Phase-Enhanced Amplitude–Phase Gated Recurrent Unit Neural Network for Digital Predistortion in RF Power Amplifiers

Biao Long, Zhuhua Hu, *Senior Member, IEEE*, and Dake Liu, *Senior Member, IEEE*

***Abstract*—To address the performance limitations and computational redundancy of neural networks in digital predistortion (DPD) for radio frequency (RF) power amplifiers (PAs), this article proposes a phase-aware enhanced amplitude–phase gated recurrent unit (PEAPGRU) model. By structuring the input features within a joint amplitude–phase representation in polar coordinates, the model provides a more cohesive and efficient framework and ensures theoretical compatibility with polar-based RF transmitter architectures. A lightweight module for dynamic feature weighting across channel and spatial dimensions is proposed to suppress redundant information, thereby improving both convergence speed and overall performance. Through a mathematical analysis of distortion mechanisms, the phase-aware mechanism detects temporal phase jumps and enables directional feature enhancement in regions of abrupt phase change, leading to superior nonlinear distortion capture and noise suppression. Compared to state-of-the-art DPD networks, the proposed algorithm achieves a relative improvement of 2.1–4.1 dB in ACLR, 1.9–3.9 dB in NMSE, and 1.1–2.5 dB in EVM. Experimental results demonstrate the superior performance of the proposed method in DPD tasks, marking a new advancement in the application of deep learning techniques to the RF front-end design.**

***Index Terms*—Digital predistortion (DPD), phase awareness, phase-aware enhanced amplitude–phase gated recurrent unit (PEAPGRU), power amplifiers (PAs).**

I. Introduction

As A core component of modern communication systems, the radio frequency (RF) power amplifier (PA) is typically deployed in the front-end chain of transmitters or receivers, where it performs the critical function of signal amplification [1]. When operating in the nonlinear region, the PA not only causes in-band signal distortion but also introduces adjacent channel power leakage, severely compromising the transmission quality of the communication system. To counteract these effects, digital predistortion (DPD) has emerged as a leading technique that preconditions the input signal to invert the PA's nonlinear transfer function [2]. This approach effectively compensates for both nonlinearities and memory effects while largely preserving power efficiency [3], making it a favored solution for linearizing PAs.

Received 23 September 2025; revised 25 November 2025; accepted 30 December 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62361024 and Grant 62161010, in part by the Key Research and Development Project of Hainan Province under Grant ZDYF2024GXJS021, and in part by Hainan Seed Industry Laboratory under Grant B23H10004. *(Corresponding author: Zhuhua Hu.)*

The authors are with the State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou 570228, China (e-mail: eagler_hu@hainanu.edu.cn).

Digital Object Identifier 10.1109/TMTT.2025.3650528

A. Research Motivation

Conventional DPD implementations frequently employ mathematical modeling based on Volterra memory polynomial series, as demonstrated in studies [4], [5], [6]. A novel particle swarm optimization (PSO) approach in [7] added dynamic nonlinear adaptive hyperparameters to DPD. For multiband scenarios, the concept of adaptive parameter configuration has also been applied to DPD. Djorić et al. [8] combined DPD compensation techniques with digital second-order methods across two distinct frequency bands. Similarly, the Pareto principle was applied by [9] to extract optimal coefficients and propose distinct configurations for multiple linearity metrics across different frequency regions. Despite these efforts, global polynomials often fail to capture the distinctive behaviors of advanced PAs that deviate from weak nonlinearity and continuous smoothness [10]. Moreover, improving model accuracy by adding more terms to the global polynomial structure is considered to have inherent limitations [11].

In recent years, the remarkable achievements of artificial neural networks in the field of communications have attracted the attention of researchers working on wireless PA modeling. Thanks to their exceptional forward and backward modeling capabilities, neural networks excel at nonlinear compensation [12], making them ideal for tackling PA distortions. Amplifier modeling has seen the application of several classical neural network architectures, including convolutional neural networks (CNNs) [13] and deep neural networks (DNNs) [14]. The real-valued time-delay neural network (RVTDNN) offers a favorable complexity–performance tradeoff, enabling high throughput in wireless transmitters while accurately tracking PA characteristics [15], [16]. Wei et al. [17] and Wang et al. [18] indicate that event-driven spiking neural networks (SNNs) can also be used for DPD, which can effectively reduce computational cost.

The capability of recurrent neural networks (RNNs) to capture temporal dependencies enables them to deliver excellent compensation performance in DPD. While fundamental RNN models like long short-term memory (LSTM) networks [19] and gated recurrent unit (GRU) networks effectively address

TABLE I
SUMMARY OF PREDISTORTION TECHNIQUES

Reference	Technical Solution	Core Method	Strength	Limitation	Objective
[4]-[9]	Volterra Memory Polynomial	High-order polynomial modeling	Strong mathematical interpretability	High complexity, difficult design	Simplify the calculation difficulty
[12]-[18]	Feedforward Static Neural Networks	Learning I/Q to predistorted mapping using DNN/CNN	Low deployment cost	Limited tracking of long-term variations	Dynamic modeling capability
[19]-[21]	Recurrent Neural Networks (RNN)	Capturing long-term memory effects via gating mechanisms	Strong temporal correlation modeling	Ineffective in tracking complex nonlinear dynamics	Enhanced Time Series Modeling
[22]-[23]	Feature Extended Models	Expanding features to Amp, Amp^2, Amp^3, etc.	Enhanced nonlinear capture capability	Significant interference from redundant features	Feature space optimization

long-term dependency issues, they often fall short in modeling the complex nonlinear characteristics of PAs. A novel RNN variant, the adaptive inference pathway-gated neural network (AIPGNN), was proposed in [20] to dynamically activate finite impulse response (FIR) filter branches depending on current configuration data. A phase-normalized simplified RNN was also introduced, significantly reducing computational cost by replacing traditional dynamic gating mechanisms with static parameters [21] However, when the input feature dimensionality is low, models often struggle to comprehensively capture the nonlinear dynamics of PAs, a key limitation in current designs. To address this, researchers have proposed feature expansion approaches that expand input features to include amplitude, squared amplitude, cubed amplitude, and so on [22], [23], enabling more comprehensive characterization of signal properties and thereby improving DPD model performance. The aforementioned studies are summarized in Table I.

Despite advancements, prevailing DPD strategies remain constrained by two fundamental limitations. First, the focus on I/Q components and their magnitudes overlooks the critical role of phase data in polar coordinates, which limits a holistic understanding of input features. Second, these neural networks lack dynamic feature weighting capabilities, resulting in computational redundancy within high-dimensional spaces and suboptimal training convergence.

B. Main Contributions

This article makes the following key contributions.

1) To address the insufficient representational capacity of traditional DPD models for input features, this article proposes a phase-aware enhanced amplitude–phase GRU (PEAPGRU) neural network. A main innovation is a joint amplitude–phase feature framework in polar coordinates, which markedly enhances training convergence and final model accuracy by comprehensively capturing nonlinear dynamics. Through the construction of a multidimensional feature space encompassing amplitude (Amp), phase (ϕ), and their higher order nonlinear terms [e.g., Amp^3 and $\sin(\Phi)$], this architecture overcomes the dimensionality limitations of conventional I/Q component analysis. The complete retention of polar coordinate information provides a theoretical foundation for compatibility with RF polar transmitter architectures.
2) A channel–spatial dual-path feature enhancement architecture is proposed to mitigate feature redundancy caused by expanded feature dimensions. This work integrates a dual-path feature weighting module, comprising both channel and spatial dimensions, with the GRU network, enabling adaptive feature selection and enhancement. In the channel dimension, global pooling dynamically prioritizes critical features by generating a weighting vector, effectively suppressing redundant dimensions. For spatial enhancement, temporal convolutional kernels accurately capture transient signal dynamics, thereby significantly improving the PA's temporal modeling accuracy.
3) To address the tradeoff between accuracy and efficiency in feature enhancement for PA modeling, an optimization scheme based on a phase dynamic awareness mechanism is proposed. Through a mathematical analysis of PA nonlinearity in polar coordinates, we hypothesized that the model behaves distinctly across different phase intervals. This understanding led to a phase-aware enhancement model based on this mathematical analysis. This model calculates the instantaneous phase change rate through phase difference computation and identifies transition regions with a slope consistency constraint. The enhanced model achieves a breakthrough performance, reaching an adjacent channel leakage ratio (ACLR) of −50.63 dB, which represents a 2.5-dB improvement over the baseline. Compared to the model without phase awareness, the proposed approach improves the performance by approximately 1.0 dB and reduces the number of parameters by 8.9%, significantly enhancing modeling efficiency.
4) We constructed a modular test platform utilizing a 150-MHz bandwidth OFDM signal. This signal was generated and then driven through an ADL5545 PA by a vector signal generator, while a spectrum analyzer was used for experimental measurement and analysis. Evaluation results confirm the proposed PEAPGRU model achieves an ACLR improvement of approximately 4.1 and 3.2 dB over conventional VDLSTM and RVT-DCNN models, respectively, and maintains a 2.1-dB performance advantage over the state-of-the-art QGRU algorithm. In terms of convergence speed, PEAPGRU converges 10–30 epochs faster in the early training

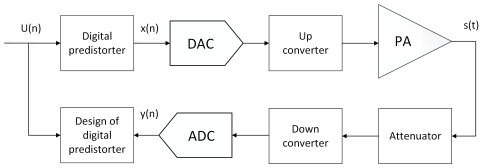

Fig. 1. Block diagram of a DPD system.

phase than APGRU (without feature weighting), validating its capability for rapid time–frequency feature extraction.

