

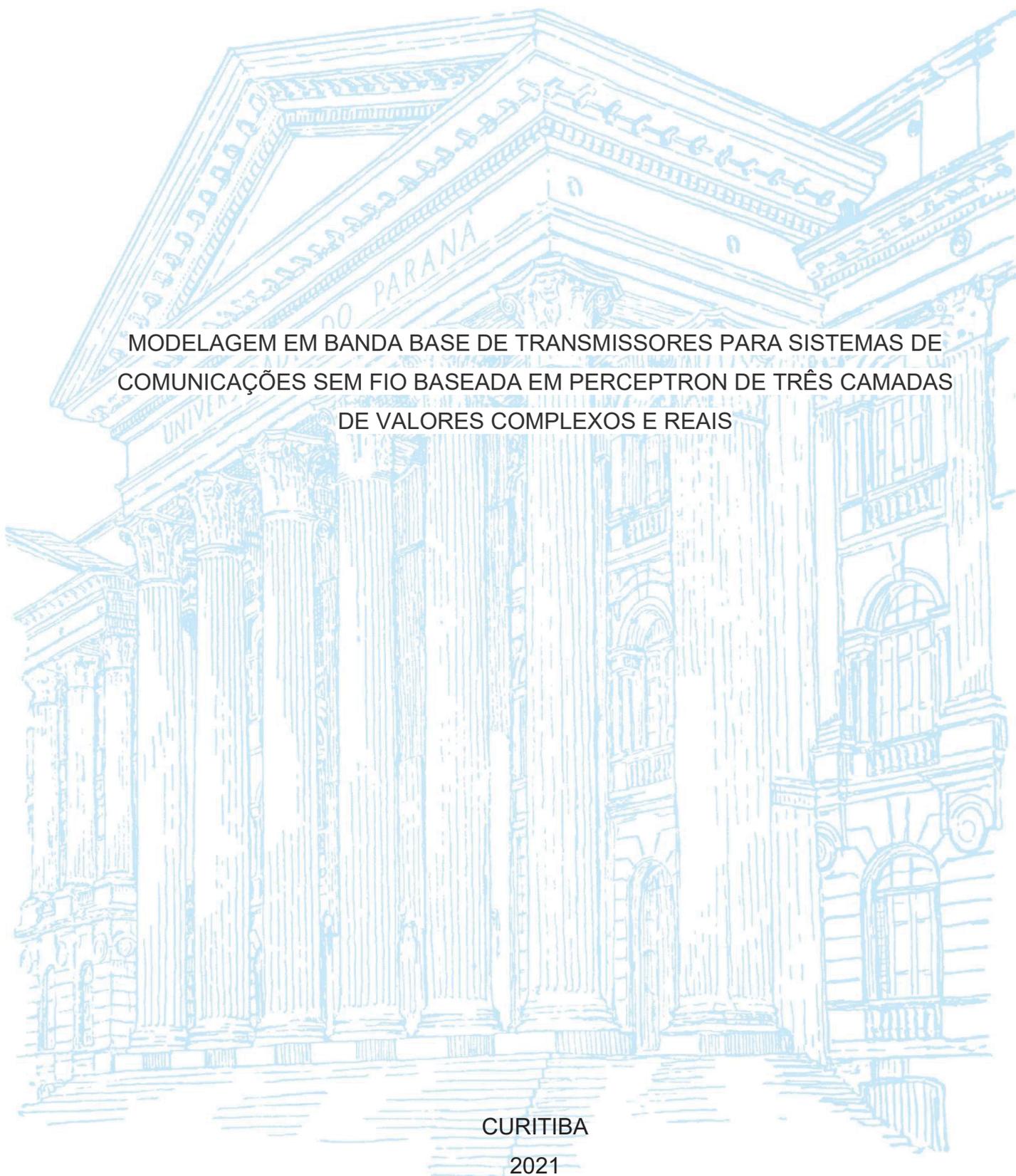
UNIVERSIDADE FEDERAL DO PARANÁ

LUIZA BEANA CHIPANSKY FREIRE

MODELAGEM EM BANDA BASE DE TRANSMISSORES PARA SISTEMAS DE
COMUNICAÇÕES SEM FIO BASEADA EM PERCEPTRON DE TRÊS CAMADAS
DE VALORES COMPLEXOS E REAIS

CURITIBA

2021



LUIZA BEANA CHIPANSKY FREIRE

MODELAGEM EM BANDA BASE DE TRANSMISSORES PARA SISTEMAS DE
COMUNICAÇÕES SEM FIO BASEADA EM PERCEPTRON DE TRÊS CAMADAS
DE VALORES COMPLEXOS E REAIS

Tese apresentada ao curso de Pós-Graduação em Engenharia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Dr. Eduardo Gonçalves de Lima

CURITIBA

2021

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

F866m Freire, Luiza Beana Chipansky
Modelagem em banda base de transmissores para sistemas de comunicações sem fio baseada em perceptron de três camadas de valores complexos e reais [recurso eletrônico] / Luiza Beana Chipansky Freire. – Curitiba, 2021.
Tese - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, 2021.
Orientador: Eduardo Gonçalves de Lima.
1. Sistemas de comunicação sem fio. 2. Amplificadores de potência. 3. Redes neurais (Computação). I. Universidade Federal do Paraná. II. Lima, Eduardo Gonçalves de. III. Título.
CDD: 384.5

Bibliotecária: Vanusa Maciel CRB- 9/1928



MINISTÉRIO DA EDUCAÇÃO
SETOR DE TECNOLOGIA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA
ELÉTRICA - 40001016043P4

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA ELÉTRICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **LUIZA BEANA CHIPANSKY FREIRE** intitulada: **Modelagem em banda base de transmissores para sistemas de comunicações sem fio baseada em perceptron de três camadas de valores complexos e reais**, sob orientação do Prof. Dr. EDUARDO GONÇALVES DE LIMA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 14 de Julho de 2021.

Assinatura Eletrônica
14/07/2021 16:59:13.0
EDUARDO GONÇALVES DE LIMA
Presidente da Banca Examinadora

Assinatura Eletrônica
14/07/2021 16:58:04.0
GLAUBER GOMES DE OLIVEIRA BRANTE
Avaliador Externo (UNIVERSIDADE TECNOLÓGICA FEDERAL DO
PARANÁ)

Assinatura Eletrônica
16/07/2021 10:34:31.0
LUIS CARLOS VIEIRA
Avaliador Externo (UNIVERSIDADE TECNOLÓGICA FEDERAL DO
PARANÁ)

Assinatura Eletrônica
15/07/2021 11:33:14.0
OSCAR DA COSTA GOUVEIA FILHO
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ -
DELT/UFPR)

RESUMO

Há um crescimento contínuo de consumo de dados por redes móveis e isso força uma adaptação das tecnologias utilizadas nas redes sem fio para atender a essa alta demanda. Alguns fatores de grande impacto nas redes sem fio são a linearidade e o consumo energético. A linearidade impacta na qualidade do sinal a ser transmitido e recebido. O consumo energético impacta no custo de operação e na autonomia de dispositivos móveis. Em se tratando dos transmissores sem fio o amplificador de potência (PA) é o dispositivo que mais influencia nesses fatores. Para o PA ser capaz de entregar um sinal o mais fidedigno possível com uma alta eficiência energética é necessário utilizar técnicas de linearização. A técnica mais utilizada é a de pré-distorção digital. Essa técnica precisa de modelos matemáticos da inversa do comportamento de um PA. Existem várias estratégias de modelagem. A que tem demonstrado grande capacidade de modelagem de sistemas altamente não lineares com menor complexidade computacional são as redes neurais, mais especificamente as redes *perceptron* de três camadas (TLP). Os modelos lidam com sinais banda base que ao interagir com a função de ativação das redes neurais acabam gerando contribuições não relacionadas a sinais físicos do PA, que são espúrios matemáticos. Se torna interessante adaptar esses modelos baseados em redes TLP a fim de convergirem mais rapidamente e gerar apenas sinais relacionados ao PA sendo modelado. Nesse estudo são abordadas as redes TLP de valor complexo, sendo analisadas quanto às suas qualidades de modelagem de sinais banda base e sua complexidade computacional. São comparados diversos modelos da literatura e propostos modelos que geram melhor resposta em termos de acurácia e quantidade de parâmetros. Também são estudados os modelos que modelam o comportamento não linear do PA e os desbalanceamentos do modulador I/Q em conjunto, analisando suas características. Da mesma forma são propostos modelos que apresentam melhor resposta que os modelos usados na literatura.

Palavras-chave: Pré-distorção digital (DPD). Amplificador de potência. Modulador em quadratura. Rede perceptron de três camadas (TLP). Funções de ativação complexa.

ABSTRACT

There is a continuous growth of data consumption by mobile networks and this forces wireless networks to adapt to meet this high demand. Some factors that have a big impact on wireless networks are linearity and energy consumption. Linearity impacts the quality of the signal to be transmitted and sent. Energy consumption impacts the operating cost and autonomy of mobile devices. When it comes to wireless transmitters, the power amplifier (PA) is the device that most influences the factors mentioned before. For the PA to be able to deliver a signal as reliable as possible, with high efficiency it is necessary to use linearization techniques. The most used technique is the digital pre-distortion. This technique uses mathematical models of the PA inverse behavior. There are several modeling strategies. One that presents great capacity for modeling highly non-linear systems with less computational complexity are neural networks, more specifically three-layer perceptron (TLP) networks. The models deal with baseband signals that, when interacting with the neural networks activation function, end up generating contributions not related to physical PA signals, which are mathematical spurious. It is interesting to adapt these models based on TLP networks in order to converge more quickly and generate only signals related to the PA being modeled. Complex valued TLP networks are analyzed regarding their qualities of baseband modeling and their computational complexity. Several models in the literature are compared. Models that generate the best response in terms of accuracy and number of parameters are proposed. Joint models of the PA nonlinear behavior and I/Q modulator unbalances are also studied, analyzing their characteristics. Likewise, models that present a better response than the models used in the literature are proposed.

Keywords: Digital pre-distortion (DPD). Power amplifier. I/Q modulator. Three layer perceptron (TLP). Complex-valued activation function.

LISTA DE FIGURAS

FIGURA 1 – TRÁFEGO DE REDES MÓVEIS GLOBAL E CRESCIMENTO ANO A ANO (EB POR MÊS).....	14
FIGURA 2 – ESQUEMÁTICO DE UM SISTEMA DPD.....	16
FIGURA 3 – ORGANIZAÇÃO DOS CAPÍTULOS	20
FIGURA 4 – DIAGRAMA DE BLOCOS DOS PRINCIPAIS COMPONENTES DE UM TRANSMISSOR.....	21
FIGURA 5 – SINAIS BANDA BASE E PASSA BANDA NO DOMÍNIO DA FREQUÊNCIA.....	24
FIGURA 6 – DIAGRAMA SIMPLIFICADO DE MODULADOR EM QUADRATURA IDEAL.....	25
FIGURA 7 – COMPONENTES DO MODULADOR EM QUADRATURA E SUAS NÃO LINEARIDADES.....	26
FIGURA 8 – DIAGRAMA SIMPLIFICADO DE MODULADOR EM QUADRATURA NÃO IDEAL.....	27
FIGURA 9 – PRINCIPAIS COMPONENTES DO AMPLIFICADOR DE POTÊNCIA RF	29
FIGURA 10 – CARACTERÍSTICA AM-AM DE AMPLIFICADORES DE POTÊNCIA	30
FIGURA 11 – EFEITO DAS DISTORÇÕES DO PA NOS CANAIS ADJACENTES ..	30
FIGURA 12 – CONTRIBUIÇÕES FORA DA BANDA.....	33
FIGURA 13 – MODELO NÃO LINEAR DE UM NEURÔNIO k	39
FIGURA 14 – TRANSFORMAÇÃO AFIM.....	40
FIGURA 15 – ARQUITETURA DIRETA E TOPOLOGIA MLP COM TRÊS CAMADAS	41
FIGURA 16 – ARQUITETURA RECORRENTE	43
FIGURA 17 – MAPEAMENTO DE DOIS SINAIS REAIS	45
FIGURA 18 – REDE NEURAL PARA MAPEAMENTO DE DOIS SINAIS REAIS.	45
FIGURA 19 – REDE NEURAL DE VALOR COMPLEXO PARA MAPEAMENTO DE DOIS SINAIS REAIS.....	46
FIGURA 20 – REDES DE VALORES COMPLEXOS DA LITERATURA.....	49
FIGURA 21 – REDES DE VALORES REAIS NA LITERATURA.....	50
FIGURA 22 – REDES COM REALIMENTAÇÃO DA LITERATURA	51

FIGURA 23 – REDE DE VALOR REAL QUE CONSIDERA A NATUREZA PASSA BAIXAS DO PA	53
FIGURA 24 – DIAGRAMA SIMPLIFICADO DA DIMENSIONALIDADE DOS SUBESPAÇOS	54
FIGURA 25 – MODELO COM DUAS REDES DE VALORES REAIS QUE CONSIDERA AS CARACTERÍSTICAS PASSA BAIXAS DO PA.....	54
FIGURA 26 – MEDIÇÃO EM DUAS ETAPAS DO SISTEMA MODULADOR I/Q E PA	55
FIGURA 27 – MEDIÇÃO EM UMA ÚNICA ETAPA DO SISTEMA MODULADOR I/Q E PA.....	56
FIGURA 28 – MODELO DE PA PROPOSTO BASEADO EM TLP DE VALORES COMPLEXOS	63
FIGURA 29 – CONFIGURAÇÃO DO SISTEMA PARA AQUISIÇÃO DOS DADOS..	63
FIGURA 30 – RESULTADOS DE NMSE	64
FIGURA 31 – RESULTADOS DE ACEPR	65
FIGURA 32 – PSD DOS SINAIS DE ERRO.....	65
FIGURA 33 – SINAIS DE SAÍDA DO MODELO E DO PA EM FUNÇÃO DO TEMPO	66
FIGURA 34 – COMPORTAMENTO DAS FUNÇÕES f_1 E f_2	67
FIGURA 35 – MODELO PROPOSTO DO INVERSO DO COMPORTAMENTO DO PA MULTIMODO	68
FIGURA 36 – FUNÇÃO DE TRANFERÊNCIA DO PA MULTIMODO	69
FIGURA 37 – SUPERPOSIÇÃO DAS AMOSTRAS DE AMPLITUDE	70
FIGURA 38 – DETALHAMENTO DAS AMOSTRAS SUPERPOSTAS	71
FIGURA 39 – ESCOLHA DOS LIMITES DOS MODOS DE OPERAÇÃO.....	72
FIGURA 40 – ALGORITMO DO MÉTODO ASCENDENTE	74
FIGURA 41 – EXTRAÇÃO DO MODELO DE PA E SEU USO NA UNIDADE DPD..	75
FIGURA 42 – VALORES DE NMSE.....	77
FIGURA 43 – DENSIDADE ESPECTRAL DE POTÊNCIA DOS SINAIS DE ERRO..	77
FIGURA 44 – MODELO SEM MEMÓRIA DO CONJUNTO MODULADOR I/Q MAIS PA DA LITERATURA	79
FIGURA 45 – MODELO SEM MEMÓRIA DO CONJUNTO MODULADOR I/Q MAIS PA DA LITERATURA	80

FIGURA 46 – MODELO PROPOSTO SEM MEMÓRIA COM DUAS REDES TLP DE VALOR REAL.....	81
FIGURA 47 – DIAGRAMA DE BLOCOS DAS ITERAÇÕES ÍMPARES.....	82
FIGURA 48 – DIAGRAMA DE BLOCOS DAS ITERAÇÕES PARES.....	83
FIGURA 49 – VALORES DE NMSE.....	84
FIGURA 50 – VALORES DE ACEPR.....	85
FIGURA 51 – PSD DOS SINAIS DE ERRO.....	85
FIGURA 52 – VALORES DE NMSE GRUPO B.....	90
FIGURA 53 - VALORES DE NMSE GRUPO C.....	90
FIGURA 54 - VALORES DE NMSE GRUPO E.....	91
FIGURA 55 - VALORES DE NMSE GRUPO F.....	91
FIGURA 56 – VALORES DE NMSE.....	93
FIGURA 57 – VALORES DE NMSE.....	94
FIGURA 58 – DIAGRAMA DE CONTRIBUIÇÕES DO TRABALHO.....	97

LISTA DE TABELAS

Tabela 1 – NMSEs dos modelos da literatura e proposto e os valores correspondentes de parâmetros complexos sem o método ascendente	76
Tabela 2 – Valores comparativos de NMSE e ACEPR	84
Tabela 3 – Definição dos sinais de entrada usados nos modelos analisados	87
Tabela 4 – Definição dos sinais de saída usados nos modelos analisados	87
Tabela 5 – Conversão de operações para FLOPs	88
Tabela 6 – Resumo dos modelos sendo analisados	89

LISTA DE ABREVIATURAS OU SIGLAS

3G	terceira geração
4G	quarta geração
5G	quinta geração
ACEPR	razão de potência do erro de canal adjacente
AM/AM	conversão modulação em amplitude/modulação em amplitude
AM/PM	conversão modulação em amplitude/modulação em fase
BP	retropropagação de erro
CDMA	acesso múltiplo por divisão de código
DPD	pré distorção digital
FoM	figura de mérito
IF	infravermelho
LFP	potencial de campo local
LO	oscilador local
MEMR	razão de modelagem do efeito de memória
MER	razão de efeito de memória
MLP	<i>perceptron</i> de múltiplas camadas
NMSE	erro quadrático médio normalizado
OFDM	multiplexação ortogonal por divisão de frequência
PA	amplificador de potência
PAPR	relação entre potência de pico e média
PD	pré distorcedor
RF	rádio frequência
TLP	<i>perceptron</i> de três camadas
WESPR	razão da potência erro-sinal ponderada

SUMÁRIO

1 INTRODUÇÃO	14
1.1 CONTEXTUALIZAÇÃO	14
1.2 MOTIVAÇÃO E OBJETIVOS	17
1.3 OBJETIVOS ESPECÍFICOS	18
1.4 ORGANIZAÇÃO DO TRABALHO	18
2 TRANSMISSOR SEM FIO	21
2.1 REPRESENTAÇÃO BANDA BASE DE SINAIS BANDA PASSANTE	22
2.2 CARACTERIZAÇÃO DAS DISTORÇÕES DO TRANSMISSOR SEM FIO	24
2.2.1 Modulador I/Q.....	24
2.2.2 Amplificador de Potência.....	27
2.3 COMPORTAMENTO PASSA BANDA DO PA	31
2.4 MITIGANDO AS DISTORÇÕES DO TRANSMISSOR	31
2.5 EXIGÊNCIAS DE MODELOS DE PA.....	32
2.6 CRITÉRIO DE AVALIAÇÃO DE MODELOS COMPORTAMENTAIS DE PA	34
2.6.1 Acurácia da modelagem.....	34
2.6.2 Complexidade de modelos comportamentais.....	36
3 REDES NEURAIS ARTIFICIAIS	38
3.1 UM MODELO DE NEURÔNIO	38
3.2 ARQUITETURAS E TOPOLOGIAS DAS REDES NEURAIS	40
3.3 REDES TLP DE VALORES COMPLEXOS	43
4 ESTADO DA ARTE DE MODELOS MATEMÁTICOS PARA USO EM SISTEMAS DE PRÉ DISTORÇÃO DIGITAL EM TRANSMISSORES BASEADOS EM REDES PERCEPTRON DE MÚLTIPLAS CAMADAS	48
4.1 MODELOS DE AMPLIFICADORES DE POTÊNCIA BASEADOS EM PERCEPTRON DE MÚLTIPLAS CAMADAS.....	48
4.2 MODELOS CONJUNTOS DE PA E MODULADOR I/Q BASEADOS EM PERCEPTRON DE MÚLTIPLAS CAMADAS	55
5 MODELOS PROPOSTOS PARA MODELAGEM COMPORTAMENTAL DE PA BASEADOS EM TLP DE VALORES COMPLEXOS	57
5.1 ANÁLISE DAS FUNÇÕES DE ATIVAÇÃO DE VALOR COMPLEXO	58
5.2 MODELO PROPOSTO COM MODIFICAÇÃO NO SINAL DE ENTRADA DO MODELO.....	62

5.3 MODELO PROPOSTO COM MODIFICAÇÃO NA FUNÇÃO DE ATIVAÇÃO DA REDE TLP DE VALOR COMPLEXO	66
5.3.1 Estudo de caso de modelo com modificação na função de ativação da rede TLP de valor complexo com PA multimodo.....	67
6 MODELO COMPORTAMENTAL CONJUNTO DO PA E DO MODULADOR I/Q PROPOSTO	78
6.1 ANÁLISE DOS MODELOS CONJUNTOS DO PA E DO MODULADOR I/Q BASEADOS EM REDES TLP DE VALOR REAL	78
6.2 ESTUDO DE CASO COM MODELO CONJUNTO DE PA E DO MODULADOR I/Q BASEADO EM REDE TLP DE VALOR REAL E COM TREINAMENTO ITERATIVO	81
6.3 ESTUDO DE CASO DE 15 MODELOS CONJUNTO DE PA E DO MODULADOR I/Q BASEADO EM REDES TLP DE VALOR REAL E COMPLEXO	86
7 CONSIDERAÇÕES FINAIS	95
REFERÊNCIAS.....	99

1 INTRODUÇÃO

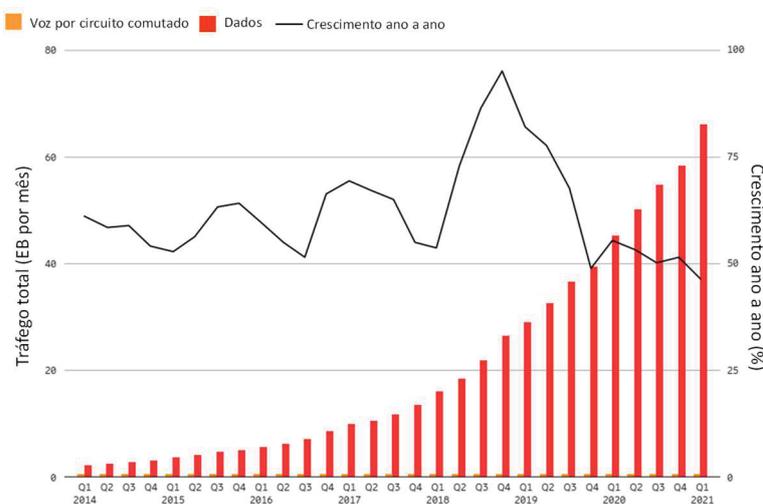
Na introdução será abordada a contextualização do trabalho, sua motivação e objetivos e por final será comentada a organização do trabalho.

1.1 CONTEXTUALIZAÇÃO

As redes de telecomunicação sem fio surgiram por volta de 1896 quando Guglielmo Marconi inventou o telégrafo sem fio (TSE e VISWANATH, 2005). O objetivo das redes de telecomunicação sem fio é transmitir informações através do ar usando ondas eletromagnéticas como rádio frequência (RF), infravermelho (IF), satélite e outras. Com a revolução das telecomunicações além de conectar pessoas através da rede de telefonia móvel o sistema de comunicação sem fio permite conectar dispositivos e processos à internet, sendo chamada de internet das coisas (IoT). Basicamente a rede interconecta objetos físicos através de sensores, softwares e diversas outras tecnologias, a fim de trocar dados entre si.

A forte tendência atual, que guia as transformações no sistema de telefonia móvel pelo mundo, é o crescente consumo de banda RF, gerada pelo aumento de consumo de dados. O gráfico na FIGURA 1 mostra o tráfego de dados nas redes móveis a cada quadrimestre, bem como seu crescimento percentual. O crescimento entre os primeiros quadrimestres de Q1 2020 e Q1 2021 foi de 46%, e o total de tráfego de dados mensal no Q1 2021 excedeu os 66 Exabyte. O crescimento do tráfego é

FIGURA 1 – TRÁFEGO DE REDES MÓVEIS GLOBAL E CRESCIMENTO ANO A ANO (EB POR MÊS)



FONTE: (ERICSSON, 2021)

derivado do crescimento de usuários de *smartphones*, e aumento do volume de dados por usuário, devido principalmente ao consumo de vídeos. Além disso a indústria de comunicação e informação usa aproximadamente 4% da eletricidade mundial e é esperado chegar de 10% a 20% em 2034, devido às novas tecnologias envolvidas nos serviços de prédios/casas inteligentes, cidades inteligentes, realidade aumentada, automação industrial, saúde etc (ERICSSON, 2021).

