34}$ and $S_{31} = S_{24}$. This results in the existing design of unequal RF couplers producing a symmetric propagation matrix. Using existing designs of unequal RF couplers to construct propagation networks encounters limitations in learning optimal propagation weights. Thus, we conduct experiments to verify these limitations: these networks are unable to fit arbitrary weight matrices $\boldsymbol{W}$ in traditional NNs. For details, please refer to Sec. 3.1.3.

In designing a propagation module with maximal adjustability, the core challenge is to develop an RF coupler capable of achieving arbitrary coupling coefficients, allowing independent control of the four complex numbers in the propagation matrix $\boldsymbol{p}$. To achieve this, we draw inspiration from the existing unequal RF coupler designs and introduce additional RF circuits that enables arbitrary power splitting and combining, as well as phase modulation. The specific implementation is illustrated in Fig. 5, which includes three sub-modules.

The first sub-module, depicted on the left in Fig. 5, comprises a power splitter and a power combiner. An unequal power splitter [19, 26] first distributes the RF signal from the input (*e.g.*, port 2) into two branches in a tunable ratio of n:(1-n), where $0 < n < 1$. Subsequently, we design an unequal power combiner [26, 43] to merge a portion of the signal from port 2 with the input port 1, enabling efficient power combining via impedance matching based on pre-determined power ratios. Notably, the combiner does not require the input signals to be phase-aligned; any out-of-phase losses are dissipated through resistive elements. This configuration yields the propagation matrix: $p_{sc} = \begin{bmatrix} 1 & n \\ 0 & 1-n \end{bmatrix}$. In the same manner, the signal from input port 1 can be divided in a ratio of n:(1-n) and combined with the signal from port 2, yielding the following propagation matrix: $p_{sc} = \begin{bmatrix} n & 0 \\ 1-n & 1 \end{bmatrix}$. To enable gradient properties in the propagation matrix, facilitating optimization via gradient descent to learn optimal parameters, we integrate the matrices of the two modes as follows:

$$p_{sc}(\beta, n) = \begin{bmatrix} \beta n + (1-\beta) & (1-\beta)n \\ (1-\beta)(1-n) & \beta + (1-\beta)(1-n) \end{bmatrix} \quad (6)$$

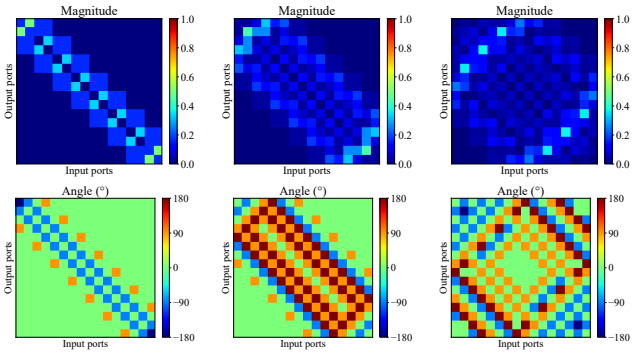
where, β is a step function: $\beta(x) = \frac{1}{1+e^{-10x}}$, n is the power ratio of the splitter. This module is designed to break the power symmetry of the RF coupler unit.

The second sub-module, illustrated in the middle of Fig. 5, comprises two independent RF phase shifters inserted into the signal paths from the previous stage to establish phase symmetry. Letting the phase delays on the two paths be θ_1 and θ_2, the resulting propagation matrix is:

$$p_{pd}(\theta_1, \theta_2) = \begin{bmatrix} e^{j\theta 1} & 0 \\ 0 & e^{j\theta 2} \end{bmatrix} \quad (7)$$

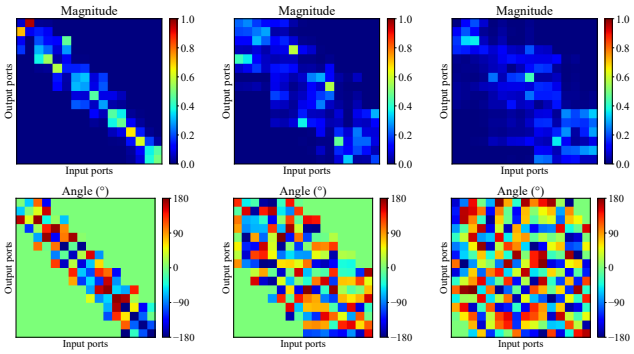
The third sub-module is an unequal RF coupler, which exhibits symmetry, as shown in the right part of Fig. 5. Its propagation matrix is Eq. 5. By integrating these three sub-modules, we create a novel RF coupler with the following propagation matrix: $\boldsymbol{p}(\beta, n, \theta_1, \theta_2, a_c, \theta_{c1}, \theta_{c2}) = \boldsymbol{p_{sc} p_{pd} p_{cp}}$, and the detailed parameters are as:

$$\begin{cases} S_{21} = (\beta n + 1 - \beta) n_c e^{j(\theta_1 + \theta_{c1})} + (1-\beta) n (1-n_c) e^{j(\theta_2 + \theta_{c2})} \\ S_{31} = (\beta n + 1 - \beta)(1-n_c) e^{j(\theta_1 + \theta_{c2})} + (1-\beta) n n_c e^{j(\theta_2 + \theta_{c1})} \\ S_{24} = (1-\beta)(1-n) n_c e^{j(\theta_1 + \theta_{c1})} + \\ \qquad [\beta + (1-\beta)(1-n)](1-n_c) e^{j(\theta_2 + \theta_{c2})} \\ S_{34} = (1-\beta)(1-n)(1-n_c) e^{j(\theta_1 + \theta_{c2})} + \\ \qquad [\beta + (1-\beta)(1-n)] n_c e^{j(\theta_2 + \theta_{c1})} \end{cases} \quad (8)$$



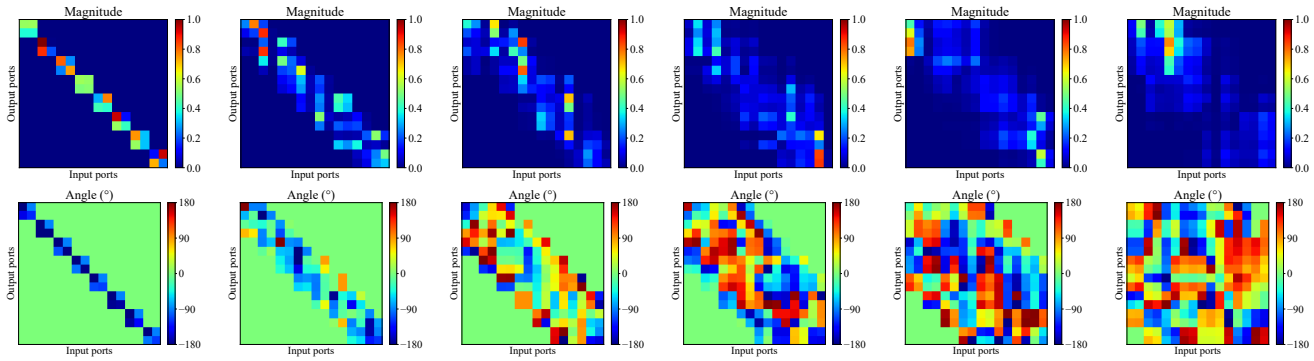
(a) M=8, N=3 (b) M=8, N=7 (c) M=8, N=13

Figure 6: Propagation matrices of the RF coupler network under various setting parameters. Each RF coupler unit is configured as an equal power splitter [18]. The propagation weights are fixed with the specific setting parameters.