C. Article's Structure

The organization of this work is as follows. Section II provides an overview of the system modeling for RFPA. Section III presents the proposed PEAPGRU neural network in detail. Section IV describes the experimental setup and analyzes the results. Section V discusses the conclusions and future directions of this work.

II. DPD System Modeling

A typical block diagram of a DPD system is shown in Fig. 1. In the digital baseband domain, the original modulated signal $u(n)$ is predistorted to generate an intermediate signal $x(n)$ with inverse distortion characteristics.This signal is then reconstructed into a time-domain waveform through a digital-to-analog converter (DAC), upconverted to the RF band by a mixer, and used to drive the PA. When the nonlinear response of the predistorter is the exact inverse of the PA's transfer function, the RF output signal $s(t)$ exhibits ideal linear amplification properties [9]. The output signal $s(t)$ from the PA is attenuated, downconverted to baseband, and sampled by a high-speed analog-to-digital converter (ADC), ultimately yielding an observed signal $y(n)$ that captures the nonlinear characteristics of the PA. From the joint analysis of $x(n)$ and $y(n)$, an adaptive linearization model for the PA is built.

Accurate modeling of PA nonlinearity is core to DPD. Typical DPD system first builds a PA behavioral model to accurately predict the output response, thereby fully characterizing its nonlinear distortion and memory effects. Subsequently, the DPD module predistorts the input signal based on this behavioral model to counteract the nonlinear distortion introduced by the amplifier. We adopt an indirect learning architecture [24] to construct the DPD system. The advantage of this approach lies in its ability to indirectly derive the predistortion function by analyzing the relationship between the PA output signal and the ideal input signal, effectively circumventing complex real-time measurement issues. A key measure is the model reuse mechanism, where the trained PA behavioral model is deployed as a virtual PA to create a controlled environment for DPD training, ensuring stability and efficiency. Specifically, the output of the PA is used as the training input, and the original signal serves as the target output. A neural network is then employed to continuously optimize the loss between them, thereby learning the inverse characteristics of the PA. This offline training

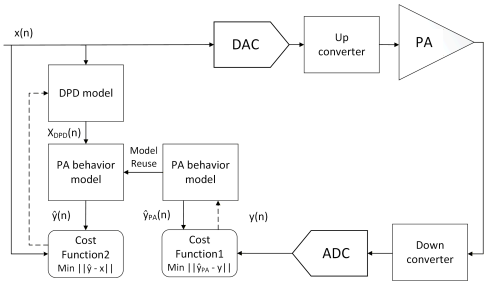

Fig. 2. DPD model building block diagram.

mechanism significantly improves implementation efficiency, enhances adaptability to various PAs, and ensures excellent nonlinear compensation performance.

Fig. 2 illustrates the network architecture employing the proposed DPD approach. The core workflow comprises three main components: complete signal chain DAC, PA behavioral modeling, and DPD parameter learning. In PA behavioral modeling, the time-series waveform of the modulated signal is used as input and the actual PA output as label. The neural network output and the measured PA output are fed into a loss function, and the prediction distortion is minimized to force the behavioral model to approximate the real hardware. The cost function for PA modeling (Cost Function 1) minimizes the difference between the predicted PA output $\hat{y}_{\mathrm{PA}}(n)$ and the actual output $y(n)$, while the cost function for DPD training (Cost Function 2) minimizes the difference between the virtual output $\hat{y}(n)$ of the reused PA model and the original input $x(n)$. Subsequently, an indirect learning methodology is adopted, which focuses on predicting the PA output rather than directly tuning hardware parameters. Next, the proposed neural DPD network is used to compensate for the distortion introduced by the PA. The DPD is typically designed from the PA behavioral model and iteratively approximates its inverse.

III. Neural Network Model

A. Research Methodology

The DPD process for PA is inherently a problem with strong temporal dependencies, as the current state of the signal is closely related to previous states. RNN-based neural networks, which naturally model time-series dependencies, represent an ideal choice for DPD tasks [25]. These networks analyze the temporal characteristics of baseband I/Q signals to establish a nonlinear mapping between input and output, enabling precise distortion prediction and compensation. This time-series modeling approach adaptively tracks changes in PA characteristics, demonstrating excellent dynamic compensation capability and offering significant advantages in the field of RFPA linearization.

In previous studies, to address the limitations of traditional RNNs in handling long-term dependencies, researchers have developed more complex network architectures, such as LSTM and GRU variants. While these networks exhibit enhanced capability in capturing temporal dependencies, their application to PA DPD demands refinements in feature engineering to ensure robust convergence and task performance gains. This

necessitates co-optimizing network architecture with adaptive feature extraction mechanisms, which collectively accelerate model convergence while advancing linearization efficacy beyond conventional approaches.

When processing multidimensional feature inputs, the computational complexity of LSTM increases significantly, leading to slower training and inference. In contrast, simplified RNN variants, such as the simple recurrent unit (SRU) and other lightweight RNN architectures, are computationally efficient but often lack the structural capacity to effectively capture the complex nonlinear characteristics of PAs. These simplified models generally underperform in dynamic modeling compared to LSTM and GRU. Therefore, we selected the GRU as the core unit for its optimal balance between capturing long-term dependencies and maintaining computational efficiency, making it particularly suitable for handling high-dimensional input features.

In DPD applications, although the original dataset may only contain basic 2-D input features such as I/Q signals, we can enhance the representational capacity of the model by combining or deriving additional features—such as amplitude, phase, and sine of the phase angle. The GRU's hidden states effectively encode the complex relationships among these input features. The introduction of phase information Φ is a key focus of our work. While traditional I/Q modeling often overlooks these dynamic variations, our work highlights that the instantaneous phase difference ($\Delta\Phi$), derived from Φ, is a critical feature for characterizing the PA's strong nonlinear regions, thereby significantly boosting model performance.

B. Nonlinear Modeling With Polar Coordinate Feature Fusion

In a typical DPD system, the nonlinear behavior of a PA is commonly described using a memory polynomial structure. Let $x(t)$ denote the input complex baseband signal and $y(t)$ represent the corresponding PA output signal. The nonlinear model of the PA can be expressed by the following equation, where a_{km} denotes the nonlinear coefficients targeted for compensation in the DPD system:

$$y(n) = \sum_{m=0}^{M} \sum_{k=0}^{K} a_{km} \cdot x(n-m) \cdot |x(n-m)|^k . \tag{1}$$

In this study, we propose an innovative approach that leverages composite input features—such as I_x(in-phase component), Q_x (quadrature component), Amp (signal amplitude), Φ (signal phase angle), Amp^3 (cubic nonlinear term of amplitude), and sin(Φ) (sine component of the phase)—to enhance the representational capacity of the model. We place particular emphasis on the utilization of the phase Φ, which represents the angle of the complex signal in polar coordinates and is critical for implementing modulation schemes. By changing the phase value, the transmitter can modulate the characteristics of the signal, ensuring effective transmission of information [26]. Assuming that the input signal is given in the form $x(t) = I_x + jQ_x$, the conversion to polar coordinates is defined by the following equation:

$$\begin{cases} A(t) = \sqrt{I_x^2 + Q_x^2} \\ \Phi(t) = \tan^{-1}\left(\dfrac{Q_x}{I_x}\right). \end{cases} \tag{2}$$

