

UNIVERSIDADE FEDERAL DO PARANÁ – UFPR
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

MATHEUS HENRIQUE SILVEIRA SANTANA

PRODUTO ENTRE PERCEPTRON DE CAMADAS MÚLTIPLAS E FILTRO DE
RESPOSTA AO IMPULSO FINITA APLICADO NA MODELAGEM
COMPORTAMENTAL DE AMPLIFICADORES DE POTÊNCIA

CURITIBA

2025

MATHEUS HENRIQUE SILVEIRA SANTANA

PRODUTO ENTRE PERCEPTRON DE CAMADAS MÚLTIPLAS E FILTRO DE
RESPOSTA AO IMPULSO FINITA APLICADO NA MODELAGEM
COMPORTAMENTAL DE AMPLIFICADORES DE POTÊNCIA

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia Elétrica,
Departamento de Engenharia Elétrica, Setor de
Tecnologia, Universidade Federal do Paraná,
como requisito parcial à obtenção do título de
Bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Eduardo Gonçalves de Lima

CURITIBA

2025

RESUMO

A modelagem de amplificadores de potência (PAs) é essencial na concepção e otimização de sistemas de comunicação sem fio. Eles são componentes fundamentais nesses sistemas, pois são responsáveis por aumentar o sinal a um nível que pode ser transmitido a longas distâncias. No entanto, os PAs são dispositivos altamente não lineares e o seu comportamento pode ser afetado por uma série de fatores. Por isso, são necessários modelos precisos dos amplificadores de potência para prever o seu comportamento e otimizar o desempenho do sistema. Entre esses métodos, existem modelos polinomiais baseados em filtros digitais de resposta ao impulso finita (FIR) e modelos baseados em redes neurais. Com base nisso, a partir de um determinado modelo polinomial, uma nova topologia foi criada para substituir parte do modelo por uma rede neural do tipo perceptron de três camadas (TLP). Obteve-se um Erro Quadrático Médio Normalizado (NMSE) de -43,8815 dB para o modelo polinomial com $P=5$ (66 coeficientes), enquanto que para o modelo com a rede neural obteve-se um NMSE de -44,0168 dB para $N=10$ neurônios na camada oculta (68 coeficientes). Isso evidencia que a substituição parcial por uma rede neural pode aumentar a precisão da modelagem sem exigir um aumento significativo no número de coeficientes.

Palavras-chave: Amplificadores de potência. Redes Neurais. Otimização não-linear

ABSTRACT

Behavioral modeling of power amplifiers (PAs) used in wireless communication systems is fundamental to ensure signal integrity and overall system efficiency. Among the traditional approaches, polynomial models with memory, such as those based on finite impulse response (FIR) digital filters, are widely employed to capture nonlinearities and memory effects inherent to these devices. Recently, neural network-based methods have emerged as promising alternatives due to their ability to represent highly nonlinear relationships. This work proposes a hybrid structure that combines a polynomial model with a three-layer perceptron (TLP) neural network, replacing part of the traditional model with a learning-based neural architecture. The performance evaluation was validated using the Normalized Mean Square Error (NMSE) metric. Results show that the hybrid model outperformed the conventional one: while the polynomial model with order $P=5$ and 66 coefficients yielded an NMSE of -43.8815 dB, the hybrid model with 10 hidden layer neurons and 68 coefficients achieved an NMSE of -44.0168 dB. These results highlight the potential of the hybrid approach for future applications in transmission systems requiring high accuracy and linearity. This shows that the partial replacement with a neural network can improve modeling accuracy without requiring a significant increase in the number of coefficients.

Keywords: Power amplifiers. Neural networks. Behavioral modeling. Nonlinear optimization.

LISTA DE FIGURAS

Figura 1 - Diagrama de blocos de uma rede de transmissão sem fio	10
Figura 2 – Características gerais de um amplificador de potência	11
Figura 3 - Diagrama em blocos da linearização de amplificadores de potência. (a) Feedforward; (b) Feedback; (c) Pré-distorção analógica.	14
Figura 4 - Diagrama de Blocos do Modelo Polinomial.....	20
Figura 5 - Neurônio biológico comparado a uma rede neural artificial (a) Neurônio biológico, (b) Neurônio artificial, (c) Neurônio e sinapse, (d) Rede neural artificial. ..	21
Figura 6 - Diagrama de Blocos do Modelo Proposto	23
Figura 7 - Diagrama de uma rede neural TLP	24
Figura 8 - NMSE em função do número de coeficientes	28
Figura 9 - Forma de onda de amplitude de saída.....	28
Figura 10 - Curva AM-AM.....	29
Figura 11 - Curva AM-PM.....	30
Figura 12 – Comparação dos erros entre os modelos	31

LISTA DE TABELAS

Tabela 1: Resultados para o modelo polinomial.....	27
Tabela 2: Resultados para o modelo proposto.....	27

LISTA DE ABREVIATURAS E SIGLAS

PA – AMPLIFICADOR DE POTÊNCIA

NMSE – NORMALIZED MEAN SQUARE ERROR

RF – RADIOFREQUÊNCIA

FIR – FINITE IMPULSE RESPONSE

TLP – THREE LAYER PERCEPTRON

MLP – MULTI LAYER PERCEPTRON

CNN – CONVOLUTIONAL NEURAL NETWORK

RNA – REDE NEURAL ARTIFICIAL

MP – MEMORY POLYNOMIAL

GMP – GENERALIZED MEMORY POLYNOMIAL

EMP – ENVELOPPE MEMORY POLYNOMIAL

LUT – LOOKUP TABLE

DPD – DIGITAL PRE-DISTORTION

HDL – HARDWARE DESCRIPTION LANGUAGE

OFDM – ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING

SUMÁRIO

1. INTRODUÇÃO	8
1.1 OBJETIVOS	8
1.2 ESTRUTURA DO TRABALHO	9
2 REVISÃO DA LITERATURA	10
2.1 AMPLIFICADORES DE POTÊNCIA	10
2.2 MODELAGEM COMPORTAMENTAL DE AMPLIFICADORES DE POTÊNCIA	12
2.3 MÉTODO DOS MÍNIMOS QUADRADOS	15
2.4 MODELO POLINOMIAL	18
2.4 REDES NEURAIS	20
2.4 PERCEPTRON DE TRÊS CAMADAS	22
3. DESENVOLVIMENTO TEÓRICO	23
3.1 MODELO PROPOSTO	23
4. RESULTADOS E DISCUSSÃO	27
5. CONCLUSÕES FINAIS	32
6. BIBLIOGRAFIA	33

1. INTRODUÇÃO

Os amplificadores de potência (PAs) são dispositivos eletrônicos concebidos para aumentar o nível de potência de um sinal. São normalmente utilizados numa vasta gama de aplicações, desde a amplificação áudio até sistemas de comunicação por radiofrequência (RF). Essencialmente, um amplificador de potência recebe um sinal de entrada de baixa potência e amplifica-o para um nível de potência mais elevado, adequado para acionar um alto-falante, transmitir um sinal a longa distância ou fornecer a potência de saída necessária.

A modelagem comportamental dos PAs é uma tarefa essencial na concepção e otimização dos sistemas de comunicação modernos. Uma abordagem comum à modelagem de amplificadores de potência é a utilização de modelos matemáticos que captam o comportamento de entrada-saída do amplificador.

Esta pesquisa começa por apresentar uma visão geral da modelagem comportamental de um amplificador de potência baseada em filtros de resposta ao impulso finita (FIR) e funções unidimensionais [1]. Em seguida, apresenta uma metodologia para combinar uma Rede Neural Three-Layer-Perceptron (TLP) e filtros FIR e avalia o desempenho do modelo proposto usando simulação com dados experimentais [2].

Uma das principais vantagens da utilização de redes neurais em vez de modelos polinomiais para a modelagem de PAs é a capacidade das redes neurais em captar relações não lineares complexas que podem ser difíceis de modelar com polinômios de ordem elevada. No entanto, é importante enfatizar que o treinamento de redes neurais pode exigir uma quantidade significativa de tempo.

1.1 OBJETIVOS

Os resultados fundamentais a serem atingidos neste estudo incluem:

a) Desenvolvimento de um modelo mais preciso de um amplificador de potência:

O objetivo principal é a criação de modelos que possam capturar de forma eficiente as não linearidades do PA. Espera-se que o modelo híbrido que combina redes neurais com filtros FIR ofereça melhor desempenho do que o modelo polinomial tradicional.

b) Redução do erro de modelagem:

Através da comparação do Erro Quadrático Médio Normalizado (NMSE) entre os modelos, espera-se alcançar um NMSE mais baixo para o modelo híbrido, indicando uma maior precisão na modelagem do PA. Além disso, será importante verificar a relação entre o número de coeficientes e a precisão do modelo.

c) Maior eficiência computacional e redução no número de coeficientes:

Um resultado desejado é que o modelo híbrido apresente uma significativa redução no número de coeficientes necessários para descrever o comportamento do PA, em comparação com o modelo polinomial. Isso pode resultar em maior eficiência computacional, essencial para aplicações em tempo real.

1.2 ESTRUTURA DO TRABALHO

Este trabalho está organizado em cinco capítulos principais de forma a apresentar, desenvolver e discutir a proposta de modelagem comportamental de amplificadores de potência utilizando uma combinação entre Perceptron de Camadas Múltiplas (MLP) e Filtros de Resposta ao Impulso Finita (FIR).