(a) M=8, N=3 (b) M=8, N=7 (c) M=8, N=13

Figure 7: Propagation matrices of the RF coupler network under various setting parameters. Each RF coupler unit is configured as an unequal power coupler [44]. The propagation weights are not fixed and have limited adjustability.



(a) M=8, N=1 (b) M=8, N=3 (c) M=8, N=5 (d) M=8, N=7 (e) M=8, N=9 (f) M=8, N=13

Figure 8: Propagation matrix of the RF coupler network illustrating various setting parameters; each RF coupler unit is configured as our proposed unequal asymmetric power coupler (see Fig. 5). The propagation weights are adjustable.

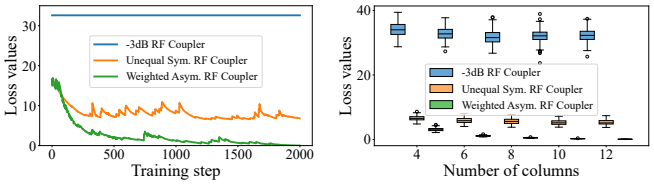
This approach enables independent adjustment of the four complex numbers in the RF coupler matrix and ensures that the proposed coupler unit exhibits near-zero reflection.

3.1.3 Propagation Performance Comparison. We consider a diffractive NN with 16 input ports, and its hidden layer also has 16 hidden RF ports, requiring approximately $8 \times N$ RF couplers to establish connections between layers. We first apply the -3dB RF couplers in the propagation module, each with transmission parameters $S_{21} = S_{34} = S_{31} = S_{24} = 0.5e^{j\frac{\pi}{2}}$, which ensure uniform energy distribution and fixed phase relations. We analyze the propagation matrices for $N = 1$ to 13, as shown in Fig. 6. The resulting weight matrices display consistent diffusion patterns as N increases. However, the fixed amplitude and phase characteristics of -3dB couplers limit the tunability of the propagation weights, thereby inhibiting their optimization via backpropagation in pla-NN.

Then, unequal RF couplers are employed in the propagation module, with each coupler assigned arbitrary values for a_{cp}, θ_{c1}, and θ_{c2}, thus enabling independent control over energy and phase. We compute the propagation matrices for $N = 1$ to 13, as presented in Fig. 7. The results show that as N increases, there are significant variations in both amplitude and phase within the diffusion region.

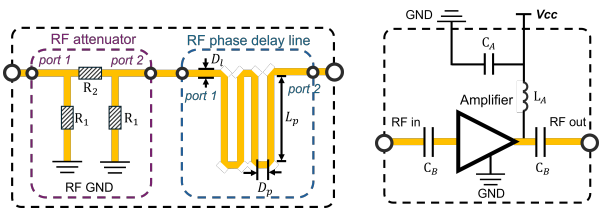
Furthermore, we examine the propagation weights of a coupler network built using our designed weighted asymmetric couplers, where each coupler is defined according to Eq. 8, allowing for arbitrary energy and phase settings. For N ranging from 1 to 13, the corresponding propagation matrices are shown in Fig. 8. With increasing N, both amplitude and phase distributions of the propagation matrix exhibit greater irregularity and randomness, preliminarily confirming the enhanced tunability and effectiveness of our design.

To further demonstrate the adjustability of our weighted asymmetric RF coupler, we construct propagation modules using various RF coupler types to fit arbitrary target weight matrices $\boldsymbol{W}$. Specifically, we randomly generate 1000 instances of 16×16 matrices $\boldsymbol{W}$ and employ different RF coupler networks for fitting, with the number of input and output ports $M = 8$, and N as a tunable hyperparameter. Fig. 9(a) gives



(a) Fitting loss curves, N=5 (b) Fitting error bars varying N

Figure 9: Comparison of the adaptability of propagation weights in networks formed by different RF couplers.



(a) Bias module (b) Activation module

Figure 10: Bias and activation unit design: each module's left is input, right is output.

an example of the loss curve for the fitting of one instance of W, with N set to 5. Fig. 9(b) shows the fitting error bar analysis. Although Fig. 7 demonstrates that networks built with unequal symmetric RF couplers exhibit increased diffusivity in the propagation matrix as N increases, both equal and unequal symmetric coupler networks fail to converge effectively. Notably, simply increasing N does not enhance the RF coupler network's ability to fit an arbitrary target matrix $\boldsymbol{W}$ for either configuration. Fundamentally, symmetric RF couplers present intrinsic limitations: from an energy exchange perspective, the energy from the $2M$ input ports of the couplers fully propagates to the input ports of the next column of couplers. Thus, the total energy of the previous column is transferred to the total energy of the next column, meaning each column sum of matrix W is the same. After multiplication by the symmetric block propagation matrix $\mathbf{P}_k$, the resulting matrix W maintains symmetry, leading to equal row sums in the final matrix W. This structural property constrains the flexibility of weight adjustment, making it impossible to realize arbitrary propagation matrices, regardless of whether the symmetric RF couplers are equal or unequal.

In contrast, the propagation modules constructed with our weighted asymmetric RF couplers demonstrate superior coverage and fitting performance. The findings indicate that the integration of power splitters, combiners, and phase shifters into otherwise symmetric but unequal RF couplers effectively breaks their inherent symmetry in both power distribution and phase alignment. This structural modification enables highly tunable propagation weights and facilitates efficient inter-layer connectivity within the diffractive NN.

3.2 Bias Module

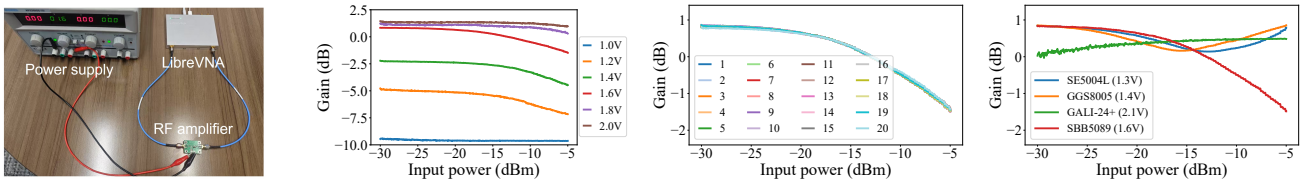
We implement a bias module via RF circuits that allow arbitrary weight adjustment. The current SOTA, pla-NN, achieves **b** solely through phase delay lines (see Fig. 1(a)), which presents significant limitations. Since the **b** vector is a complex vector, its elements comprise not only phase delay but also attenuation. Therefore, we incorporate an RF attenuator [23] with phase delay functionality to achieve the complex-valued bias, $A_b e^{j\theta_b}$, for each RF neuron. The detailed design is shown in Fig. 10(a). It is important to note that the RF bias modules are restricted to attenuation, meaning $|b| \leq 1$.