Nas telecomunicações existem as gerações de padrões de tecnologia para redes celulares de banda larga. A terceira geração (3G) abriu as portas para os serviços de internet e multimídia nas redes de telefonia móvel. A quarta geração (4G) ainda no seu começo já estava pronta para ser substituída pela quinta geração (5G). E de uma geração à outra a taxa de crescimento de transferência de dados segue aumentando drasticamente (ERICSSON, 2021).

Os sistemas atuais são multi-portadores e multi-bandas e a largura de banda é um quesito crítico nesse cenário, pois são necessárias maiores larguras de banda para sustentar a crescente taxa de dados.

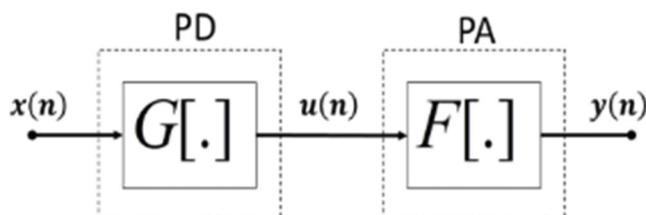
Juntamente com a eficiência energética e a largura de banda outro quesito importante é a linearidade do sistema de comunicação. Deve-se diminuir as distorções no sistema, geradas principalmente pelo circuito analógico do transmissor de rádio frequência, visando preservar a qualidade do sinal transmitido evitando perdas e interferências em outras bandas.

O sistema de transmissão sem fio é constituído de vários estágios em cascata incluindo a conversão digital-analógica, modulação e amplificação. O componente que mais impacta no consumo energético no transmissor é o amplificador de potência (PA). O PA é o responsável por amplificar o sinal a ser enviado ao receptor. Ele também é conhecido por possuir um antagonismo entre eficiência energética e comportamento linear, ou seja, para se alcançar um PA com alta eficiência é necessário que ele opere em regiões não lineares. Dessa forma, para se aumentar a linearidade do transmissor, sem comprometer sua eficiência é imperativo o uso de um método de linearização. A técnica de pré-distorção digital (DPD) é um dos métodos mais adotados (WOOD, 2017).

O sistema de pré-distorção, FIGURA 2, possui um distorcedor digital (PD) que é responsável por aplicar na entrada do PA a inversa do comportamento não linear, tornando a característica geral do sistema linear. A função $F[.]$ designa o

comportamento não linear do PA e a função $G[.]$ é o comportamento inverso do PA, $G[.] = F^{-1}[.]$.

FIGURA 2 – ESQUEMÁTICO DE UM SISTEMA DPD



FONTE: O autor (2021)

LEGENDA: Esquemático do sistema DPD, onde o sistema contendo a inversa do comportamento do PA, o sistema PD, fica em cascata com o PA.

As novas tecnologias trouxeram também esquemas de modulação com alta eficiência espectral, porém por possuírem envoltória variante produzem altos valores de relação entre potência de pico e potência média (PAPR), além de maior largura de banda (na ordem de centenas de MHz). Essas características, além de forçarem um comportamento não desejado no PA, dificultam o processo de linearização demandando maior número de coeficientes para modelar adequadamente o comportamento não linear, como também exigem mais dos conversores analógico-digital e digital-analógico por terem que processar um sinal banda larga e suas intermodulações (DESGREYS, NARASIMHA MANYAM, *et al.*, 2017).

De acordo com o princípio básico do DPD, conseguir modelar de forma precisa a inversa do comportamento não linear do amplificador é essencial. São vários os métodos estudados para tanto, porém cada método possui desvantagens um em relação ao outro. E dentro desse cenário de evolução das tecnologias, modelar com eficiência o comportamento não linear do PA é um desafio até mesmo para os métodos já utilizados (JARAUT, ABDELHAFIZ, *et al.*, 2021).

A fim de se obter a inversa do comportamento do PA são utilizados os chamados modelos comportamentais. Esses modelos são empíricos e se assume não ter nenhum conhecimento prévio da composição interna do circuito a ser modelado e são, portanto, construídos a partir de amostras dos sinais de entrada e saída medidos (abordagem caixa preta). A acurácia de um modelo depende de sua estrutura e do processo de extração de parâmetros. Diferentes conjuntos de dados e topologias de modelos levam a distintos resultados. Não é possível garantir a generalidade e as

capacidades preditivas de tais modelos comportamentais, pois um modelo que utiliza os parâmetros extraídos de um certo PA em circunstâncias específicas pode não gerar respostas aceitáveis se utilizado sob circunstância adversa ou para diferentes PAs da mesma família ou de tecnologias diferentes.

Duas abordagens de modelos de PA se sobressaem, os filtros polinomiais (RUGH, 1981), (BOYD e CHUA, 1985), (MATHEWS e SICURANZA, 2000) e as redes neurais artificiais (ZHANG e GUPTA, 2000), (HORNIK, STINCHCOMBE e WHITE, 1989).

Os filtros polinomiais têm um suporte matemático formal e sua teoria é bem consolidada na literatura. Algumas características negativas impactam seu uso para modelagem de PAs. Eles sofrem degradação na acurácia ao modelar sistemas sob regimes de operação fortemente não lineares, sendo necessária uma quantidade muito grande de parâmetros que torna impeditivo sua utilização prática (ZHU e BRAZIL, 2004). Além disso, alguns modelos de filtros polinomiais possuem pouca generalização ficando presos à zona de treinamento (MORGAN, MA, *et al.*, 2006). As redes neurais são uma opção de modelo que possuem boas respostas quanto a esses quesitos conseguindo representar sistemas fortemente não lineares com poucos parâmetros e possuem melhor extrapolação além da zona de treinamento.

1.2 MOTIVAÇÃO E OBJETIVOS

Visto que as tecnologias da rede de telecomunicações sem fio seguem evoluindo e as técnicas atuais de linearização não são adequadas à crescente complexidade dessas novas tecnologias, se faz necessário estudos acerca de novas técnicas de linearização, bem como de melhorias nas técnicas que já existem.

Em (JARAUT, ABDELHAFIZ, *et al.*, 2021) pode-se observar que quanto mais complexa a tecnologia envolvida no sistema de transmissão mais os modelos baseados em redes neurais se mostram um diferencial, apresentando resultados superiores em relação a outros modelos da literatura.

Esta tese visa estudar as redes neurais para modelos comportamentais do PA e do transmissor, embasando conceitos sobre suas características e como a interrelação entre elas pode influenciar a acurácia dos modelos baseados em redes neurais.

Mais especificamente serão discutidas as características das redes neurais de valor complexo, bem como a da modelagem conjunta do PA e do modulador em quadratura.

1.3 OBJETIVOS ESPECÍFICOS

Nesse trabalho serão analisados quatro estudos de casos, sendo dois no âmbito dos modelos baseados em redes *perceptron* de três camadas (TLP) de valor complexo e dois no dos modelos conjuntos PA mais modulador em quadratura com TLP de valor real e complexo.

Sobre os estudos baseados nos modelos com TLP de valor complexo, são analisados os modelos da literatura e é verificado se esses modelos geram apenas contribuições úteis, ou seja, que estão dentro da banda de operação do PA. Para tanto são analisadas as funções de ativação utilizadas na literatura nas redes TLP de valor complexo para modelagem de PAs. No primeiro estudo de caso é proposto um novo modelo que apresenta melhor acurácia que os modelos da literatura através de mudança na decomposição do sinal apresentado à rede neural. O segundo estudo de caso propõe uma nova função de ativação para TLPs de valor complexo utilizados para a modelagem de PAs.

Quanto aos estudos de caso baseados nos modelos conjuntos de PA e modulador I/Q, o primeiro estudo de caso visa avaliar se os modelos da literatura também respeitam as restrições impostas pela característica banda passante do PA, e propor um modelo que utilize uma técnica de treinamento iterativo. Como até agora existe apenas um modelo utilizado na literatura, o segundo estudo avalia as diferentes composições que uma rede TLP pode ter e como isso influencia na acurácia da modelagem do sistema conjunto PA mais modulador I/Q.

1.4 ORGANIZAÇÃO DO TRABALHO

O trabalho é dividido em 7 capítulos, como pode ser visto na FIGURA 3, sendo que os Capítulos 2 e 3 abordam aspectos teóricos para embasar o estudo e o Capítulo 4 apresenta o estado da arte de modelos baseados em redes neurais de três camadas. As contribuições do trabalho se encontram nos Capítulos 5 e 6. No Capítulo 7 é feita a conclusão do trabalho.

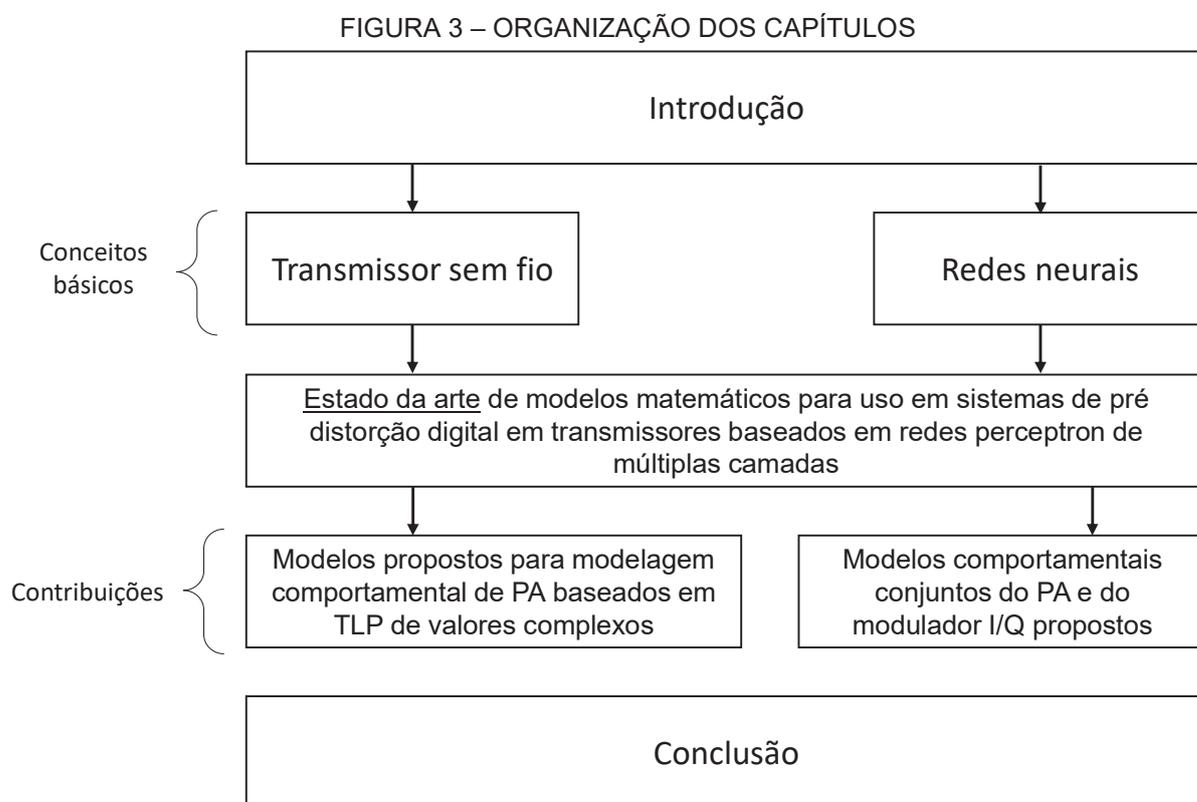
O Capítulo 2 define conceitos teóricos sobre os sinais processados pelo transmissor, bem como analisa as distorções que ocorrem dentro do transmissor e os efeitos dessa distorção nesses sinais. As técnicas pelas quais o transmissor pode ser linearizado também são abordadas.

O Capítulo 3 discorre sobre a teoria das redes neurais relevantes para essa tese. Além de embasar os princípios básicos das redes neurais, analisa as redes de valores complexos.

O Capítulo 4 apresenta o estado da arte sobre modelos baseados em redes neurais de três camadas.

O Capítulo 5 discute os modelos baseados em redes neurais de valor complexo analisando os modelos existentes da literatura no tocante às funções de ativação. São propostos dois modelos baseados em redes neurais de valor complexo validados através da modelagem de dois amplificadores de potência.

O Capítulo 6 aborda os modelos conjuntos de amplificador e modulador I/Q. É feita uma análise de modelos da literatura, além de variações de modelos com arquitetura e topologias diversas. Também são propostos modelos que possuem melhor acurácia que os modelos já existentes de redes neurais de três camadas.



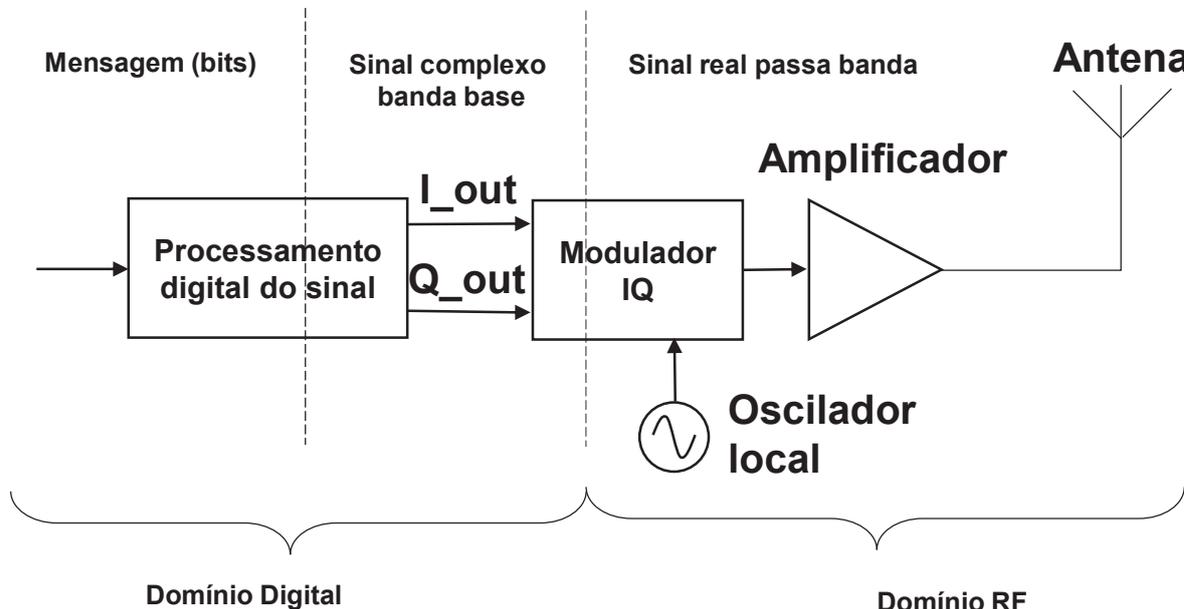
FONTE: O autor (2021)

2 TRANSMISSOR SEM FIO

O transmissor é um subsistema em um sistema sem fio. Ele tem a função de transformar um fluxo de *bits* em uma forma de onda adequada à transmissão através de uma antena (SHANNON, 1948). As especificações que o transmissor necessita atender dependem da aplicação onde ele está inserido. Para sistemas de comunicação são essenciais baixo ruído e boa estabilidade. Para transmissões de longas distâncias é ideal alta potência e baixo ruído. E para sistemas operando com bateria (em exemplo os celulares), alta eficiência energética (CHANG, 2000).

Os transmissores mais utilizados são os de conversão digital direta. O transmissor geralmente é composto por um oscilador, um modulador, um conversor ascendente, filtros e amplificadores de potência. A FIGURA 4 exemplifica os principais componentes de um transmissor. O modulador faz parte dos componentes analógicos. Ele gera um sinal real passa banda modulado pelo sinal complexo banda base. A saída do modulador então alimenta o PA que amplifica o sinal a ser transmitido pela antena (SOLTANI TEHRANI, 2012).

FIGURA 4 – DIAGRAMA DE BLOCOS DOS PRINCIPAIS COMPONENTES DE UM TRANSMISSOR



FONTE: Modificado de (SOLTANI TEHRANI, 2012).

2.1 REPRESENTAÇÃO BANDA BASE DE SINAIS BANDA PASSANTE

Na prática, para modelagem ou realização de análises de sinais modulados em relativa baixa frequência (ordem de MHz), são utilizados sinais banda base. Eles possuem o espectro de frequência em volta da frequência nula. No entanto, para o sistema de comunicação sem fio, na maior parte do tempo, sinais passa banda são considerados e usados nas simulações de sistemas sem fio. Os sinais passa banda, em teoria, possuem a frequência da portadora (f_c) relativamente alta (em torno de GHz) e possuem o espectro da frequência concentrado acerca dessa frequência. Sinais banda base podem ser convertidos em sinais passa banda através da conversão ascendente e vice-versa através da conversão descendente (VUOLEVI e RAHKONEN, 2002), (VUOLEVI, 2001).

No sistema de comunicação sem fio o modulador tem o papel de converter sinais banda base em sinais banda passante a fim de serem transmitidos.

O sinal modulado em frequência e em amplitude é um sinal real descrito por:

$$S(n) = A(n)\cos[\omega_c n + \theta_n], \quad (1)$$

onde $\omega_c = 2\pi f_c$ é a frequência angular da portadora e $A(n)$ e θ_n são respectivamente a amplitude e a fase do sinal de envoltória. Essa representação do sinal se chama representação polar. O sinal em (1) tem uma largura de banda de envoltória muito menor que a frequência da portadora f_c , sendo chamado de sinal de banda estreita (MORGAN, MA, *et al.*, 2006).

A representação da sua forma complexa $\tilde{s}(n)$ pode ser escrita como:

$$\begin{aligned} S(n) &= \mathcal{R}[A(n)e^{j(\omega_c n + \theta_n)}] \\ &= \mathcal{R}[A(n)e^{j\theta_n}e^{j\omega_c n}] \\ &= \mathcal{R}[\tilde{s}(n)e^{j\omega_c n}], \end{aligned} \quad (2)$$

onde $\tilde{s}(n)$ é chamado sinal banda base ou envoltória complexa, e contém a mesma informação que $S(n)$. O sinal $\tilde{s}(n)$ pode ser chamado então de equivalente em banda base de $S(n)$.

Usando trigonometria o sinal pode ser decomposto em suas componentes em fase $s_I(n)$ e em quadratura $s_Q(n)$:

$$\begin{aligned}\tilde{s}(n) &= A(n) \cos(\omega_c n) \cos(\theta_n) - A(n) \sin(\omega_c n) \sin(\theta_n) \\ &= s_I(n) \cos(\omega_c n) - s_Q(n) \sin(\omega_c n).\end{aligned}\quad (3)$$

Os componentes $s_I(n)$ e $s_Q(n)$ são definidos por:

$$\begin{aligned}s_I(n) &= \mathcal{R}\{\tilde{x}(t)\} = A(n) \cos(\theta_n) \\ s_Q(n) &= \mathcal{I}\{\tilde{x}(t)\} = A(n) \sin(\theta_n)\end{aligned}\quad (4)$$

Os sinais $s_I(n)$ e $s_Q(n)$ são sinais banda base enquanto $S(n)$ é um sinal passa banda que possui o espectro situado em torno de f_c . Essa representação é chamada representação cartesiana.

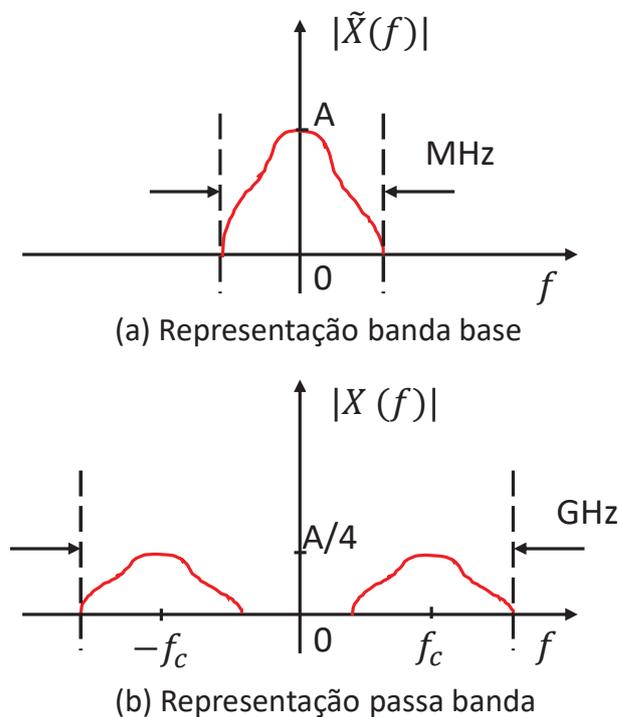
As componentes instantâneas de magnitude $A(n)$ e fase θ_n são definidas como:

$$A(n) = \sqrt{s_I(n)^2 + s_Q(n)^2} \quad \theta_n = \tan^{-1} \left(\frac{s_I(n)}{s_Q(n)} \right) \quad (5)$$

Denotando as densidades espectrais de potência de $S(n)$ e $\tilde{s}(n)$ como $X(f)$ e $\tilde{X}(f)$ obtêm-se a FIGURA 5. A parte real de $X(f)$ é uma função par e a sua parte imaginária uma função ímpar. Isso quer dizer que é possível reconstruir $S(n)$ unicamente de $X(f)$ usando $f > 0$ (BENEDETTO e BLIGLIERI, 1999).

A descrição da natureza dos sinais aqui demonstrada será útil na posterior análise dos componentes do transmissor e para os modelos baseados em redes neurais.

FIGURA 5 – SINAIS BANDA BASE E PASSA BANDA NO DOMÍNIO DA FREQUÊNCIA.



FONTE: O autor (2021)

2.2 CARACTERIZAÇÃO DAS DISTORÇÕES DO TRANSMISSOR SEM FIO

Em geral as distorções se referem às alterações do sinal devido a imperfeições nos componentes do circuito analógico do transmissor. Essa seção tem por objetivo repassar as diversas fontes de distorções no transmissor focando primariamente nas distorções geradas pelos componentes do modulador e do amplificador. Quando o transmissor for referenciado nesse trabalho entende-se o conjunto modulador-PA.

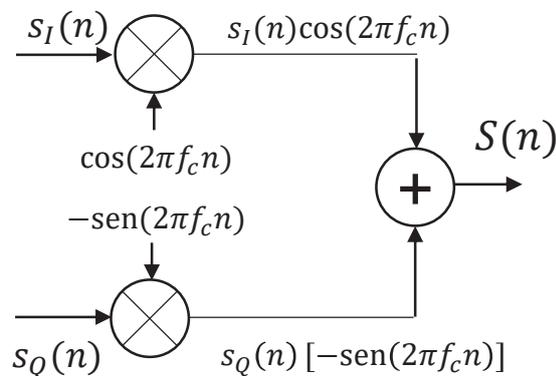
2.2.1 Modulador I/Q

O processo de modulação traduz uma informação que está em um sinal de frequência baixa para um sinal de frequência alta. O sinal de frequência mais baixa se chama envoltória e o sinal de frequência alta, portadora. O sinal de saída do modulador se chama sinal modulado. Além do benefício de se poder usar menores antenas devido à maior frequência do sinal, a modulação é necessária para transmitir

simultaneamente diversos canais banda base em diferentes portadoras sem que interferiram uma na outra (CHANG, 2000).

A FIGURA 6 mostra o diagrama simplificado de um modulador I/Q ideal (sem ganho no oscilador e erro de fase). Supondo um sinal banda base $\tilde{s}(n)$. A parte real é a componente $s_I(n)$ e a parte imaginária a componente $s_Q(n)$. Para modular a envoltória complexa $\tilde{s}(n) = s_I(n) + js_Q(n)$ da banda base para um sinal portadora passa banda $S(n)$, o fator $e^{j\omega_c n}$ com frequência f_c foi usado. O sinal assume, então, a forma da equação (1).