No primeiro capítulo, são apresentados o contexto geral da pesquisa, a motivação que impulsiona o estudo e os objetivos propostos. Em seguida, o segundo capítulo reúne os fundamentos teóricos necessários, abordando conceitos relacionados aos amplificadores de potência, técnicas de modelagem comportamental, o método dos mínimos quadrados, modelos polinomiais, redes neurais artificiais e a estrutura do perceptron de três camadas.

O terceiro capítulo se dedica à formulação teórica da proposta, detalhando o modelo desenvolvido, que integra técnicas de redes neurais e filtros digitais. No quarto capítulo, são apresentados os resultados obtidos a partir da implementação do modelo proposto, juntamente com análises comparativas avaliando o desempenho e a eficácia da abordagem.

Por fim, o quinto capítulo traz as conclusões finais, destacando os principais achados do trabalho, suas contribuições para a área de telecomunicações e sugestões para possíveis estudos futuros.

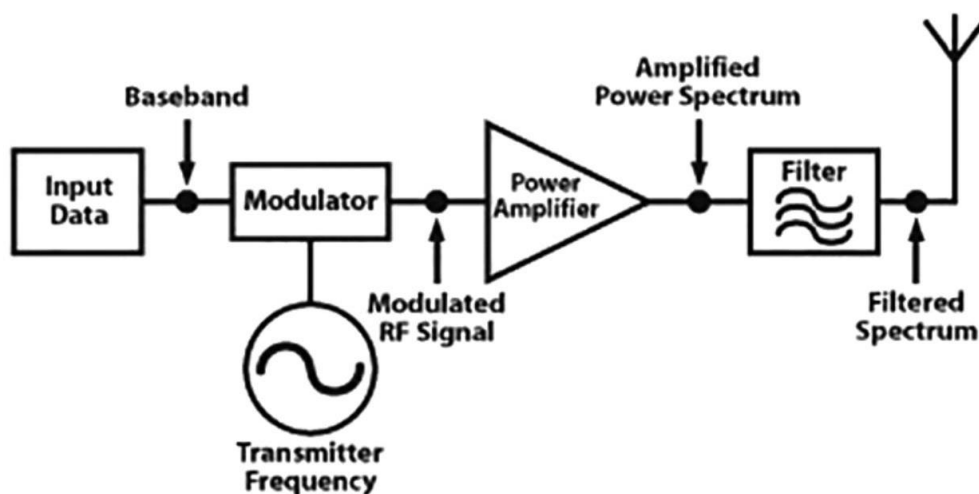
2 REVISÃO DA LITERATURA

2.1 AMPLIFICADORES DE POTÊNCIA

Os amplificadores de potência (PAs, do inglês *Power Amplifiers*) são dispositivos fundamentais em sistemas de comunicação analógica e digital, especialmente na transmissão dos sinais. Sua principal função é elevar o nível de potência do sinal modulado, tornando-o capaz de ser propagado eficientemente por meio de antenas e atingir distâncias maiores sem perda significativa de qualidade [3].

A Figura 1 apresenta um diagrama de blocos típico de uma rede de transmissão sem fio, onde o amplificador de potência aparece como um dos blocos do canal de transmissão. Ela ilustra o papel do PA ao receber um sinal modulado, gerado e processado por blocos anteriores como modulador, misturador e filtro, e amplificá-lo antes que seja enviado à antena para propagação.

Figura 1 - Diagrama de blocos de uma rede de transmissão sem fio



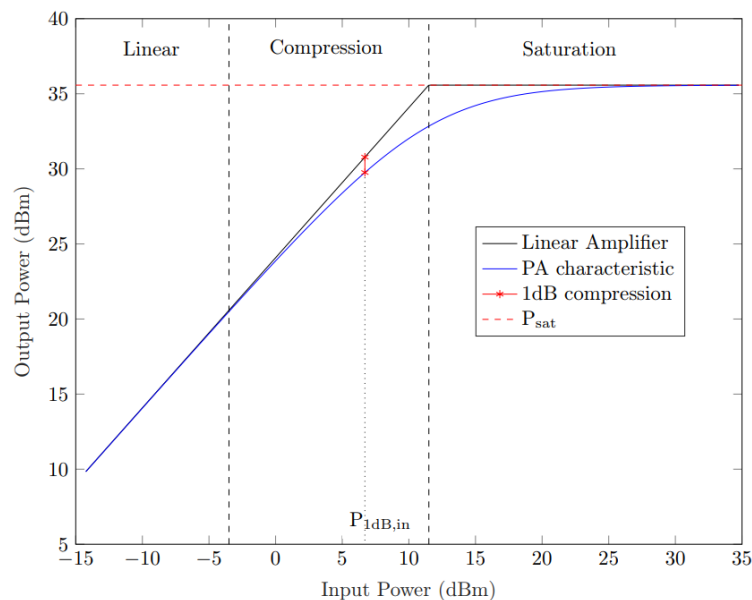
FONTE: AnalogICTips. Disponível em: <https://www.analogictips.com/rf-power-amplifier-part-1-functions-faq/>. Acesso em: 17/05/2025

Amplificadores de potência não se comportam de forma perfeitamente linear, pois à medida que o sinal de entrada aumenta, a saída do amplificador nem sempre cresce na mesma proporção. Classicamente, o PA costuma operar em três regiões:

1. Região Linear: a saída é proporcional à entrada, ou seja, se o sinal de entrada aumenta, a saída aumenta na mesma proporção. Esta é a região segura para operação, com pouca ou nenhuma distorção.
2. Região de Compressão: quando a entrada começa a ficar forte demais, o PA não consegue mais amplificar o sinal com o mesmo ganho. O ganho começa a diminuir mesmo com o aumento da entrada. Isso gera distorções, e um ponto de referência importante aqui é o Ponto de Compressão de 1 dB ($P_{1\text{ dB}}$), que corresponde ao nível de entrada no qual o ganho cai 1 dB em relação ao valor ideal.
3. Região de Saturação: ao aumentar ainda mais o sinal de entrada, o PA chega ao seu limite físico. A saída para de crescer mesmo com mais potência na entrada. É a região com mais distorção e, portanto, a menos desejada.

As regiões apresentadas anteriormente são mostradas na Figura 2.

Figura 2 – Características gerais de um amplificador de potência



FONTE: Falempin, 2002

A operação dos PAs ocorre frequentemente em regimes de alta potência, próximos da saturação, com o objetivo de maximizar a eficiência energética do sistema. Porém, operar nesta zona de não linearidades, afeta tanto a amplitude quanto a fase do sinal de saída, em relação ao sinal de entrada.

Essas distorções são descritas, respectivamente, como AM/AM (Amplitude-to-Amplitude) e AM/PM (Amplitude-to-Phase), e resultam em degradação da qualidade do sinal transmitido, aumento da interferência em canais adjacentes e violação de requisitos de espectro definidos por normas [4].

Além das distorções puramente estáticas, os amplificadores também apresentam efeitos de memória, ou seja, a resposta do sistema em um dado instante depende não apenas do valor atual do sinal de entrada, mas também de valores passados. Esses efeitos estão relacionados à presença de elementos reativos no circuito (capacitâncias e indutâncias), variações térmicas e características do dispositivo ativo [5].

Deste modo, a modelagem e compensação desses efeitos se faz necessária em aplicações modernas de comunicação, como 4G, 5G e Wi-Fi, que utilizam modulações complexas (OFDM) e exigem alta linearidade espectral. Sem essa compensação, a fidelidade do sinal e a eficiência espectral do sistema fica comprometida, dificultando o cumprimento de padrões regulatórios e a convivência com outros sistemas no espectro de frequências [6].

2.2 MODELAGEM COMPORTAMENTAL DE AMPLIFICADORES DE POTÊNCIA

Segundo [7], três principais técnicas de linearização de amplificadores de potência são amplamente estudadas: feedforward, feedback e pré-distorção, sendo esta última a mais utilizada atualmente. Estas três técnicas estão ilustradas de maneira simplificada na Figura 3.

A técnica feedforward foi uma das primeiras propostas na literatura, ainda na década de 1920, como forma de compensar externamente as distorções introduzidas pelo PA. Seu funcionamento baseia-se na extração do erro de distorção, obtido por meio da comparação entre o sinal de entrada original e uma versão atenuada da saída amplificada. Esse erro, que representa as não linearidades introduzidas, é então amplificado por um caminho auxiliar linear e reinjetado, com fase invertida, na saída do PA. O resultado é uma tentativa de cancelamento ativo da distorção. Apesar de esta abordagem ter uma boa capacidade de linearização, sua implementação prática exige uma sincronização precisa de atrasos e ganhos entre os dois caminhos de sinal,

o que aumenta a complexidade. Além disso, o fato de serem necessários dois amplificadores (um principal e outro auxiliar) impacta negativamente a eficiência energética do sistema [8].