By leveraging both amplitude and phase information in polar coordinates, our method significantly enhances the model's ability to capture complex signal nonlinearities, thereby reducing PA distortion. This approach offers great flexibility by streamlining compatibility and conversion between IQ and polar transmitter architectures, which facilitates interoperability across different transmitter systems. After transforming the signal into its complex form, the input to the PA can be expressed as shown in the following equation:

$$x(n) = I_x(n) + jQ_x(n) = A(n)\, e^{j\Phi(n)}. \tag{3}$$

Substituting (3) into (1) yields an expression with both the amplitude nonlinear term $H(\Phi)$ and the phase nonlinear term $\Delta\Phi(n)$, as shown in the following equation:

$$\begin{aligned} y(n) &= \sum_{m=0}^{M} \sum_{k=0}^{K} a_{km} \cdot A(n-m)\, e^{j\Phi(n-m)} \cdot |A(n-m)|^k \\ &= \sum_{m=0}^{M} \sum_{k=0}^{K} a_{km} \cdot A(n-m)^{k+1}\, e^{j[\Phi(n-m)-\Phi(n)]} \cdot e^{j\Phi(n)} \\ &= |H(\Phi)| \cdot e^{j(\Phi(n)+\Delta\Phi(n))}. \end{aligned} \tag{4}$$

The combined effects of memory depth m and nonlinear order k are incorporated in the phase nonlinearity term here, $\Delta\Phi(n) = \arg\left(\sum_{m,k} a_{km} A^k(n-m) e^{j\Phi(n-m)}\right) - \Phi(n)$. In practical applications, directly computing $\Delta\Phi(n)$ is highly complex and challenging. Therefore, a feature engineering approach can be adopted to construct a set of features that approximate the influence of $\Delta\Phi(n)$. These features include the current amplitude, phase, and their nonlinear combinations. Based on the properties of complex arithmetic, the dominant term of $\Delta\Phi(n)$ can be expanded into: $\Delta\Phi(n) \propto \sum_{m,k} a_{km} A^k(n-m) \cdot \sin\Phi(n-m) + H(\Phi)$. Thus, their nonlinear combinations can be approximated by a sinusoidal series. Therefore, a sine function can be introduced in the feature engineering process to enhance the model's learning capability. In this way, the nonlinear variations in phase can be effectively learned by the subsequent neural network.

C. Design of Phase-Aware Dynamically Enhanced GRU Architecture

However, the increase in input feature dimensionality leads to heightened feature redundancy, which adversely affects both the convergence efficiency and nonlinear modeling capability of the model. To address this challenge, we designed a lightweight phase-aware feature enhancement module placed before the GRU. The lightweight nature of this module is architecturally achieved by employing a minimalist design. Specifically, compared to standard attention mechanisms like the convolutional block attention module (CBAM) [27], [28], our approach replaces fully connected layers with 1×1

convolutions for channel weighting and utilizes a compact 3×3 kernel for spatial weighting. This streamlined structure significantly reduces both the parameter count and computational cost. By employing convolutional kernels and phase-aware control, the module learns temporal correlations within the signal. As a result, continuous variations of the signal across adjacent time steps are accurately captured, thereby enhancing the modeling accuracy of time-varying characteristics in polar coordinate signals.

The nonlinear response of the PA shows significant phase dependency. Traditional analysis in the I/Q plane struggles to capture the distortion mechanisms in polar coordinates while directly interpreting the Volterra series expression in (4) does not easily or intuitively reveal the phase distortion mechanism. To simplify the discussion, it is necessary to construct a phase-sensitive kernel function based on (4). The function we select is presented in the following equation:

$$H_{\phi}^{(m,k)} = \frac{\partial \Delta\Phi(n)}{\partial \phi(n-m)}. \tag{5}$$

Its purpose is to quantify the contribution of the phase perturbation $\phi(n-m)$ at historical time m to the current distortion $\Delta\Phi(n)$, which reflects the nonlinear intensity of phase compression in the PA. When $m=0$, it corresponds to instantaneous self-disturbance, while $m>0$ characterizes memory effects. Based on this, the following equation can be derived:

$$\begin{aligned} H_{\phi}^{(m,k)} &= \frac{\partial \Delta\Phi(n)}{\partial \phi(n-m)} \\ &= \frac{\partial}{\partial \phi(n-m)} \left\{ \arg\left(\sum_{m'=0}^{M} \sum_{k'=0}^{K} a_{k'm'} A^{k'}(n-m') \right. \right. \\ &\qquad \left. \left. e^{j\phi(n-m')} \right) - \phi(n) \right\}. \end{aligned} \tag{6}$$

The expression defines $H_0 = \sum\sum a_{k'm'} A^{k'} e^{j\phi(n-m')}$, which represents a complex-valued function constructed through double summation. It is then decomposed into its real and imaginary parts as $H_0 = u(\phi) + jv(\phi)$, where $u(\phi) = \mathrm{Re}(H_0)$ and $v(\phi) = \mathrm{Im}(H_0)$. Thus, we can further derive

$$\begin{aligned} H_{\phi}^{(m,k)} &= \frac{\partial}{\partial \phi(n-m)} \left(\arg H_0 - \phi(n) \right) \\ &= \frac{\partial}{\partial \phi(n-m)} \left(\tan^{-1} \frac{v}{u} \right) \\ &= \frac{1}{1+\left(\frac{v}{u}\right)^2} \cdot \frac{\partial}{\partial \phi} \left(\frac{v}{u} \right) = \frac{u\frac{\partial v}{\partial \phi} - v\frac{\partial u}{\partial \phi}}{u^2+v^2}. \end{aligned} \tag{7}$$

Although (7) is accurate, its physical interpretation is not sufficiently intuitive. By defining position and velocity vector $R=(u,v)$ and $V=((\partial u)/(\partial\phi), v(\partial v)/(\partial\phi))$, the numerator of (7) is analogous to the cross product $R\times V$, similar to angular momentum. This measures the component of motion orthogonal to the instantaneous position direction, corresponding to the rotational distortion caused by phase variations.

To extract this physical quantity systematically, we construct an auxiliary function $T = \overline{H_0}(\partial H_0)/(\partial\phi)$. The conjugate term $\overline{H_0}$ provides a phase reference, while $(\partial H_0)/(\partial\phi)$ represents the response velocity. We expand the product

$$\begin{aligned} T &= (u-jv)\left(\frac{\partial u}{\partial \phi} + j\frac{\partial v}{\partial \phi} \right) \\ &= u\frac{\partial u}{\partial \phi} + ju\frac{\partial v}{\partial \phi} - jv\frac{\partial u}{\partial \phi} - j^2 v\frac{\partial v}{\partial \phi} \\ &= u\frac{\partial u}{\partial \phi} + v\frac{\partial v}{\partial \phi} + j\left(u\frac{\partial v}{\partial \phi} - v\frac{\partial u}{\partial \phi} \right). \end{aligned} \tag{8}$$

Thus, the following expression can be derived:

$$H_{\phi}^{(m,k)} = \frac{u\frac{\partial v}{\partial \phi} - v\frac{\partial u}{\partial \phi}}{u^2+v^2} = \frac{\mathrm{Im}\left(\overline{H_0}\frac{\partial H_0}{\partial \phi} \right)}{|H_0|^2}. \tag{9}$$

Equation (9) provides a physical interpretation. The kernel represents the normalized tangential component. The imaginary part of T, $\mathrm{Im}(T) = R\times V$, is analogous to angular momentum, capturing the inertial effect of phase rotation.