FIGURA 6 – DIAGRAMA SIMPLIFICADO DE MODULADOR EM QUADRATURA IDEAL

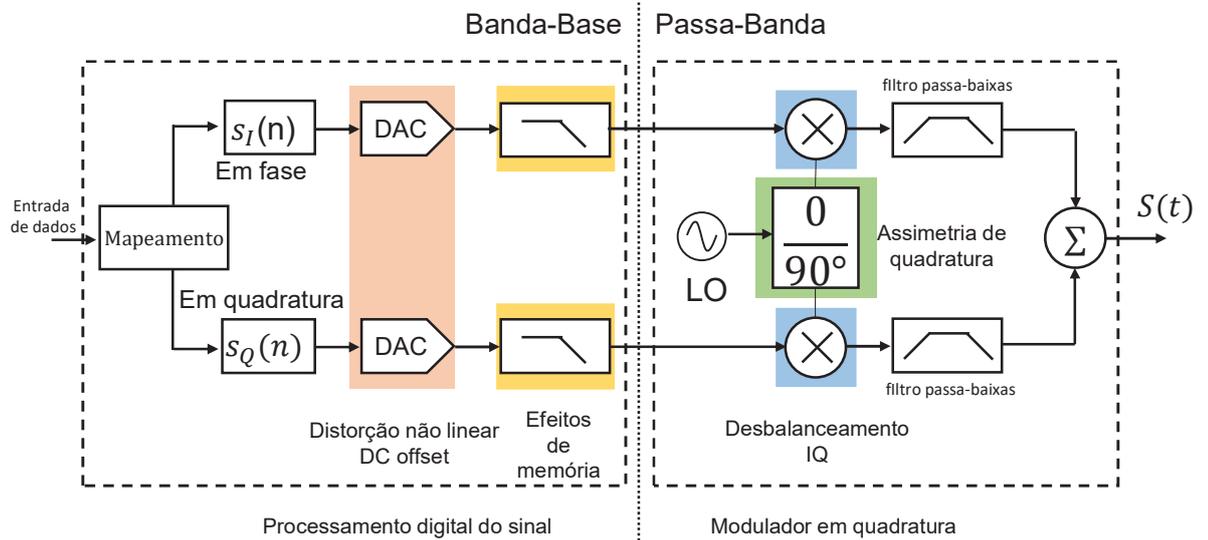


FONTE: O autor (2021)

A FIGURA 7 mostra um diagrama de um modulador não ideal e os efeitos não lineares que podem existir em um modulador. Conversores digitais para analógicos (DAC) transformam o sinal digital em analógico. A parte real e imaginária do sinal complexo banda base, $\tilde{s}(n)$, também chamados de componentes em fase (s_I) e em quadratura (s_Q), passam por caminhos ortogonais. No processo distorções não lineares são geradas, aparecendo um deslocamento da corrente contínua (*DC offset*). O sinal analógico passa então por filtros de reconstrução que introduzem atrasos no tempo e deslocamentos de fase. Esse fenômeno se mostra como efeitos de memória, onde amostras da saída dependem não só das amostras atuais de entrada como também de amostras passadas. Os sinais analógicos em banda base $s_I(n)$ e $s_Q(n)$ passam pelo conversor ascendente (*upconverter*) que os transforma em sinais passa banda na frequência da portadora (ω_c). Os misturadores de frequência multiplicam $s_I(n)$ e $s_Q(n)$ com um componente em fase e de quadratura ($+90^\circ$) do oscilador local (LO) respectivamente. As saídas do misturador são somadas para gerar o sinal RF

modulado $S(n)$. Desde que os sinais $s_I(n)$ e $s_Q(n)$ são modulados ortogonalmente eles efetivamente ocupam dois canais I e Q independentemente.

FIGURA 7 – COMPONENTES DO MODULADOR EM QUADRATURA E SUAS NÃO LINEARIDADES



FONTE: Modificado de (SOLTANI TEHRANI, 2012)

LEGENDA: O diagrama apresenta os principais componentes de um processo de modulação e as principais distorções causadas pelos respectivos componentes. São identificados os componentes do bloco de processamento digital de sinais e do modulador em quadratura.

O processo de conversão pode gerar um desbalanceamento de ganho entre os sinais I e Q, chamado desbalanceamento I/Q (FAULKNER e MATTSSON, 1992). Outro fenômeno que pode ocorrer é o sinal do oscilador local (LO) interferir no sinal original, e é chamado de *LO leakage*. Devido às perturbações elétricas, térmicas ou mecânicas, o oscilador pode ainda deteriorar a ortogonalidade dos dois sinais os tornando assimétricos.

Até agora foi descrita a resposta de um modulador ideal. Será analisado agora um cenário não-ideal onde o modulador I/Q é corrompido pelo desbalanceamento em ambos os ramos do modulador, I e Q, conforme apresentado na FIGURA 8. No ramo I os desbalanceamentos de amplitude e fase são caracterizados pelos termos $(1 + \alpha)$ e $\Delta\varphi$ respectivamente. No ramo Q os desbalanceamentos de amplitude e fase são caracterizados pelos termos $(1 - \alpha)$ e $-\Delta\varphi$ respectivamente.

A saída do modulador em banda base com desbalanceamentos em fase e amplitude é dada por:

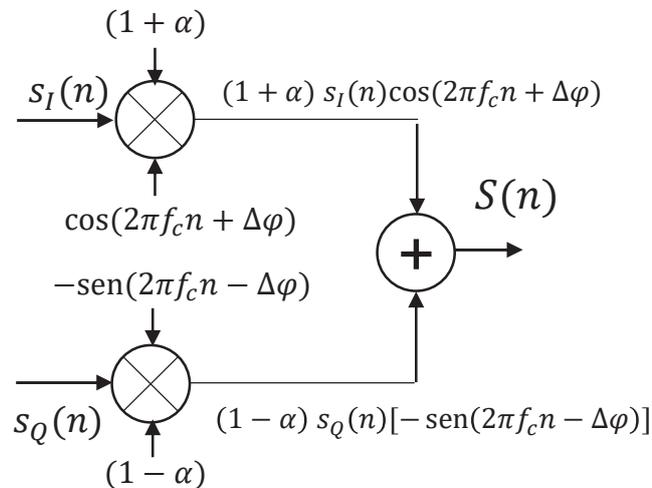
$$\begin{aligned}
 S_{\text{não ideal}}(n) = & (1 + \alpha)\cos\Delta\varphi\mathcal{R}\{\tilde{s}(n)\} - j(1 - \alpha)\text{sen}\Delta\varphi\mathcal{R}\{\tilde{s}(t)\} \\
 & + j(1 - \alpha)\cos\Delta\varphi\mathcal{I}\{\tilde{s}(n)\} - (1 + \alpha)\text{sen}\Delta\varphi\mathcal{I}\{\tilde{s}(n)\}
 \end{aligned} \tag{6}$$

$$= (\cos\Delta\varphi - j\alpha\text{sen}\Delta\varphi)\tilde{s}(n) - (\alpha\cos\Delta\varphi - j\text{sen}\Delta\varphi)\tilde{s}(n)^* \quad (7)$$

$$= k_1\tilde{s}(n) - k_2\tilde{s}(n)^*, \quad (8)$$

onde k_1 e k_2 são termos relacionados aos desbalanceamentos de amplitude e fase. O símbolo (*) representa o conjugado de um número complexo.

FIGURA 8 – DIAGRAMA SIMPLIFICADO DE MODULADOR EM QUADRATURA NÃO IDEAL



FONTE: O autor (2021)

2.2.2 Amplificador de Potência

O amplificador de potência tem como função amplificar linearmente o sinal de comunicação até certo nível a fim de evitar perdas na transmissão.

Um amplificador ideal deve reproduzir em sua saída uma réplica do sinal de entrada multiplicado por um valor escalar, ou seja:

$$P_{out}(t) = GP_{in}(t), \quad (9)$$

onde $P_{out}(t)$ e $P_{in}(t)$ são a potência de saída e de entrada do PA respectivamente. E G é o ganho em potência do amplificador. Pode-se definir matematicamente a linearidade de um operador L quando os princípios de homogeneidade, $L(\alpha x) = \alpha L(x)$, e adição, $L(x + y) = L(x) + L(y)$, são satisfeitos, sendo válida para qualquer par de funções x , y e escalar α (EDWARDS, 1995). A linearidade implica que o

amplificador não afeta a forma de onda do sinal transmitido nem introduz componentes na frequência.

Os amplificadores reais possuem diversos graus de não linearidades. As características das não linearidades do PA dependem principalmente de sua classe de operação e de sua topologia. Em termos de classe de operação o PA pode apresentar comportamento linear, classe A; comportamento de média não linearidade classe AB; e comportamento altamente não linear, classes C, D e E. Quanto à topologia é observado se a arquitetura do PA é *single-ended* (apenas um transistor) ou mais complexa tais como, Doherty, amplificação linear usando componentes não lineares (*LINC*), etc (M. GHANNOUCHI, HAMMI e HELAOUI, 2015).

O comportamento do PA é visualizado através das características de transferência da amplitude e da fase. A característica de transferência da amplitude é chamada de conversão modulação em amplitude/modulação em amplitude (AM-AM) e a da fase é chamada conversão modulação em amplitude/modulação em fase (AM-PM).

Os tipos de modulação no transmissor também impactam nas não linearidades do PA, pois elas dependem do nível de potência da entrada, ou em outros termos, da amplitude do sinal de entrada. Dessa forma, sinais modulados em fase por possuírem envoltória constante não são afetados pelas não linearidades do PA. Por outro lado, os esquemas de modulação que são utilizados atualmente precisam utilizar avançadas técnicas de multiplexação, como exemplo podem-se citar a multiplexação ortogonal por divisão de frequência (OFDM) e a acesso múltiplo por divisão de código (CDMA), que utilizam sinais modulados em amplitude com forte variação da envoltória (GANGWAR e BHARDWAJ, 2012).

Os sinais modulados podem ser caracterizados pelo PAPR dado por:

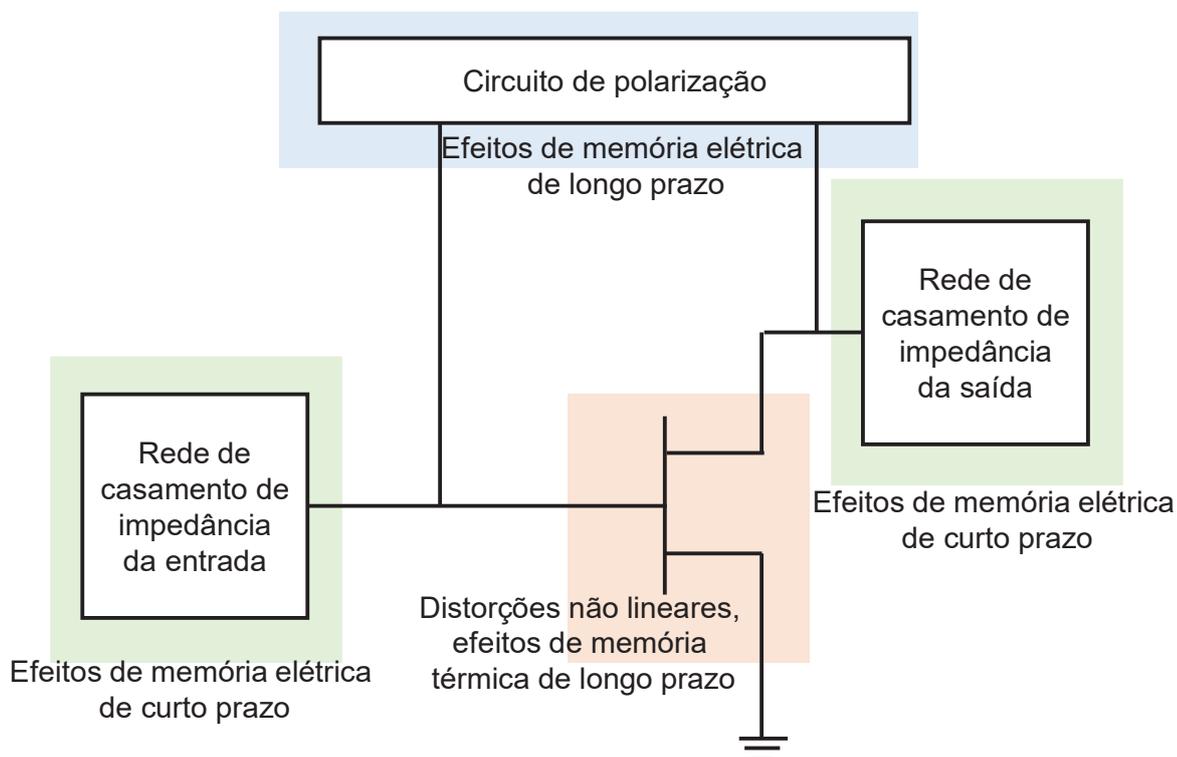
$$\text{PAPR}_{dB} = 10 \times \log_{10} \left(\frac{P_{max,W}}{P_{med,W}} \right) = P_{max,dB} - P_{med,dB}, \quad (10)$$

onde PAPR_{dB} é o valor do PAPR expresso em dB, $P_{max,W}$ e $P_{med,W}$ são as potências máxima e média do sinal em watts, respectivamente. Da mesma forma $P_{max,dB}$ e $P_{med,dB}$ são as potências máxima e média do sinal em dB, respectivamente. O PAPR de um sinal é um fator crítico quando se considera as não linearidades dos PAs.

O diagrama de blocos da FIGURA 9 mostra uma arquitetura simplificada do PA. Nesse diagrama estão destacadas as diferentes distorções criadas pelo PA. Essas distorções podem ser classificadas como:

- Distorções não lineares, que provêm primariamente da característica não linear do transistor;
- Efeitos de memória de curto prazo, derivados de atrasos no tempo ou desvios de fase, relacionados à rede de casamento de impedâncias e aos circuitos dos elementos do PA;
- Efeitos de memória de longo prazo, causados pelo circuito de polarização da rede, dependência térmica e outros.

FIGURA 9 – PRINCIPAIS COMPONENTES DO AMPLIFICADOR DE POTÊNCIA RF



FONTE: Modificado de (SOLTANI TEHRANI, 2012).

LEGENDA: Representação dos principais componentes de um amplificador RF e suas respectivas contribuições para o comportamento não linear do mesmo.

Na FIGURA 10 é mostrada a característica AM-AM de um PA classe AB. Observa-se desse gráfico que a relação entrada-saída não possui uma característica linear, e a resposta ideal difere da resposta real do PA. Do comportamento típico do PA mostrado na FIGURA 10 pode-se concluir que sinais de comunicação com amplitude variante são severamente impactados, pois o ganho não é uniforme.

O efeito das distorções do PA nos canais adjacentes pode ser observado na FIGURA 11, e é denominado crescimento espectral (do inglês *spectral regrowth*). Ele se caracteriza por produtos de intermodulação que interferem nos sinais do canal adjacente que impedem de serem satisfeitas as exigências impostas pelas organizações reguladoras (TAKAQI e WALKE, 2008). No exemplo da FIGURA 11 as distorções fora da banda são tão fortes que mascaram boa parte do sinal do usuário adjacente e distorce a sua comunicação.

FIGURA 10 – CARACTERÍSTICA AM-AM DE AMPLIFICADORES DE POTÊNCIA

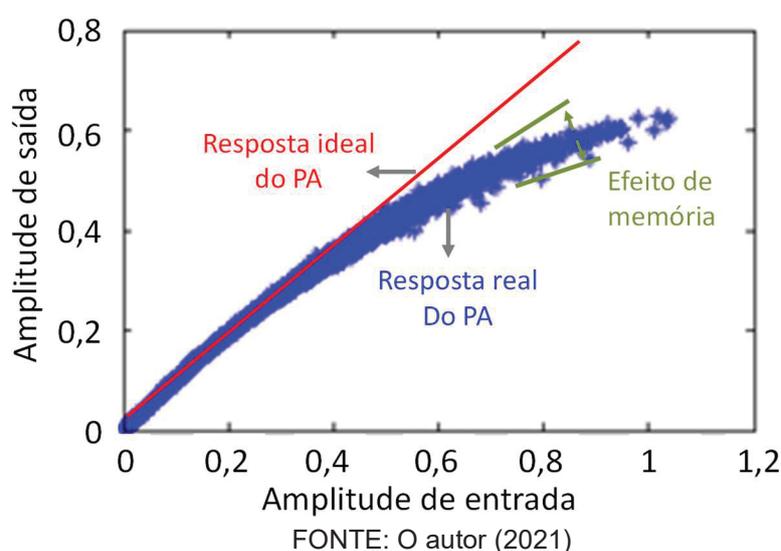
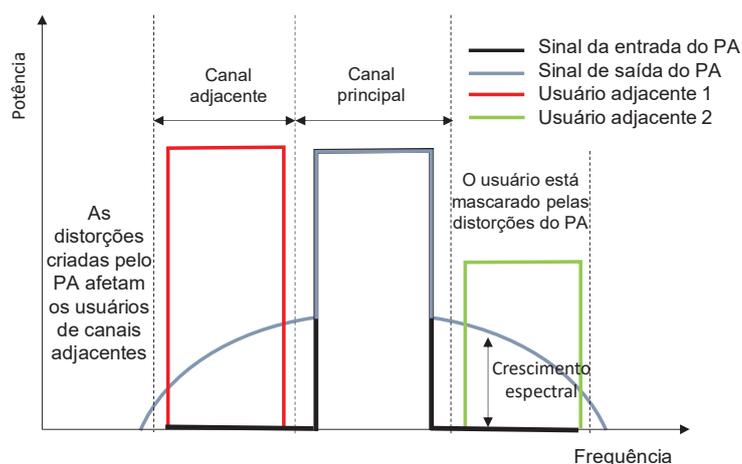


FIGURA 11 – EFEITO DAS DISTORÇÕES DO PA NOS CANAIS ADJACENTES



FONTE: Modificado de (SOLTANI TEHRANI, 2012)

2.3 COMPORTAMENTO PASSA BANDA DO PA

As redes de telecomunicações transmitem dados a partir de um sinal RF, com frequência na ordem de GHz, modulado por uma envoltória complexa com largura de banda na faixa de MHz. Como a largura de banda desses sinais é muito menor do que a frequência central, eles são classificados como sinais de banda estreita. O sinal de entrada de um PA é o sinal de saída do modulador e pode ser escrito matematicamente como:

$$x(n) = \mathcal{R}[\tilde{x}(n)e^{j\omega_c n}] = |\tilde{x}(n)| \cos[\omega_c n + \theta_n], \quad (11)$$

onde ω_c é a frequência central da portadora, θ_n é o ângulo de fase da envoltória complexa de entrada, n é o instante de tempo atual e \tilde{x} é o valor complexo da envoltória de entrada no tempo. Então, o sinal de saída do PA pode ser escrito como:

$$y(n) = \mathcal{R}[\tilde{y}(n)e^{j\omega_c n}] = |\tilde{y}(n)| \cos[\omega_c n + \theta_n + \varphi_n], \quad (12)$$

onde φ_n é a diferença entre as fases das envoltórias complexas de saída e de entrada.

2.4 MITIGANDO AS DISTORÇÕES DO TRANSMISSOR

A fim de alcançar os padrões de qualidade definidos para a comunicação sem fio foram desenvolvidas diversas técnicas que diminuem consideravelmente as distorções geradas no transmissor. Essas técnicas podem ser categorizadas em analógicas e digitais.

As técnicas analógicas mais comuns são *feedforward* (P. STAPLETON, FRASER e BURNABY, 1999), *feedback* (KIM, PARK, *et al.*, 2010) e pré-distorção (YI, YANG, *et al.*, 2000), (NEIL BRAITHWAITE, 2013). Esses modelos possuem a qualidade de serem capazes de lidar com sinais de banda larga, porém são sensíveis à frequência e são complexas e dispendiosas para implementar (LANDIN, 2012), (YI, YANG, *et al.*, 2000), (LEE, LEE, *et al.*, 2009).

A principal técnica digital para diminuir as distorções no transmissor é a pré distorção digital (DPD). Essa técnica mostrou possuir vários benefícios: diminuição da

complexidade e dos custos e ser independente da frequência de operação (ROBLIN, MYOUNG, *et al.*, 2008). Como a principal fonte de não linearidades do transmissor é o PA, usualmente se utiliza a DPD para linearizar o comportamento do PA. Primeiramente é necessário modelar a inversa do comportamento do PA. A DPD insere essa informação no sinal de entrada do PA. Quando o sinal distorcido então passa pelo PA, a inversa do comportamento anula as distorções para assim obter um comportamento linear em sua saída. Mais recentemente, a DPD está sendo utilizada para dirimir não só as distorções trazidas pelo PA como também pelo modulador I/Q (RAWAT e GHANNOUCHI, 2012), (GHANNOUCHI, YOUNES e RAWAT, 2013), (FU, LAW e THEIN, 2013), (RAWAT, RAWAT, *et al.*, 2013) e (FREIRE, MARCONDES e LIMA, 2018).

Com a expectativa das novas tecnologias da era 5G a modelagem do inverso do comportamento do PA não é trivial, menos ainda a modelagem do conjunto modulador-PA (DESGREYS, NARASIMHA MANYAM, *et al.*, 2017).

2.5 EXIGÊNCIAS DE MODELOS DE PA

Existem duas classes de modelos comportamentais: o passa banda, que lida com o sinal RF completo e o modelo equivalente banda base que processa apenas a informação da envoltória do sinal (PEDRO e MAAS, 2005). Por sua menor complexidade computacional são mais usados os modelos banda base. Dessa forma, os modelos aqui mencionados dizem respeito a modelos banda base a não ser que mencionado de outra forma.

Modelos que relacionam o sinal de envoltória complexa da entrada e saída do PA, geralmente são chamados de modelos comportamentais envoltória, banda base ou passa baixa (PEDRO e MAAS, 2005).

Modelos comportamentais banda base são insensíveis à informação da frequência da portadora, como pode ser deduzido de (12). Como consequência, se é submetido a uma operação arbitrária f , contribuições fora da banda são facilmente geradas pelo modelo comportamental banda base (BENEDETTO e BIGLIERI, 1979). Essas harmônicas podem ser chamadas de espúrios matemáticos já que são geradas pela natureza do modelo e não têm correlação com o comportamento direto do PA. Esses espúrios matemáticos comprometem a exatidão do modelo, além de aumentar sua complexidade computacional.

Uma forma de visualizar esse fenômeno é submeter um sinal passa-banda de (11) a um operador polinomial de 3º ordem, de acordo com:

$$f(x) = a_1x + a_2x^2 + a_3x^3, \quad (13)$$

onde a_1 , a_2 e a_3 são coeficientes constantes. Em (14) é possível ver que de todas as contribuições, apenas duas se encontram dentro da banda:

$$f(x) = a_1|\tilde{x}| \cos(\omega_c n + \theta) + \frac{3}{4}a_3|\tilde{x}|^3 \cos(\omega_c n + \theta), \quad (14)$$

uma vez que preservam um valor escalar unitário multiplicando ω_c . Mais detalhadamente, na FIGURA 12 podem ser observados todos os termos gerados e suas características. Os termos pares sempre geram contribuições não físicas (fora do canal principal). E apenas alguns termos ímpares geram contribuições dentro do canal principal ($1\omega_c$).

FIGURA 12 – CONTRIBUIÇÕES FORA DA BANDA

$$f(x) = a_1x \cos(1\omega_c n + \theta) + \frac{1}{2}a_2x^2 \cos(0\omega_c n + 0\theta) + \frac{1}{2}a_2x^2 \cos(2\omega_c n + 2\theta) + \frac{1}{4}a_3x^3 \cos(3\omega_c n + 3\theta) + \frac{3}{4}a_3x^3 \cos(1\omega_c n + \theta)$$

○ - termos ímpares
○ - termos pares
□ - contribuições físicas
↗ - termos fora do canal principal

FONTE: O autor (2021)

LEGENDA: Gráfico detalhado das contribuições do sinal de (11) aplicado a uma função arbitrária f : um polinômio de terceiro grau.

Deduz-se então que além de apresentarem um valor escalar unitário multiplicando ω_c , o sinal de envoltória complexa de saída estimado por um operador f , do modelo comportamental passa baixas, deve ter paridade ímpar com respeito ao sinal de envoltória complexa aplicado na entrada do PA (LIMA, CUNHA e PEDRO, 2011). É recomendada a adoção de modelos comportamentais banda base que não gerem contribuições fora da banda, visto que essas contribuições não trazem informação útil sobre o comportamento do PA dificultando a extração de um modelo de PA. Dessa forma, as restrições acima demonstradas, e que são associadas ao

comportamento passa banda dos PAs quando excitados por banda estreita, precisam ser impostas para garantir que apenas contribuições físicas dentro da banda sejam estimadas pelo modelo comportamental banda base (BENEDETTO e BIGLIERI, 1979).

Outro requisito que os modelos equivalentes banda base têm que respeitar para modelar de maneira completa o comportamento dos PAs é ser dinâmico. Ou seja, devem ser capazes de modelar as fortes características de memória que o PA apresenta na presença de modulações com banda larga.

2.6 CRITÉRIO DE AVALIAÇÃO DE MODELOS COMPORTAMENTAIS DE PA

Existem diversos modelos de PA baseados nos mais diversos métodos. Ainda não se têm um consenso de qual método é mais indicado, pois existem diversas variáveis a serem levadas em consideração, por exemplo a tecnologia do PA, a natureza do sinal de excitação, o consumo energético disponível dentre outros. Alguns trabalhos visam comparar modelos dentro de um cenário de estudo, porém basta modificar uma característica do sistema e não se tem a certeza das respostas dos modelos a esse novo sistema. O que se pode observar são as características gerais dos modelos e se conseguem mapear todas as características do sistema a que se propõe modelar e sua dificuldade de implementação.