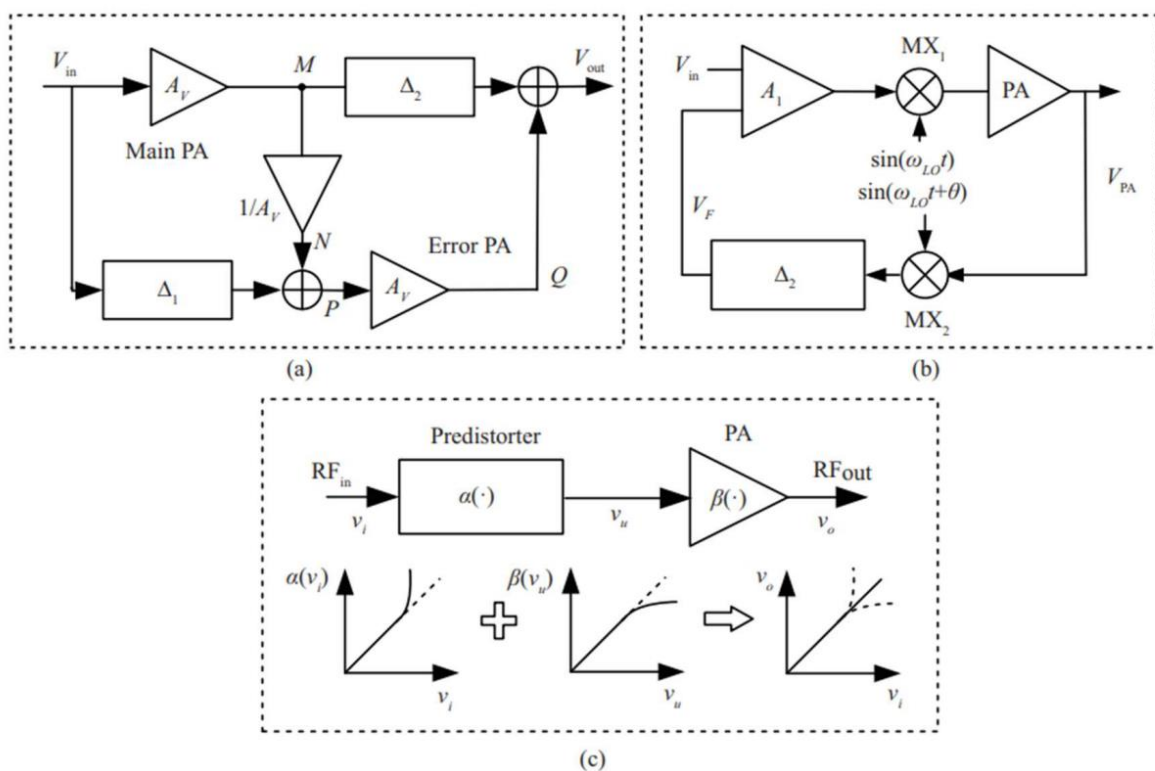
A técnica de feedback surgiu como um aprimoramento conceitual da abordagem anterior. Inspirada nos princípios dos sistemas de controle, essa estratégia utiliza um laço de realimentação no qual a saída do PA é constantemente comparada com a entrada. A diferença entre os dois sinais, interpretada como erro, é usada para ajustar dinamicamente o sinal de entrada de modo a reduzir esse erro. Deste modo, o sistema busca se autorregular e manter uma resposta mais próxima da linearidade. Este feedback pode ser implementado em diferentes domínios (cartesiano ou polar), dependendo da arquitetura envolvida. Contudo, sua eficácia é afetada com a presença de atrasos na malha de realimentação, o que se torna crítico em aplicações de banda larga, onde os sinais possuem altas frequências e grande largura de banda. Esses atrasos dificultam a atuação para corrigir distorções rápidas, reduzindo a aplicabilidade do método em sistemas modernos de comunicação [9].

A abordagem de pré-distorção é atualmente a mais amplamente adotada na indústria de telecomunicações, especialmente em sistemas que exigem alta eficiência espectral e linearidade, como os padrões LTE, 5G e Wi-Fi. Diferente das abordagens anteriores, a pré-distorção atua antes do amplificador de potência, adicionando ao sinal de entrada uma forma específica de distorção que é intencional e calculada. O objetivo é que essa distorção pré-aplicada seja anulada pela não linearidade do PA, de forma que o sinal final na saída do sistema (pré-distorção + PA) seja linear. Para isso, é necessário um modelo que represente com precisão o comportamento inverso do PA. Com o avanço do processamento digital de sinais, essa técnica evoluiu para a chamada *Digital Pre-Distortion* (DPD), em que algoritmos baseados em modelos polinomiais, estruturas com memória e até redes neurais são usados para ajustar dinamicamente os coeficientes de pré-distorção em tempo real. Apesar de sua flexibilidade e alto desempenho, esta técnica apresenta desafios relacionados à modelagem precisa do PA, ao consumo computacional e à necessidade de adaptação constante frente a fatores como temperatura e variações elétricas [7], [10].

De maneira geral, a eficácia dessas técnicas depende diretamente da qualidade dos modelos utilizados para representar o comportamento do PA. Deste

modo, a modelagem comportamental tem se mostrado como abordagem eficiente e prática nesse contexto. Contrariamente à modelagem física, que exige conhecimento detalhado da estrutura interna do dispositivo e alto custo computacional, a modelagem comportamental busca descrever a relação entre o sinal de entrada e de saída com base em observações experimentais, através de estruturas matemáticas adequadas [11].

Figura 3 - Diagrama em blocos da linearização de amplificadores de potência. (a) Feedforward; (b) Feedback; (c) Pré-distorção analógica.



FONTE: UROOJ, A.; AL ABSI, M. A (2022)

Diversos modelos comportamentais clássicos são empregados na representação de sistemas não lineares, especialmente naqueles com características de memória. Um dos modelos mais tradicionais é o modelo de Volterra, que utiliza uma expansão funcional composta por séries de integrais convolucionais para descrever o comportamento dinâmico de sistemas não lineares com memória. Apesar de sua generalidade e fundamentação teórica sólida, sua aplicação prática é um pouco limitada devido à complexidade computacional, que cresce com a ordem do sistema e a profundidade da memória.

Por outro lado, os modelos polinomiais com memória, como o *Memory Polynomial* (MP) e o *Generalized Memory Polynomial* (GMP), surgem como alternativas mais viáveis. Além disso, abordagens baseadas em tabelas de consulta, conhecidas como *Look-Up Tables* (LUTs), em conjunto com filtros digitais do tipo FIR (Finite Impulse Response), também podem ser utilizadas. As LUTs são eficazes na representação de não linearidades estáticas por meio de mapeamentos diretos de entrada-saída, enquanto os filtros FIR contribuem para a caracterização de comportamentos dinâmicos, oferecendo uma solução com boa eficiência computacional, especialmente em sistemas embarcados ou de tempo real.

Entretanto, mesmo os modelos polinomiais podem apresentar limitações quando confrontados com comportamentos fortemente não lineares e efeitos de memória de longo alcance [1]. Para contornar esses desafios, modelos híbridos são explorados exhaustivamente na literatura, seja combinando estruturas clássicas ou combinando estas estruturas clássicas com métodos de aprendizado de máquina, como redes neurais do tipo perceptron de múltiplas camadas (*Multi-Layer Perceptron* – MLP) ou redes neurais convolucionais (*Convolutional Neural Network* - CNN). Estes modelos aproveitam a capacidade de generalização das redes neurais e a representação temporal dos filtros FIR, permitindo obter modelos mais eficientes.

Essa abordagem híbrida, foco deste trabalho, busca aliar precisão de modelagem, eficiência computacional e potencial de aplicação em tempo real, sendo altamente promissora para futuras aplicações em sistemas de transmissão com requisitos cada vez mais exigentes.

2.3 MÉTODO DOS MÍNIMOS QUADRADOS

O método dos mínimos quadrados serve para encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados. Este método só pode ser utilizado em modelos que sejam lineares nos parâmetros, ou seja, as variáveis apresentam uma relação linear entre si. Caso contrário, deveria ser usado um modelo de regressão não-linear.

Considerando um vetor de entradas $X = [x(1) \dots x(n)]^T$ e um vetor de saídas $Y = [y(1) \dots y(N)]^T$, deseja-se encontrar uma relação entre estas entradas e saídas para cada amostra $1 \leq n \leq N$. Um exemplo de modelo é o polinômio com memória dado pela expressão

$$y(n) = \sum_{p=1}^P \sum_{m=0}^M h_{pm} [x(n-m)]^p, \quad (1)$$

onde P é a ordem do polinômio, M é a profundidade da memória e h_{pm} são coeficientes do modelo. Como é um modelo linear nos coeficientes, é possível encontrar os coeficientes que minimizam o erro através do método dos mínimos quadrados.

Expandido o somatório tem-se:

$$\begin{aligned} y(n) = & h_{10}[x(n)]^1 + h_{11}[x(n-1)]^1 + h_{12}[x(n-2)]^1 + \dots + h_{1M}[x(n-M)]^1 + \\ & + h_{20}[x(n)]^2 + h_{21}[x(n-1)]^2 + h_{22}[x(n-2)]^2 + \dots + h_{2M}[x(n-M)]^2 + \dots + \\ & + h_{P0}[x(n)]^P + h_{P1}[x(n-1)]^P + h_{P2}[x(n-2)]^P + \dots + h_{PM}[x(n-M)]^P. \end{aligned} \quad (2)$$

Os coeficientes h_{pm} podem ser associados em um único vetor, bem como os polinômios $[x(n-m)]^p$ conforme mostrado a seguir:

$$H^T = [h_{10} \ h_{11} \ h_{12} \ \dots \ h_{1M} \ h_{20} \ h_{21} \ h_{22} \ \dots \ h_{2M} \ \dots \ h_{P0} \ h_{P1} \ h_{P2} \ \dots \ h_{PM}], \quad (3)$$

$$X(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ x(n-2) \\ \vdots \\ x(n-M) \\ x^2(n) \\ x^2(n-1) \\ x^2(n-2) \\ \vdots \\ x^2(n-M) \\ \vdots \\ x^P(n) \\ x^P(n-1) \\ x^P(n-2) \\ \vdots \\ x^P(n-M) \end{bmatrix}. \quad (4)$$