3.3 Nonlinear Activation Module

To enhance RF data processing, we need an activation function for $a = \tau(y)$. Given that y is complex, the function must introduce nonlinearity in amplitude and/or phase. In RF circuits, passive components are linear within their operating range, maintaining constant loss regardless of input power, while active components exhibit nonlinear characteristics [8]. Amplitude nonlinearity is relatively easy to achieve using an RF amplifier, as amplifiers are active devices with inherent nonlinear properties. Thus, we utilize an RF amplifier as the activation module in the WDNN, as detailed in Fig. 10(b).

Leveraging the nonlinear properties of RF amplifiers as activation functions in WDNN requires careful design, as nonlinearity mainly arises in the saturation and cutoff regions. Firstly, we focus on the saturation region. At low input power, the amplifier operates linearly with constant gain, but as input power increases, gain compression occurs and output power saturates, resulting in significant nonlinearity. However, this nonlinearity cannot be directly exploited since typical RF signals have low input power (*e.g.*, -30 to -20 dBm), insufficient to drive the amplifier into saturation.

When an RF amplifier enters the cutoff region, the collector current is nearly zero, causing it to be unable to output voltage or amplify signals. One way to force the amplifier into the cutoff region is to provide a supply voltage significantly lower than its normal operating voltage. We assume WDNN operates at 5 GHz and use a LibreVNA [48] to measure the RF amplifier's characteristics at this frequency. Fig. 11(a) shows the experimental setup. We adjust the RF power at one port, and measure the amplifier's gain through the other port. Fig. 11(b) shows the amplification characteristics of the SBB5089z RF amplifier [37] under different supply voltages. When the voltage is in the 1~1.4V range, the RF amplifier is indeed in the cutoff region, with a gain of around -10~-2.5dB. Although the curve shows slight nonlinearity, there is significant signal attenuation. At 1.6V, the amplification curve is around 0 dB and exhibits strong nonlinearity, similar to the ReLU and ELU activation functions [13]. As the voltage increases further, the amplifier moves out of the cutoff region, and the amplification curve becomes smoother with less noticeable nonlinearity. We also test multiple SBB5089Z RF amplifiers at a supply voltage of 1.6V to evaluate the



(a) Setup for testing amplifier performance at different voltages (b) Different supply voltages cause SBB5089z abnormal operation (c) Stability across RF amplifiers of the same SBB5089z model (d) Different RF amplifiers and their abnormal operation characteristic

Figure 11: Adjusting the supply voltage of the RF amplifier to operate in an nonlinear region for activation function.

stability of their nonlinear activation capability. As shown in Fig. 11(c), the stability of this chip is satisfactory.

While many other RF amplifiers can meet the design requirements of the WDNN's nonlinear activation module, we test several amplifier chips, such as SE5004L [40], GGS8005 [14], and GALI-24+ [32]. As shown in Fig. 11(d), the results reveal that the SBB5089Z exhibites the best nonlinear performance under conditions of approximately 0dB gain and operation in the cutoff region. Therefore, in this study, we select the SBB5089Z RF amplifier and set the supply voltage to 1.6V to implement the nonlinear activation module for a WDNN operating at 5 GHz.

3.4 Inverse Design for WDNN Implementation

We obtain the weights for each RF coupler using the gradient descent method (see Fig. 5): $\beta, n, \theta_1, \theta_2, a_c, \theta_{c1}, \theta_{c2}$, the same applies to the RF bias module: A_b and θ_b. Next, we select suitable substrate materials and thicknesses to facilitate the design of the geometric parameters and component values for each RF module. In the unequal power splitter and combiner, when β approaches 1, input port 1 performs power splitting and combines with input port 2; when β approaches 0, the power splitting is performed by input port 2. Based on the power splitter ratio n, we can then determine the specific parameters: D_{s1}, D_{s2}, D_{m1}, D_{m2}, D_{c1}, D_{c2}, D_{m3}, D_{m4}, and the resistance values of the resistors: R_s, and R_c. The specific formulas can be referenced in [19]. For the phase delays θ_1 and θ_2, we calculate the required line length based on transmission line theory [41]. For the RF unequal coupler sub-module, we apply the optimized n_c, θ_{c1}, and θ_{c2} to calculate the same even- and odd-mode characteristic impedances (z_{e1}, z_{e2}, z_{o1}, and z_{o2}) with the closed-form design formulas (equations 8-11 in [44]). Then we use the coupled microstrip transmission line theory [17] to determine the geometric parameters inversely: $L_1, L_2, D_1, D_2, W_1, W_2, S_1, S_2$ with these characteristic impedances. Finally, we use HFSS [16] to fine-tune these geometric parameters to address errors caused by parasitic capacitance and coupling between lines. For the bias modules, to implement the $A_b e^{j\theta_b}$, we use the Pi attenuator calculator [34] to calculate the resistance values R_1, R_2, and R_3, and we also use transmission line theory [41] to calculate the required phase delay length.

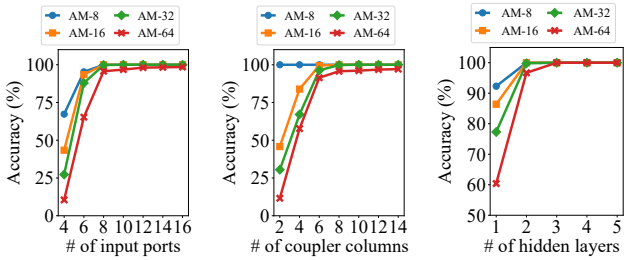
4 Evaluation

To assess the capability of WDNN in RF signal processing at the physical layer, we design three representative tasks: (1) amplitude-modulated (AM) wireless communication decoding; (2) phase-based angle-of-arrival (AoA) estimation; and (3) human fall detection using WiFi channel state information (CSI). The first two tasks use a signal frequency of 5 GHz. In the third task, we use the 2.4 GHz WiFi and select the CSI of one sub-carrier to perform the sensing task. Given that the diffractive NN model is a supervised learning-based method, we provide a detailed description of the training procedures for the WDNN and the SOTA. Furthermore, comprehensive comparative experiments are conducted to evaluate their respective performances.

4.1 Training Dataset Preparation

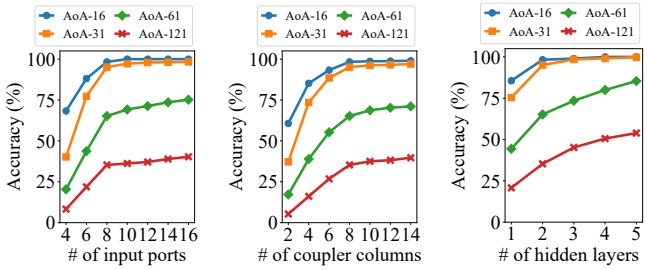
AM Decoding. Consider that an AM signal, $s(t) = [A_m b(t) + A_c]cos(2\pi f_c t + \theta_c)$, where A_c is the carrier amplitude, f_c is the carrier frequency, θ_c is the initial phase, and A_m is the amplitude of the modulating signal. AM decoding systems typically include an antenna and an automatic gain control (AGC) module to receive RF signals. Since the AGC compensates for A_c, we only need to generate normalized amplitude information for the training dataset. With only a single receiving (RX) antenna, phase information can be ignored, allowing the network's amplitude decoding to remain unaffected by θ_c. We configure the system to use 32-level AM, meaning each level encodes 5 bits. Therefore, we generate an amplitude-value dataset and corresponding labels. Additionally, we incorporate the simulated noise with the RF noise modeling tool [20] to enhance the complexity of the dataset.