Dessa forma, existem dois critérios que se deve levar em consideração ao avaliar um modelo comportamental de PA. O primeiro é a acurácia. Ela mede o quão bem o modelo consegue representar o sinal de saída do PA. Porém, esse critério por si só não é suficiente, pois um modelo pode possuir incontáveis parâmetros e ser capaz de modelar qualquer função não linear, inviabilizando sua aplicação prática. Então o segundo critério a se levar em consideração é o esforço computacional necessário para a construção do modelo.

2.6.1 Acurácia da modelagem

Na literatura existem muitas métricas para se avaliar a predição de um modelo em comparação com a resposta medida de um PA. Comparações visuais podem ser feitas ao se traçar gráficos das respostas AM-AM e AM-PM do modelo e do PA. Da mesma forma podem ser utilizados gráficos dos sinais de saída do modelo e do PA

no domínio do tempo ou da frequência para comparação. O método auxiliar visual que será utilizado nesse trabalho será a densidade espectral de potência (PSD). O PSD de um sinal permite analisar sua distribuição energética em toda faixa de frequência.

Esses métodos, porém, não substituem os métodos quantitativos servindo como ferramentas auxiliares.

As figuras de mérito (FoM) são utilizadas para definir a acurácia de um modelo. Na literatura diversas FoM foram analisadas e comparadas (LANDIN, 2012). As métricas razão de efeito de memória (*memory effect ratio* - MER), razão de modelagem do efeito de memória (*memory effect modeling ratio* - MEMR) e erro quadrático médio normalizado (*normalized mean square error* - NMSE) apresentam características quantitativas equivalentes. De fato MER e MEMR são NMSEs em diferentes escalas. Essas métricas podem ser utilizadas para mensurar o erro dentro da banda de operação do PA, isso porque utilizam medidas de potência e a soma dos erros ao quadrado. E como a maior parte da energia se encontra dentro da banda, os erros dentro da banda são dominantes.

Será utilizado ao longo do trabalho a FoM NMSE e é comumente definida como:

$$NMSE = 10 \log_{10} \frac{\sum_n |out_{medido}(n) - out_{modelo}(n)|^2}{\sum_n |out_{medido}(n)|^2}, \quad (15)$$

onde n é a amostra atual, out_{medido} é o sinal de saída medido do PA e out_{modelo} é o sinal de saída de seu modelo. Essa métrica consta em vários trabalhos e pode se adquirir uma noção de valores adequados para uma boa acurácia de modelos de PA (ISAKSSON, WISSEL e RONNOW, 2006), (TEHRANI, CAO, *et al.*, 2010), (MORGAN, MA, *et al.*, 2006), (JARAUT, ABDELHAFIZ, *et al.*, 2021). Desses trabalhos pode ser observado que ao se modificar o sistema a ser modelado os valores de acurácia dos modelos se alteram de maneira significativa. A escolha do modelo mais adequado é inerente às características do sistema a ser modelado. Observa-se variações nos valores de NMSE dentro de uma faixa de -30 dB até -50 dB.

As métricas utilizadas para os erros fora da banda são a razão de potência do erro de canal adjacente (*adjacent channel error power ratio* - ACEPR) e razão da potência erro-sinal ponderada (*weighted error-to-signal power ratio* - WESPR), e se

assemelham nas respostas obtidas para análise. A métrica ACEPR será utilizada para avaliar os modelos deste trabalho e é definida como:

$$\text{ACEPR} = 10 \log_{10} \left[\frac{\int_{adj} E(f)^2 df}{\int_{ch} |Y_{ref}|^2 df} \right], \quad (16)$$

onde $E(f)$ é a transformada discreta de Fourier do sinal de erro e $Y_{ref}(f)$ é a transformada discreta de Fourier do sinal de referência. A integral no denominador é sobre o canal principal, enquanto a integral no numerador é sobre o canal adjacente superior ou inferior de mesma largura de banda que o canal principal. Em (SCHREURS e ET AL., 2008) o ACEPR é descrito como a métrica de baixa complexidade que melhor descreve desbalanceamentos não lineares. De acordo com os estudos da literatura pode ser avaliado que valores adequados de ACEPR variam em torno de -40 dB e -60 dB.

Quanto mais similar o sinal de saída calculado do modelo é do sinal experimental, menores os valores de NMSE e de ACEPR, isto é, melhor a acurácia de um determinado modelo. Porém, nem sempre um modelo que possui os melhores valores de NMSE possui também os melhores valores de ACEPR.

A acurácia de um modelo não é uma característica suficiente para determinar sua aplicabilidade. Deve se levar em conta também a complexidade computacional do mesmo.

2.6.2 Complexidade de modelos comportamentais

Existem diferentes complexidades dependendo do aspecto do modelo a que se está referindo. As complexidades podem ser classificadas em complexidade de identificação, complexidade de adaptação e complexidade de operação.

Complexidade de identificação se refere ao processo de treinamento e aquisição de parâmetros do modelo. Como é feito *a priori*, desconectado do sistema de linearização, não tem um valor tão relevante.

A complexidade de adaptação deve ser levada em conta em sistemas práticos onde devido a variações pequenas no comportamento do PA (em razão de mudança na temperatura por exemplo), o modelo deve ser atualizado em certos intervalos de

tempo. Quando essas variações são pequenas e gradativas essa complexidade é de menor importância.

Por último, a complexidade de operação é definida pelo número de operações que é feita para cada amostra quando o modelo é utilizado. Essa complexidade limita o sistema, desde que é um problema de tempo real. A máxima complexidade aceitável varia de acordo com o sistema. Em uma estação rádio base pode ser utilizado modelos com maiores complexidades enquanto que para aparelhos móveis os requisitos são estritos.

Para a comparação de modelos nesse trabalho, a fim de analisar certas características teóricas independentes da complexidade de operação, será utilizada a complexidade de identificação caracterizada pela quantidade de parâmetros das redes neurais. No estudo de caso apresentado na Seção 6.3 os modelos também serão avaliados quanto à sua complexidade de operação utilizando a quantidade de operações em vírgula flutuante (FLOP).

3 REDES NEURAIS ARTIFICIAIS

O interesse por redes neurais artificiais se deu inicialmente após a introdução do conceito simplificado dos neurônios artificiais dada por McCulloch e Pitts em 1943 (MCCULLOCH e PITTS, 1943). Os neurônios artificiais foram apresentados como modelos dos neurônios biológicos e como possíveis componentes para circuitos que poderiam realizar tarefas computacionais.

O assunto teve um declínio nas pesquisas quando o livro *perceptrons* foi lançado em 1969 (MINSKY e PAPERT, 1969), onde várias deficiências do modelo foram analisadas.

Com a introdução do conceito de retro propagação do erro e o avanço das tecnologias de processamento de dados na década de 80 as redes neurais têm sido objeto de estudo em várias áreas (KRÖSE e VAN DER SMAGT, 1996).

Nas próximas seções a parte teórica sobre redes neurais será abordada. Posteriormente será introduzida a aplicação das redes neurais para modelagem do PA e do transmissor.

3.1 UM MODELO DE NEURÔNIO

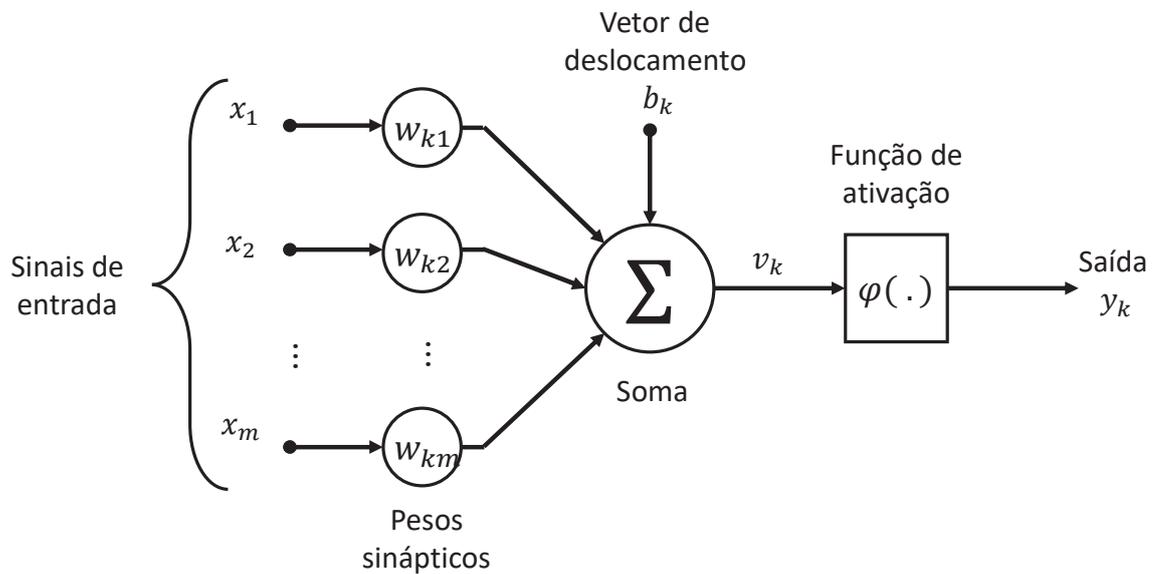
Um neurônio é uma unidade fundamental das redes neurais responsável pelo processamento de informação. Na FIGURA 13 observa-se as partes essenciais de um neurônio:

- a) Um conjunto de sinapses, ligações entre os neurônios onde cada uma é caracterizada por um peso excitatório ou inibitório. Diferentemente de uma sinapse de um neurônio biológico, os pesos sinápticos de um neurônio artificial podem possuir valores negativos bem como positivos; mais especificamente, um sinal x_j na entrada da sinapse j conectada ao neurônio k é multiplicada pelo peso sináptico w_{kj} ;
- b) Um somador, que soma os sinais de entrada ponderados pelo respectivo peso da sinapse. Essa operação é descrita como uma combinação linear;

- c) É uma função de ativação que limita a amplitude do sinal do neurônio. Normalmente a variação da amplitude do sinal de saída do neurônio está entre $[0,1]$ ou $[-1,1]$.

O modelo da FIGURA 13 também inclui um vetor de deslocamento b_k externamente aplicado. O deslocador tem a função de aumentar o grau de liberdade do modelo, o que melhora a adaptação do modelo aos dados a ele fornecidos tornando-o menos dependente da amostra (GUARIENTO, 2014).

FIGURA 13 – MODELO NÃO LINEAR DE UM NEURÔNIO k



FONTE: Modificado de (HAYKIN, 2009)

A combinação linear da FIGURA 13 pode ser escrita como:

$$u_k = \sum_{j=1}^m w_{kj} x_j. \quad (17)$$

Dessa forma a saída y_k pode ser descrita matematicamente como:

$$y_k = \varphi(u_k + b_k), \quad (18)$$

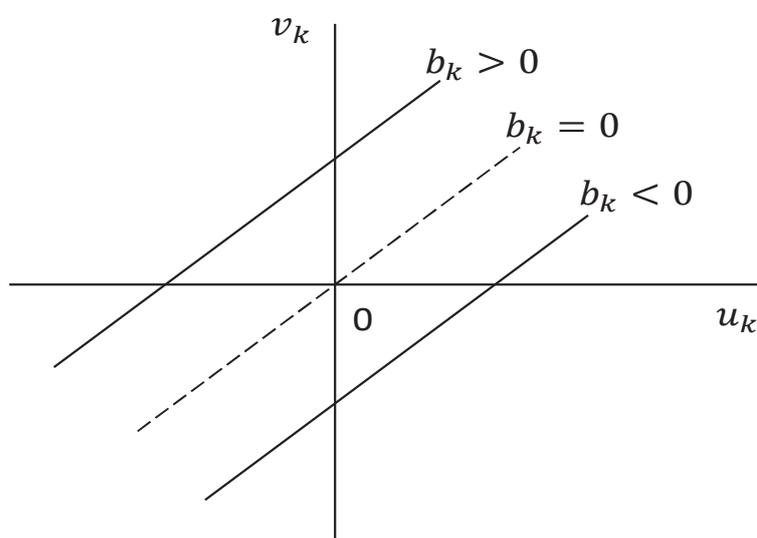
onde x_1, x_2, \dots, x_m são os sinais de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos do neurônio k ; u_k é a saída da combinação linear (não mostrada na FIGURA 13); b_k é o vetor de deslocamento; $\varphi(\cdot)$ é a função de ativação; e $y(k)$ é o sinal de saída do neurônio.

O vetor de deslocamento tem a função de aplicar uma *transformação afim* (WEISSTEIN, 2018) à saída $u(k)$ da combinação linear como definido em:

$$v_k = u_k + b_k, \quad (19)$$

onde v_k pode ser definido como potencial de campo local (LFP) (termo advindo do neurônio biológico), ou potencial de ativação. A FIGURA 14 demonstra a relação entre v_k e u_k quando b_k é negativo ou positivo (HAYKIN, 2009).

FIGURA 14 – TRANSFORMAÇÃO AFIM



FONTE: Modificado de (HAYKIN, 2009)

LEGENDA: Transformação afim causada pelo vetor de deslocamento b_k . Notar que para $u_k = 0$, $v_k = b_k$.

3.2 ARQUITETURAS E TOPOLOGIAS DAS REDES NEURAIS

Essa seção foca no padrão de interconexão dos neurônios e a propagação da informação. A arquitetura define a forma com que os neurônios artificiais se interconectam e como são dispostos. Existem diversos tipos de arquiteturas que se diferenciam de acordo com algumas características: disposição dos neurônios, sua interconexão e a forma como suas camadas são arranjadas.

Dentro de uma mesma arquitetura as redes neurais podem se diferenciar de acordo como o número de neurônios, o tipo de função de ativação, etc. O termo para descrever esse tipo de diversidade estrutural é topologia. Por exemplo, uma rede direta que possui 2 neurônios possui uma topologia diferente de uma rede direta que

possui 4 neurônios, porém as duas redes compartilham a mesma arquitetura. Um compêndio de diversas arquiteturas e topologias pode ser visto em (VAN VEEN, 2016).

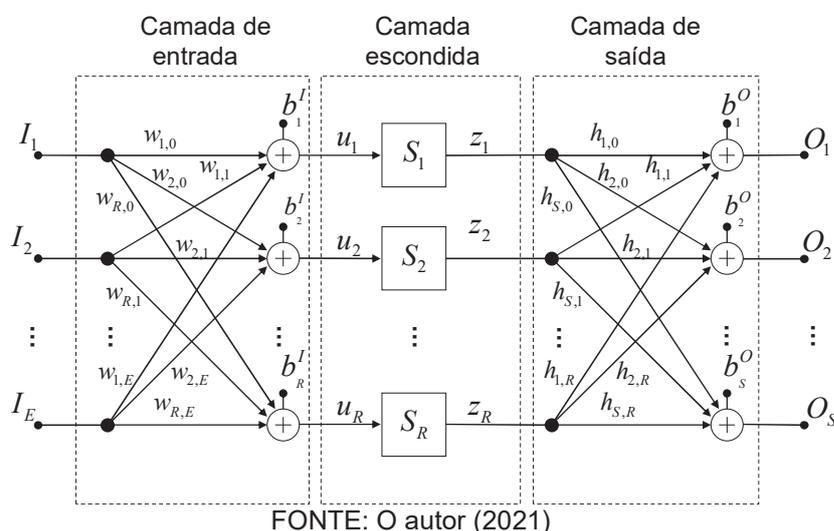
Existem dois tipos principais de arquitetura em termos de interconexões neurais (KRÖSE e VAN DER SMAGT, 1996):

(i) Rede direta

Arquitetura com uma camada de entrada, uma de saída e uma ou várias camadas escondidas com fluxo de informação em apenas um sentido (as conexões são feitas sempre com o neurônio da camada seguinte). O termo “escondidas” se refere ao fato de que parte da rede neural não é vista diretamente pela entrada ou pela saída da rede (CHURCHLAND e SEJNOWSKI, 1992). A camada escondida é responsável pelo processamento da informação (GHANNOUCHI, YOUNES e RAWAT, 2013). Ao se adicionar mais camadas se tem um conjunto extra de conexões sinápticas e de dimensão de interações neurais permitindo extrair informações de maneira mais eficiente do conjunto de dados à que é exposto.

Na FIGURA 15 é mostrado o diagrama de blocos de uma arquitetura típica MLP de três camadas (TLP) tendo E entradas, R neurônios na camada escondida e S saídas. E os nós de cada camada são totalmente conectados, significando que um neurônio de qualquer camada está conectado a todos os neurônios da camada anterior. Em um TLP, as E entradas são inicialmente multiplicadas pelos pesos de entrada (w) e somadas aos $bias$ de entrada (b^I) para então serem aplicadas à função de ativação (S_R) e finalmente serem multiplicadas pelos pesos de saída (h) e somadas aos $bias$ de saída (b^O). Essa arquitetura é a mais utilizada na literatura.

FIGURA 15 – ARQUITETURA DIRETA E TOPOLOGIA MLP COM TRÊS CAMADAS



Os parâmetros da rede (w, b^I, h, b^O) são normalmente identificados pelo algoritmo de retro propagação do erro (BP) (HAYKIN, 2001). Os valores iniciais dos parâmetros são de escolha randômica ou otimizados por outro algoritmo externo. O algoritmo BP trabalha de forma iterativa realizando 2 processos a cada iteração. No primeiro passo os sinais de entrada são propagados pela rede a fim de calcular o sinal de erro definido pela diferença entre o sinal estimado e o desejado. No segundo passo, o sinal de erro é retro propagado pela rede atualizando os valores de todos os parâmetros.

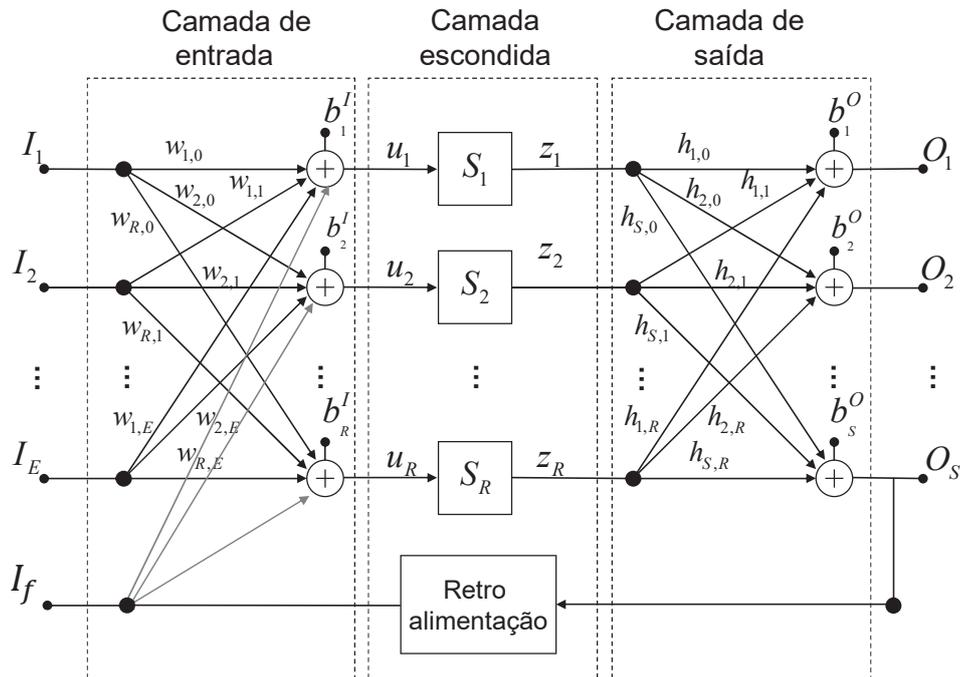
(ii) Rede recorrente

As redes recorrentes são caracterizadas por possuírem pelo menos um laço de realimentação. Os laços de realimentação são ramificações que possuem elementos de atraso de unidade de tempo (Z^{-1}), que agregam comportamentos não lineares dinâmicos à rede (se a rede possuir elementos não lineares) (HAYKIN, 2009). Dessa forma, a rede pode processar sinais de sistemas variantes no tempo, como predição de séries temporais, identificação e otimização de sistemas, etc (NUNES DA SILVA, HERNANE SPATTI, *et al.*, 2017).

Essa arquitetura pode ou não possuir camadas escondidas, bem como pode ter realimentação própria ou não. Realimentação própria se diz quando a saída de um neurônio é enviada para alimentar a sua própria entrada (HAYKIN, 2009). No caso da FIGURA 16, os sinais de saída dos neurônios da camada de saída são enviados para alimentar as entradas dos neurônios da camada de entrada.

As topologias mais utilizadas desse tipo de arquitetura são a *Hopfield* (HOPFIELD, 1982) e a *perceptron* com realimentação (NUNES DA SILVA, HERNANE SPATTI, *et al.*, 2017).

FIGURA 16 – ARQUITETURA RECORRENTE



FONTE: modificado de (NUNES DA SILVA, HERNANE SPATTI, *et al.*, 2017).

3.3 REDES TLP DE VALORES COMPLEXOS

Antes de discutir as redes neurais de valores complexos é necessário entender alguns conceitos sobre números complexos.

O plano complexo \mathbb{C} é uma extensão ao plano real \mathbb{R} . Define-se um número complexo como $\tilde{x} = x_r + jx_i$, onde $j = \sqrt{-1}$ é a unidade imaginária e x_r, x_i são a parte real e imaginária respectivamente. Usualmente se utilizam os símbolos j ou i para designar a unidade imaginária. A representação cartesiana descreve um número complexo com suas partes reais e imaginárias, e em um sistema cartesiano representa as coordenadas de um vetor. É também possível descrever um número complexo usando componentes de amplitude e ângulo, chamando assim de representação polar (HIROSE, 2011).

A unidade imaginária j não tem o papel apenas de representar a parte imaginária de um sinal, mas também de conectar e combinar as componentes de amplitude e de fase, além de gerar uma dependência estatística. Por exemplo, ao se multiplicar números complexos o operador j converte a parte imaginária em parte real, mas colocando um sinal negativo. Focando na operação aritmética de multiplicação de números complexos, é possível representar um número complexo como uma

matriz 2x2. Isto quer dizer que, para todo número complexo $c = a + jb$ e $z = x + jy$ onde a, b, x e y são números reais, pode-se associar uma transformação linear $F = \mathbb{C} \rightarrow \mathbb{C}$, onde $F(a, b, x, y) = (a + jb, x + jy)$

$$f: z \mapsto cz = ax - by + j(bx + ay), \quad (20)$$

segue que,

$$F = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix} = \begin{pmatrix} a - b \\ b \ a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (21)$$

Pode-se expressar as funções de atenuação ou amplificação da amplitude e de rotação da fase específicas dos números complexos como:

$$\begin{pmatrix} a - b \\ b \ a \end{pmatrix} = r \begin{pmatrix} \cos\theta - \text{sen}\theta \\ \text{sen}\theta \ \cos\theta \end{pmatrix}, \quad (22)$$

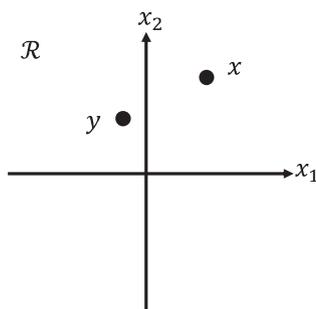
onde $r \equiv \sqrt{a^2 + b^2}$ e $\theta \equiv \arctan b/a$ são a modulação de amplitude e rotação de ângulo do sinal de valor complexo z , respectivamente.

Em uma rede neural o peso w de valor complexo é multiplicado paralelamente com as entradas, misturando as partes real e imaginária. Essa iteração tem dependência direta com os valores de w . Assim a saída da rede mantém uma relação vetor-direção no plano complexo (HIROSE, 2012).

Uma rede neural de valor real, onde os sinais de entrada são as componentes reais de um sinal complexo, não tem as mesmas características estatísticas que uma rede neural totalmente complexa (HIROSE e YOSHIDA, 2012).

Em (HIROSE, 2012) é sugerido um exemplo para demonstrar as diferenças entre as redes de valor complexo e valor real e como a natureza dos números complexos implica em um menor grau de liberdade. Suponha um mapeamento de um sinal de entrada $x = \{x_1, x_2\}$ para um sinal de saída $y = \{y_1, y_2\}$, FIGURA 17, feito por uma rede neural direta de valores reais com uma camada, 2 entradas e 2 saídas, como mostrado na FIGURA 18 (a).