Deste modo, o modelo pode ser reescrito como:

$$Y(n) = X^T(n)H. \quad (5)$$

Pode-se então, para um conjunto com N amostras montar uma matriz de regressão XX que segue a seguinte lei de formação:

$$\begin{bmatrix} x(M+1) & x(M) & \dots & x(1) & x^2(M+1) & x^2(M) & \dots & x^2(1) & x^P(M+1) & x^P(M) & \dots & x^P(1) \\ x(M+2) & x(M+1) & \dots & x(2) & x^2(M+2) & x^2(M+1) & \dots & x^2(2) & x^P(M+2) & x^P(M+1) & \dots & x^P(2) \\ x(M+3) & x(M+2) & \dots & x(3) & x^2(M+3) & x^2(M+2) & \dots & x^2(3) & x^P(M+3) & x^P(M+2) & \dots & x^P(3) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x(N) & x(N-1) & \dots & x(N-M) & x^2(N) & x^2(N-1) & \dots & x^2(N-M) & x^P(N) & x^P(N-1) & \dots & x^P(N-M) \end{bmatrix}. \quad (6)$$

Para uma dada memória M o menor valor de amostra que podemos colocar na matriz de regressão é $n = M + 1$, pois deste modo é garantido que haja valores para as M amostras anteriores. Assim, o vetor das saídas deve iniciar também em $n = M + 1$:

$$Y' = [y(M+1) \quad y(M+2) \quad \dots \quad y(N-1) \quad y(N)]. \quad (7)$$

Temos então que

$$XX_{(N-M)x(P(M+1))} \cdot H_{(P(M+1))x(1)} = Y'_{(N-M)x(1)}, \quad (8)$$

onde axb subscrito indica a ordem da matriz. O vetor H com os coeficientes que minimizam o erro pode ser calculado por:

$$H = (XX^T XX) \cdot (XX^T Y'). \quad (9)$$

Esta operação pode ser feita com o auxílio do Matlab utilizando o comando “\”. Neste caso temos $H = XX \backslash Y$.

2.4 MODELO POLINOMIAL

Há vários modelos polinomiais propostos na literatura [1], [12], [13], [14]. Entre estas abordagens, existem duas topologias capazes de modelar amplificadores de potência de forma eficiente. Essas topologias são conhecidas como *Memory Polynomial* (MP) e *Envelope Memory Polynomial* (EMP). Nas seguintes equações, P é a ordem do polinômio, M é a profundidade da memória e a e b são coeficientes ajustáveis.

Em um modelo MP [9], o sinal de entrada é processado por uma rede de elementos polinomiais considerando apenas produtos de sinais avaliados no mesmo instante de tempo e pode ser representado por

$$y(n) = \left[\sum_{m=0}^M \sum_{p=1}^P b_{pm} x(n-m) |x(n-m)|^{p-1} \right]. \quad (10)$$

A saída do modelo é obtida combinando amostras passadas do sinal de entrada com suas respectivas magnitudes elevadas a potências, ponderadas por coeficientes.

No modelo EMP [12], a ideia é que a rede polinomial seja multiplicada por uma função de envoltória, concebida para modelar o comportamento da amplitude do sistema. Este modelo pode ser representado por

$$y(n) = x(n) \left[\sum_{m=0}^M \sum_{p=1}^P b_{pm} |x(n-m)|^{p-1} \right], \quad (11)$$

e utiliza apenas a magnitude das amostras passadas do sinal de entrada, ignorando informações de fase anteriores. O modelo calcula um ganho complexo com base nessas magnitudes e o aplica à amostra atual do sinal de entrada para estimar a saída. Deste modo, o EMP depende da envoltória complexa atual combinada com a história de magnitudes passadas, resultando em uma estrutura mais simples e eficiente para certas aplicações.

Os modelos MP e EMP são particularmente úteis porque são flexíveis e podem capturar uma ampla gama de comportamentos não lineares. Como forma de combinar os efeitos destes dois modelos, o seguinte modelo foi proposto em [1-]:

$$y(n) = \left[\sum_{m=0}^M a_{m+1} x(n-m) \right] \cdot \left[\sum_{m=0}^M \sum_{p=1}^P b_{pm} |x(n-m)|^{p-1} \right]. \quad (12)$$

Este modelo não é linear nos seus coeficientes, uma vez que existe uma multiplicação entre os coeficientes a e b . No entanto, fixando os coeficientes a , este modelo torna-se linear, permitindo encontrar os coeficientes b pelo método dos mínimos quadrados. Para ajustar os coeficientes a , pode ser utilizada uma otimização não linear.

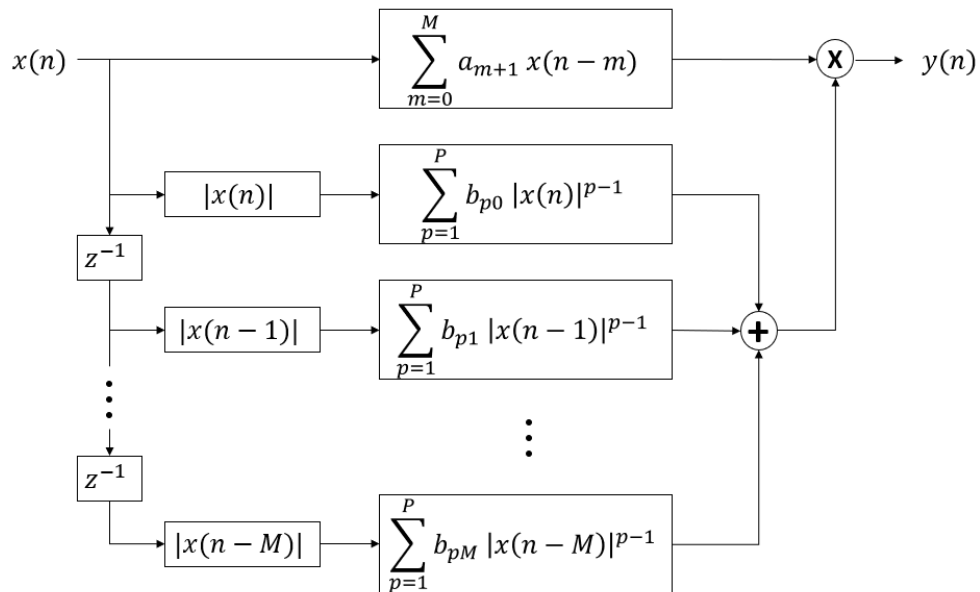
É possível descrever o comportamento deste modelo como um produto de filtros digitais de resposta ao impulso finita e funções unidimensionais. Um filtro FIR é um filtro digital de processamento de sinais que opera num número finito de amostras de entrada para produzir um número finito correspondente de amostras de saída [10]. A equação de um filtro FIR é:

$$y_{FIR}(n) = \sum_{i=0}^{N_f} a_i x_{FIR}(n-i), \quad (13)$$

onde N_f é a ordem do filtro, i é o número de amostras disponíveis, a_i são os coeficientes do filtro e $x_{FIR}(n)$ e $y_{FIR}(n)$ são os sinais de entrada e saída do filtro. Uma das vantagens de um filtro FIR é a sua resposta de fase linear, o que significa que não introduz qualquer distorção de fase no sinal filtrado.

A Figura 4 mostra o diagrama de blocos do modelo polinomial. Há um único filtro FIR que será multiplicado pela soma de funções unidimensionais $f(|x(n-m)|)$, com $0 \leq m \leq M$. O bloco z^{-1} indica a mudança para uma amostra anterior.

Figura 4 - Diagrama de Blocos do Modelo Polinomial



FONTE: O autor, 2025.

2.4 REDES NEURAIS

As redes neurais artificiais (RNAs) são modelos matemáticos inspirados na estrutura e no funcionamento do cérebro humano. Elas foram concebidas a partir da analogia com os neurônios biológicos, unidades fundamentais do sistema nervoso que recebem sinais pelos dendritos, integram essa informação no corpo celular, e transmitem um impulso elétrico ao longo do axônio, caso a intensidade do sinal ultrapasse um certo limiar. A comunicação entre neurônios ocorre nas sinapses, onde o sinal pode ser reforçado ou inibido.