AoA Estimation. We use $2M$ separate antennas to collect the RF signal for AoA estimation, and we set both input and output channels to $2M$. These data are modeled using a Uniform Linear Array (ULA) with an antenna spacing of half a wavelength. We set the incident angle range between -60°and 60°. For each incident angle θ_{inc}, we compute the array steering vector $a(\theta_{inc})$ as: $[e^{-j0}, e^{-j\lambda 2\pi d sin(\theta_{inc})}, e^{-j2\lambda 2\pi d sin(\theta_{inc})}, ..., e^{-j(P-1)\lambda 2\pi d sin(\theta_{inc})}]$, where P indicates the antenna number, λ is the wavelength, and d is the antenna spacing (here is 3 cm). We assume the noise vector to be n and the signal



(a) Number of the input/output ports ($2M$) (b) Number of columns in coupler network (N) (c) Number of the hidden layers (H)

Figure 12: Micro-benchmark of the AM decoding tasks with different modulation levels.



(a) Number of the input/output ports ($2M$) (b) Number of columns in coupler network (N) (c) Number of the hidden layers (H)

Figure 13: Micro-benchmark of the AoA estimation task with different categorizations.

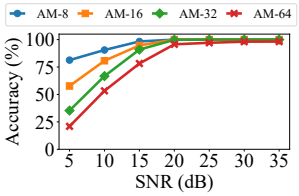


Figure 14: AM decoding classification performance under different noise levels.

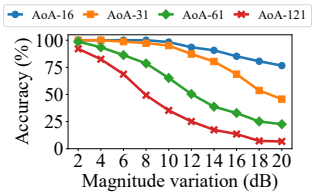


Figure 15: AoA estimation performance under different incident signal intensity.

source to be s; thus, the received signal is given by $y(\theta_{inc}) = a(\theta_{inc})s + n$. Note that both s and n are complex numbers. For each incident angle θ_{inc}, we define its classification label as $(\theta_{inc} + 60°)//res$, and res is the defined resolution. This results in a total of 31 classes.

Wi-Fi CSI Based Fall Detection. We directly utilize the fall detection dataset provided in [2] as our training dataset. This dataset is collected using a TP-LINK AC1750 access point and an Intel 5300 network card, with the experimental setup comprising three TX antennas and three RX antennas. Given the unique characteristics of diffractive NNs, which currently can only process instantaneous sequential signals, we select the signals corresponding to periods when the user is in a stable posture in the dataset. Specifically, we select one sub-carrier in the 20 MHz band and extract the initial and final frames of each activity sequence, representing upright (standing) and fallen postures, respectively. We partition the dataset based on the three TX antennas, treating the signals received by each of the three RX antennas from each TX antenna as an independent sample.

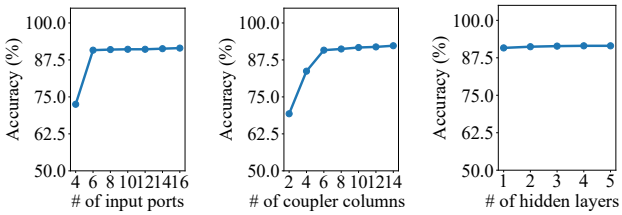
4.2 WDNN Model Training

We configure the input and output of WDNN with $2M$ ports. In addition, the parameter N (the number of columns in the RF coupler network) in the propagation modules serves as a hyperparameter that affects propagation performance. The number of hidden layers, H, is another key hyperparameter for network performance. Overall, the WDNN model contains $H+1$ propagation modules. We construct a mathematical model of the WDNN based on the hyperparameters: M, N and H. We use the Adam solver [21] to obtain the optimal weights and the NNI [31] hyperparameter tuning tool to achieve optimal settings for the WDNN.

AM Decoding: The amplitude categories are assigned to the $2M$ output ports, following an output scheme similar to [15], where the signal energy at each output port is measured by RF power detectors and used as features. For example, with $M = 4$ and a 32-class classification task, the 8 output ports each represent category groups 0–3, 4–7, ..., 28–31; the signal energy at each port is divided into 4 levels for classification within each group. Considering that the output of WDNN is analog and exhibits continuous classification characteristics (similar to a regression task), the overall network can also achieve 8-, 16-, 32-, and 64-class classification based on the voltage classification accuracy at each output port.

To this end, we perform hyperparameter optimization experiments for WDNN with the NNI tool [31] in four classification tasks under a simulated SNR of 20 dB. To determine the values of the three hyperparameters, we conduct three sets of experiments. In each experiment, two hyperparameters are fixed at their optimal values while the third is varied, and WDNN is retrained for each variation. As illustrated in Fig. 12, for the 32-level amplitude classification task, the optimal values for M, N, and H are 4, 8, and 2, respectively. Furthermore, as shown in Fig. 14, the system demonstrates robust noise immunity when employing the optimal hyperparameter settings and trained WDNN weights under various simulated SNR conditions.

AoA Estimation: This WDNN is to classify input signals from $2M$ RX antennas to determine the AoA. We take a 31-class AoA classification task (4° resolution) as an example, the output categories are distributed across 8 ports (*i.e.*, M=4), with each output corresponding to a category group,*e.g.*, 0-2,



(a) Number of the input/output ports ($2M$) (b) Number of columns in coupler network (N) (c) Number of the hidden layers (H)

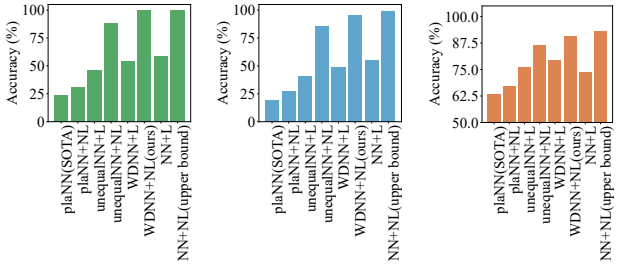
Figure 16: Micro-benchmark of the Wi-Fi CSI-based gesture recoginition task.

3-6, 7-10, ..., 27-30. The signal energy level of each output is used to classify within each sub-category group, and the overall model can also achieve 16-, 31-, 61-, and 121-class classification based on the voltage classification accuracy at each output port. We also employ NNI for hyperparameter tuning and conduct experiments, including training and testing AoA classification at varying resolution requirements, to confirm the optimal hyperparameters. The results are shown in Fig. 13, resulting in the values of M, N, and H as 4, 8, 2. Additionally, we test WDNN's AoA classification effectiveness across various incident signal intensities , as illustrated in Fig. 15. The results demonstrate that WDNN efficiently isolates amplitude effects, enabling AoA prediction based solely on phase differences between antennas.

Wi-Fi CSI Based Fall Detection. The proposed WDNN model is designed to detect user fall events by leveraging WiFi CSI collected from three RX antennas, thereby distinguishing between standing and lying states. In this setup, both the input and output ports are configured to be $2M$. Signals from the three independent incident antennas are distributed to three of the input ports, while the remaining input ports are terminated with a 50 Ω load. For the output channels, we designate port 1 and the last port as the output label channels for binary classification. Hyperparameter optimization is performed using NNI and results are illustrated in Fig. 16. The optimal hyperparameter values are determined to be $M = 3$, $N = 6$, and $H = 2$. These experimental results demonstrate that our WDNN is indeed capable of directly determining whether a user in the environment is standing or lying down, solely based on the CSI obtained from three RX antennas.