FIGURA 17 – MAPEAMENTO DE DOIS SINAIS REAIS



FONTE: (HIROSE e YOSHIDA, 2012)

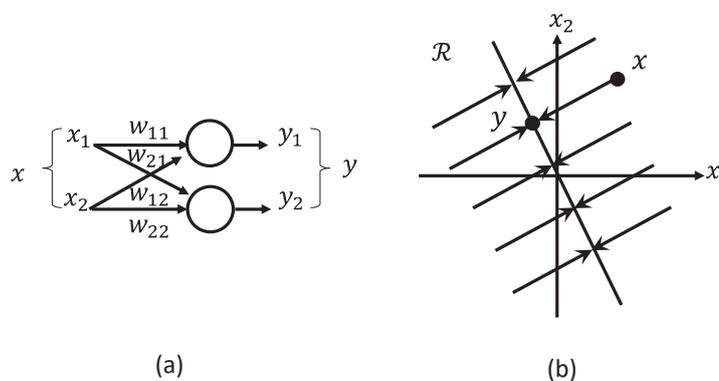
LEGENDA: Gráfico para exemplificação de mapeamento de dois sinais reais, x_{in} e x_{out} .

A função de ativação é a função identidade. A relação entrada-saída pode então ser representada por:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \quad (23)$$

Dessa forma se tem uma variedade alta de possibilidades de mapeamentos porque o número de parâmetros a serem determinados é maior que as condições, ou seja, esse mapeamento é um problema mal condicionado. Um possível mapeamento é a solução degenerada dada na FIGURA 18 (b), que não é uma solução viável. Sabe-se também da teoria de aprendizagem de máquinas que quanto mais complexa uma rede em termos de número de parâmetros maior a tendência de ocorrer sobre ajuste

FIGURA 18 – REDE NEURAL PARA MAPEAMENTO DE DOIS SINAIS REAIS.



FONTE: (HIROSE e YOSHIDA, 2012)

LEGENDA: (a) rede neural de valor real com 2 camadas, uma de entrada e uma de saída; (b) a linha que passa pelo ponto y é uma possível resposta da rede neural.

do modelo ao conjunto de dados de treinamento. O sobre ajuste de um modelo ocorre quando seus parâmetros se ajustam de forma exacerbada ao sinal de treinamento, e dessa forma perdem a capacidade de traduzir de forma adequada a resposta de um sistema quando outro sinal de entrada é aplicado. Para avaliar um modelo então é necessário testá-lo com um conjunto de dados de validação para avaliar sua generalização.

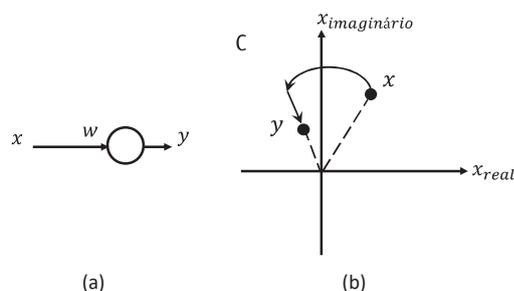
A FIGURA 19 (a) mostra a rede de valor complexo com um neurônio de valor complexo para realizar o mesmo mapeamento da FIGURA 17. A operação da rede pode ser expressa como:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} |w|\cos\theta & -|w|\sin\theta \\ |w|\sin\theta & |w|\cos\theta \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad (24)$$

onde $\theta \equiv \arg(w)$. A FIGURA 19 (b) mostra a resolução do problema. O grau de liberdade e a arbitrariedade da solução é diminuída. O mapeamento é uma combinação de rotação de fase e atenuação de amplitude. Apesar da exemplificação acima ser simplista e a dinâmica das redes neurais dependerem de vários outros fatores, as características de rotação de fase e modulação da amplitude estão embutidas nas redes neurais de valores complexos como um processo elementar e universal de ponderação.

Então, usar operações de valor real para aproximar parâmetros complexos pode resultar em um modelo com generalização indesejada. De outra forma, no domínio complexo, a representação em matriz é similar à matriz de rotação, ou seja, metade das entradas está fixa quando a outra metade é conhecida. Essa restrição diminui o grau de liberdade e melhora a capacidade de generalização do modelo.

FIGURA 19 – REDE NEURAL DE VALOR COMPLEXO PARA MAPEAMENTO DE DOIS SINAIS REAIS



FONTE: (HIROSE e YOSHIDA, 2012)

LEGENDA: (a) rede neural de valor complexo com 1 camada de entrada e 1 camada de saída; (b) visualização do mapeamento da rede neural complexa.

A história das redes neurais complexas começou no advento da formulação do algoritmo de mínimos quadrados (LMS – *least mean square*) de valores complexos (WIDROW e HOFF, 1960). Recentemente foi despertado intenso interesse pelas redes neurais de valores complexos (BASSEY, QIAN e LI, 2021), (TRABELSI e ET AL., 2018). Algumas aplicações se beneficiam das características que as redes de valores complexos possuem, mais especificamente as que envolvem sinais senoidais, ou que tenham uma natureza ondulatória. É esperado que as redes neurais de valores complexos consigam refletir melhor o mundo natural composto por partículas que interagem entre si através da amplitude (energia) e da fase (tempo) (SARROFF, 2018).

4 ESTADO DA ARTE DE MODELOS MATEMÁTICOS PARA USO EM SISTEMAS DE PRÉ DISTRIBUIÇÃO DIGITAL EM TRANSMISSORES BASEADOS EM REDES *PERCEPTRON* DE MÚLTIPLAS CAMADAS

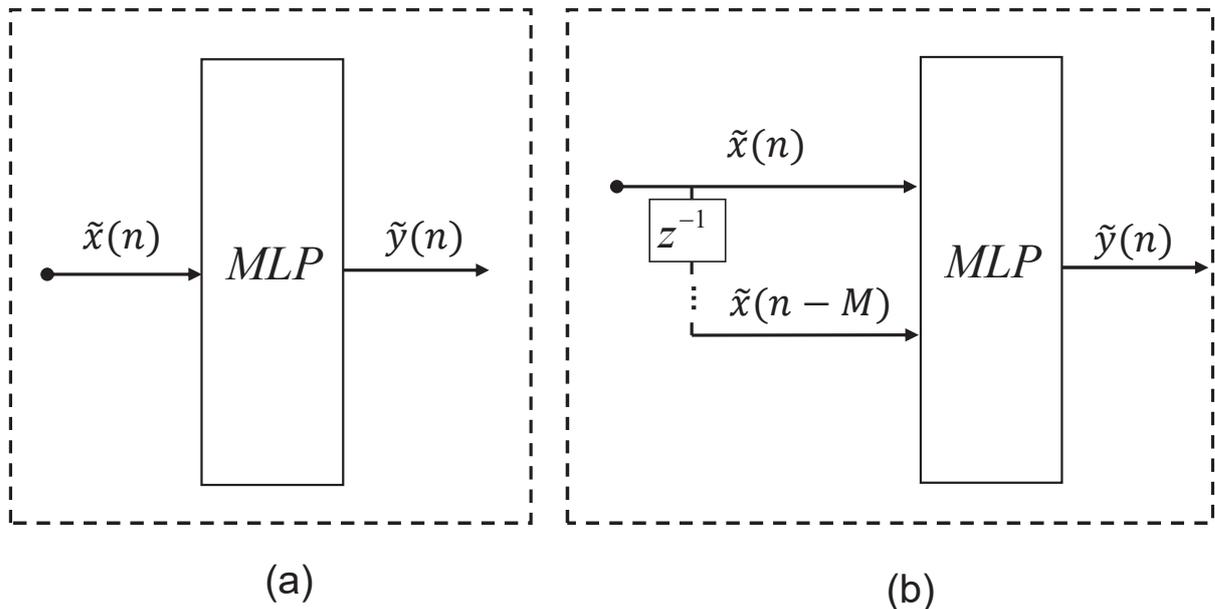
Esse capítulo aborda os modelos baseados em redes neurais artificiais de múltiplas camadas para uso em sistemas DPD. Na Seção 4.1 estão os principais modelos de PAs; e na Seção 4.2 os modelos conjuntos de PAs e moduladores I/Q. A rede MLP utilizada nos modelos das seções seguintes tem como estrutura base a rede MLP descrita no Capítulo 3.

4.1 MODELOS DE AMPLIFICADORES DE POTÊNCIA BASEADOS EM *PERCEPTRON* DE MÚLTIPLAS CAMADAS

As primeiras redes utilizadas para modelagem de PAs eram redes de múltiplas camadas treinadas com o algoritmo de retropropagação de erro, (ZHANG, 1990), (CHEN, GIBSON, *et al.*, 1990), (BENVENUTO, PIAZZA e UNCINI, 1993) e (IBNKAHLA, SOMBRIN, *et al.*, 1997) e (NASKAS e PAPANANOS, 2002). Inicialmente se utilizavam redes complexas. Pode-se especular que isso se deva ao fato de os PAs (HIROSE e YOSHIDA, 2012) lidarem com valores complexos, além de não existirem na época ferramentas computacionais prontas para se lidar apenas com sinais reais como se tem atualmente. Novas bibliotecas de redes neurais de valores reais facilitaram e incentivaram seu uso, pois não era preciso entender o funcionamento interno das redes neurais e apenas se preocupar com o processamento do sinal que iria ser inserido na rede. Dessa forma, ainda hoje vários estudos tratam as redes neurais como caixas pretas (RUDIN, 2019) e (DUKE, 2020).

A FIGURA 20 mostra dois modelos utilizando redes de valores complexos mais utilizados, o modelo (a) é proposto em (IBNKAHLA, SOMBRIN, *et al.*, 1997) e o modelo (b) em (BENVENUTO, PIAZZA e UNCINI, 1993). No modelo (a) da FIGURA 20 a entrada da rede é a amostra n do sinal banda base de entrada do PA, $\tilde{x}(n)$. O sinal de saída do modelo que a rede deve simular é o sinal banda base de saída do PA da amostra n , $\tilde{y}(n)$. Essa rede não considera os efeitos de memória do PA.

FIGURA 20 – REDES DE VALORES COMPLEXOS DA LITERATURA



FONTE: (IBNKAHLA, SOMBRIN, *et al.*, 1997) e (BENVENUTO, PIAZZA e UNCINI, 1993)

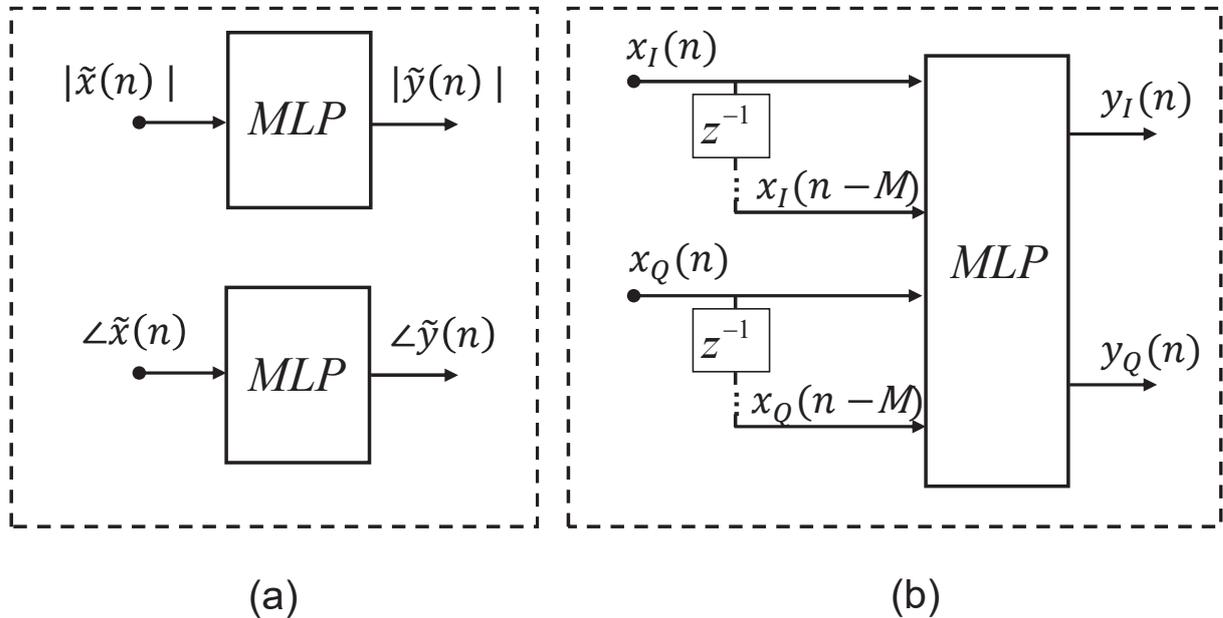
LEGENDA: A rede em (a) é sem memória e processa os sinais de saída e entrada complexos de um PA, e dessa forma possui parâmetros e função de ativação de valores complexos. O modelo em (b) possui rede igual a rede em (a) com o diferencial de adicionar memória através da inserção das amostras passadas do sinal de entrada.

O modelo (b) da FIGURA 20 possui como entradas da rede a amostra atual e M amostras passadas do sinal $\tilde{x}(n)$. A saída desejada nesse modelo continua sendo o sinal de saída do PA, $\tilde{y}(n)$. Ao se introduzir as amostras passadas, o modelo ganha a capacidade de modelar os efeitos de memória de um sistema.

A FIGURA 21 mostra dois modelos que utilizam redes de valores reais. O modelo (a) foi proposto em (NASKAS e PAPANANOS, 2002), e o modelo (b) em (LIU, BOUMAIZA e GHANNOUCHI, 2004). No modelo (a) da FIGURA 21 são utilizadas 2 redes reais. Cada rede é responsável por modelar cada componente polar do sinal banda base de entrada do PA, $\tilde{x}(n) = |\tilde{x}(n)| \exp j \angle \tilde{x}(n)$. Então a entrada de uma das redes é a amostra n da componente $|\tilde{x}(n)|$ e da outra rede a amostra n da componente $\angle \tilde{x}(n)$. O sinal de saída de cada rede é a amostra n das componentes polares do sinal banda base de saída do PA, $\tilde{y}(n) = |\tilde{y}(n)| \exp j \angle \tilde{y}(n)$. Esse modelo não insere os efeitos de memória do PA. Além disso essas redes modelam de forma desvinculada ambas as componentes. É interessante ressaltar que as componentes do sinal do PA possuem dependência uma da outra e haverá perda de precisão do modelo ao não se considerar essa característica. Todos os modelos analisados incluirão como

entradas das redes TLP ambas as componentes do sinal do PA, seja na forma polar ou cartesiana.

FIGURA 21 – REDES DE VALORES REAIS NA LITERATURA



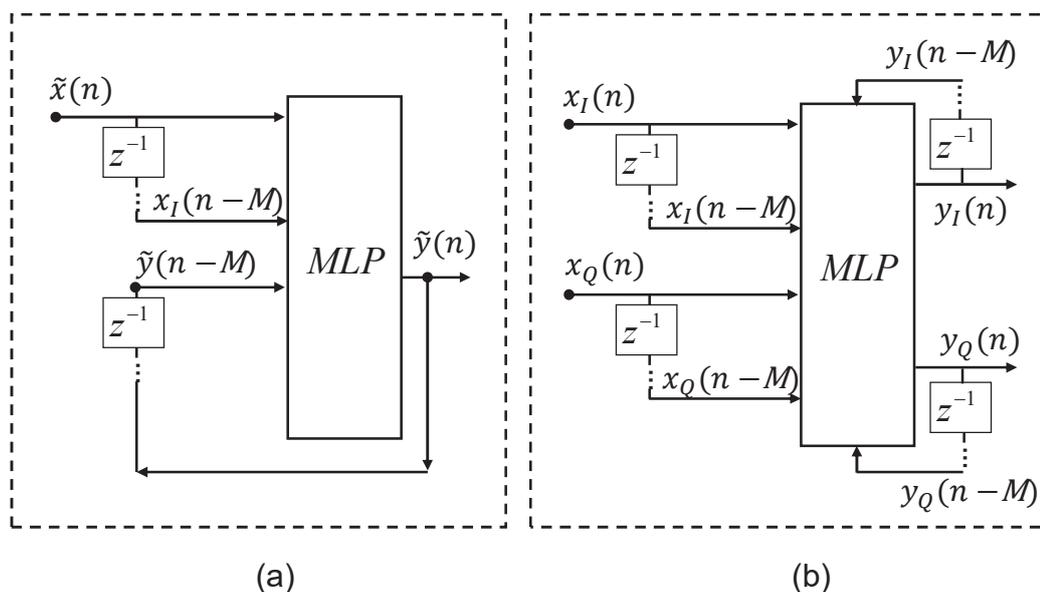
FONTE: (NASKAS e PAPANANOS, 2002) e (LIU, BOUMAIZA e GHANNOUCHI, 2004)

LEGENDA: As duas redes realizam a decomposição do sinal complexo em componentes reais. A rede em (a) utiliza duas redes, cada rede modela uma das componentes polares do sinal de entrada e não apresenta memória. O modelo em (b) utiliza uma rede para modelar ambas as componentes cartesianas do sinal de entrada e é um modelo com memória.

No modelo (b) da FIGURA 21 as entradas da rede são as componentes cartesianas da amostra n do sinal banda base de entrada do PA. Diferentemente do modelo (a) é utilizada uma única rede para simular os dois componentes cartesianos do sinal de entrada do PA. Nota-se também que se inserem as M amostras passadas de cada amostra n , adicionando assim ao modelo o efeito de memória do PA. Os sinais de saída da rede são as componentes cartesianas da amostra n do sinal banda base da saída do PA.

A FIGURA 22 mostra dois modelos que utilizam redes com realimentação. O modelo (a) da FIGURA 22, proposto em (LUONGYINH e KWON, 2005), utiliza uma rede de valores complexos. A entrada da rede é o sinal banda base de entrada do PA da amostra n , $\tilde{x}(n)$, bem como os sinais da realimentação do sinal de saída da rede da amostra n , $\tilde{y}(n)$. Essa rede também tem como entradas M amostras passadas para cada amostra n dos sinais a serem inseridos na rede. O modelo (b) da FIGURA 22,

FIGURA 22 – REDES COM REALIMENTAÇÃO DA LITERATURA



FONTE: (LUONGYINH e KWON, 2005) e (RAWAT, RAWAT e GHANNOUCHI, 2010)

LEGENDA: Ambos os modelos possuem memória. O modelo (a) utiliza uma rede de valores complexos enquanto a rede (b) utiliza uma rede de valores reais com decomposição cartesiana.

proposto em (RAWAT, RAWAT e GHANNOUCHI, 2010), utiliza uma rede de valores reais com decomposição cartesiana. As entradas da rede são as componentes cartesianas do sinal banda base de entrada do PA da amostra n , bem como os sinais da realimentação dos sinais de saída da rede da amostra n , $y_I(n)$ e $y_Q(n)$.

Considerando a aplicação de modelagem de equipamentos eletrônicos, existem duas abordagens: a modelagem física, onde se gera um modelo de acordo com o circuito interno do equipamento; e a modelagem comportamental onde se gera um modelo a partir da relação dos sinais de entrada e saída do equipamento. Na modelagem comportamental o equipamento eletrônico é uma caixa preta, onde sua constituição interna é desconhecida. De forma oposta a nomenclatura utilizada para a modelagem física é caixa branca.

Na maioria dos casos a abordagem da caixa preta é mais direta, pois não exige conhecimentos sobre o sistema a ser avaliado, mas apenas sobre suas entradas e saídas (RUDIN, 2019), (DUKE, 2020). Por outro lado, se houver prévio conhecimento da natureza do sistema ele pode ser inserido na rede neural. Certas dinâmicas dos amplificadores não podem ser modeladas pela rede a não ser que essas informações sejam consideradas na modelagem pela rede neural. Dois exemplos de características dos PAs que a rede não consegue reproduzir sem que

haja intervenção externa à rede são: a memória resultante das características de seus circuitos; e sua característica passa baixas que dificulta a modelagem pelas redes neurais usualmente utilizadas. Esse tipo de abordagem é uma mistura da caixa branca e da caixa preta recebendo a nomenclatura de modelagem caixa cinza (PEARSON e POTTMAN, 2000), (PEDRO, CARVALHO e LAVRADOR, 2003).

A fim de facilitar o estudo e entendimento dos vários modelos de PAs baseados em redes neurais algumas características podem ser utilizadas para classificá-los:

- a presença ou não de memória;
- se utiliza valores reais ou complexos;
- modificações na estrutura da rede neural (quantidade de redes e/ou quantidade de camadas etc.);
- se possui ou não realimentação;
- se considera ou não a característica passa baixas do amplificador.

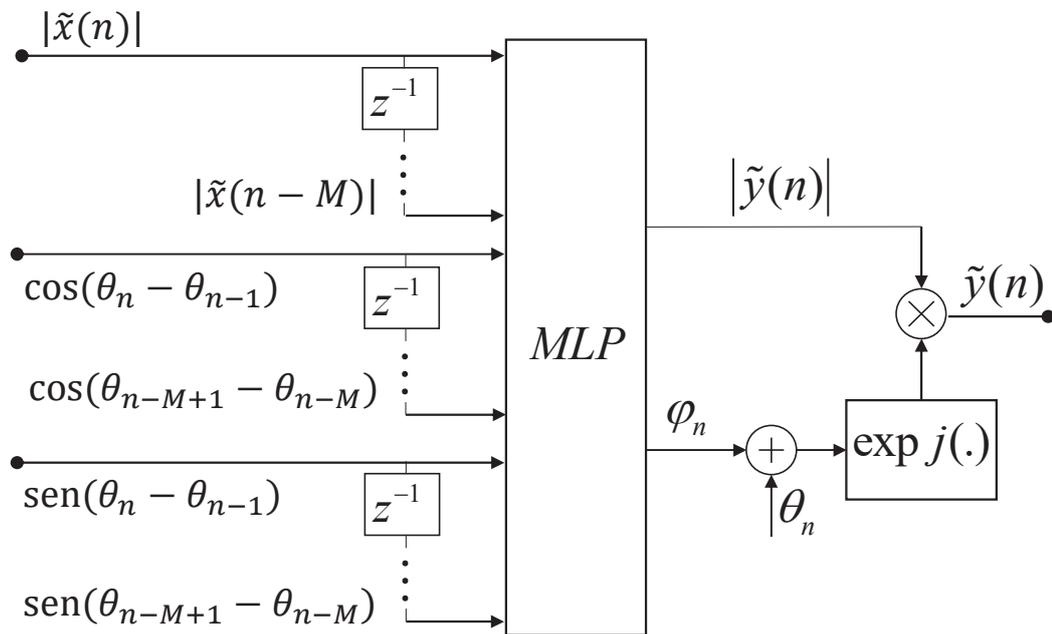
Além das pesquisas relacionadas à estrutura dos modelos os estudos acompanham as tecnologias mais recentes em termos de modulação e tecnologias de PAs, então uma mesma estrutura de modelo pode ser analisada para diversas tecnologias.

Devido às restrições computacionais é necessário o uso do modelo passa baixas do PA. Porém isso traz limitações no uso das redes neurais, pois ao processar a informação de fase do sinal de entrada do PA ela acaba gerando sinais fora da banda de operação do transmissor, pois essa informação de fase está diretamente relacionada à fase da portadora na forma $\cos(\omega_c n + \theta_n)$. Ou seja, a rede gera informação não relacionada ao sinal do PA o que aumenta a complexidade computacional do modelo. Em (LIMA, CUNHA e PEDRO, 2011) é sugerida uma forma de retirar a fase do sinal de entrada para que o sinal seja processado pelas redes neurais sem gerar espúrios matemáticos. Esse modelo é apresentado na FIGURA 23. Os modelos que consideram essas limitações das redes neurais quanto à modelagem de PAs conseguem uma acurácia melhor utilizando menos parâmetros.

Na FIGURA 23 o modelo utiliza uma rede de valor real. As suas entradas buscam inserir a informação da amplitude do sinal banda base de entrada do PA, $|\tilde{x}(n)|$, e a informação da fase do sinal de entrada do PA sem a informação da fase

da portadora, $\cos(\theta_n - \theta_{n-1})$ e $\sin(\theta_n - \theta_{n-1})$. Esses termos da fase são a parte real e imaginária, respectivamente, da subtração de duas amostras de fase em momentos distintos, $\exp j(\theta_n - \theta_{n-1}) = \exp(j\theta_n)\exp(-j\theta_{n-1})$. Além das amostras atuais são utilizadas M amostras passadas de cada entrada da rede. Os sinais de saída da rede são a amplitude do sinal banda base de saída do PA, $|\tilde{y}(n)|$, e a sua fase, φ_n . Como a informação da fase de entrada do PA é retirada do sinal de entrada da rede ela deve ser reinserida no sinal de saída.