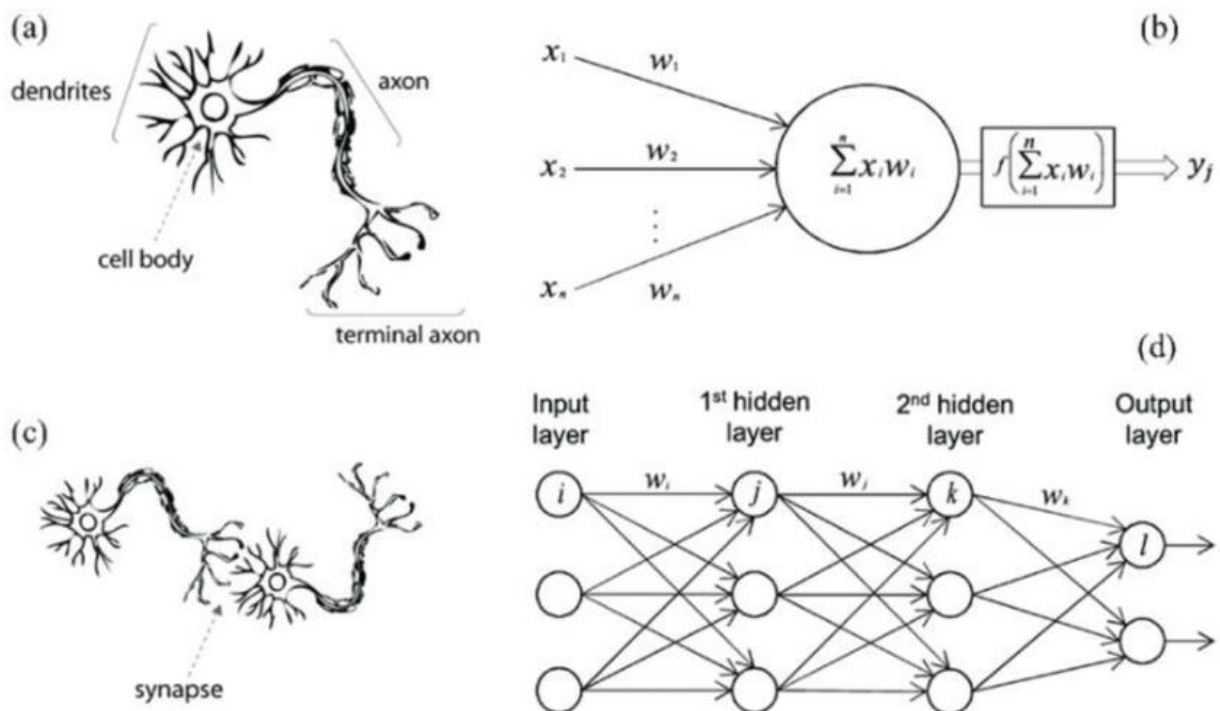
Inspirando-se nesse funcionamento, McCulloch e Pitts propuseram em 1943 o primeiro modelo matemático de um neurônio formal. Esse modelo simplificado recebe diversas entradas numéricas associadas a pesos, realiza uma soma ponderada dessas entradas, subtrai um valor chamado *bias* e aplica uma função de ativação não linear que determina a saída do neurônio. Assim como no cérebro, essa ativação representa uma decisão: transmitir ou não transmitir o sinal. Uma comparação entre o neurônio biológico e o neurônio artificial é apresentada na Figura 5.

As redes neurais artificiais consistem na interconexão de vários desses neurônios organizados em camadas. Cada conexão possui um peso ajustável, e a

rede aprende por meio de um processo de ajuste desses pesos, com base em exemplos de entrada e saída, o que permite capturar relações complexas e não lineares entre dados [15*].

Devido à sua alta capacidade de adaptação e generalização, as redes neurais têm sido amplamente empregadas em diferentes áreas. Entre as aplicações mais conhecidas, destacam-se o reconhecimento de voz, imagem e escrita, além de previsão de séries temporais, diagnóstico médico, sistemas de controle e classificação de dados. No contexto da engenharia e das telecomunicações, as RNAs também têm se mostrado eficazes na modelagem comportamental de sistemas físicos, como amplificadores de potência, onde sua habilidade de capturar características não lineares e dependentes do histórico do sinal é particularmente interessante.

Figura 5 - Neurônio biológico comparado a uma rede neural artificial (a) Neurônio biológico, (b) Neurônio artificial, (c) Neurônio e sinapse, (d) Rede neural artificial.



FONTE: Suzuki, 2013.

2.4 PERCEPTRON DE TRÊS CAMADAS

Uma rede neural artificial do tipo *Three-Layer-Perceptron* é composta por três camadas de neurônios interligadas: uma camada de entrada, uma camada oculta e uma camada de saída. Cada neurônio de uma determinada camada está ligado a todos os outros neurônios da camada seguinte e a cada uma destas ligações está associado um peso w [1].

Os neurônios da camada oculta e da camada de saída recebem como entrada a soma ponderada dos valores de saída dos neurônios da camada anterior. A saída é produzida pela passagem desses valores de entrada por uma função de ativação $\varphi(x)$. Alguns exemplos muito recorrentes são a função sigmoide hiperbólica

$$\varphi(x) = \frac{2}{1 + e^{-2x}} - 1, \quad (14)$$

que gera saídas suavemente limitadas no intervalo $[-1,1]$, e a função gaussiana

$$\varphi(x) = e^{-x^2}, \quad (15)$$

comum em redes de base radial, que responde com mais intensidade quando o sinal de entrada está próximo do centro do neurônio e decai rapidamente à medida que se afasta. São estas funções que introduzem a não linearidade essencial para que a rede consiga modelar comportamentos não lineares.

Além disso, cada neurônio carrega uma constante chamada *bias*, que é adicionada à soma ponderada antes de passar pela função de ativação. A saída v_k de um neurônio na k -ésima posição de uma determinada camada pode ser escrita matematicamente por:

$$v_k = \varphi \left(\sum_{i=1}^{N_p} w_{ik} z_i + b_k \right), \quad (16)$$

em que z_i é a saída e N_p é o número de neurônios da camada anterior.

A camada de saída geralmente utiliza uma função de transferência linear, nesse caso, o sinal de saída é obtido apenas como uma combinação linear dos sinais provenientes das funções de ativação e dos *biases*.

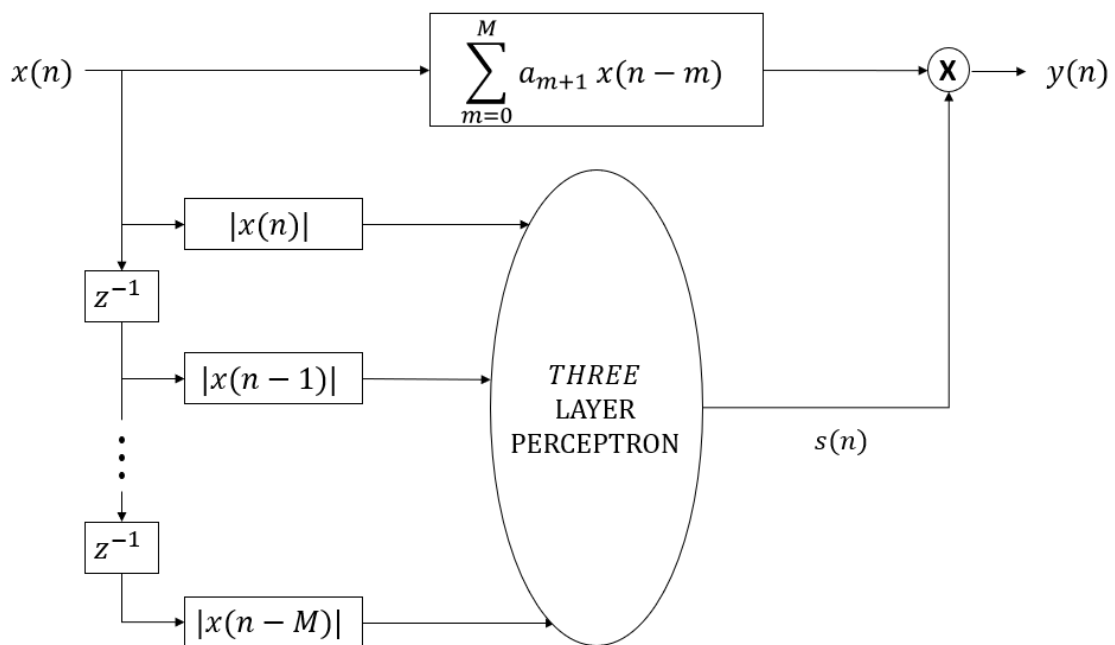
Na literatura, existem abordagens que utilizam exclusivamente redes neurais do tipo perceptron para modelagem e controle de sistemas similares [11], [16], [17]. Contudo, o objetivo deste trabalho é propor uma topologia híbrida que combine um perceptron de três camadas (TLP) com filtros digitais, aproveitando as vantagens de ambos para aprimorar a modelagem dos amplificadores de potência.

3. DESENVOLVIMENTO TEÓRICO

3.1 MODELO PROPOSTO

A ideia principal deste trabalho é, a partir do modelo polinomial, substituir o conjunto de $M + 1$ polinômios por uma rede neural do tipo TLP, como mostra o diagrama da Fig. 6.

Figura 6 - Diagrama de Blocos do Modelo Proposto

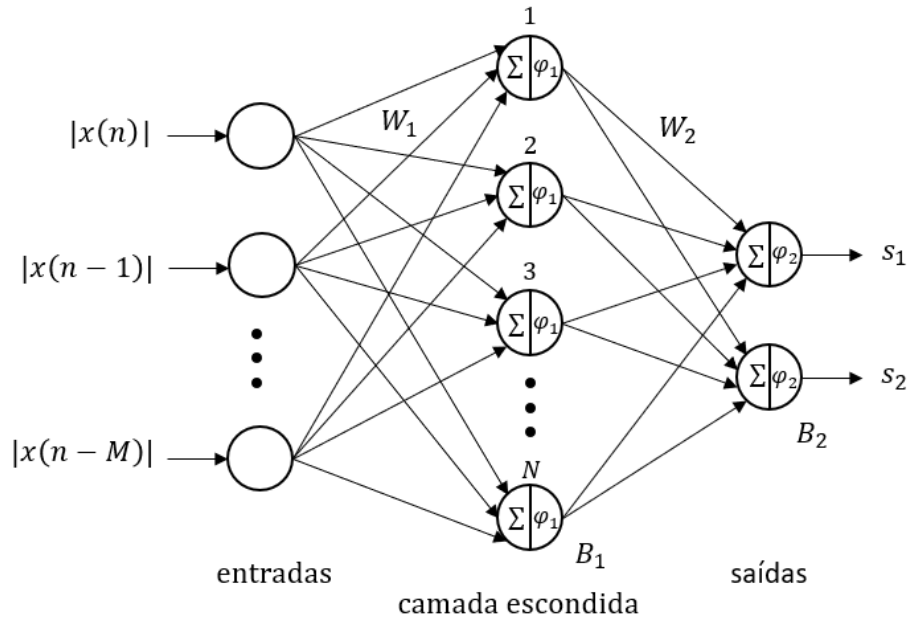


FONTE: O autor, 2025.