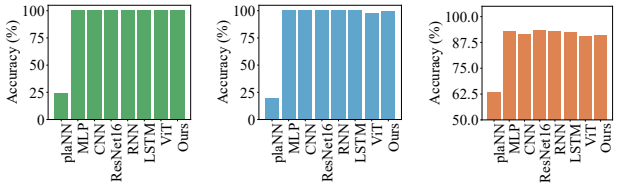
4.3 Ablation Study

To validate the effectiveness of the propagation module composed of our proposed weighted asymmetric RF couplers, and the nonlinear activation module implemented using RF amplifiers, we conduct an ablation study. The results are shown in Fig. 17, we detail the x-axis categories as follows. "plaNN" is the current SOTA with an equal RF coupler network, but no activation. "plaNN+NL" adds nonlinear activation. "unequalNN+L" uses unequal symmetric RF couplers with linear



(a) AM decoding (b) AoA estimation (c) Fall detection

Figure 17: Ablation study results showing performance variation with the proposed propagation modules and nonlinear activation modules.



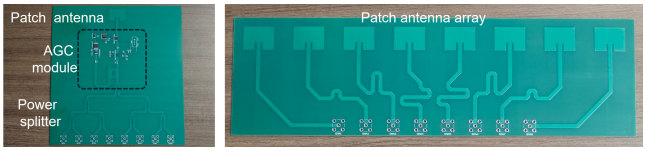
(a) AM decoding (b) AoA estimation (c) Fall detection

Figure 18: Performance comparison with the baseline diffractive NNs and digital deep learning methods.

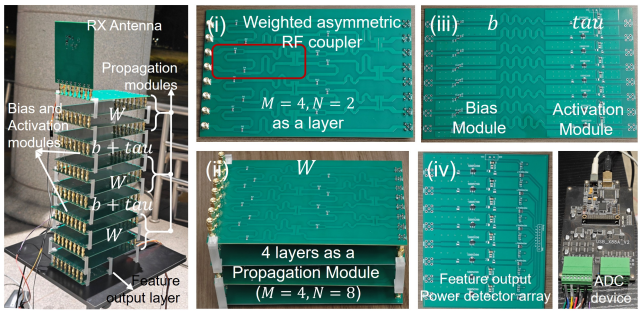
activation, while "unequalNN+NL" uses our nonlinear activation modules. "WDNN+L" features our weighted asymmetric RF coupler with linear activation, and "WDNN+NL" includes nonlinear activation, completing our design. "NN+L" is a digital NN with linear activation, and "NN+NL" includes nonlinear activation, serving as the NN's upper bound. Notably, "plaNN", "unequalNN", and "WDNN" represent models optimized through hyperparameter tuning to achieve their best classification performance. Additionally, "NN" shares the same network parameters as "WDNN", such as two hidden layers. The experimental results demonstrate that our proposed propagation module and nonlinear activation significantly enhance the performance of diffractive NNs, closely approaching the performance of digital NNs with similar settings. These results validate that our method enables NN computations in the analog domain to achieve performance comparable to those in the digital domain.

4.4 Comparison with Digital SOTA Models

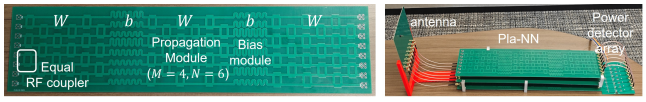
We systematically compare the performance of our physical-layer-based WDNN approach with baseline fully digital machine learning methods on the three RF data processing tasks. Specifically, we select MLP, 1D CNN, ResNet, RNN, LSTM, and vision transformer (ViT), as baseline digital methods, and utilize the dataset described in Sec. 4.1 for training. In all experiments, the models process data in the time domain. Considering that ResNet and ViT require two-dimensional input, the original 1×2M antenna data is reshaped into 2×M. Experimental results, shown in Fig. 18, demonstrate that WDNN



(a) Left: AM communication RX antenna; right: AoA estimation RX antennas



(b) Details of WDNN prototypes. Our optimized two WDNN have the same hyperparameters: $M = 4$, $N = 8$, with 3 propagation modules, 2 bias and activation modules, both having sizes of $12cm \times 22cm \times 42cm$



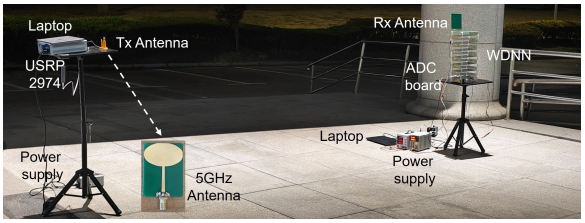
(c) Details of the pla-NN (SOTA) prototypes. The optimized pla-NNs for the two tasks share the same hyperparameters: $M = 4$, $N = 6$, with 8 propagation modules and 7 bias modules, both having sizes of $13cm \times 52cm \times 6cm$

Figure 19: Details of our proposed WDNN prototypes and the existing SOTA (pla-NN [15]) prototypes.

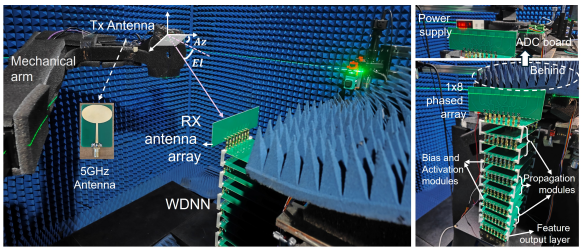
achieves a classification accuracy comparable to that of digital MLP and other digital SOTA methods, and even outperforms ViT on certain tasks. This can be attributed to the fact that ViT is less effective when handling input data of relatively small size. These experimental results indicate that our WDNN is well-suited for processing RF data and can achieve performance comparable to digital machine learning algorithms on certain tasks.

4.5 Real-world Experiments

4.5.1 Prototype Implementation and Experiment Setup. Following the implementation outlined in Sec. 3.4, we prototype the WDNNs with the trained weights using PCB technology for two tasks: AM decoding and AoA estimation. The details are as shown in Fig. 19. We use an 8-channel RF power detector as the output layer, paired with an 8-channel ADC (CS4382A [29]) device to retrieve feature data or processing results, as shown in Fig. 19(c)-iv. This output module is shared with two prototyped WDNNs. To facilitate performance comparison with the SOTA (pla-NN), we also conduct micro-benchmark experiments using a hyperparameter tuner to determine the optimal hyperparameters for the pla-NN models. For both AM decoding and AoA estimation tasks,

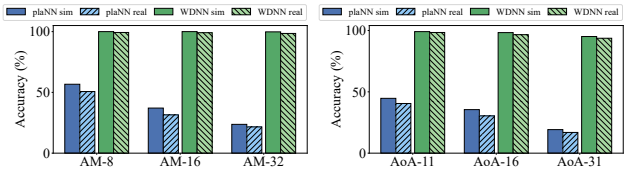


(a) AM communication decoding



(b) Azimuth AoA estimation

Figure 20: Experimental setups of two testing tasks.