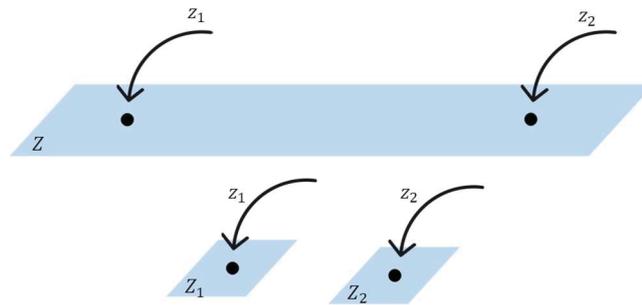
FIGURA 23 – REDE DE VALOR REAL QUE CONSIDERA A NATUREZA PASSA BAIXAS DO PA



FONTE: (LIMA, CUNHA e PEDRO, 2011)

Em (FREIRE, FRANÇA e LIMA, 2014) é sugerido um modelo com duas redes que retira a fase do sinal de entrada a ser processado pela rede e gera um modelo com menor NMSE que o modelo de (LIMA, CUNHA e PEDRO, 2011). Ao se pensar em modelos de duas redes intuitivamente se espera dobrar o número de parâmetros do modelo, porém como os sinais que a rede deve modelar são muito distintos o subespaço necessário para projetar os dois sinais é maior, ou seja, faz uso de mais neurônios do que os subespaços separados, onde serão projetados cada sinal, vide FIGURA 24. Essas redes conseguem uma mesma acurácia com menos neurônios, e por conseguinte menor número de parâmetros.

FIGURA 24 – DIAGRAMA SIMPLIFICADO DA DIMENSIONALIDADE DOS SUBESPAÇOS



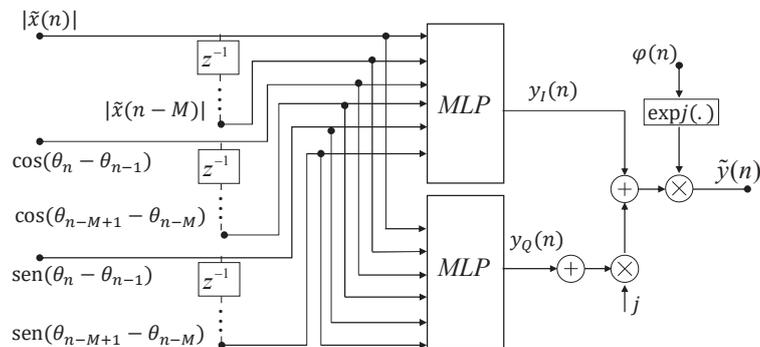
FONTE: (FREIRE, FRANÇA e LIMA, 2015)

LEGENDA: Diagrama simplificado da dimensionalidade dos subespaços dos modelos analisados em (FREIRE, FRANÇA e LIMA, 2014). O subespaço Z da rede neural com duas saídas e suas duas projeções. E os subespaços das duas redes separadas com 1 saída cada uma e as projeções das saídas das redes. Como os sinais a serem projetados possuem valores muito diferentes o subespaço Z possui uma dimensionalidade desnecessária.

Em (FREIRE, FRANÇA e LIMA, 2015) é sugerida uma rede modificada baseada no modelo de (FREIRE, FRANÇA e LIMA, 2014), FIGURA 25, onde se modelam as componentes em quadratura e em fase do sinal de saída, em vez das componentes polares usuais.

O modelo de (FREIRE, FRANÇA e LIMA, 2015) consegue modelar o comportamento do PA com menos neurônios porque com a nova característica de transferência apresentada à rede ela fica menos suscetível a encontrar mínimos locais. Ao treinar uma rede neural é realizada a busca pelo valor mínimo do erro, diferença entre o sinal de saída da rede e do sinal desejado, através de métodos de otimização. Esse valor é chamado mínimo global. Em um determinado espaço amostral podem existir vários pontos de mínimo e qualquer ponto que não seja o mínimo global é um mínimo local.

FIGURA 25 – MODELO COM DUAS REDES DE VALORES REAIS QUE CONSIDERA AS CARACTERÍSTICAS PASSA BAIXAS DO PA

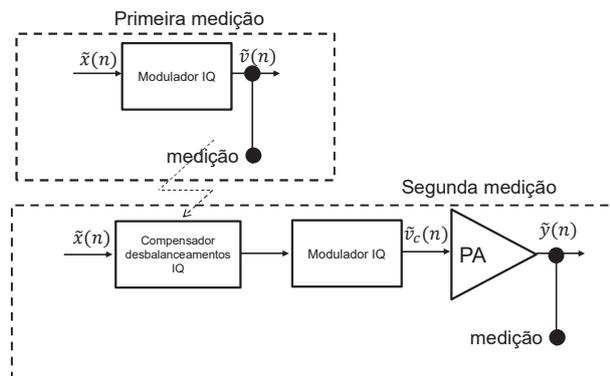


FONTE: (FREIRE, FRANÇA e LIMA, 2015)

4.2 MODELOS CONJUNTOS DE PA E MODULADOR I/Q BASEADOS EM PERCEPTRON DE MÚLTIPLAS CAMADAS

Um conceito recentemente abordado é modelar o PA em conjunto com o modulador I/Q (RAWAT e GHANNOUCHI, 2012), (FU, LAW e THEIN, 2013), (RAWAT, RAWAT, *et al.*, 2013), (GHANNOUCHI, YOUNES e RAWAT, 2013). É necessário um dispositivo para compensar os desbalanceamentos das componentes em fase (I) e em quadratura (Q), além do deslocamento em DC, que o modulador causa. Ao se realizar a compensação das não linearidades do modulador de forma separada é necessário realizar a medição do sistema duas vezes. Na FIGURA 26 pode-se ver o esquemático da medição em duas etapas. Primeiro se mede a saída do modulador a fim de criar um compensador para os desbalanceamentos causados por ele. Na segunda etapa pode ser realizada a medição da saída do PA em cascata com o compensador de desbalanceamentos e o modulador.

FIGURA 26 – MEDIÇÃO EM DUAS ETAPAS DO SISTEMA MODULADOR I/Q E PA



FONTE: (TEMPORAL, 2019)

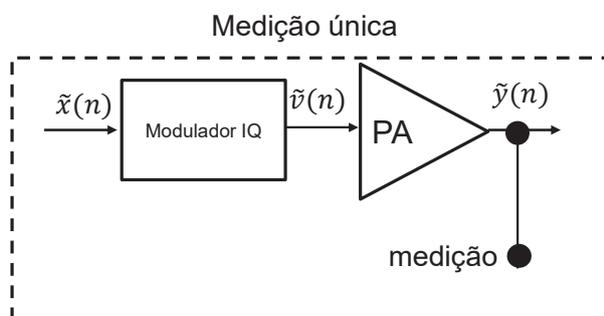
LEGENDA: A primeira medição define os parâmetros do compensador de desbalanceamentos do modulador. A segunda medição obtém o sinal de saída do PA em cascata com o compensador.

Uma das vantagens de se realizar a linearização do PA e do modulador em conjunto é eliminar essa etapa a mais de compensação das não linearidades do modulador. Além disso, a identificação de parâmetros para a modelagem pode ser feita em apenas uma etapa, como na FIGURA 27.

Na literatura são reportados dois modelos baseados em TLP de valores reais. Na literatura o modelo mais usual é descrito na FIGURA 21 (b) (RAWAT, RAWAT, *et al.*, 2013), (RAWAT e GHANNOUCHI, 2012), (FU, LAW e THEIN, 2013). A autora não tem ciência de outros modelos baseados em redes TLP, dessa forma é considerada

contribuição deste trabalho a análise de modelos conjuntos PA mais modulador I/Q com arquiteturas e topologias diversas.

FIGURA 27 – MEDIÇÃO EM UMA ÚNICA ETAPA DO SISTEMA MODULADOR I/Q E PA



FONTE: (TEMPORAL, 2019)

LEGENDA: O sinal medido na saída do PA possui os desbalanceamento do modulador I/Q. E a compensação das não linearidades do PA é feita em conjunto com a compensação dos desbalanceamentos do modulador I/Q.

5 MODELOS PROPOSTOS PARA MODELAGEM COMPORTAMENTAL DE PA BASEADOS EM TLP DE VALORES COMPLEXOS

Esse capítulo apresenta os trabalhos realizados no estudo e adequação das redes TLP de valores complexos para a modelagem de amplificadores de potência em RF, (FREIRE e LIMA, 2017), (FREIRE, SCHUARTZ, *et al.*, 2021). Partindo do princípio de que todos os modelos de PA baseados em TLP de valores reais devem respeitar as restrições do comportamento passa baixas do PA será atestado nas próximas seções que essa afirmação também é válida para os modelos de TLP de valores complexos e serão propostos modelos que satisfaçam essas restrições. Uma vez que operadores lineares não geram energia em frequências que não possuem excitação pelo sinal de estímulo, a única fonte de perturbação que pode existir provém da operação não linear que ocorre na rede TLP. É necessário dessa forma analisar apenas a função de ativação de valor complexo na camada escondida. Na Seção 5.1 é feita uma análise das funções de ativação de valor complexo da literatura e será mostrado que apenas uma cumpre com a restrição de comportamento banda passante do PA.

Considerando as limitações impostas pela característica passa baixas do PA, descrita na Seção 2.5, existem duas abordagens para evitar que o modelo gere contribuições não relacionadas ao sinal medido do PA: tratar os sinais que serão inseridos na rede TLP ou adequar a rede para receber os sinais do PA. Alguns trabalhos já foram realizados nesse âmbito, porém para redes TLP de valores reais (LIMA, CUNHA e PEDRO, 2011), (FREIRE, FRANÇA e LIMA, 2015), (FREIRE, FRANÇA e LIMA, 2014), (FREIRE, MARCONDES e LIMA, 2018). Dentro dessas duas abordagens em (FREIRE e LIMA, 2017) é feita uma análise para mostrar que modificar a forma como o sinal do PA deve ser apresentado à rede também se aplica aos modelos TLP de valor complexo. A Seção 5.2 discute os resultados do estudo de caso desse trabalho. Já em (FREIRE, SCHUARTZ, *et al.*, 2021) é proposta uma nova função de ativação de valor complexo para modelagem de PA. Nesse trabalho o estudo de caso para avaliação da nova função de ativação é feito utilizando um PA multimodo e será apresentado na Subseção 5.3.1. Devido às peculiaridades da tecnologia desse PA, o modelo baseado em TLP de valor complexo, além de utilizar a função de ativação proposta, foi otimizada para a modelagem de PAs multimodos.

Que a autora tenha conhecimento não existem trabalhos no âmbito de modelagem de PAs multimodo com redes TLP de valor complexo, sendo uma contribuição original deste trabalho.

5.1 ANÁLISE DAS FUNÇÕES DE ATIVAÇÃO DE VALOR COMPLEXO

As funções de ativação de valor complexo utilizadas nos modelos da literatura são baseadas em modificações efetuadas na função de ativação de valor real convencional da literatura, a tangente hiperbólica sigmoide:

$$\text{tansig}(u_r) = \frac{2}{1 + \exp(-2u_r)} - 1, \quad (25)$$

onde u_r é um sinal de valor real. O sinal de envoltória banda base de valor complexo relacionado à entrada do PA é $\tilde{x}(n) = a_n e^{j\theta_n}$, a_n e θ_n são as componentes de amplitude e ângulo polar de $\tilde{x}(n)$ respectivamente.

Sem perda de generalidade o sinal de valor complexo aplicado à função de ativação da rede TLP pode ser descrito como:

$$\tilde{u}_r(n) = w_{r,e} \tilde{x}(n - m) + b = u_r^I + ju_r^Q, \quad (26)$$

onde $w_{r,e}$ e b se referem aos parâmetros peso e *bias* de uma rede TLP de valor complexo.

A entrada de uma função de ativação F é definida por:

$$\tilde{u}_r(n) = \sum_{r,e=1}^{R,E} \tilde{x}(n - m) w_{r,e}. \quad (27)$$

TLPs de valor complexo devem possuir funções de ativação capazes de lidar com números de valores complexos. As funções de ativação de valor complexo dadas em (BENVENUTO, MARCHESI, *et al.*, 1991), (BENVENUTO, PIAZZA e UNCINI, 1993) e (GEORGIU e KOUTSOUGERAS, 1992), irão ser analisadas a fim de verificar se geram contribuições que não são úteis para a modelagem comportamental

do modelo equivalente passa baixa do PA e se tem a capacidade de mapear todas as características não lineares do PA.

Na literatura existem três modificações em (25) a fim de que lide com números complexos. Na primeira, introduzida em (BENVENUTO, MARCHESI, *et al.*, 1991), o operador exponencial de (25) lida com sinais de valores complexos. A nomenclatura que será usada posteriormente para essa função será tansig de valor complexo (*TVC*). A expansão de Taylor dessa primeira função de ativação é:

$$TVC(\tilde{u}_r) = \tilde{u}_r - \frac{1}{3}\tilde{u}_r^3 + \frac{2}{15}\tilde{u}_r^5 + \dots \quad (28)$$

Todas as contribuições de (28) são inúteis para estimar os sinais medidos no PA. Qualquer contribuição de (28) – para qualquer ordem ($3^\circ, 5^\circ, 7^\circ$) e para qualquer tempo amostral ($n, n-1, \dots, n-M$) – não mantém a constante multiplicativa do ângulo polar em um.

Independente dos parâmetros da rede, todas as contribuições não lineares de (28) são respostas fora da banda, ou seja, são harmônicas da frequência ω_c . Essa afirmação pode ser analisada expandindo \tilde{x}^3 :

$$\tilde{x}^3(n-m) = a_{n-m}^3 [\cos(3\theta_{n-m}) + j\text{sen}(3\theta_{n-m})]. \quad (29)$$

Note que não há nenhum termo que envolva $1(\omega_c t)$.

Na segunda forma, proposta em (BENVENUTO, PIAZZA e UNCINI, 1993), (25) é aplicada duas vezes, uma para a componente em fase e outra para a componente em quadratura, sendo ambas as componentes de valores reais. A nomenclatura que será usada posteriormente para essa função será tansig separada (*TS*). A expansão da série de Taylor é dada por:

$$\begin{aligned} TS(\tilde{u}_r) &= \text{tansig}(u_r^I) + j\text{tansig}(u_r^Q) = \\ &= \left(u_r^I - \frac{1}{3}[u_r^I]^3 + \frac{2}{15}[u_r^I]^5 + \dots \right) + j \left(u_r^Q - \frac{1}{3}[u_r^Q]^3 + \frac{2}{15}[u_r^Q]^5 + \dots \right). \end{aligned} \quad (30)$$

Analisando (30), pode ser visto que para cada contribuição dentro da banda outra contribuição fora da banda é formada. Para $\tilde{x} = x_I + jx_Q$, se x_I^3 é expandido então:

$$[x^I(n-m)]^3 = a_{n-m}^3 \left[\frac{3}{4} \cos(\theta_{n-m}) + \frac{1}{4} \cos(3\theta_{n-m}) \right]. \quad (31)$$

Na contribuição com coeficiente de 3/4, o ângulo polar é multiplicado por 1. Na contribuição com coeficiente de 1/4, o ângulo polar é multiplicado por 3. Dessa forma, junto com uma contribuição física é gerada uma contribuição não física fora da banda.

O requisito de manter a constante 1 multiplicando o ângulo polar é atendido quando a função de ativação é incapaz de modificar essa componente. A função de ativação em (GEORGIU e KOUTSOUGERAS, 1992) processa o número complexo sem modificar o ângulo polar de nenhuma forma. Essa função de ativação apenas modifica a amplitude. A nomenclatura que será usada posteriormente para essa função será tansig de amplitude (*TA*). A função de ativação de (GEORGIU e KOUTSOUGERAS, 1992) é descrita como:

$$TA(\tilde{u}_r) = \{ \text{tansig}[|\tilde{u}_r(n)|] \} \exp[j\angle\tilde{u}_r(n)]. \quad (32)$$

Em (32) apenas a amplitude de valor real $|\tilde{u}_r(n)|$ é sujeita ao operador sigmoide não linear. O ângulo polar $\angle\tilde{u}_r(n)$ passa sem modificações pela função de ativação em (32).

O modelo deve possuir paridade ímpar em relação ao sinal de envoltória complexa \tilde{x} , porém apesar de essa ser uma condição necessária ela não é suficiente. A paridade ímpar é satisfeita ao se usar as funções de ativação de (28) e (30) apenas se o parâmetro $b = 0$, pois é uma constante e possui paridade par.

Se (25) é expandido na série de Taylor nas proximidades de $u_r = 0$ então:

$$\text{tansig}(u_r) = \sum_{p=1}^{\infty} \frac{(-1)^{p-1} 2^{2p} (2^{2p} - 1) B_p}{(2p)!} u_r^{2p-1}, \quad (33)$$

onde B_p é o p -ésimo número de Bernoulli (ZUCKER, 1972).

Ao substituir (33) em (32) se tem:

$$\begin{aligned}
 TA(\tilde{u}_r) &= \left[\sum_{p=0}^{\infty} \frac{(-1)^p 2^{2(p+1)} [2^{2(p+1)} - 1] B_{p+1}}{[2(p+1)]!} |\tilde{u}_r(n)|^{2(p+1)-1} \right] \exp[j\angle\tilde{u}_r(n)] \\
 &= \left[\sum_{p=0}^{\infty} \frac{(-1)^p 2^{2(p+1)} [2^{2(p+1)} - 1] B_{p+1}}{[2(p+1)]!} |\tilde{u}_r(n)|^{2p} \right] \tilde{u}_r(n) \\
 &= G_{real}[|\tilde{u}_r(n)|^2] \tilde{u}_r(n).
 \end{aligned} \tag{34}$$

Para simplificar o entendimento, em (34) uma nova função G_{real} é definida e toma como argumento $|\tilde{u}_r(n)|^2$. G_{real} realiza um mapeamento não linear entre números de valores reais. A equação (34) auxilia a visualizar a função descrita em (33) de outra forma. Essa função de ativação pode ser descrita como um produto entre um sinal complexo $\tilde{u}_r(n)$ e uma função par em termos de sua amplitude $|\tilde{u}_r(n)|$. Para $p = [0,1,2]$, (34) é reduzida a:

$$\begin{aligned}
 TA(\tilde{u}_r) &= |\tilde{u}_r(n)| \exp[j\angle\tilde{u}_r(n)] - \frac{1}{3} |\tilde{u}_r(n)|^3 \exp[j\angle\tilde{u}_r(n)] \\
 &\quad + \frac{2}{15} |\tilde{u}_r(n)|^5 \exp[j\angle\tilde{u}_r(n)]. \\
 &= \tilde{u}_r(n) - \frac{1}{3} |\tilde{u}_r(n)|^2 \tilde{u}_r(n) + \frac{2}{15} |\tilde{u}_r(n)|^4 \tilde{u}_r(n).
 \end{aligned} \tag{35}$$

Fica claro que a função de ativação de (32) cumpre com o requisito de não gerar contribuições não físicas. Um estudo de caso em (FREIRE, SCHUARTZ, *et al.*, 2020) demonstra que a rede TLP de valor complexo de (GEORGIU e KOUTSOUGERAS, 1992) apresenta desempenho superior quando comparada com as redes TLP de valor complexo de (BENVENUTO, MARCHESI, *et al.*, 1991), (BENVENUTO, PIAZZA e UNCINI, 1993).

Vale salientar que as funções mencionadas até agora utilizam como base a função tangente hiperbólica sigmoide de (25). Se para o treinamento da rede forem utilizados algoritmos com descida de gradiente existem outros fatores a serem levados em consideração. Em (BENVENUTO e PIAZZA, 1992) é demonstrado que ao estender a função de (25) da linha real para o plano complexo essa função se torna sem limites. Para um treinamento adequado a função de ativação deve ser limitada, pois o valor de saída depende da entrada e dos pesos e pode tomar qualquer valor

no plano complexo dificultando o treinamento. Em (BENVENUTO e PIAZZA, 1992) é explicitado o problema da singularidade ao se utilizar a equação (25) no domínio complexo. Essa afirmação é facilmente verificada, pois se u_r se aproximar de qualquer valor no conjunto $[0 \pm j((2k + 1)/2)\pi]$, sendo k um inteiro qualquer, $|\text{tansig}(u_r)| \rightarrow \infty$, então é uma função sem limites definidos. Ainda em (BENVENUTO e PIAZZA, 1992) é sugerido escalonar os valores de entrada para uma região no plano complexo. Porém, sem restringir os valores que os pesos da rede podem gerar, e por conseguinte os de u_r , não se tem como controlar a resposta da função de ativação, não sendo a solução proposta válida para todos os casos.

Outra questão a se levar em consideração com relação às funções de ativação é o teorema de Liouville, que afirma que se uma função é inteira e limitada no plano complexo, então é uma função constante. As funções de ativação de valores reais precisam ser deriváveis em todo o plano \mathbb{R} , no plano \mathbb{C} essas funções são chamadas funções inteiras. Dessa forma o uso das funções providas das redes neurais reais que se traduzem em funções inteiras e possuem a característica de serem limitadas, de acordo com Liouville, são funções constantes. E por conseguinte, não são adequadas para se utilizar no treinamento de redes neurais.

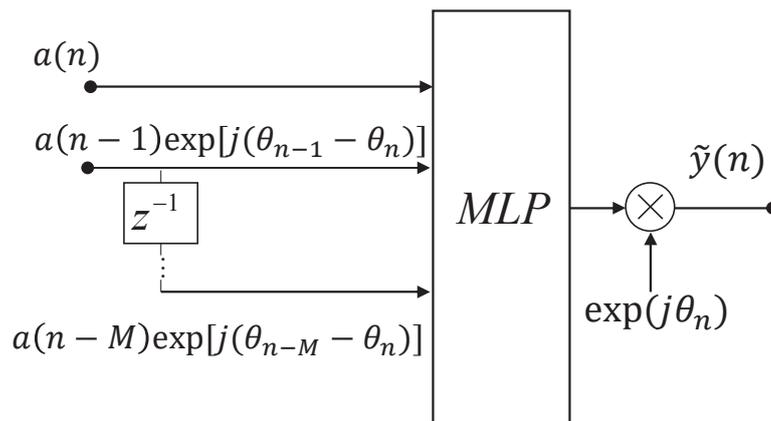
Visto que os modelos usuais utilizam as funções de ativação (28) e (30) e que invariavelmente acabam gerando contribuições fora da banda, se faz necessário ou trocar a função de ativação por uma que não gere contribuições fora da banda ou outro artifício que elimine a informação da fase da portadora do sinal a ser processado pela rede. Mesmo a função em (32), que não modifica o ângulo polar do sinal de entrada, possui contraindicações ao ser utilizada no plano complexo, por ser uma função singular no plano complexo.

5.2 MODELO PROPOSTO COM MODIFICAÇÃO NO SINAL DE ENTRADA DO MODELO

A fim de eliminar a dependência com a informação da fase da portadora é proposto um modelo que modifica o sinal do PA que será inserido na rede (FREIRE e LIMA, 2017). Esse modelo pode ser visto na FIGURA 28. As entradas do modelo são os valores atuais, $a(n)$, e passados, $a(n - M)$, do módulo do sinal de envoltória de valor complexo de entrada do PA multiplicados por uma componente de diferenças de fases no formato $\exp[j(\theta_{n-m} - \theta_n)]$, sendo M a profundidade de memória. Esse

artifício elimina a relação entre o ângulo polar de entrada θ e ω_c , uma vez que $\exp[j(\omega_c t + \theta_{n-m})] \exp[-j(\omega_c t + \theta_n)] = \exp[j(\theta_{n-m} - \theta_n)]$, para m variando de 0 a M . Uma vez que a rede agora processa um sinal em banda base (sem a informação da frequência da portadora, ω_c) é necessário na saída da rede reaver a relação com ω_c multiplicando o sinal de saída da rede pelo termo $\exp[j\theta_n]$. Essa forma de reaver a informação de ω_c é a utilizada na literatura.