In (9), the symbol $(\partial H_0)/(\partial\phi)$ actually denotes the partial derivative with respect to the specific variable $\phi(n-m)$ since $H_{\phi}^{(m,k)}$ is defined as the sensitivity kernel for historical time index m and nonlinear order k. Thus, during substitution, we can directly use the expression from (10). The summation sign in (10) is eliminated because, when computing the partial derivative, only the term with $m'=m$ explicitly contains $\phi(n-m)$. The contributions of other terms are zero, reducing the summation to a single term

$$\begin{aligned} \frac{\partial H_0}{\partial \phi(n-m)} &= j a_{km} A^k (n-m) e^{j\phi(n-m)} \\ &\quad \times \left(\text{where } m' = m \text{is nonzero} \right). \end{aligned} \tag{10}$$

Bring (10) into (9) and get

$$\begin{aligned} H_{\phi}^{(m,k)} &= \frac{\mathrm{Im}\left(\overline{H_0}\frac{\partial H_0}{\partial \phi} \right)}{|H_0|^2} = \frac{\mathrm{Im}\left(j a_{km} A^k e^{j\phi(n-m)} \overline{H_0} \right)}{|H_0|^2} \\ &= \frac{a_{km}A^k}{|H_0|^2} \mathrm{Re}\left(e^{j\phi(n-m)} \overline{H_0} \right). \end{aligned} \tag{11}$$

Let $\theta_0 = \arg(H_0)$, $\Delta\phi = \phi(n-m) - \theta_0$, and then, $H_0 = |H_0|e^{j\theta_0}$ and $\overline{H_0} = |H_0|e^{-j\theta_0}$

$$\begin{aligned} H_{\phi}^{(m,k)} &= \frac{a_{km}A^k}{|H_0|^2} \mathrm{Re}\left(e^{j\phi(n-m)} |H_0| e^{-j\theta_0} \right) = \frac{a_{km}A^k}{|H_0|} \cos(\Delta\phi) \\ &\approx \frac{a_{km}A^k}{|H_0|} \cdot \begin{cases} 1, & \text{if } |\Delta\phi| \le \theta_a \\ 1 - \dfrac{\Delta\phi^2}{2}, & \text{if } \theta_a < |\Delta\phi| \le \theta_c \\ \cos\Delta\phi, & \text{if } |\Delta\phi| > \theta_c. \end{cases} \end{aligned} \tag{12}$$

As shown in (12), the proposed phase-sensitive kernel function is a decreasing function within its defined range. Its amplitude decreases as $|\Delta\phi|$ increases, revealing the inertial response mechanism of the PA to phase variations. The more abrupt the phase change of the signal, the more difficult it is for the PA to respond instantaneously, leading to reduced efficiency in responding to input features. Based on the segmented expression in (12), the system exhibits significantly different nonlinear characteristics in distinct phase difference regions. Let θ_a be a very small angular threshold and θ_c be the critical angular threshold. When $|\Delta\phi| \le \theta_c$, the second derivative of

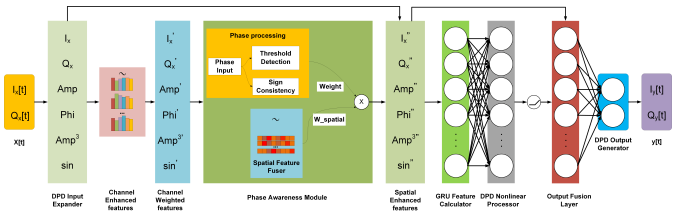

Fig. 3. Framework of the phase-aware enhanced amplitude–phase GRU neural network.

the phase-sensitive kernel remains constant, indicating that the system is in a stable and predictable state. When $|\Delta\phi| > \theta_c$, the second derivative exhibits time-varying oscillatory behavior, marking the system's entry into a strongly nonlinear region. This fundamental difference in stability behavior stems from the cumulative memory effects of the PA. Within small phase difference ranges, signal distortion patterns remain stable, whereas beyond θ_c, the nonlinear coupling of historical phase perturbations manifests chaotic characteristics.

The architecture proposed in this article demonstrates multiple adaptive advantages for DPD tasks, enabling more comprehensive utilization of phase information to enhance the model's capability to capture signal characteristics. The channel-wise feature fusion mechanism is inherently suited for collaborative processing of multiple physical quantities, effectively identifying nonlinear coupling relationships among amplitude-related features and strengthening the distinct representation of the phase angle Φ and its sinusoidal components. Meanwhile, the spatially constructed phase-aware sparse processing submodule incorporates an innovative dual-condition discrimination mechanism based on phase partition detection and slope consistency detection. This enables efficient modeling of temporal nonlinear characteristics and significantly improves feature processing efficiency. Experimental validation confirms that the lightweight dual-path GRU network significantly enhances compensation capability for complex signal nonlinearity and accelerates training convergence through optimized feature coupling. The proposed overall neural network architecture is shown in Fig. 3.

The network first expands the 2-D I/Q input into a 6-D feature vector. The features are then recalibrated by a channel-weighting module (utilizing global average pooling and a 1×1 convolution), after which they are further enhanced by an integrated module. This module combines spatial weighting based on average pooling, max pooling, and 3×3 convolution with piecewise phase-aware enhancement derived from the instantaneous phase difference. Then, the features are modeled by a single-layer GRU (hidden size 9), then undergo a nonlinear transformation through a fully connected layer, and are fused with the input features. Finally, they output through a fully connected layer with two neurons.

According to the system design, the input signals $I_x[t]$ and $Q_x[t]$ are first expanded into a multidimensional input feature vector $x_t = \{I_x, Q_x, \text{Amp}, \phi, \text{Amp}^3, \sin(\phi)\}$. Dynamic weights are then computed for each feature, and the feature weight parameters are updated. The dynamic feature weighting mechanism used in this work is given by the following equation:

$$
\begin{cases}
M_c = \sigma\left(\text{Conv}\left(\text{ReLU}\left(\text{Conv}\left(\text{AvgPool}\left(x\right)\right)\right)\right)\right) \\
x_c = x \odot M_c \\
\Delta\phi_t = \phi_t - \phi_{t-1} \\
\Delta\widetilde{\phi}_t = \left[\left(\Delta\phi_t + \pi\right) \bmod 2\pi\right] - \pi \\
\Gamma_t = \begin{cases}
2 - \cos\Delta\widetilde{\phi}_t, & \text{if } \left|\Delta\widetilde{\phi}_t\right| > \theta_c \\
& \quad \& \ \text{sgn}\left(\Delta\widetilde{\phi}_t\right) = \text{sgn}\left(\Delta\widetilde{\phi}_{t-1}\right) \\
1 + \left|\Delta\widetilde{\phi}_t\right|^2/2, & \text{if } \theta_a < \left|\Delta\widetilde{\phi}_t\right| \le \theta_c \\
& \quad \& \ \text{sgn}\left(\Delta\widetilde{\phi}_t\right) = \text{sgn}\left(\Delta\widetilde{\phi}_{t-1}\right) \\
1, & \text{otherwise}
\end{cases} \\
M_s = \sigma\left(\text{Conv}\left(\left[\text{AvgPool}\left(x_c\right), \text{MaxPool}\left(x_c\right)\right]\right)\right) \\
W' = x_c \odot \left(\Gamma_t \odot \left(1 + M_s\right)\right)
\end{cases}
\tag{13}
$$

where M_c denotes the channel modulation weight matrix. It is generated by first performing global average pooling over the spatial–temporal dimensions using the AvgPool() operation. This is followed by two convolutional layers that perform significant dimension reduction and expansion through linear transformations to extract key features in a computationally efficient manner. The process is then activated by ReLU, and finally, the weight distribution is produced by the sigmoid function σ. The modulated feature X_c is obtained through the element-wise multiplication ($\odot$) of the original input feature x and the channel weight matrix M_c. This operation adaptively enhances critical physical quantities while suppressing redundant feature dimensions.

The core design of (13) builds upon the threshold partitioning strategy of (12), adapting its decreasing function characteristic into an increasing function form via a linear transformation, making it more suitable for neural network mechanisms. Theoretical analysis indicates that the phase distortion threshold can demarcate the linear and nonlinear operation zones of the PA. However, directly calculating this threshold relies on the PA's implicit nonlinear parameters, which are inaccessible through external measurements. Therefore, this work employs the phase difference $\Delta\phi_t$ between adjacent time steps to indirectly sense the nonlinear intensity. This approach inherits the core theoretical concept of "identifying nonlinearity via phase variation" while significantly improving computational efficiency, thereby paving the way for practical system deployment. Here, $\Delta\phi_t$ represents the original phase difference. It is normalized via modulo-2π operation into $\Delta\widetilde{\phi}_t \in [-\pi, \pi]$ to maintain phase continuity. The dynamic enhancement factor Γ_t is determined based on the absolute value of $\Delta\widetilde{\phi}_t$ relative to the thresholds θ_a and θ_c. When $|\Delta\widetilde{\phi}_t| > \theta_c$ and the direction of adjacent phase changes is consistent, $\Gamma_t = 2 - \cos(\Delta\widetilde{\phi}_t)$ is applied to compensate for strong nonlinear effects. In the transition region $(\theta_a, \theta_c]$ (where $\theta_a = 5°$) with consistent phase change direction, $\Gamma_t = 1 + |\Delta\widetilde{\phi}_t|^2/2$ is used to match second-order nonlinear characteristics. A unit gain ($\Gamma_t = 1$) is retained in other regions to prevent overcompensation.