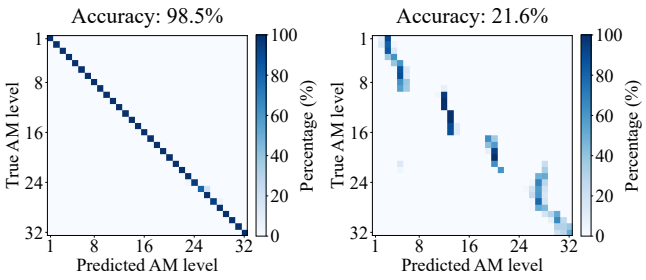
(a) AM decoding (b) AoA estimation

Figure 21: Real-world results for two tasks. AM-8, AM-16, and AM-32 denote using the prototype for 8, 16, and 32-level classification, similar to AoA.

the optimal settings are as follows: $M = 4$, $N = 6$, with 8 propagation modules and 7 hidden layers. The best-performing hyperparameters and their corresponding trained weights are then used to fabricate the pla-NN prototype.

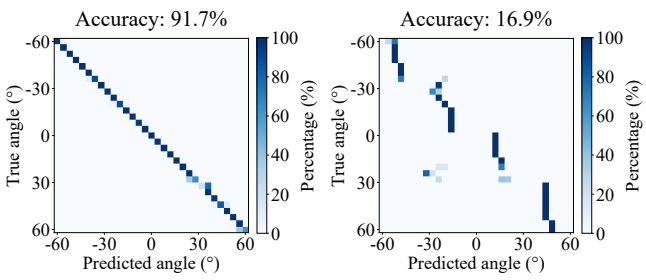
Fig. 20 illustrates our experimental setups. For the AM communication, we use a USRP2974 [38] to generate 32-level AM-modulated signals at a 48 Kbps rate, with a 5GHz patch antenna as the transmitter; the WDNN serves as the receiver for signal reception and amplitude classification. For the AoA estimation, we also use the USRP2974 and a 5GHz antenna to transmit continuous waves. A laser-calibrated robotic arm precisely controls the incidence angles. The WDNN, as the receiver, facilitates signal reception and AoA estimation.

4.5.2 AM decoding performance. We conduct real-world tests on the 32-level amplitude classification task. As shown in Fig. 21(a), our system achieves a classification accuracy of 98.5%, closely matching the simulated accuracy of 99.8%. In contrast, pla-NN only reaches accuracies of 21.6% in real-world tests and 23.7% in simulations. This demonstrates that our system significantly outperforms the SOTA in data processing capabilities at the physical layer. The classification



(a) WDNN (Ours) (b) pla-NN (SOTA)

Figure 22: Confusion matrices on the 32-level amplitude classification task in the real-world experiment.



(a) WDNN (Ours) (b) pla-NN (SOTA)

Figure 23: Confusion matrices on the 31-class AoA estimation task in the real-world experiment.

confusion matrices for both systems are shown in Fig. 22, indicating that pla-NN struggles with accurate decoding, whereas our WDNN can nearly perfectly decode the original data, achieving stable AM communication with simple redundancy code design.

4.5.3 AoA estimation performance. We test the performance for 31-class AoA estimation in an anechoic chamber. As shown in Fig. 21, our system achieves 93.7% classification accuracy, closely matching the simulation accuracy of 95.1%. In contrast, pla-NN achieves only 16.9% accuracy in real-world and 19.2% in simulations. The confusion matrices in Fig. 23 show that pla-NN struggles with AoA estimation, having up to 4-type classification ability. Our WDNN effectively achieves 4° resolution in AoA estimation within the range of -60° to 60°. This task demonstrates WDNN's powerful capability to process multi-channel phase information and perform multi-class classification.

4.5.4 Cost of WDNN. We measure the power consumption of the system prototype, including all propagation, bias, and activation modules in Fig. 19(c), totaling approximately 51 mW. The primary power-consuming component is the RF amplifiers, which operate in the nonlinear region with a gain of about -3 to 0 dB, resulting in minimal power usage. This is why our WDNN consumes very little power during analog domain computation. The cost of our WDNN prototype is around 34.5 dollars, including PCB boards at ~ $20, electronic components at $1, RF amplifiers at $4.5, and RF connectors at $9. Although the prototype is not cheap, we believe that employing analog IC technologies for WDNN can reduce manufacturing costs.

5 Related Work

5.1 Diffractive neural networks

Diffractive neural networks have become a promising approach for NN-based signal processing at the physical layer of wireless systems A primary implementation involves metasurface-based diffractive NNs, where each metasurface layer serves as a hidden layer in the NNs. For instance, [27, 28, 46] employ diffraction for parallel analog computing, with weights learned offline. Although tunable metasurfaces allow limited real-time adjustments, diffraction inherently constrains weight adaptability, as propagation characteristics are dictated by fixed structural parameters. Furthermore, these designs lack nonlinear activation, as diffraction-based transformations are fundamentally linear [47, 50], limiting their ability to perform complex inference tasks. Additionally, metasurface-based implementations require precise calibration and occupy significant physical space, further restricting their scalability.

Analog circuit-based diffractive NNs reduce spatial constraints and calibration complexity by enabling direct signal transmission through electrical pathways. Planar diffractive neural networks (pla-NNs) [15] have demonstrated feasibility in tasks such as AoA estimation at 10 GHz, with terahertz implementations [12] extending their applicability to 265 GHz. However, these designs rely on fixed propagation weights with limited adaptability and only phase delay mechanisms, lacking the nonlinear activation needed for advanced learning tasks. Therefore, they remain constrained to linear operations.

5.2 Feature enhancement in analog domain

5.2.1 Radio-frequency signals. Metasurfaces have been widely explored for enhancing RF signal processing at the physical layer, particularly in millimeter-wave (mmWave) applications. [35] optimized mmWave metasurfaces to enable small phased-array antennas to achieve dynamic steering and focusing capabilities comparable to large arrays. Similarly, [30] extended the coverage of 60 GHz WiFi through optimized metasurface designs, while [42] improved indoor GNSS signal reception using similar enhancements. Beyond passive designs, active metasurfaces have been introduced to dynamically adapt to environmental variations, offering greater flexibility in RF signal control. [9, 24, 25] developed reconfigurable metasurfaces capable of adjusting their configurations in real time, enhancing feature extraction and signal amplification in the analog domain and facilitating more efficient RF signal processing.

5.2.2 Acoustic signals. 3D-printed passive acoustic metasurfaces have been widely used for efficient feature extraction and signal enhancement at the physical layer. [49] optimized metasurfaces to convert digital beamforming into an analog process. Expanding on this capability, [11] designed a metasurface to enhance incident angle feature extraction, improving angle sensing accuracy. Similarly, [10] refined channel features to enhance acoustic imaging, achieving high-resolution imaging with minimal hardware. Complementing these advancements, [1] developed a passive filter for spatial encoding, improving spatial perception through multipath encoding and robotic-assisted imaging.