Na topologia adotada neste trabalho, a primeira camada tem $M+1$ neurônios, a camada oculta tem N neurônios e a última camada tem dois neurônios,

correspondentes à parte real e imaginária do coeficiente $s = s_1 + js_2$, como representado na Fig. 7.

Figura 7 - Diagrama de uma rede neural TLP



FONTE: O autor, 2025.

A função de ativação escolhida para a segunda camada foi a tangente hiperbólica $\varphi_1(x) = \tanh(x)$, enquanto que para a última camada foi a função identidade $\varphi_2(x) = x$.

Com apenas quatro matrizes é possível representar completamente esta rede neural. As matrizes W_1 e W_2 correspondem aos pesos w da primeira para a segunda camada e da segunda para a terceira camada, de acordo com:

$$W_1 = \begin{bmatrix} w_{11} & w_{21} & \cdots & w_{(M+1)1} \\ w_{12} & w_{22} & \cdots & w_{(M+1)2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1N} & w_{2N} & \cdots & w_{(M+1)N} \end{bmatrix} \quad e \quad W_2 = \begin{bmatrix} w_{11} & w_{21} & \cdots & w_{N1} \\ w_{12} & w_{22} & \cdots & w_{N2} \end{bmatrix}. \quad (17)$$

As matrizes B_1 e B_2 correspondem aos *biases* b da segunda e terceira camadas, dadas por:

$$B_1 = [b_1 \quad b_2 \quad \dots \quad b_N] \quad e \quad B_2 = [b_1 \quad b_2]. \quad (18)$$

Então, o modelo proposto neste trabalho pode ser resumido pela equação:

$$y(n) = \left[\sum_{m=0}^M a_{m+1} x(n-m) \right] s(n). \quad (19)$$

De modo análogo ao modelo polinomial, os coeficientes deste modelo proposto também estão relacionados de forma não linear. Assim, os coeficientes do modelo são extraídos por otimização não linear. O número de coeficientes depende apenas da profundidade de memória M e do número de neurônios na camada oculta N .

Para validar o modelo foram utilizados dados extraídos de um PA composto por um transistor HEMT, excitado por duas portadoras moduladas por sinais WCDMA de larguras de banda iguais a 3,84 MHz, separadas entre si por 5 MHz e com o sinal resultante centrado em 900 MHz. A frequência de amostragem utilizada para as medições dos sinais de entrada e saída foi de 61,44 MHz. Este conjunto de dados foi o mesmo que o utilizado em [1].

O Matlab foi utilizado para implementar e comparar os modelos [18]. No modelo polinomial, foi efetuada uma otimização não linear utilizando o comando 'lsqnonlin' do Matlab para extrair os coeficientes a . Em seguida, foi utilizado o método dos mínimos quadrados para obter os coeficientes b . No modelo proposto, foi utilizada apenas uma otimização não linear.

A exatidão dos modelos é medida pelo Erro Quadrático Médio Normalizado (NMSE), calculado pela equação

$$NMSE = 10 \log_{10} \frac{\sum_{n=1}^{N_s} |y(n) - y_{ref}(n)|^2}{\sum_{n=1}^N |y_{ref}(n)|^2}, \quad (20)$$

onde N_s é o número total de amostras, $y_{ref}(n)$ é a saída desejada e $y(n)$ é a saída estimada [19].

O Erro Quadrático Médio (MSE), base da métrica NMSE, é bastante utilizado em modelagem estatística para ajustar os parâmetros de um modelo de modo

que as diferenças entre os valores previstos e os valores medidos sejam minimizadas. Essa técnica é útil para estruturas complexas e não lineares, como é o caso dos modelos híbridos apresentados que envolvem múltiplas interações entre variáveis.

A principal vantagem do NMSE em relação ao MSE simples é que ele normaliza o erro em relação à variância dos dados reais, fornecendo uma medida relativa da qualidade do ajuste, permitindo comparar o desempenho de diferentes modelos, mesmo quando os sinais de entrada têm escalas ou magnitudes diferentes.

Para inicializar o código para ambos os modelos, primeiro foi criado um vetor com coeficientes aleatórios. Desta forma, o resultado é sempre diferente para cada vez que o código é executado. Em seguida, para cada valor de N ou P , foram efetuadas várias simulações e tabulados os melhores resultados.

Para comparar os modelos, o valor da memória foi fixado em $M = 2$. O valor da ordem polinomial P foi variado para o primeiro modelo. Já para o segundo modelo foi variado o número de neurônios na camada oculta N .

Para o primeiro modelo, todos os coeficientes a e b são complexos, enquanto no segundo modelo, os coeficientes a são complexos e os coeficientes b e w são reais. Para comparar o NMSE em função do número de coeficientes, será utilizado apenas o número de coeficientes reais, considerando que um número complexo tem dois coeficientes reais.

Para o modelo polinomial, o número de coeficientes C_{POL} é função da profundidade de memória M e da ordem polinomial P . No modelo proposto, o número de coeficientes C_{TLP} é função da profundidade de memória M e do número de neurônios N na camada escondida. As expressões para o cálculo destes coeficientes são apresentadas a seguir:

$$C_{POL} = 2(M + 1)(P + 1) , \quad (21)$$

$$C_{TLP} = (M + 1)(N + 2) + 3N + 2. \quad (22)$$

4. RESULTADOS E DISCUSSÃO

As tabelas I e II apresentam os resultados NMSE para os modelos polinomial e os modelos propostos, respetivamente.

Tabela I - Resultados para o modelo polinomial

P	$NMSE (dB)$	C_{POL}
3	-41,0111	24
4	-42,8654	30
5	-43,5705	36
6	-43,4667	42
7	-43,7914	48
8	-43,7958	54
9	-43,8572	60
10	-43,8815	66

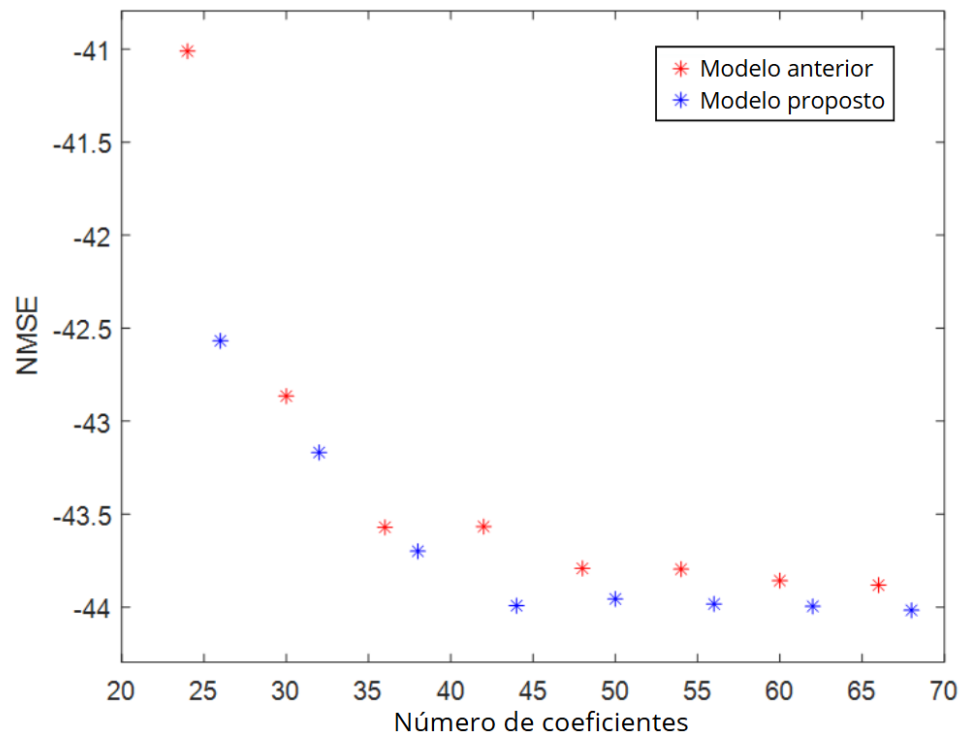
Tabela II - Resultados para o modelo proposto

N	$NMSE (dB)$	C_{TLP}
3	-42,5681	26
4	-43,1681	32
5	-43,6992	38
6	-43,9921	44
7	-43,9556	50
8	-43,9825	56
9	-43,9956	62
10	-44,0168	68

Em ambas as tabelas é possível ver que ao aumentar a ordem polinomial P ou o número de neurônios na camada oculta N , o modelo tem melhor desempenho, pois ao aumentar esses parâmetros há consequentemente um aumento no número de coeficientes que representam o modelo. Para comparar os dois modelos, o gráfico do NMSE em função do número de coeficientes dos modelos é apresentado na Figura 8.