Leveraging the adaptive capabilities of the neural network, the complex calculations in (12) are transformed into a dynamic enhancement factor generation mechanism Γ_t based on partition thresholds. This approach utilizes the nonlinear

fitting ability of the neural network to circumvent complex operations such as cosine and Taylor expansions. The spatial modulation vector $\mathbf{M}_s$ is generated by concatenating features obtained from average pooling (AvgPool) and max pooling (MaxPool) along the time dimension. This concatenated feature is then fused through a convolutional layer (Conv) and activated by a sigmoid function σ. The final output $\mathbf{W}'$ is the product of x_c and the composite enhancement term $\Gamma_t \odot (1 + \mathbf{M}_s)$, which serves as the input feature for the downstream GRU network. This strategic placement establishes the module as a dedicated preprocessing stage that performs feature refinement before temporal modeling. This division of labor allows the subsequent GRU to focus its complex recurrent computations solely on learning temporal dependencies from the already enhanced features, thereby improving overall efficiency. The GRU computation for multidimensional input features is given by the following equation:

$$\begin{cases} z_t = \sigma\left(W_z \cdot \left(W' x_t\right) + U_z \cdot h_{t-1} + b_z\right) \\ r_t = \sigma\left(W_r \cdot \left(W' x_t\right) + U_r \cdot h_{t-1} + b_r\right) \\ \tilde{h}_t = \tanh\left(W_h \cdot \left(W' x_t\right) + U_h \cdot (r_t \odot h_{t-1}) + b_h\right) \\ h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \end{cases} \tag{14}$$

where W_z, W_r, and W_h are the weight matrices corresponding to the hidden state of the current input; U_z, U_r, U_h are the weight matrices corresponding to the hidden state of the previous input; and b_z, b_r, and b_h are the bias terms. The term x_t corresponds to the multidimensional input features at the current time step, W' denotes the feature enhancement weight applied to these multidimensional input features, h_{t-1} represents the hidden state from the previous time step, and h_t is the new candidate hidden state computed based on the current input and the control of the reset gate. Finally, σ refers to the sigmoid activation function and tanh denotes the hyperbolic tangent activation function. When the input is a 6-D feature vector, the weight matrix of GRU (with dimensions [hidden_size = 9, 6]) maps the input x_t at each time step from the 6-D space to the hidden space through matrix multiplication. This operation essentially learns the dynamic correlation patterns of the six features in the context of the time series.

Within the PEAPGRU framework, a feedforward network dynamically computes feature enhancement weights for multidimensional inputs, prioritizing features based on their task relevance through training. These weights are applied to the input features to generate a weighted feature representation, enabling the model to focus on critical characteristics. The weighted features then flow into the GRU network, where update and reset gates model long-term temporal dependencies, enhancing dynamic feature tracking. The GRU output is processed through a fully connected layer and an activation function to generate the final prediction. The proposed PEAPGRU architecture enhances both the convergence rate and ultimate performance while only moderately increasing model complexity. The model achieves a favorable complexity–performance balance; its marginally increased complexity substantially boosts expressive power, enabling it to capture more intricate patterns and ultimately enhancing overall learning outcomes.

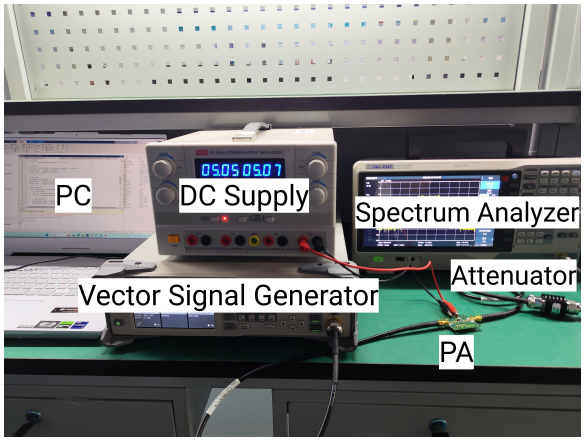

Fig. 4. Diagram of the experimental equipment.

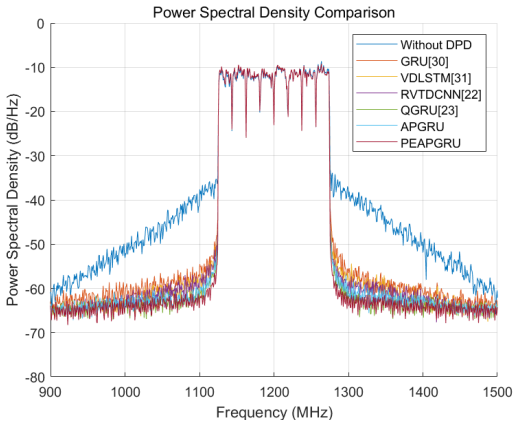

Fig. 5. Output spectra of the PA with several DPD algorithms.

IV. Experimental Results and Analysis

A. Experimental Setup

This study employed the modular test platform illustrated in Fig. 4 to validate the proposed DPD model. The test signal was a 150-MHz orthogonal frequency division multiplexing (OFDM) signal with a PAPR of 10.6 dB, generated using MATLAB software on a personal computer (PC). The test signal was then loaded into a vector signal generator (SSG-5040X-V). The RF signal produced by the vector signal generator at a carrier frequency of 1.2 GHz was connected via cable to the input of the PA (ADL5545). A dc power supply (MCH-3205II) provided the required dc voltage to the PA.

The amplified RF signal was transmitted via test cable to a Spectrum Analyzer SSA1015X-C for spectral monitoring. The PA output was passed through an attenuator to appropriately reduce the signal power and protect the spectrum analyzer. The attenuated signal was then fed into the spectrum analyzer for measurement and analysis. The PC used software to control the vector signal generator for signal production while simultaneously monitoring and recording test data. The training protocol leverages a cohesive configuration that processes

TABLE II
COMPARATIVE ANALYSIS OF PA PERFORMANCE FOR SEVERAL DPD ALGORITHMS

DPD Model	Input Feature Set	Hidden Size	Number of coefficients	FLOPs	NMSE (dB)	EVM (dB)	ACLR (dB)
Without DPD	-	-	-	-	-20.804421	-21.435288	-26.304027
GRU [29]	Ix,Qx	10	442	1020	-36.373234	-35.850413	-45.199333
VDLSTM [30]	Ix,Qx,sin(Φ),cos(Φ)	9	638	1400	-38.113533	-35.941953	-46.537270
RVTDCNN [22]	Ix, Qx, Amp, Φ, Amp^3, sin(Φ)	10	582	1300	-38.221323	-36.373357	-47.455457
QGRU [23]	Ix, Qx, Amp, Φ, Amp^3, sin(Φ)	10	684	1484	-40.063093	-37.299783	-48.515952
APGRU	Ix, Qx, Amp, Φ, Amp^3, sin(Φ)	9	581	1266	-40.005847	-37.161415	-48.147468
PEAPGRU	Ix, Qx, Amp, Φ, Amp^3, sin(Φ)	9	630	1366	-42.011374	-38.447812	-50.634451

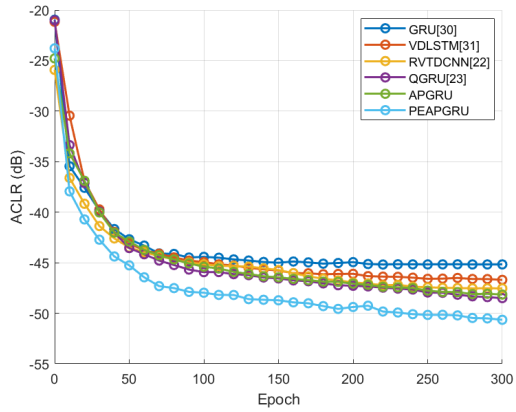

Fig. 6. ACLR comparison of the PA for different DPD algorithms.

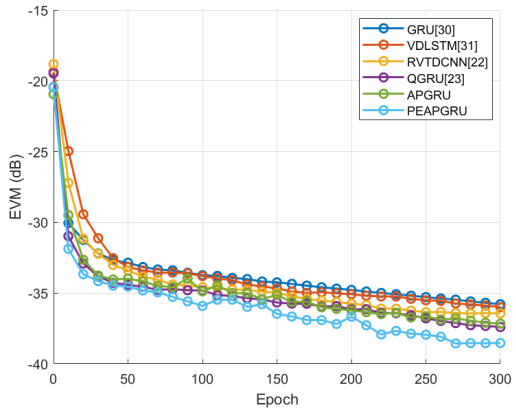

Fig. 7. EVM comparison of the PA for different DPD algorithms.