5.2.3 Light signals. [4, 22, 36] integrated optical masks with low-resolution CMOS sensors to develop the minimalist visual system for performing visual tasks. The optical mask functions as the feature extraction layer of a neural network, with subsequent layers handling inference. [22] further optimized pixel shapes through task-driven training, enhancing its adaptability. Despite relying on only a few pixels, the system achieves performance comparable to high-resolution cameras in tasks such as indoor monitoring, illumination measurement, and traffic estimation, highlighting the effectiveness of optical-domain feature extraction.

Our work: While these approaches introduce external physical-layer hardware to enhance signal processing, our work directly integrates adaptive feature extraction and robust signal processing within the physical layer, enabling more efficient and scalable analog-domain computation.

6 Discussion

Supporting Higher Frequency. By scaling the size of components such as couplers and phase shifters, and selecting suitable amplifiers, we can adapt WDNN to any desired band. [12] demonstrates a diffractive NN using plasmonic circuits at 265 GHz, confirming the feasibility of analog circuit-based networks at higher frequencies.

Supporting Wideband Wireless System. WDNN can be extended to support wideband applications, such as Frequency-Modulated Continuous Wave (FMCW) radar systems. This capability relies on the wideband design of RF components, including RF couplers, phase shifters, attenuators, and amplifiers. The frequency response characteristics of these RF components are integrated into the wideband WDNN weight optimization process, thereby ensuring that the constructed WDNN efficiently supports wideband signal processing. Existing works [3, 45] have proposed coupler designs suitable for multiple frequency bands, while [39] introduces RF couplers that can adapt to various frequency offsets. Therefore, by employing advanced RF circuit design and wideband weight optimization techniques, WDNN enables flexible allocation of network weights across different frequencies, leading to improved wideband adaptability.

Impact of the temperature. The gain parameter curve of amplifiers used in nonlinear RF modules may fluctuate with large changes in ambient temperature. To enable our WDNN to automatically compensate for the impact of outdoor temperature variations, we can extract the amplifier's gain parameters at different temperatures from its datasheets. Furthermore, by incorporating a temperature sensor, the supply voltage of the amplifier chips can be dynamically adjusted to ensure they consistently operate within the nonlinear region.

Reusability of WDNN. The activation modules in the WDNN are reusable, meaning that they can be used in different WDNN models for various tasks. However, the existing weight modules like feature extraction and classification layers are not reusable, as the parameters need to be retrained for each new task. If a general-purpose NN can be developed as a foundational model for RF feature extraction, it would facilitate the design of a universal extraction module based on WDNN. Such a module would exhibit significant reusability and modularity, enabling it to serve as a standardized add-on component that could be commercialized and seamlessly integrated into a wide variety of RF hardware platforms.

Analog RF IC Design. Analog integrated circuit (IC) technology is highly mature and widely utilized in modern mobile devices. For instance, RF front-end components such as low-noise amplifiers, power amplifiers, and mixers are typical analog chips that are integrated within RF modules and communication system-on-chips (SoCs). The implementation of WDNN can also be realized by the analog IC technology, as it allows a substantial reduction in hardware footprint and supports efficient deployment of advanced neural networks directly at the physical layer.

7 Conclusion

In this paper, we introduce WDNN, a weighted diffractive neural network designed for direct RF signal processing at the physical layer. Our designed RF coupler unit implements arbitrary asymmetric propagation weights, creating a diffractive propagation module with adjustable connections. We also integrate RF amplifiers that reliably operate in the nonlinear region for activation functionality. These features enable our WDNN to efficiently and robustly process RF signals, surpassing the performance of current SOTA diffractive NNs.

Acknowledgments

We thank the anonymous reviewers for their constructive comments. This work is supported in part by NSFC (No. 62432004 and 62072306) and by a grant from the Guoqiang Institute, Tsinghua University.

References

[1] Yang Bai, Nakul Garg, and Nirupam Roy. 2022. Spidr: Ultra-low-power acoustic spatial sensing for micro-robot navigation. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. 99–113.

[2] J Klein Brinke. 2019. Channel state information (wifi traces) for 6 activities.

[3] Yuan Cao, Xiang Chao Jin, and Peng He Xu. 2025. A highly flexible dual-band directional coupler with arbitrary phase difference and power division ratio. *AEU-International Journal of Electronics and Communications* (2025), 155719.

[4] Yitong Chen, Maimaiti Nazhamaiti, Han Xu, Yao Meng, Tiankuang Zhou, Guangpu Li, Jingtao Fan, Qi Wei, Jiamin Wu, Fei Qiao, et al. 2023. All-analog photoelectronic chip for high-speed vision tasks. *Nature* 623, 7985 (2023), 48–57.

[5] F Depasse, MA Paesler, D Courjon, and JM Vigoureux. 1995. Huygens–Fresnel principle in the near field. *Optics letters* 20, 3 (1995), 234–236.

[6] Altium Designer. 2025. What Are PCB Vias and How They Affect High Speed Signals. https://resources.altium.com/p/how-vias-can-affect-very-high-speed-signals.

[7] DigiKey. 2025. The Fundamentals of RF Power Dividers and Combiners. https://www.digikey.com/en/articles/the-fundamentals-of-rf-power-dividers-and-combiners.

[8] EverythingRF. 2025. Non-Linear Devices. https://www.microwaves101.com/encyclopedias/non-linear-devices.

[9] Chao Feng, Xinyi Li, Yangfan Zhang, Xiaojing Wang, Liqiong Chang, Fuwei Wang, Xinyu Zhang, and Xiaojiang Chen. 2021. RFlens: metasurface-enabled beamforming for IoT communication and sensing. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 587–600.

[10] Yongjian Fu, Yongzhao Zhang, Yu Lu, Lili Qiu, Yi-Chao Chen, Yezhou Wang, Mei Wang, Yijie Li, Ju Ren, and Yaoxue Zhang. 2024. Adaptive metasurface-based acoustic imaging using joint optimization. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services*. 492–504.

[11] Yongjian Fu, Yongzhao Zhang, Hao Pan, Yu Lu, Xinyi Li, Lili Chen, Ju Ren, Xiong Li, Xiaosong Zhang, and Yaoxue Zhang. 2024. Pushing the limits of acoustic spatial perception via incident angle encoding. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–28.

[12] Xinxin Gao, Ze Gu, Qian Ma, Bao Jie Chen, Kam-Man Shum, Wen Yi Cui, Jian Wei You, Tie Jun Cui, and Chi Hou Chan. 2024. Terahertz spoof plasmonic neural network for diffractive information recognition and processing. *Nature Communications* 15, 1 (2024), 6686.

[13] ML Glossary. 2025. Activation Functions. https://ml-cheatsheet.readthedocs.io/en/latest/activation_functions.html.

[14] GPowerTek. 2025. GGS8005: DC to 6000MHz HBT UltraWideband Amplifier. https://item.szlcsc.com/datasheet/GGS8005/43218042.html.

[15] Ze Gu, Qian Ma, Xinxin Gao, Jian Wei You, and Tie Jun Cui. 2024. Direct electromagnetic information processing with planar diffractive neural network. *Science Advances* 10, 29 (2024), eado3937.

[16] Ansys HFSS. 2025. Best-In-Class 3D High Frequency Structure Simulation Software. https://www.ansys.com/products/electronics/ansys-hfss.

[17] The MathWorks Inc. 2025. coupledMicrostripLine. https://www.mathworks.com/help/rfpcb/ref/coupledmicrostripline.html.