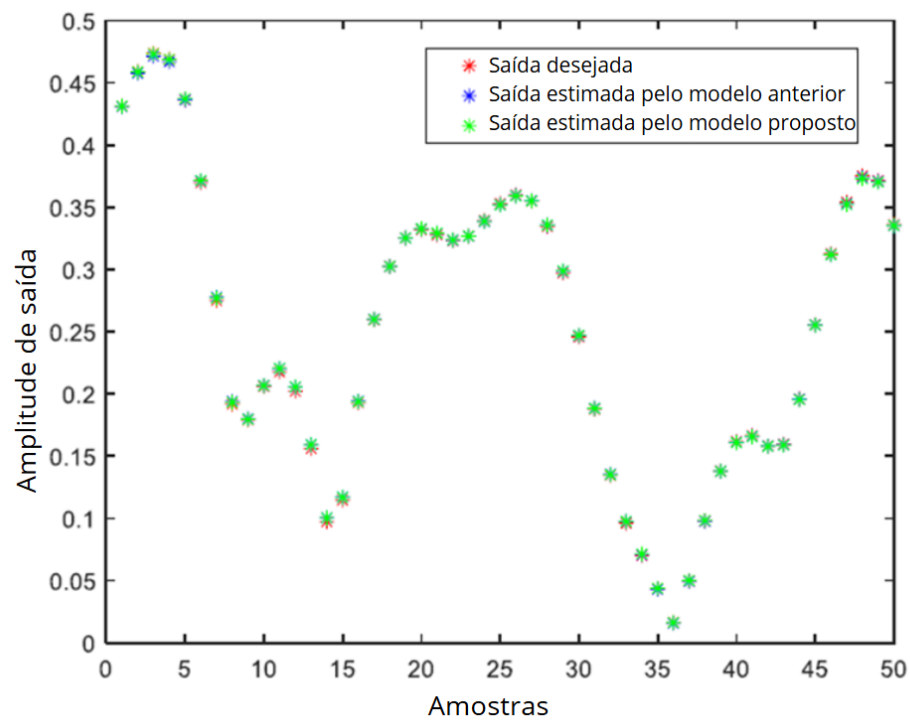
Como a métrica NMSE é uma medida em decibéis, quanto mais baixo o seu valor, mais pequeno é o erro cometido. Na Figura 8, os pontos do novo modelo estão abaixo dos pontos do modelo antigo, indicando que o modelo proposto neste trabalho tem um desempenho melhor com menos coeficientes. A Figura 9 mostra o gráfico da amplitude de saída em função das amostras para $N = 10$ e $P = 10$.

Figura 8 - NMSE em função do número de coeficientes



FONTE: O autor, 2025.

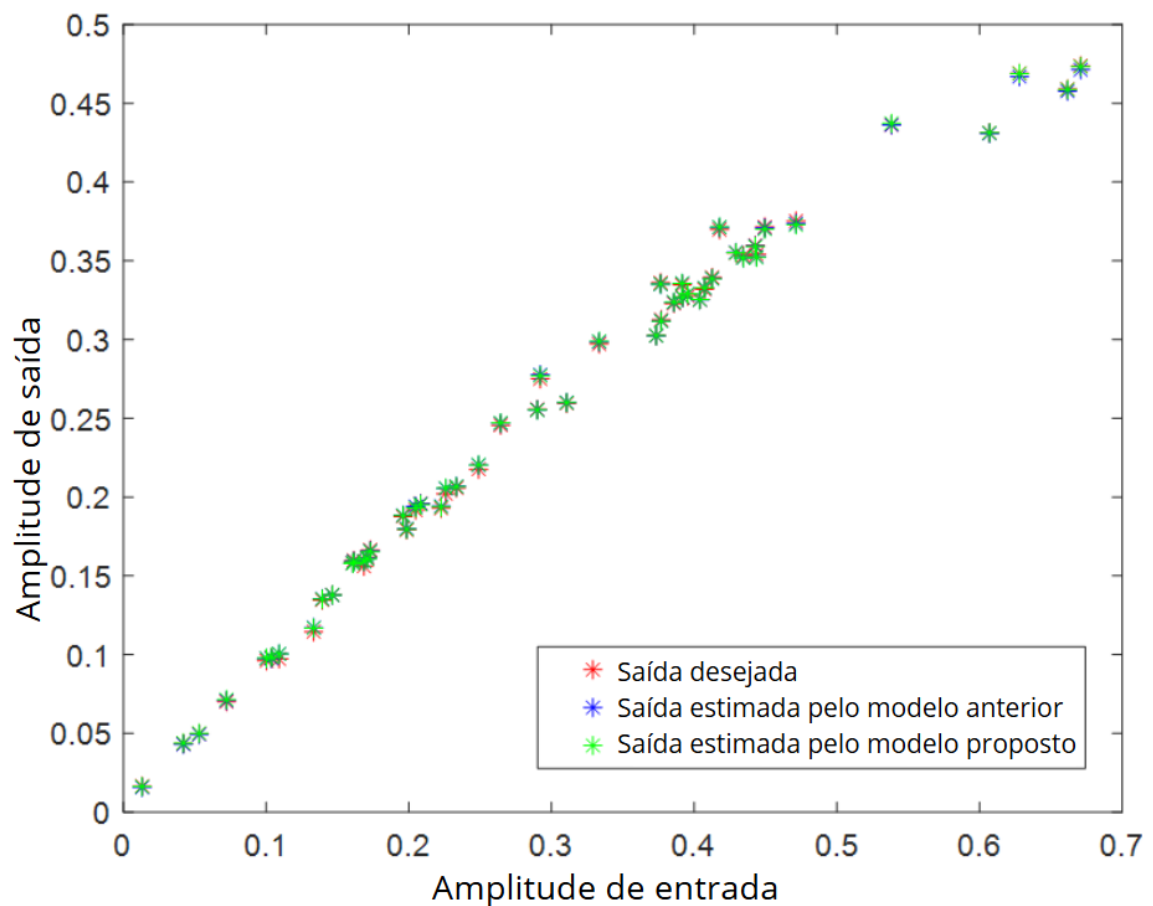
Figura 9 - Forma de onda de amplitude de saída



FONTE: O autor, 2025.

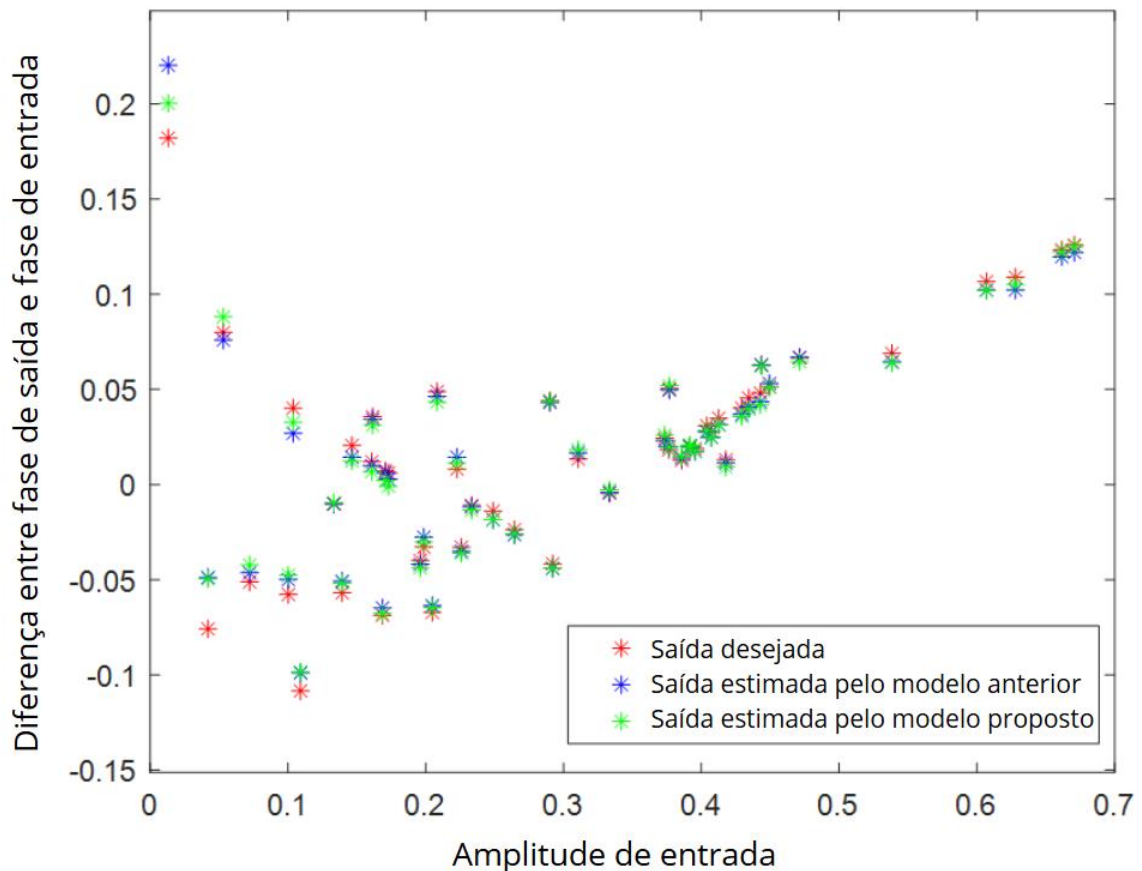
Ambos os modelos representam corretamente este amplificador de potência. Os pontos de ambos os modelos aparecem quase sobrepostos aos pontos das saídas do PA real. Existem dois gráficos importantes para avaliar a qualidade do amplificador em termos de distorção e não-linearidade. O gráfico AM-AM indica como a amplitude do sinal de saída do amplificador varia em relação à amplitude do sinal de entrada, enquanto o gráfico AM-PM indica como a diferença de fase entre o sinal de saída e o sinal de entrada varia em relação à amplitude do sinal de entrada [4]. As curvas características de transferência AM-AM e AM-PM para o PA real e modelado são mostradas na Figura 10 e Figura 11.