TABLE III
DPD PERFORMANCE OF THE PEAPGRU ALGORITHM UNDER DIFFERENT PHASE THRESHOLDS

Phase Threshold	NMSE (dB)	EVM (dB)	ACLR (dB)
15°	-41.839352	-38.200742	-49.317444
20°	-41.916876	-38.263210	-49.389144
25°	-41.484810	-37.913325	-50.290355
30°	-42.011374	-38.447812	-50.634451
40°	-41.349335	-37.873611	-49.634965
60°	-41.481518	-38.075061	-49.916031
90°	-41.669630	-38.199378	-49.429555

input sequences of 28 800 samples through a batch size of 128, optimizes weights via the AdamW optimizer with a learning rate of 0.001.

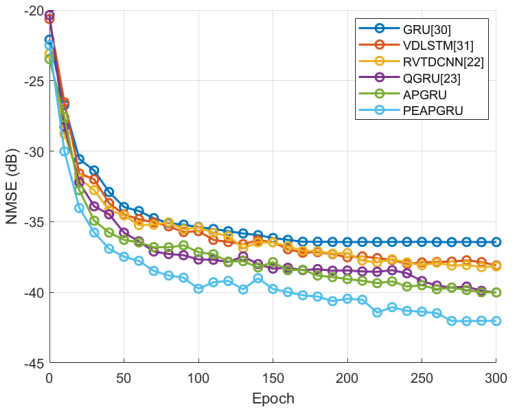

Fig. 8. NMSE comparison of the PA for different DPD algorithms.

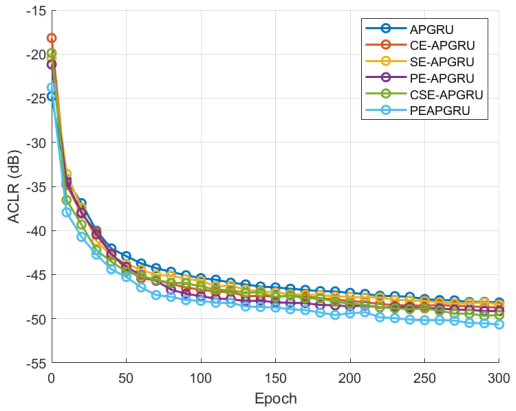

Fig. 9. ACLR convergence curves of the APGRU algorithm under different enhanced module configurations.

Using this setup, the output spectra of the PA under several DPD modeling methods were obtained, as shown in Fig. 5. The PC also recorded the ACLR, EVM, and NMSE metrics of multiple DPD modeling algorithms from 1 to 300 epochs, as illustrated in Figs. 6–8, respectively. The final performance results after iterative training are summarized in Table II. To validate the core contribution and parameter sensitivity of the proposed phase-aware enhancement mechanism in DPD modeling, Table III analyzes the influence of phase detection thresholds on nonlinear correction performance. Furthermore, Table IV presents an ablation study examining the dynamic impact of different module combinations on ACLR. For a controlled and isolated analysis of each module's contribution, all configurations in this ablation study share the same

TABLE IV
DPD PERFORMANCE OF THE APGRU ALGORITHM UNDER DIFFERENT MODULE CONFIGURATIONS

Channel Feature Enhancement Module	Spatial Awareness Enhancement Module	Phase Awareness Enhancement Module	Number of coefficients	FLOPs	NMSE (dB)	EVM (dB)	ACLR (dB)
×	×	×	581	1266	-40.005847	-37.161415	-48.147468
✓	×	×	609	1280	-40.245676	-37.362769	-48.697399
×	✓	×	600	1302	-40.104023	-37.243022	-48.376100
×	✓	✓	604	1352	-40.313864	-37.452186	-49.105243
✓	✓	×	686	1474	-41.408123	-38.092644	-49.574650
✓	✓	✓	630	1366	-42.011374	-38.447812	-50.634451

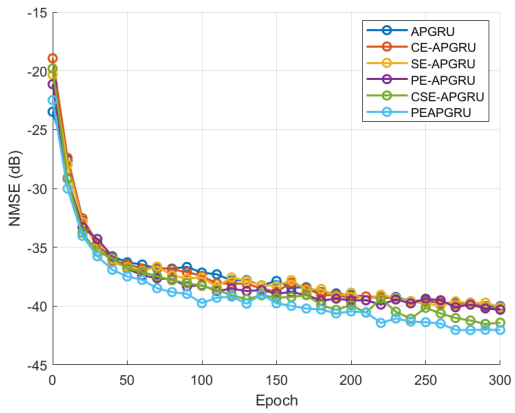

Fig. 10. NMSE convergence curves of the APGRU algorithm under different enhanced module configurations.

foundational settings: the input feature set is Ix, Qx, Amp, Φ, Amp^3, sin(Φ) and the hidden size is 9. Figs. 9 and 10, respectively, show the convergence trajectories of ACLR and NMSE under various module configurations, revealing the synergistic optimization mechanism between channel selection and phase dynamic awareness. The experimental study focused on validating the efficacy of model efficacy within a conventional IQ transmitter framework. This provides a foundational verification of its core principles, paving the way for future applications in polar transmitter systems.

B. Analysis of Experimental Results

To ensure a fair evaluation of our network architecture's innovation, we configured the proposed PEAPGRU model, the benchmark APGRU, and the RVTDNN and QGRU models all used exactly the same 6-D feature set [Ix, Qx, Amp, Φ, Amp^3, and sin(Φ)]. We respected and replicated the original feature engineering designs of the other baseline models. Specifically, VDLSTM was implemented using its characteristic 4-D feature set, and the basic GRU uses standard I/Q components. We also adjusted the hidden size of each model to align their total parameter countsto a similar magnitude.

The proposed PEAPGRU model demonstrates significant advantages in PA linearization performance. As shown in Fig. 5, the spectral suppression capability of PEAPGRU is notably superior to that of conventional algorithms, effectively controlling out-of-band leakage. According to Table II, PEAPGRU ultimately achieves an ACLR of −50.6 dB, which is the optimal value obtained after multiple adjustments of the phase-aware threshold in Table III. This result significantly outperforms traditional GRU (−45.20 dB), VDLSTM (−46.54 dB), and RVTDCNN (−47.46 dB). The baseline system without DPD achieves only −26.30-dB ACLR, confirming the importance of DPD for PA linearization. In terms of dynamic performance, the training dynamics in Figs. 6–8 show that PEAPGRU reaches an ACLR of −45.0 dB by epoch 50, which is 40 epochs faster than APGRU (amplitude–phase GRU without enhancement). Critically, PEAPGRU performance is achieved with a computational complexity of 1366 FLOPs, which is at the median among compared models, enabling a fair efficiency assessment. Compared with GRU, PEAPGRU achieves an ACLR gain of over 5.4 dB with only about a 34% increase in FLOPs, far outweighing its cost. Versus RVTDCNN, an increase of only about 5% in FLOPs can bring a 3.2-dB improvement in ACLR. Notably, against state-of-the-art QGRU, PEAPGRU reduces FLOPs by about 8% while leading by 2.1 dB in ACLR. The results demonstrate that PEAPGRU does not enhance performance simply by increasing complexity. Its innovative design achieves performance breakthroughs at similar or even lower computational costs, highlighting exceptional architectural efficiency. Regarding signal quality, PEAPGRU also excels in both error vector magnitude (EVM) and normalized mean square error (NMSE) metrics. The NMSE and EVM of the PEAPGRU algorithm reach −42.0 and −38.4 dB, respectively, indicating a substantial improvement in signal quality. The NMSE of PEAPGRU is improved by 1.9–3.9 dB relative to the comparison algorithms, while EVM is improved by 1.1–2.5 dB. Collectively, the proposed algorithm achieves at least a 1–3-dB improvement over other advanced algorithms.

The effect of the phase-aware threshold is systematically validated in Table III. It can be observed that the setting of the phase-aware threshold significantly influences the DPD performance of the PEAPGRU algorithm. The optimal phase threshold $\theta_c = 30°$ was determined through a systematic parameter sweep using a validation set, with ACLR as the primary optimization metric. This process involved evaluating thresholds from 10° to 90° in 5° increments, with $\theta_c = 30°$ achieving best performance as summarized in Table III. When the phase threshold is set to 30°, the model achieves its peak performance across all three key metrics: NMSE, EVM, and ACLR, with values of −42.01, −38.45, and −50.63 dB, respectively. If the threshold is lower than 30°, the model becomes overly sensitive to phase fluctuations caused by noise, leading to degradation in ACLR. Conversely, if the threshold is too high, some phase transitions may go

undetected, resulting in a degradation of approximately 1 dB in ACLR.