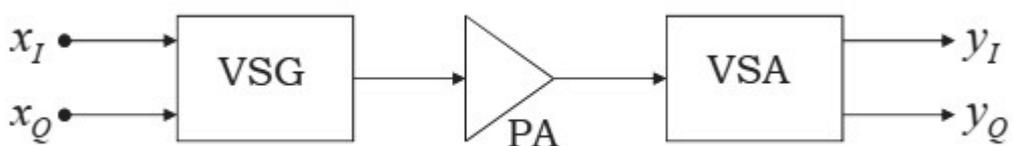
FIGURA 28 – MODELO DE PA PROPOSTO BASEADO EM TLP DE VALORES COMPLEXOS



FONTE: (FREIRE e LIMA, 2017)

Para a análise do modelo são utilizados dados experimentais gerados por um PA de classe AB GaN, excitado por um sinal de envoltória de valor complexo WCDMA com 8,84 MHz de largura de banda. O diagrama de blocos da configuração da aquisição dos dados do PA usado para a modelagem é mostrado na FIGURA 29. O sinal banda base sofre uma modulação ascendente para 900 MHz por um gerador de sinal vetorial (VSG – *vector signal generator*) Rohde & Schwartz SMU200A. O sinal de saída sofre modulação descendente e é medido através de um analisador de sinal vetorial (VSA – *vector signal analyzer*) a uma frequência de amostragem de 61,44 MHz, quando o PA operava a uma potência de saída média de 26 dBm. Dois subconjuntos de dados foram usados na rede TLP, um para treinamento e outro para

FIGURA 29 – CONFIGURAÇÃO DO SISTEMA PARA AQUISIÇÃO DOS DADOS

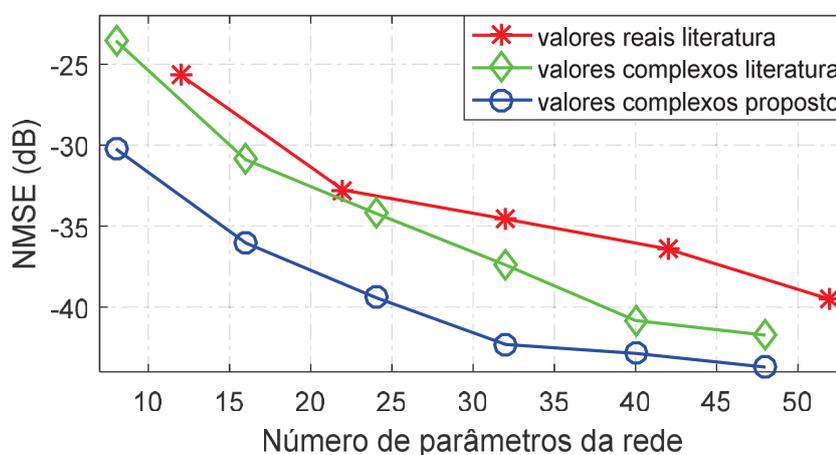


FONTE: (FREIRE e LIMA, 2017)

validação no programa Matlab. Para identificação dos pesos foi utilizado o método quase-Newton (SALEH, 1981). Para avaliar a acurácia dos métodos são empregados o NMSE (MUHA, CLARK, *et al.*, 1999) e o ACEPR (ISAKSSON, WISSEL e RONNOW, 2006). A acurácia do modelo baseado em TLP de valor complexo proposto é comparada com duas abordagens da literatura: um modelo TLP de valor real de (LIMA, CUNHA e PEDRO, 2011) tendo como função de ativação a função sigmoide dada em (25); e o modelo TLP de valor complexo de (BENVENUTO, MARCHESI, *et al.*, 1991) tendo a função de ativação de (30), a tansig separada. O modelo proposto utiliza a função de ativação de (30).

O número de parâmetros da rede, os pesos e *bias*, variam de acordo com o número de neurônios R utilizados. A profundidade de memória utilizada foi $M = 2$. Um parâmetro de valor complexo foi considerado como equivalente a dois parâmetros de valor real. Os resultados de NMSE e ACEPR em função do número de parâmetros são dados na FIGURA 30 e na FIGURA 31, respectivamente. Com respeito ao modelo que utiliza TLP de valores complexos, comparando o número de parâmetros para valores de acurácia semelhantes, o modelo proposto obteve menores números de parâmetros e convergiu mais rapidamente, corroborando a teoria proposta de que ao se retirar a geração de termos não físicos pela rede se diminui a complexidade computacional do modelo. Dessa forma, um NMSE de -42,3 dB pode ser conseguido pela rede proposta com 32 parâmetros complexos enquanto o modelo complexo da literatura alcança um NMSE de -37,3 dB.

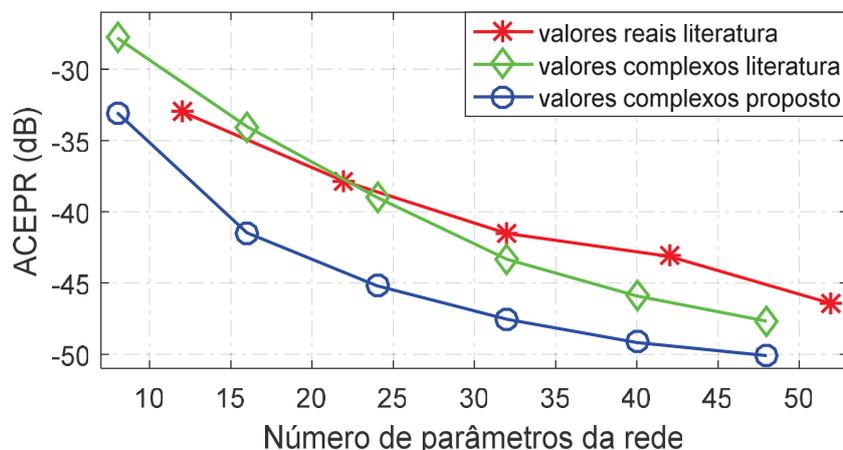
FIGURA 30 – RESULTADOS DE NMSE



FONTE: (FREIRE e LIMA, 2017)

LEGENDA: Valores de NMSE em função do número de parâmetros da rede para os modelos de valor real e complexo da literatura e do modelo proposto de valor complexo.

FIGURA 31 – RESULTADOS DE ACEPR

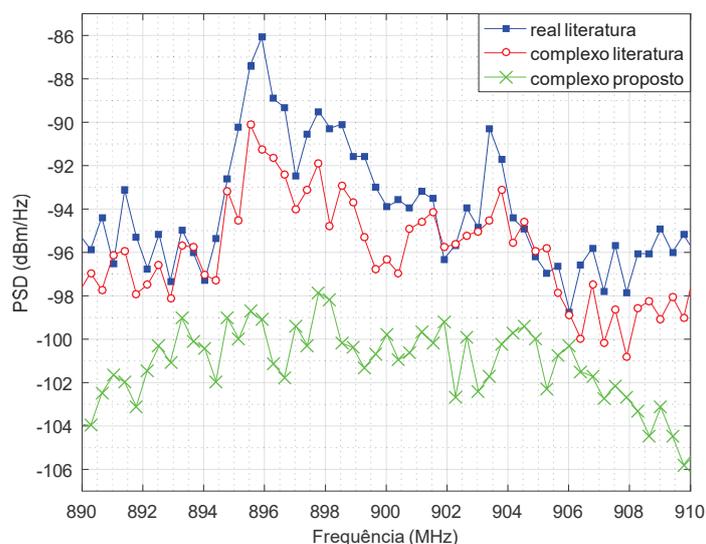


FONTE: (FREIRE e LIMA, 2017)

LEGENDA: Valores de ACEPR em função do número de parâmetros da rede para os modelos de valor real e complexo da literatura e do modelo proposto de valor complexo.

A FIGURA 32 mostra o PSD dos sinais de erro dos modelos sendo avaliados. Os sinais de erro são calculados fazendo a diferença entre o sinal de saída do modelo e o sinal medido do PA. O modelo proposto obteve melhor resposta em toda a faixa de frequência de operação do PA. A FIGURA 33 mostra os sinais de saída do PA e do modelo proposto no tempo. Pode-se visualizar que as amostras se sobrepõem indicando que o modelo conseguiu mapear de forma correta o comportamento do PA.

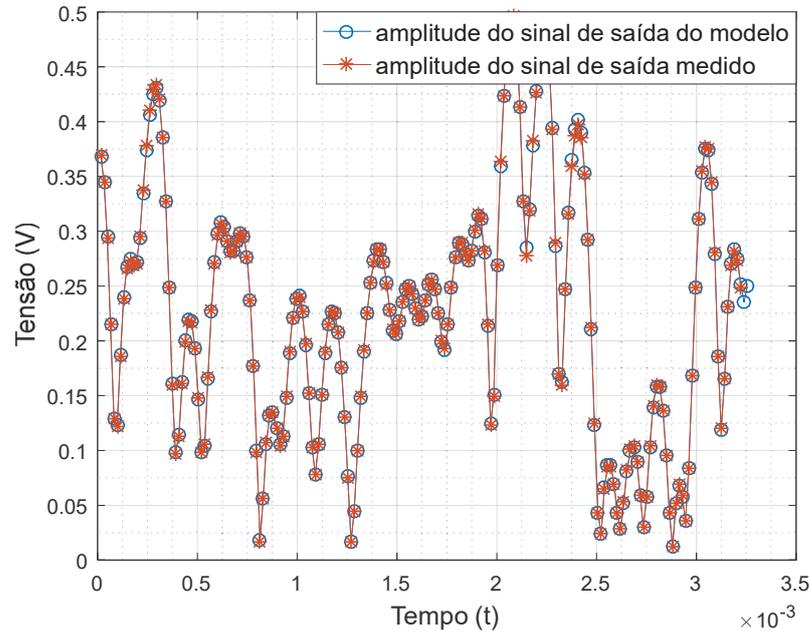
FIGURA 32 – PSD DOS SINAIS DE ERRO



FONTE: (FREIRE e LIMA, 2017)

LEGENDA: PSDs dos sinais de erro: modelo real da literatura com 32 parâmetros reais; e modelos complexos da literatura e proposto com 16 parâmetros complexos.

FIGURA 33 – SINAIS DE SAÍDA DO MODELO E DO PA EM FUNÇÃO DO TEMPO



FONTE: (FREIRE e LIMA, 2017)

LEGENDA: Gráfico das 200 primeiras amostras da amplitude dos sinais de saída do modelo e do PA em função do tempo.

5.3 MODELO PROPOSTO COM MODIFICAÇÃO NA FUNÇÃO DE ATIVAÇÃO DA REDE TLP DE VALOR COMPLEXO

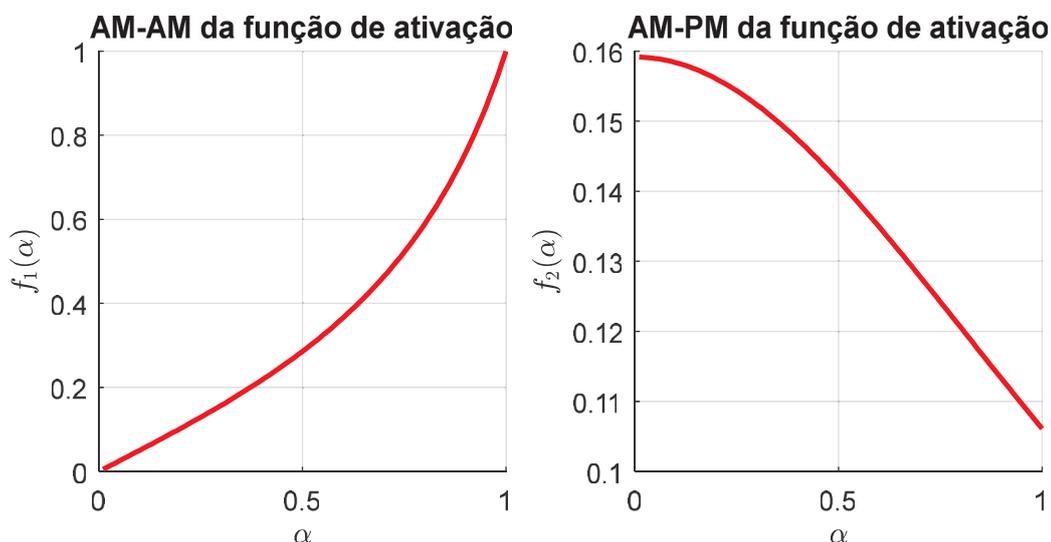
Focando em modelos de PA, uma função de ativação deve ter a capacidade de englobar as curvas características de um PA, mais especificamente os comportamentos AM-AM e AM-PM. Com isso em mente e levando em consideração as discussões da Seção 5.1 foi proposta a função:

$$F = (f_1(|\tilde{u}|)e^{jf_2(\angle\tilde{u})})e^{j\angle\tilde{u}}, \quad (36)$$

onde \tilde{u} é o sinal de valor complexo processado pela função de ativação F , $|\tilde{u}|$ é sua componente de amplitude e $\angle\tilde{u}$ sua componente de fase. A função (36) é composta ainda pelas funções f_1 que representa a característica AM-AM e pela função f_2 que representa a característica AM-PM. A informação de fase do sinal de entrada é inserida à parte das funções não lineares. As funções f_1 e f_2 podem tomar formas variadas desde que consigam representar as curvas características do PA.

Certas particularidades podem ser isoladas da curva característica AM-AM inversa do PA na FIGURA 34: o comportamento linear com inclinação positiva para baixos valores de amplitude e comportamento expansivo para valores grandes de

FIGURA 34 – COMPORTAMENTO DAS FUNÇÕES f_1 E f_2



FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

LEGENDA: A função f_1 visa mapear o comportamento AM-AM e a função f_2 , visa mapear o comportamento AM-PM do PA multimodo sob estudo.

amplitude e sua origem no zero. Analisando agora a curva característica AM-PM inversa da mesma maneira pode-se observar: um comportamento constante sem inclinação para baixas amplitudes, e um comportamento compressivo para grandes amplitudes. A função de ativação não leva em consideração os efeitos de memória, e para representar esse fenômeno são inseridas na rede as amostras passadas do sinal de entrada. Na Subseção 5.3.1 um modelo que utiliza essa função será avaliado e, inspirado no modelo de Saleh (SALEH, 1981), o uso de equações racionais é proposto para formar as funções f_1 e f_2 . Elas permitem facilmente moldar curvas que apresentam os comportamentos compressivos e expansivos para amplitudes maiores, além de adequar as inclinações para menores amplitudes.

5.3.1 Estudo de caso de modelo com modificação na função de ativação da rede TLP de valor complexo com PA multimodo

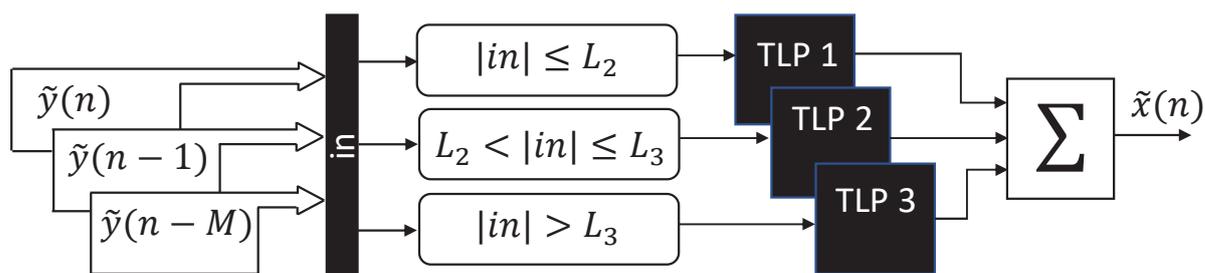
Sendo o PA um componente do transmissor que precisa de um alto consumo de energia a fim de suportar a transmissão o mais linear possível de um sinal, em

sistemas que tem como característica a restrição de consumo de energia os PAs multimodo são de particular interesse, pois focam na redução do consumo de energia. Para avaliar a função de ativação o PA multimodo proposto em (SANTOS, RIOS, *et al.*, 2017) será modelado com um TLP de valor complexo que utiliza a função de ativação de (36).

O PA multimodo de (SANTOS, RIOS, *et al.*, 2017) foi projetado em tecnologia CMOS Globalfoundries 130 nm para operar em diferentes modos de ganho de potência e com sinais na banda de 2,4 GHz. O PA deve operar na maior parte do tempo com recuo de potência devido à sua alta PAPR. Nos PAs típicos ao operar na região linear a potência média de saída é limitada além de diminuir a eficiência do PA. O PA multimodo, na região de recuo de potência comuta entre modos de acordo com a potência de entrada a fim de operar por exemplo em um modo de menor ganho e ser mais eficiente. Porém, ao operar nos modos com altos valores de potência possui um comportamento não linear que precisa de técnicas de linearização.

Devido ao seu modo de operação os PAs multimodos apresentam discontinuidades em suas curvas características AM-AM e AM-PM, o que dificulta sua sua modelagem. O modelo proposto nesse trabalho utiliza uma rede TLP de valor complexo para modelar cada modo do PA e pode ser visto na FIGURA 35.

FIGURA 35 – MODELO PROPOSTO DO INVERSO DO COMPORTAMENTO DO PA MULTIMODO



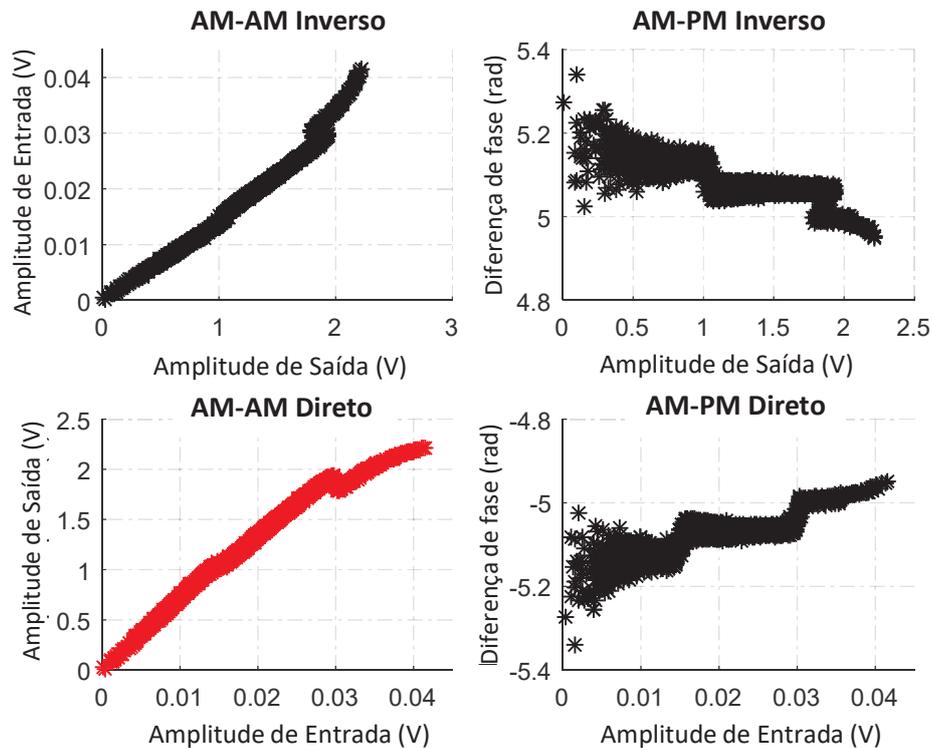
FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

LEGENDA: O modelo possui uma rede para cada modo de operação do PA multimodo.

Na FIGURA 36 são apresentadas as características AM-AM e AM-PM direta e inversa do PA multimodo simuladas. O PA comuta em tempo real entre três modos distintos. Cada modo atua em um intervalo de amplitude, de tamanhos semelhantes, do sinal de entrada. Os diversos tipos de PA possuem curvas características com semelhanças umas com as outras, mas também mostram certas especificidades

referente ao seu circuito e modo de operação. Exemplos de outras curvas características podem ser vistas em (WANG, HUSSEIN, *et al.*, 2017) e (HAMMI, 2014).

FIGURA 36 – FUNÇÃO DE TRANFERÊNCIA DO PA MULTIMODO



FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

Uma possível função de ativação para a modelagem inversa do PA multimodo que é proposta é:

$$F = \left[\frac{0,5|\tilde{u}|}{1 - 0,5|\tilde{u}|^2} e^{j\frac{0,5}{1+0,5|\tilde{u}|^2}} \right] e^{j\angle\tilde{u}}. \quad (37)$$

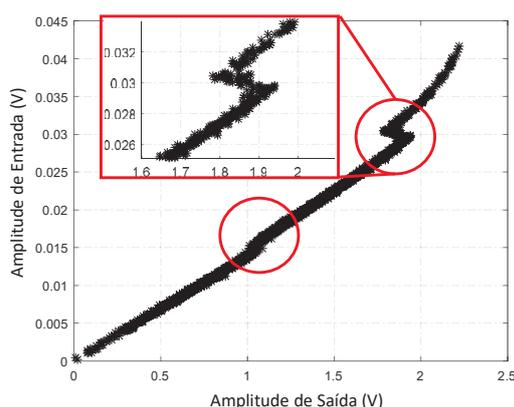
A função de ativação de (37) engloba todas as características necessárias para mapear a função de transferência inversa do PA. A FIGURA 34 mostra o contorno das funções f_1 e f_2 . A função f_1 inicia em zero, é uma linha com inclinação positiva para baixos valores de entrada e tem um comportamento expansivo para valores de entrada maiores. A função f_2 inicia com um valor constante que permanece o mesmo para baixas amplitudes e apresenta comportamento compressivo para maiores valores. Os comportamentos expansivos e compressivos de f_1 e f_2 são ditados pelo

operador no denominador. Quando o comportamento compressivo é desejado, o denominador é $1 + 0,5|\tilde{u}|^2$, então para comportamento expansivos, o denominador é $1 - 0,5|\tilde{u}|^2$.

O número de redes do modelo depende do número de modos do PA multimodo. No PA multimodo que será modelado serão habilitados 3 modos, dessa forma o modelo usará três redes TLP de valor complexo para modelar cada modo e no final a saída de cada modo será somada para recuperar o sinal completo.

No caso de modelar o comportamento inverso do PA multimodo, as entradas da rede são as amostras instantâneas e passadas dos sinais de saída medidas do PA, $\tilde{y}(n), \tilde{y}(n-1), \dots, \tilde{y}(n-M)$. A entrada de cada rede TLP está contida dentro do intervalo $L_k < |\text{in}| \leq L_{k+1}$, onde L_k e L_{k+1} são os limites superior e inferior do valor da amplitude do sinal de entrada que é definido pelo modo de operação do PA. Para esse estudo de caso $k = 3$, $L_1 = 0$ e $L_4 = +\infty$ são conhecidos. Como se trata de um modelo inverso L_2 e L_3 não são definidos e precisam ser determinados. Isso é devido à natureza multimodo do PA e ocorre porque existe uma superposição de algumas amostras de amplitude como pode ser visto na FIGURA 37 demarcadas pelos círculos vermelhos e ampliadas para melhor visualização. Nessa região para cada valor de amplitude de saída estão associados dois valores de amplitude de entrada. Ou seja, dois valores de entrada distintos geram um mesmo valor de saída.

FIGURA 37 – SUPERPOSIÇÃO DAS AMOSTRAS DE AMPLITUDE

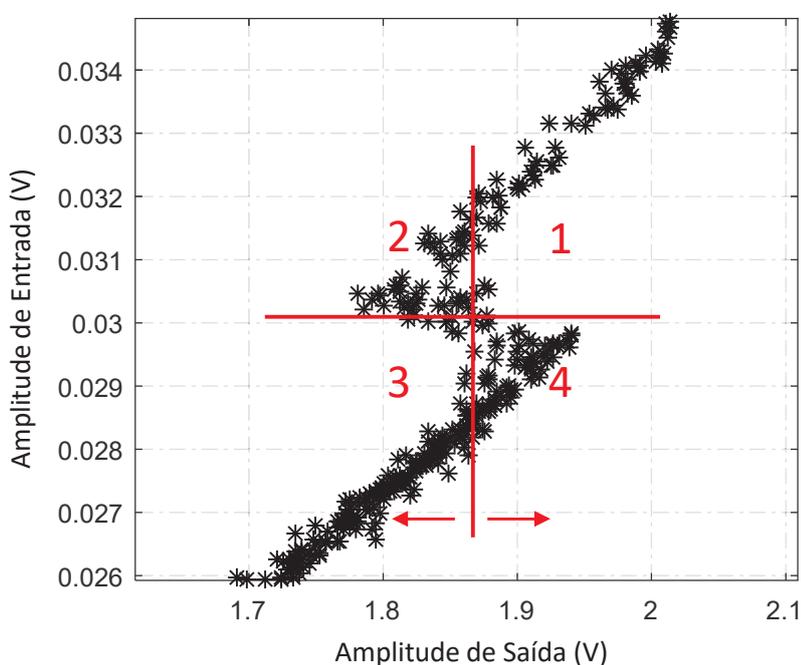


FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

LEGENDA: No comportamento AM-AM inverso do PA multimodo aparecem amostras que se superpõem e exige-se alguma técnica que elimine as amostras replicadas para facilitar o mapeamento pela rede neural.