[18] The MathWorks Inc. 2025. Introduction to 4-Port Couplers. https://www.mathworks.com/help/rfpcb/ug/introduction-to-4-port-couplers.html.

[19] The MathWorks Inc. 2025. Introduction to Equal and Unequal Split Wilkinson Power Splitter. https://ww2.mathworks.cn/help/rfpcb/ug/design-and-analyze-equal-and-unequal-split-wilkinson-power-splitter.html.

[20] The MathWorks Inc. 2025. RF Noise Modeling. https://www.mathworks.com/help/simrf/ug/rf-noise-modeling.html.

[21] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[22] Jeremy Klotz and Shree K Nayar. 2024. Minimalist vision with freeform pixels. In *European Conference on Computer Vision*. Springer, 329–346.

[23] GPID Lab. 2025. RF Pi Attenuator Kit Design and Build your own RF Attenuator. https://gpio.com/products/rf-pi-attenuator-kit-design-and-build-your-own-rf-attenuator.

[24] Xinyi Li, Chao Feng, Xiaojing Wang, Yangfan Zhang, Yaxiong Xie, and Xiaojiang Chen. 2023. {RF-Bouncer}: A Programmable Dual-band Metasurface for Sub-6 Wireless Networks. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. 389–404.

[25] Xinyi Li, Gaoteng Zhao, Ling Chen, Xinyu Zhang, and Ju Ren. 2024. RFMagus: Programming the Radio Environment With Networked Metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 16–30.

[26] Jeng-Sik Lim, Sung-Won Lee, Chul-Soo Kim, Jun-Seek Park, Dal Ahn, and Sangwook Nam. 2001. A 4.1 unequal Wilkinson power divider. *IEEE microwave and wireless components letters* 11, 3 (2001), 124–126.

[27] Xing Lin, Yair Rivenson, Nezih T Yardimci, Muhammed Veli, Yi Luo, Mona Jarrahi, and Aydogan Ozcan. 2018. All-optical machine learning using diffractive deep neural networks. *Science* 361, 6406 (2018), 1004–1008.

[28] Che Liu, Qian Ma, Zhang Jie Luo, Qiao Ru Hong, Qiang Xiao, Hao Chi Zhang, Long Miao, Wen Ming Yu, Qiang Cheng, Lianlin Li, et al. 2022. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nature Electronics* 5, 2 (2022), 113–122.

[29] Cirrus Logic. 2025. 114 dB, 24-bit, 192 kHz, 6 and 8-Channel D/A Converters with DSD Support and Low-Latency Digital Filtering. https://www.cirrus.com/products/cs4362a-82a/.

[30] Ruichun Ma, Shicheng Zheng, Hao Pan, Lili Qiu, Xingyu Chen, Liangyu Liu, Yihong Liu, Wenjun Hu, and Ju Ren. 2024. Automs: Automated service for mmwave coverage optimization using low-cost metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 62–76.

[31] Microsoft. 2025. Neural Network Intelligence Documentation. https://nni.readthedocs.io/.

[32] Mini-Circuits. 2025. Monolithic Amplifier DC-6000MHz. https://www.minicircuits.com/pdfs/GALI-24+.pdf.

[33] Aravind Nagulu and Harish Krishnaswamy. 2021. Non-magnetic nonreciprocal microwave components—State of the art and future directions. *IEEE Journal of Microwaves* 1, 1 (2021), 447–456.

[34] OneSDR. 2025. RF Pi Attenuator Calculator. https://www.onesdr.com/rf-pi-attenuator-calculator/.

[35] Hao Pan, Lili Qiu, Bei Ouyang, Shicheng Zheng, Yongzhao Zhang, Yi-Chao Chen, and Guangtao Xue. 2023. Pmsat: Optimizing passive metasurface for low earth orbit satellite communication. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–15.

[36] Parita Pooj, Michael Grossberg, Peter N Belhumeur, and Shree K Nayar. 2018. The Minimalist Camera.. In *BMVC*. 141.

[37] qorvo. 2025. 50 - 6000 MHz Cascadable InGaP HBT Active Bias MMIC Amplifier. https://www.qorvo.com/products/p/SBB5089Z.

[38] Ettus Research. 2025. USRP-2974 High Performance Embedded SDR, 10 MHz-6 GHz, 160 MHz bandwidth. https://www.ettus.com/all-products/usrp-2974/.

[39] Abinash Singh and Mrinal Kanti Mandal. 2019. Arbitrary coupling arbitrary phase couplers with improved bandwidth. *IET Microwaves,*

Antennas & Propagation 13, 6 (2019), 748–755.

[40] skyworks. 2025. SE5004L 5 GHz, 26dBm Power Amplifier with Power Detector. https://www.skyworksinc.com/Products/Amplifiers/SE5004L.

[41] ARK RF Systems. 2025. Transmission Line calculator. https://www.arkrfsystems.com/Linecalc.htm.

[42] Yezhou Wang, Hao Pan, Lili Qiu, Linghui Zhong, Jiting Liu, Ruichun Ma, Yi-Chao Chen, Guangtao Xue, and Ju Ren. 2024. GPMS: Enabling Indoor GNSS Positioning using Passive Metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 1424–1438.

[43] Andreas Wentzel, Viswanathan Subramanian, Ahmed Sayed, and Georg Boeck. 2006. Novel broadband Wilkinson power combiner. In *2006 European Microwave Conference*. IEEE, 212–215.

[44] Yongle Wu, Jun-Yu Shen, Yuanan Liu, Sai-Wing Leung, and Quan Xue. 2013. Miniaturized arbitrary phase-difference couplers for arbitrary coupling coefficients. *IEEE transactions on microwave theory and techniques* 61, 6 (2013), 2317–2324.

[45] Shuo Yang, Xindong Zhang, Xiaolong Wang, He Zhu, and Geyu Lu. 2024. Wideband Hybrid Coupler With Large Power Division Ratios/Arbitrary Output Phases and Low In-Band Amplitude/Phase Deviation. *IEEE Transactions on Microwave Theory and Techniques* (2024).

[46] Xueyuan Yang, Zhenlin An, Qingrui Pan, Lei Yang, Dangyuan Lei, and Yulong Fan. 2024. Binary Optical Machine Learning: Million-Scale Physical Neural Networks with Nano Neurons. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 603–617.

[47] Jaeyeon Yu, Seongjin Park, Inyong Hwang, Daeik Kim, Frederic Demmerle, Gerhard Boehm, Markus-Christian Amann, Mikhail A Belkin, and Jongwon Lee. 2022. Electrically tunable nonlinear polaritonic metasurface. *Nature Photonics* 16, 1 (2022), 72–78.

[48] Zeenko. 2025. LibreVNA. https://www.zeenko.tech/librevna.

[49] Yongzhao Zhang, Yezhou Wang, Lanqing Yang, Mei Wang, Yi-Chao Chen, Lili Qiu, Yihong Liu, Guangtao Xue, and Jiadi Yu. 2023. Acoustic sensing and communication using metasurface. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. 1359–1374.

[50] Ruizhe Zhao and Lingling Huang. 2024. High efficiency and dynamic modulation of nonlinear metasurfaces. *Light: Science & Applications* 13, 1 (2024), 245.