Figura 10 - Curva AM-AM



FONTE: O autor, 2025.

Figura 11 - Curva AM-PM



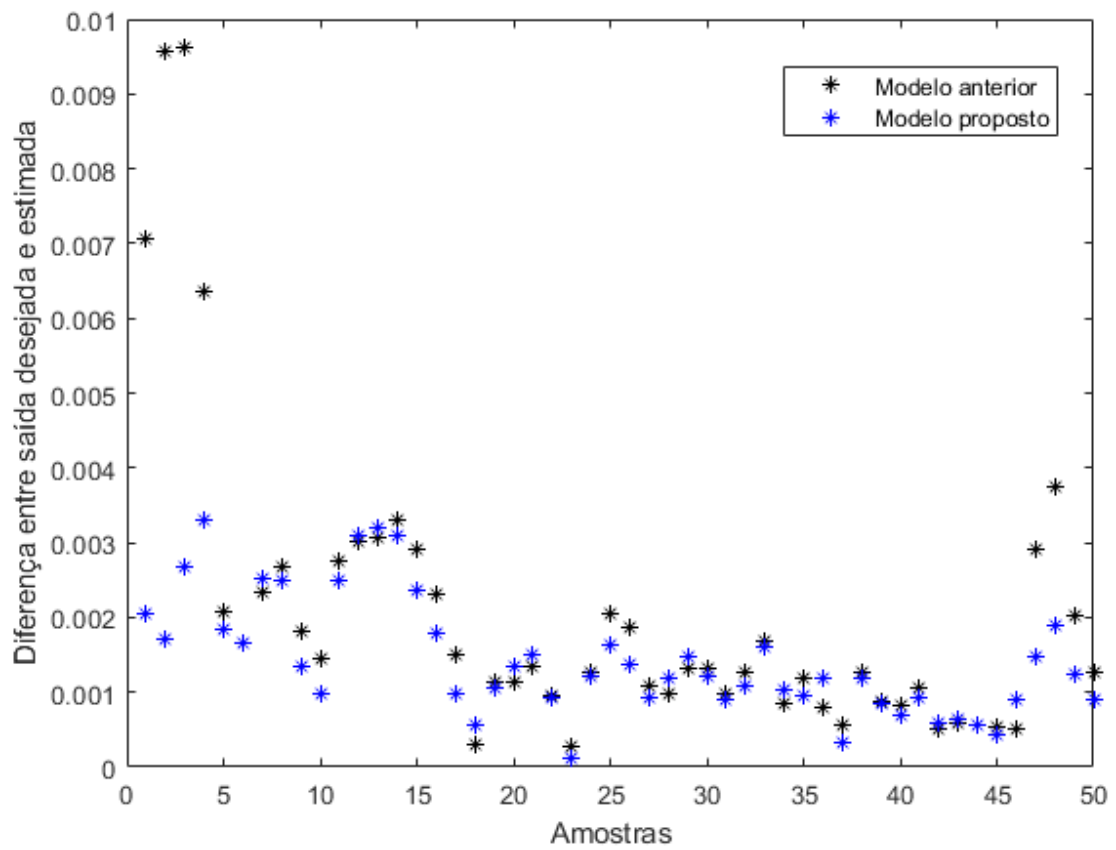
FONTE: O autor, 2025.

Utilizando o comando `tic toc` no Matlab, é possível medir o tempo de treinamento. É importante ressaltar que para cada nova simulação o tempo é diferente devido ao fato que os valores iniciais dos coeficientes são aleatórios. O tempo de treinamento do modelo polinomial é em torno de 25 segundos, enquanto que para o modelo proposto, esse tempo é de cerca de 20 minutos. Em ambos os casos, o tempo de simulação não depende da ordem do polinômio, nem do número de neurônios na camada oculta. Verifica-se que, embora o modelo proposto apresente melhores valores de NMSE, ele possui um tempo de treinamento mais longo.

A Figura 12 apresenta os valores dos erros para ambos os modelos. Verifica-se que os erros são em média inferiores a 0,003 indicando que ambos os modelos são precisos, contudo, devido aos erros estarem em sua forma linear e não em decibéis, é menos evidente notar a performance um pouco melhor do modelo

proposto, também devido ao fato de que os gráficos foram gerados apenas para as 50 primeiras amostras, enquanto a métrica leva em consideração todos os valores.

Figura 12 – Comparação dos erros entre os modelos



FONTE: O autor, 2025.

5. CONCLUSÕES FINAIS

Neste trabalho, foi apresentado um modelo híbrido para modelagem comportamental de amplificadores de potência, que combina elementos de um modelo polinomial baseado em filtros de resposta ao impulso finita com uma rede neural perceptron de três camadas. O modelo proposto foi avaliado em termos da sua capacidade de representar o comportamento não linear dos amplificadores de potência e a sua eficácia na previsão da saída do amplificador em resposta a uma dada entrada.

Os resultados da simulação mostraram que o modelo híbrido proposto teve melhor desempenho do que o modelo polinomial tradicional em termos de precisão de modelagem e redução do NMSE, no entanto, requer um tempo de simulação mais longo. Além disso, o modelo híbrido apresentou uma redução significativa no número de coeficientes necessários para descrever o comportamento do amplificador de potência. O modelo híbrido também pode ser aplicado em outros campos, como na modelagem de sistemas não lineares em geral.

Como sugestões para trabalhos futuros, destaca-se a implementação do modelo em linguagem de descrição de hardware (HDL), visando sua síntese em circuitos digitais para aplicações em tempo real. Além disso, recomenda-se a extensão da metodologia para outros tipos de amplificadores de potência, como amplificadores de modos múltiplos ou de bandas múltiplas concorrentes, ampliando assim o alcance e a aplicabilidade da técnica proposta.

6. BIBLIOGRAFIA

- [1] Machado, Carolina Luiza Rizental, “Modelagem Comportamental de Amplificadores de Potência Usando Soma de Produtos Entre Filtros Digitais de Resposta ao Impulso Finita e Tabelas de Busca Unidimensionais”. Dissertação de Mestrado, Universidade Federal do Paraná, Curitiba, Paraná, Brasil, 2016.
- [2] M. H. S. Santana and E. G. de Lima, "Product between multi-layer perceptron and finite impulse response filter applied to behavioral modeling of power amplifiers," *EMicro/SIM 2023*, Porto Alegre, Brasil, 2023.
- [3] Razavi, B. (2016). RF Microelectronics. Prentice Hall. Chapter 13: Power Amplifiers, Sections 13.1-13.5.
- [4] Cripps, Steve. RF Power Amplifiers for Wireless Communications. 2nd ed. Norwood, MA: Artech House, 2006. Section 4.4.2
- [5] R. T. Compton and A. A. M. Saleh, "Behavioral Modeling of RF Power Amplifiers," *IEEE Microwave Magazine*, vol. 3, no. 4, pp. 44-50, Dec. 2002.
- [6] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microwave Magazine*, vol. 10, no. 7, pp. 52–64, Dec. 2009.
- [7] FALEMPIN, Alexis. *Adaptive Pre-Distortion for Power Amplifier Linearization based on Neural Networks*. 2022. Tese (Doutorado em Processamento de Sinais e Imagens) – Université Grenoble Alpes, 2022.
- [8] F. Raab et al., “Power amplifiers and transmitters for RF and microwave,” *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 3, pp. 814–826, Mar. 2002.
- [9] A. Katz, “Linearization: reducing distortion in power amplifiers,” *IEEE Microw. Mag.*, vol. 2, no. 4, pp. 37–49, Dec. 2001.
- [10] A. Kaye, D. George, and M. Eric, “Analysis and compensation of bandpass nonlinearities for communications,” *IEEE Trans. Commun.*, vol. 20, no. 5, pp. 965–972, Oct. 1972.

- [11] L. B. C. Freire, L. Schuartz, E. L. Santos, B. Leite, A. Mariano, J. S. Dias and E. G. Lima, "Behavioral Modeling of a Multimode Power Amplifier Using Perceptrons with AM-AM and AM-PM Complex-Valued Activation Function", in *Journal of Integrated Circuits and Systems*, vol. 16, pp. 1-11, 2021.
- [12] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, Nov. 2001.
- [13] Hammi, Oualid, Fadhel M. Ghannouchi, and Bill Vassilakis. "A compact envelope-memory polynomial for RF transmitters modeling with application to baseband and RF-digital predistortion." *IEEE Microwave and Wireless Components Letters* 18.5 (2008): 359-361.
- [14] F. Mkadem, A. Islam and S. Boumaiza, "Multi-Band Complexity-Reduced Generalized-Memory-Polynomial Power-Amplifier Digital Predistortion," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 6, pp. 1763-1774, June 2016.
- [15] Haykin, Simon. *Neural Networks and Learning Machines*. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2009. Chapter 4, Section 4.2.
- [16] N. Benvenuto, F. Piazza, and A. Uncini, "A neural network approach to data pre-distortion with memory in digital radio systems," in *Proc. IEEE Int. Conf. Comm. ICC 93*, May 1993, pp. 232–236.
- [17] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.
- [18] *Numerical Computing with MATLAB*. C. Moler Philadelphia: SIAM, 2004.
- [19] M. S. Muha, C. J. Clark, A. Moulthrop, and C. P. Silva, "Validation of power amplifier nonlinear block models," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Anaheim, CA, Jun. 1999, pp. 759–762.
- [20] Urooj, Ahtisham & Ahmed, Muneer & Absi, Al & Al-Absi, Munir. (2024). Review on Solid-state Narrow and Wide- band Power Amplifier.