As shown in Table IV and Figs. 9 and 10, the dual-enhanced PEAPGRU configuration (channel + phase awareness) achieves an ACLR of −50.63 dB and an NMSE of −42.01 dB, representing an improvement of approximately 2.0 dB over the baseline APGRU. This confirms the synergistic effect between the phase-aware module and channel enhancement. Moreover, the phase-enhanced module (PE-APGRU) outperforms channel enhancement (CE-APGRU) and spatial enhancement (SE-APGRU), demonstrating the effectiveness of the phase threshold and slope consistency detection introduced in (13) in enhancing RF distortion modeling. Notably, Table IV shows that the fifth configuration (✓, ✓, and ×) uses a standard attention mechanism with a fully connected layer, resulting in FLOPs rising to 1474. In contrast, other configurations use lightweight modules, replacing fully connected layers with global average pooling and 1×1 convolutions for channel weighting, and using smaller 3×3 convolution kernels for spatial weighting, reducing FLOPs while maintaining performance. Analysis shows that enabling only the channel module increases FLOPs by 1.1% and improves ACLR by 0.5 dB; enabling only the spatial module results in an additional 2.8% increase in FLOPs and a 0.2-dB improvement in ACLR; when combining all three modules, FLOPs reaches 1366 and ACLR improves by 2.5 dB, demonstrating a synergistic effect. Despite added computation, the modules contribute only 7.3% to total FLOPs, indicating most computation lies in the multifeature GRU. The additional computational cost from the performance improvement is acceptable.

This study systematically validates the innovative value of the PEAPGRU architecture. The data in Table III and the ablation experiments in Figs. 9 and 10 collectively confirm the central role of the phase-aware component, with the optimal phase threshold accurately aligning with the nonlinear partitioning theory of the PA described in (12). Comparative experiments in Table IV show that removing the phase-aware mechanism results in a 2-dB degradation in ACLR, confirming its indispensability. The systematic experimental validation comprehensively covers key technical aspects such as threshold optimization, module synergy, and convergence dynamics, each quantitatively analyzed in Tables II–IV. Overall, the experiments verify that the gain compensation via the phase-sensitive kernel function effectively emulates the polynomial effects of the PA. By synergistically integrating phase kernel transformation, dynamic partitioning, and slope detection, PEAPGRU offers a novel approach to DPD—one that is driven by physical modeling within a neural network framework.

C. Discussion

The partitioned structure of (12) directly reveals the critical transition characteristics of the nonlinear behavior of the PA. When the phase offset $\Delta\phi$ crosses the 30° threshold, the system response transitions from a quadratic region dominated by third-order nonlinearities to a cosine oscillation region governed by higher order nonlinear effects. PEAPGRU maps this physical transition process into the neural network architecture by employing a dynamic weight allocation strategies that focuses learning on critical nonlinear regions, thereby improving training efficiency.

Compared to the multifeature and quantization gate control mechanism of QGRU, PEAPGRU embeds the physical model of polar coordinate nonlinear partitioning directly into the neural network architecture. This integration facilitates a hybrid approach that harnesses data-driven tuning and physics-based constraints. In contrast to VDLSTM fixed-window phase reconstruction and RVTDCNN time-delay tap compression, PEAPGRU utilizes phase jump detection to pinpoint nonlinear distortion sources with higher precision. This approach leverages the advantages of dynamism and reduces computational requirements and achieves an ACLR of −50.6 dB, yielding a 2.1-dB gain over leading alternatives.

Ablation studies in Table IV robustly demonstrate the architecture's strong cross-scenario applicability. The phase-aware mechanism locates nonlinear distortions through temporal phase jump detection, eliminating the dependency on energy analysis at specific frequency points. This frequency-independent design philosophy fundamentally enhances the model's generalization capability, providing scalable technical support for applications such as 5G-Advanced multistandard base stations and millimeter-wave fronthaul systems. The phase jump detection threshold θ serves as a core tunable parameter. Adjusting it allows the system to adapt to the varying AM–PM characteristic changes of different PA devices. These tuning parameters provide a flexible framework for multistandard use without core modifications, enabling adaptation via predefined settings.

The optimal phase threshold ($\theta_c = 30°$) identified in this study prompts consideration of its generalizability. Our analysis suggests that phase threshold is influenced by both the inherent nonlinearity of PA and the statistical properties of the input signal. First, the threshold serves to quantify the "inertial" effect in the PA's phase response, which is an intrinsic dynamic property of the specific PA. Consequently, the optimal, acting as a critical marker that distinguishes readily trackable phase variations from those dominated by nonlinear inertia, is not a universal constant but need to be calibrated according to the PA's inertial strength. Simultaneously, the threshold is also sensitive to signal characteristics and can be analyzed from the perspective of the activation frequency of the phase-aware module. Signals with greater bandwidth induce more rapid phase variations, altering the distribution of instantaneous phase jumps. To maintain an effective activation rate that is sufficient to capture nonlinearity without overreacting to noise, the threshold may need to be calibrated for signals of different bandwidths. It suggests that phase threshold is not a mere empirical constant but a parameter that can be tuned based on key system characteristics, paving the way for adaptive configuration strategies in future work.

The current work still faces several unresolved issues. For instance, the phase threshold in (12) lacks a quantitative relationship with the physical characteristics of the PA device, requiring empirical calibration. Furthermore, the present study lacks experimental validation in higher frequency bands. To

overcome these constraints, we will develop an empirical lookup table for phase thresholds, streamlining implementation by bypassing complex theoretical calculations. We will also deepen the study of phase sensitive kernel functions in polar coordinates to analyze PA characteristics from different perspectives. Validation will also be extended to higher frequencies to facilitate the design of RF systems with enhanced bandwidth and operational range.

V. Conclusion

This article innovatively proposes a PEAPGRU neural network. The model incorporates amplitude, phase, and high-order harmonics in polar coordinates, markedly enhancing PA nonlinear modeling accuracy. A mathematically grounded phase-aware module enables the neural network to converge faster and cut training costs. Experimental results demonstrate that the proposed model achieves an ACLR of −50.63 dB at a center frequency of 1.2 GHz, outperforming conventional VDLSTM and RVTDCNN by approximately 4.1 and 3.1 dB, respectively. It also maintains a 2.1-dB performance advantage over the current state-of-the-art QGRU algorithm. Furthermore, relative to other algorithms, the proposed method achieves improvements of 1.9–3.9 dB in NMSE and 1.1–2.5 dB in EVM.

Compared to existing studies, PEAPGRU achieves a systematic integration of polar coordinate features with deep learning architectures for the first time. Its core innovation lies in establishing a theoretical bridge between the phase-sensitive kernel function and neural network mapping, capturing the critical transition points of PA nonlinearity through partition thresholds, and developing a dual-path channel-phase enhanced cooperative architecture that breaks through the dimensional limitations of traditional I/Q component analysis. This outcome provides a key technical pathway for developing polar digital RF transmitters, with a flexible configuration that enables rapid adaptation to diverse PA hardware.

Future work will develop a high-fidelity, system-level simulation platform in our subsequent research. This platform will explicitly model separate amplitude and phase processing paths to emulate nonideal characteristics of polar transmitters, thereby enabling a reliable and controllable environment for assessing the compatibility of PEAPGRU model. Then, we will focus on constructing an empirical mapping table between phase threshold, the characteristics of the PA, and the signal bandwidth, extending validation to millimeter-wave frequencies, and developing a multichannel joint optimization framework to meet the demands of RF systems operating at higher frequency bands and with larger bandwidths.

Acknowledgment

The authors would like to thank the referees for their constructive suggestions.

References

[1] R. Aruna, P. M. S, and K. Kulkarni, "A composite approach for a multi-layer perceptron implementation of digital predistortion in power amplifiers," in *Proc. Int. Conf. Intell. Innov. Technol. Comput., Electr. Electron. (IITCEE)*, Bangalore, India, Jan. 2024, pp. 1–5.

[2] F. Zhang, Y. Wang, and B. Ai, "Variable step-size MLMS algorithm for digital predistortion in wideband OFDM systems," *IEEE Trans. Consum. Electron.*, vol. 61, no. 1, pp. 10–15, Feb. 2015.

[3] A. Ali and O. Hammi, "Bandwidth, power and carrier configuration resilient neural networks digital predistorter," *IEEE Access*, vol. 11, pp. 63126–63135, 2023.

[4] R. N. Braithwaite, "Digital predistortion of an RF power amplifier using a reduced Volterra series model with a memory polynomial estimator," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3613–3623, Oct. 2017.

[5] K. Shi and G. Wang, "Application of improved AHTVSSLMS algorithm in digital predistortion system," in *Proc. 5th Int. Conf. Electron. Commun., Netw. Comput. Technol. (ECNCT)*, Aug. 2023, pp. 344–349.

[6] J. Lv, G. Xu, C. Zhang, H. Zhang, and T. Liu, "Full-wave digital predistortion linearization based on memory polynomials for HF power amplifiers," *IEEE Access*, vol. 12, pp. 29362–29372, 2024.

[7] Z. Fan, Z. Hao, B. Jin, C. Yu, and J. Cai, "Design of RF PAs using PSO algorithm with dynamic nonlinear self-adaptive hyperparameters," *IEEE Trans. Microw. Theory Techn.*, vol. 73, no. 4, pp. 2157–2169, Apr. 2025.

[8] A. Đorić, N. M. Ilić, A. Atanasković, and D. Budimir, "Analysis and application of DPD compensation technique and combined linearization approach on dual band PA for 5G signals," in *Proc. 4th Int. Conf. Artif. Intell. Signal Process. (AISP)*, Vijayawada, India, Oct. 2024, pp. 1–5.

[9] H. Yin and A. Zhu, "Pareto-optimal multimetric model extraction for digital predistortion of RF power amplifiers for error spectrum redistribution," *IEEE Trans. Microw. Theory Techn.*, vol. 73, no. 4, pp. 2230–2241, Apr. 2025.

[10] A. Fischer-Bühner et al., "Sparsely gated mixture of experts neural network for linearization of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 72, no. 7, pp. 4367–4382, Jul. 2024.

[11] J. Cai, C. Yu, L. Sun, S. Chen, and J. B. King, "Dynamic behavioral modeling of RF power amplifier based on time-delay support vector regression," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 2, pp. 533–543, Feb. 2019.

[12] Y. Wang et al., "Residual real-valued time-delay convolutional neural network for behavioral modeling and digital predistortion," in *Proc. Int. Conf. Microw. Millim. Wave Technol. (ICMMT)*, May 2024, pp. 1–3.

[13] P. Jaraut et al., "Augmented convolutional neural network for behavioral modeling and digital predistortion of concurrent multiband power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 9, pp. 4142–4156, Sep. 2021.

[14] R. Hongyo, Y. Egashira, T. M. Hone, and K. Yamaguchi, "Deep neural network-based digital predistorter for Doherty power amplifiers," *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 2, pp. 146–148, Feb. 2019.

[15] S. Yesil, C. Sen, and A. Ö. Yilmaz, "Experimental analysis and FPGA implementation of the real valued time delay neural network based digital predistortion," in *Proc. 26th IEEE Int. Conf. Electron., Circuits Syst. (ICECS)*, Genoa, Italy, Nov. 2019, pp. 614–617.

[16] Z. He, "Time-delay/advance neural networks based digital predistorters: Enabling high efficiency and high throughput transmitter," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2023, pp. 1–5.

[17] Y. Wei, S. Wang, F. Nait-Abdesselam, and A. Benlarbi-Delai, "Study of spiking neural network-based regressor on applications in digital predistortion for power amplifiers," in *Proc. 6th Int. Conf. Adv. Commun. Technol. Netw. (CommNet)*, Rabat, Morocco, Dec. 2023, pp. 1–7.

[18] S. Wang, P. M. Ferreira, and A. Benlarbi-Delai, "Physics informed spiking neural networks: Application to digital predistortion for power amplifier linearization," *IEEE Access*, vol. 11, pp. 48441–48453, 2023.

[19] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348–10356, Nov. 2019.

[20] Q. Zhang et al., "Adaptive inference pathway-gated neural network model for digital predistortion with varying transmission configurations," *IEEE Trans. Microw. Theory Techn.*, vol. 73, no. 1, pp. 436–447, Jan. 2025.

[21] A. Fischer-Bühner, L. Anttila, M. Turunen, M. Dev Gomony, and M. Valkama, "Augmented phase-normalized recurrent neural network for RF power amplifier linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 73, no. 1, pp. 412–422, Jan. 2025.

[22] X. Hu et al., "Convolutional neural network for behavioral modeling and predistortion of wideband power amplifiers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3923–3937, Aug. 2022.

[23] Y. Wu et al., "MP-DPD: Low-complexity mixed-precision neural networks for energy-efficient digital predistortion of wideband power amplifiers," *IEEE Microw. Wireless Technol. Lett.*, vol. 34, no. 6, pp. 817–820, Jun. 2024.
[24] D. Wang, M. Aziz, M. Helaoui, and F. M. Ghannouchi, "Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 242–254, Jan. 2019.
[25] L. A. M. Pereira, L. L. Mendes, C. J. A. Bastos-Filho, and S. Arismar Cerqueira, "Machine learning applied to 6G radio over fiber systems linearization," in *Proc. Int. Conf. Electr., Comput., Commun. Mechatronics Eng. (ICECCME)*, Maldives, Maldives, Nov. 2022, pp. 1–6.
[26] Y. Zhang et al., "5.6 a power-efficient CORDIC-less digital polar transmitter using 1b DSM-based PA supporting 256-QAM," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2025, pp. 100–103.
[27] J. Sun, X. Chen, and Q. Hou, "Radar object detection based on multiscale spatiotemporal features fusion and attention network," in *Proc. China Autom. Congr. (CAC)*, Xiamen, China, Nov. 2022, pp. 2973–2978.
[28] D. Lu, N. Cui, and C. Li, "A novel transfer learning framework combining attention mechanisms and random forest regression for state of health estimation of lithium-ion battery with different formulations," *IEEE Trans. Ind. Appl.*, vol. 60, no. 4, pp. 5726–5736, Jul. 2024.
[29] J. Feng, X. Tang, B. Zhang, and Y. Ren, "A lightweight deep learning RF fingerprint recognition method," in *Proc. 4th Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, 2022, pp. 452–457.
[30] H. Li, Y. Zhang, G. Li, and F. Liu, "Vector decomposed long short-term memory model for behavioral modeling and digital predistortion for wideband RF power amplifiers," *IEEE Access*, vol. 8, pp. 63780–63789, 2020.

Biao Long received the B.S. degree from Beijing Jiaotong University, Beijing, China, in 2019, and the M.S. degree from Hainan University, Haikou, China, in 2022, where he is currently pursuing the Ph.D. degree at the School of Information and Communication Engineering.

His current research interests include digital radio frequency transmitter architectures and power amplifier linearization.

Zhuhua Hu (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from Jilin University, Changchun, China, in 2002 and 2005, respectively, and the Ph.D. degree from Hainan University, Haikou, China, in 2019.

He was a Software Engineer at Ningbo BIRD Research Institute of China, Ningbo, China, from 2005 to 2006. He was a Software Engineer at Nanjing Research Institute of ZTE, Nanjing, China, from 2006 to 2007. He was a Minister of the Software Department at Shanghai Aoxun Information Technology Company Ltd., Shanghai, China, from 2007 to 2009. He is a Professor and a Doctorial Tutor at the School of Information and Communication Engineering, Hainan University since 2020. He is currently a high-level talent in Hainan Province. He led the "Multimodal information intelligent processing and decision control" innovation team, and has published more than 140 academic papers in journals such as TMM, TIP, TVT, TAES, TGRS, TITS, TCAD, OE, and COMPAG, authorized 19 patents, and hosted more than ten large-scale commercial projects that have been successfully implemented. He acted as the Reviewer for TIP, TGRS, T-ITS, TVT, TETC, Ocean Engineering, Measurement, KNOSYS, IMAVIS, ICASSP2023&2024, and ICME 2024. His current research interests include artificial intelligence, signal and information processing.

Dr. Hu is a Senior Member of CCF.

Dake Liu (Senior Member, IEEE) received the Ph.D. degree (Technology Doctor) from Linköping University, Linköping, Sweden, in 1995.

He has experience in the design of communication systems and application-specific instruction set processors (ASIP). He is currently a Professor and the Academy Chairperson with the School of Information and Communication Engineering, Hainan University, Haikou, China. He is the Founder and the CTO at Freehand DSP AB Ltd., Stockholm, Sweden, acquired by VIA, and the Founder at Coresonic AB, Linköping, acquired by MediaTek. He is a China Nation Expert. He has authored over 200 articles. His current research interests include 5G/6G baseband processors.