Para não comprometer o mapeamento da rede neural é necessário modificar os sinais de cada modo excluindo as amostras superpostas e definir uma limitação para a transição de modos. Para tanto foi desenvolvido um método para definir os intervalos das entradas da rede TLP para a modelagem inversa de PAs multimodos. Na FIGURA 38 amostras no primeiro e quarto quadrantes se sobrepõem. A mesma coisa ocorre no segundo e terceiro quadrantes. As amostras a serem retiradas são as dos quadrantes pares. A linha horizontal é um suporte para definir os limites L do comportamento inverso.

FIGURA 38 – DETALHAMENTO DAS AMOSTRAS SUPERPOSTAS



FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

LEGENDA: Na transição de dois modos de operação do PA observa-se as amostras superpostas. A linha vertical realiza a varredura para achar o valor de L_3 que minimiza as amostras excluídas. As amostras a serem excluídas se encontram nos quadrantes pares.

As amostras excluídas são dadas pela subtração entre as amostras totais do sinal N e as amostras do sinal modificado N_m :

$$N_{exc} = N - N_m. \quad (38)$$

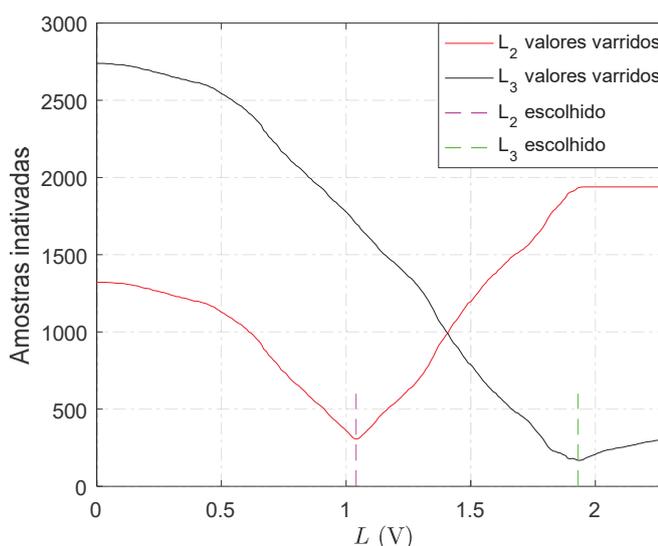
Os limites L dos modos são representados pelas linhas verticais. Os valores das linhas horizontais são L_2 e L_3 da modelagem direta (0,015 V e 0,03 V, respectivamente). Para calcular os novos L_2 e L_3 para a modelagem inversa,

primeiramente L_3 é fixado em um valor que inclui todas as amostras, ou seja, L_3 é fixado no valor da máxima amplitude do sinal de saída do PA. Para o PA sendo avaliado esse valor é igual a 2,26 V. Então se efetua uma varredura de valores de 0 a 2,26 V com intervalos de 0,01 V a fim de escolher um valor de L_2 que minimiza N_{exc} .

O mesmo processo ocorre para definir L_3 . Primeiro L_2 é fixado no valor previamente achado, então é feita uma varredura de 0 a 2,26 V com intervalos de 0,01 V, e é escolhido um valor de L_3 que minimiza N_{exc} .

Esse processo pode ser visualizado na FIGURA 39, onde é mostrado o gráfico de amostras excluídas pelos valores utilizados na varredura para se achar os limites L_2 e L_3 . Os valores dos limites se encontram nos vales e indicam o mínimo de amostras excluídas para cada modo.

FIGURA 39 – ESCOLHA DOS LIMITES DOS MODOS DE OPERAÇÃO



FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

LEGENDA: Variação de L_2 e L_3 em relação às amostras excluídas. Os vales indicam os limites superior e inferior que apresentam o menor número de amostras excluídas.

Após se calcular os limites dos modos para o comportamento inverso do PA multimodo, as amostras instantâneas e passadas do sinal de saída medido do PA são aplicadas à entrada da rede TLP de valor complexo, de acordo com a FIGURA 35.

As funções utilizadas para modelagem de PA, a tansig de valor complexo e a tansig separada, descritas em (27) e (30), respectivamente, não são satisfatórias para a modelagem de PAs como demonstrado na Seção 5.1. Além disso, em (FREIRE, SCHUARTZ, *et al.*, 2020) é feita uma comparação dessas funções com a função tansig de amplitude, descrita em (32), e é demonstrado que essa última apresenta

acurácia superior ao modelar PAs. Porém, como discutido na Seção 5.1, a função tansig de amplitude ainda possui pontos desfavoráveis. Para validar a teoria discutida na Seção 5.1 e avaliar a função de ativação proposta em contrapartida da função tansig de amplitude, será usado o modelo TLP de valor complexo proposto para PA multimodo.

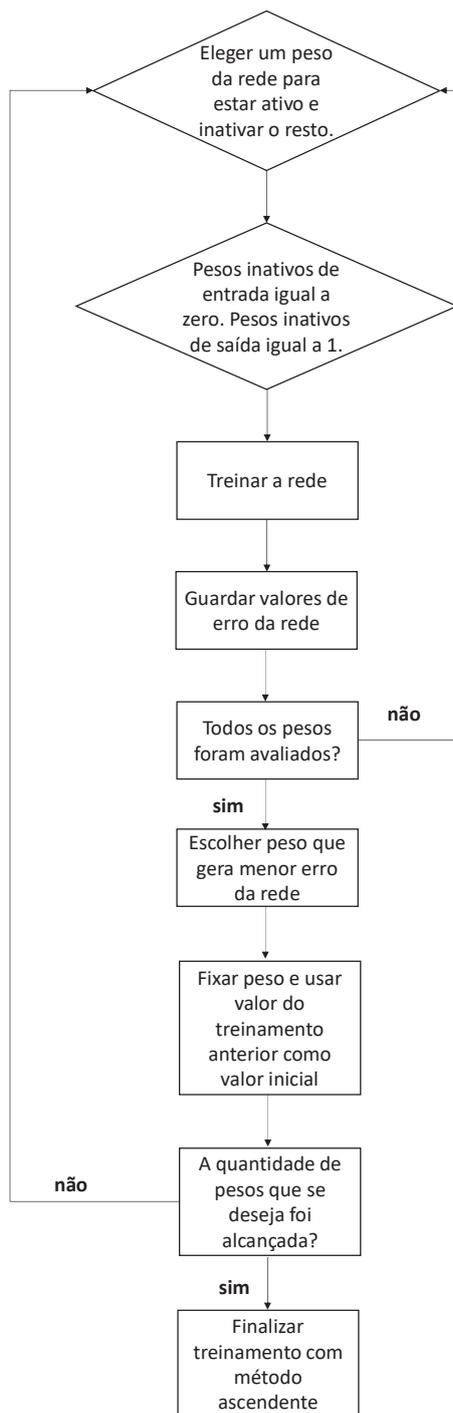
Esse trabalho também se propõe a utilizar o método ascendente para simplificação do modelo para implementação. O método ascendente foi inicialmente proposto em (SCHUARTZ, SANTOS, *et al.*, 2019) para ser utilizado com modelos polinomiais de memória para reduzir a quantidade de parâmetros. A simplificação do modelo é realizada no processo de treinamento. Esse método visa manter os parâmetros mais relevantes e cancelar os outros parâmetros. Cancelar (ou remover) um parâmetro na rede TLP é o mesmo que fazer seu valor ser igual a zero se for um peso de entrada da rede e ser igual ao valor um se for um peso de saída da rede.

O processo de treinamento ocorre em N_d passos, onde N_d é o número escolhido de parâmetros para estarem ativos. Em cada passo n parâmetros são ativados, com $n = 1, 2, \dots, N_d$, e a acurácia do modelo é analisada. O passo atual ativa um parâmetro e mantém os parâmetros que geram a melhor resposta do modelo no passo anterior. Os valores dos parâmetros do passo anterior servem como valor inicial para o novo treinamento do passo atual. O algoritmo desse método é descrito na FIGURA 40.

Exemplificando, para simplificar um modelo de 12 para 3 parâmetros ativos, $N_d = 3$, serão necessários 3 passos. No primeiro passo todos os pesos de entrada são zerados e todos os pesos de saída tem valor um. Então um parâmetro qualquer é escolhido para ser ativo, a rede é treinada e a acurácia do modelo é avaliada. Esse processo continua até que todos os 12 parâmetros sejam avaliados. O parâmetro que gerar a melhor resposta do modelo segue para o próximo passo, vamos chamá-lo w_{r1} . No segundo passo a rede é treinada com dois parâmetros. Todos os parâmetros são cancelados com exceção de w_{r1} e novamente todos os 11 parâmetros restantes serão testados para ver qual gera a melhor resposta do modelo, o parâmetro mais relevante do segundo passo é w_{r2} . Para o último passo, w_{r1} e w_{r2} são mantidos ativos e o último parâmetro relevante será escolhido dentre os 10 parâmetros restantes. Os valores iniciais para os parâmetros sendo testados variam de acordo com o conjunto de dados utilizado e do método de treinamento da rede. Para esse estudo de caso se utilizou o

valor 1×10^{-3} para valor inicial de todos os parâmetros que não haviam sido selecionados no passo anterior.

FIGURA 40 – ALGORITMO DO MÉTODO ASCENDENTE



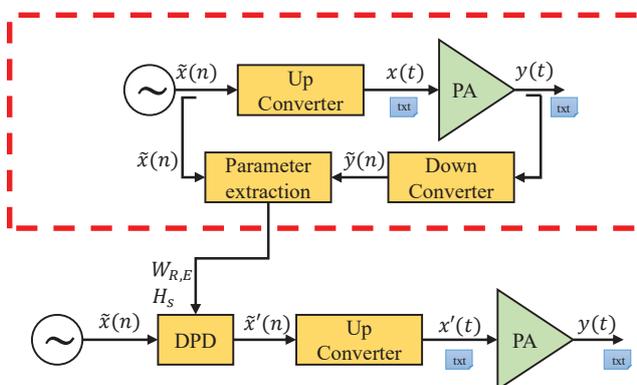
FONTE: O autor (2021)

Essa simplificação diminui o número de operações de números complexos. Quando um peso de entrada é zerado se diminui uma multiplicação complexa e uma soma complexa para cada entrada mais uma soma complexa para cada parâmetro.

Quando é designado o valor 1 a um peso de saída se diminui uma multiplicação complexa. É importante ressaltar que o número de operações depende do número de entradas bem como do número de pesos de entrada e saída que estão ativos após a simplificação.

Sobre a configuração do sistema para a obtenção dos resultados de simulação: o processamento dos dados, incluindo a modulação, demodulação, extração de parâmetros e implementação do DPD, foi realizado pelo *software* Matlab; as simulações em transitório foram realizadas no *software* Cadence Virtuoso para a avaliação do PA. A FIGURA 41 mostra o diagrama de coleta de dados e a linearização do PA. O retângulo em vermelho indica a coleta de dados de entrada e saída para uso posterior na extração de parâmetros. Os blocos em amarelo são processos efetuados no Matlab e os blocos verdes representam o processamento do PA no Cadence Virtuoso. A informação trocada entre *softwares* é realizada através de arquivos de texto. Os parâmetros extraídos são utilizados no bloco DPD.

FIGURA 41 – EXTRAÇÃO DO MODELO DE PA E SEU USO NA UNIDADE DPD



FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

Os parâmetros da rede são identificados no *software* Matlab usando o método Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton. O tamanho da memória é truncado em $M = 2$. O número de neurônios na camada escondida varia de 1 a 3. Além desses valores de neurônios a rede não apresenta melhoria significativa, ou seja, o modelo converge a valores em torno de um NMSE de -41 dB, na Tabela 1 e no gráfico da FIGURA 42.

A diferença entre o sinal estimado pela rede e o sinal de entrada medido do PA é tratada como um sinal de erro. A acurácia do modelo é analisada utilizando as métricas NMSE e ACEPR. Para melhor entendimento e visualização um gráfico com as densidades espectrais de potência (PSDs) também é apresentado.

Primeiramente os modelos baseados em TLP de valor complexo sem o método ascendente são avaliados. A Tabela 1 mostra o resultado de NMSE bem como os números de parâmetros correspondentes para os dois modelos avaliados. O modelo proposto, com o mesmo número de parâmetros, mostrou redução substancial nos valores de erro de modelagem. Para 24 parâmetros de rede, o modelo proposto apresentou o valor de -41,15 dB e o modelo da literatura -35,53 dB, então o modelo proposto teve um melhor resultado de NMSE em 5,62 dB. Isso comprova que o modelo que utiliza a função de ativação proposta tem mais facilidade de mapear o comportamento inverso do PA.

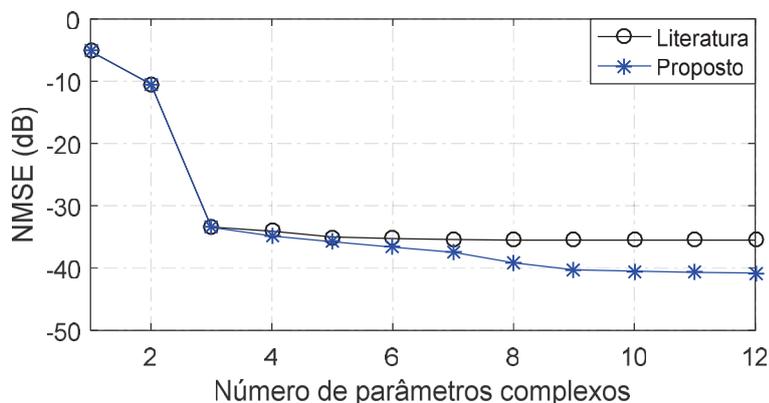
Tabela 1 – NMSEs dos modelos da literatura e proposto e os valores correspondentes de parâmetros complexos sem o método ascendente

Número de neurônios (R)	Número de parâmetros complexos	NMSE (dB)	
		Proposto	Literatura
1	12	-39,29	-35,54
2	24	-41,15	-35,53
3	36	-41,26	-35,53

FONTE: O autor (2021)

O método ascendente foi aplicado a ambos os modelos com $R = 3$ (36 parâmetros). O número de parâmetros ativos, N_d , foi escolhido como 12. A FIGURA 42 mostra os valores de NMSE para os passos do método ascendente com parâmetros crescente. O modelo proposto simplificado apresentou melhores resultados em termos de NMSE do que o modelo da literatura simplificado. O modelo proposto com 12 parâmetros alcançou um NMSE de -40,8 dB e o modelo da literatura apresentou um NMSE de -35,53 dB. Deve-se notar que a curva NMSE do modelo da literatura estabilizou com 5 parâmetros ativos enquanto o proposto manteve a melhora nos resultados até 12 parâmetros ativos. Isso demonstra a dificuldade do modelo da literatura em mapear o modelo inverso do PA multimodo.

FIGURA 42 – VALORES DE NMSE

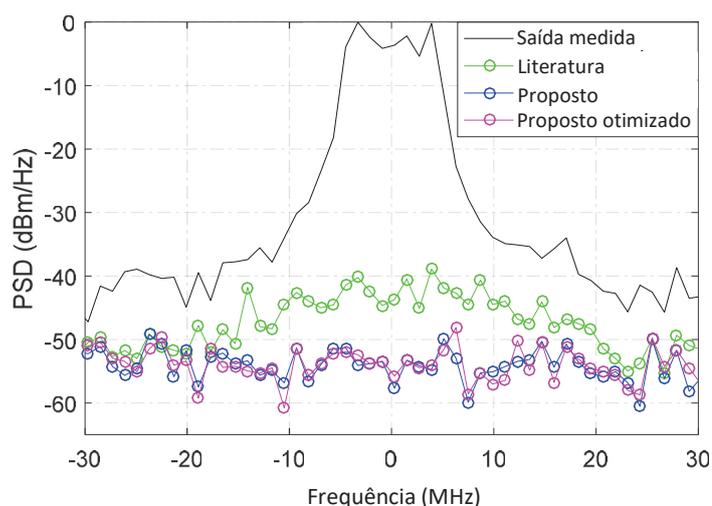


FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

LEGENDA: NMSE em função do número de parâmetros complexos para os modelos da literatura e proposto, ambos com o método ascendente. O número escolhido de parâmetros para o método ascendente é 12.

A FIGURA 43 mostra o PSD dos sinais de erro para os diferentes modelos. A energia dos sinais de erro do modelo proposto teve uma diminuição significativa em todas as frequências. Além disso o método ascendente se mostrou efetivo, pois não houve grandes discrepâncias entre os sinais de erro do modelo sem ascendente e com ascendente, indicando que a acurácia do modelo não se alterou de forma significativa e se conseguiu uma diminuição nos valores de parâmetros da rede.

FIGURA 43 – DENSIDADE ESPECTRAL DE POTÊNCIA DOS SINAIS DE ERRO



FONTE: (FREIRE, SCHUARTZ, *et al.*, 2021)

LEGENDA: No gráfico é apresentado: em linha contínua o PSD do sinal de saída do PA multimodo; o PSD dos sinais de erro dos modelos da literatura — com a função de ativação (32) — e proposto — com a função de ativação de (37) — sem otimização. E PSD do sinal de erro do modelo proposto com otimização através do método ascendente.

6 MODELO COMPORTAMENTAL CONJUNTO DO PA E DO MODULADOR I/Q PROPOSTO

Esse capítulo discutirá dois trabalhos, (FREIRE, MARCONDES e LIMA, 2018) e (FREIRE, MARCONDES e LIMA, 2019), onde são propostos e analisados modelos conjuntos de PA e modulador I/Q baseados em rede TLP de valor real e valor complexo. Na Seção 6.1 os conceitos iniciais sobre modelos conjuntos baseados em TLP serão abordados. O modelo proposto em (FREIRE, MARCONDES e LIMA, 2018) que cumpre com os requisitos da Seção 2.5 será abordado na Seção 6.2. Na Seção 6.3 serão estudados 15 modelos do sistema conjunto PA mais modulador I/Q baseados tanto em rede TLP de valor real quanto de valor complexo.

6.1 ANÁLISE DOS MODELOS CONJUNTOS DO PA E DO MODULADOR I/Q BASEADOS EM REDES TLP DE VALOR REAL

Na Subseção 2.2.1 é descrita a resposta de um modulador não ideal. Um modulador corrompido pelo desbalanceamento de fase e amplitude tem sua resposta descrita em (8).

O PA, ao elevar a potência de um sinal, gera um mecanismo de compressão. Esse comportamento pode ser descrito como funções não lineares deformando a amplitude do sinal que passa pelo PA. Dessa forma, em banda base, o sinal de saída do PA pode ser relacionado à sua entrada como:

$$\tilde{y}(n) = [f_1(|\tilde{s}(n)|) + jf_2(|\tilde{s}(n)|)]e^{j\theta_{s(n)}}, \quad (39)$$

onde f_1 e f_2 representam funções não lineares, $|\tilde{s}(n)|$ e $\theta_{s(n)}$ são os componentes de amplitude e fase de $\tilde{s}(n)$, respectivamente.

Um modelo que represente as imperfeições do modulador e as distorções do PA deve relacionar a saída do PA, $\tilde{y}(n)$, como função da entrada do modulador, $\tilde{x}(n)$. O modelo inverso considera a troca dos papéis dos sinais, e pode ser definido como:

$$\tilde{x}(n) = [g_1(|\tilde{y}(n)|) + jg_2(|\tilde{y}(n)|)]e^{j\theta_{y(n)}} + [g_3(|\tilde{y}(n)|) + jg_4(|\tilde{y}(n)|)]e^{-j\theta_{y(n)}}, \quad (40)$$

onde g_1 , g_2 , g_3 e g_4 são funções não lineares relacionadas às imperfeições do PA. $|\tilde{y}(n)|$ e $\theta_{y(n)}$ são os componentes de amplitude e fase do sinal $\tilde{y}(n)$ respectivamente.

Serão analisados dois modelos inversos baseados em redes TLP de valores reais utilizados na literatura. O intuito é verificar se esses modelos estão aptos à modelar todos os aspectos do sistema conjunto modulador I/Q mais PA.

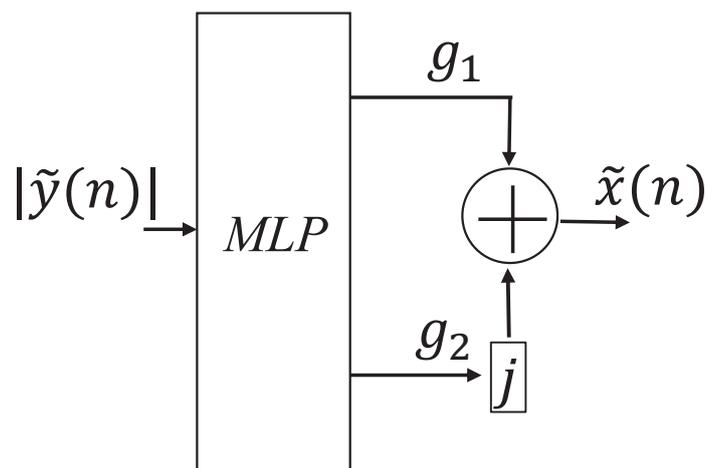
A função de ativação utilizada nas redes de valor real dos modelos é a sigmoide descrita em (25).

O primeiro modelo analisado utilizado em (ISAKSSON, WISSEL e RONNOW, 2006) é mostrado na FIGURA 44. A entrada da rede é a amplitude $|\tilde{y}(n)|$. As duas saídas da rede são as componentes real g_1 e imaginária g_2 formando um sinal complexo:

$$\tilde{x}(n) = [g_1(|\tilde{y}(n)|) + jg_2(|\tilde{y}(n)|)]e^{j\theta_{y(n)}}. \quad (41)$$

O modelo da FIGURA 44 não é capaz de modelar os fenômenos relacionados às funções g_3 e g_4 de (40).

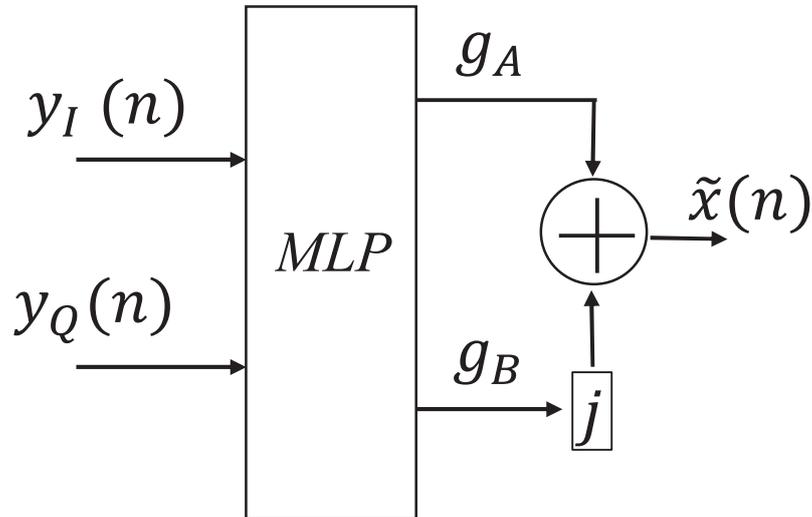
FIGURA 44 – MODELO SEM MEMÓRIA DO CONJUNTO MODULADOR I/Q MAIS PA DA LITERATURA



FONTE: (ISAKSSON, WISSEL e RONNOW, 2006).

A FIGURA 45 mostra o modelo baseado em TLP de valor real de (LIU, BOUMAIZA e GHANNOUCHI, 2004). As entradas do modelo são as componentes real $y_I(n)$ e imaginária $y_Q(n)$ do sinal de envoltória de entrada $\tilde{y}(n)$ aplicadas de maneira independente.

FIGURA 45 – MODELO SEM MEMÓRIA DO CONJUNTO MODULADOR I/Q MAIS PA DA LITERATURA



FONTE: (LIU, BOUMAIZA e GHANNOUCHI, 2004).

Depois de algumas manipulações trigonométricas a saída da envoltória estimada pode ser escrita como:

$$\begin{aligned}
 \tilde{x}(n) = [g_A + g_B] = & [g_1(|\tilde{y}(n)|) + jg_2(|\tilde{y}(n)|)]e^{j\theta_{y(n)}} \\
 & + [g_3(|\tilde{y}(n)|) + jg_4(|\tilde{y}(n)|)]e^{-j\theta_{y(n)}} \\
 & + [g_5(|\tilde{y}(n)|) + jg_6(|\tilde{y}(n)|)]e^{j3\theta_{y(n)}} \\
 & + [g_7(|\tilde{y}(n)|) + jg_8(|\tilde{y}(n)|)]e^{-j3\theta_{y(n)}} \dots
 \end{aligned} \tag{42}$$

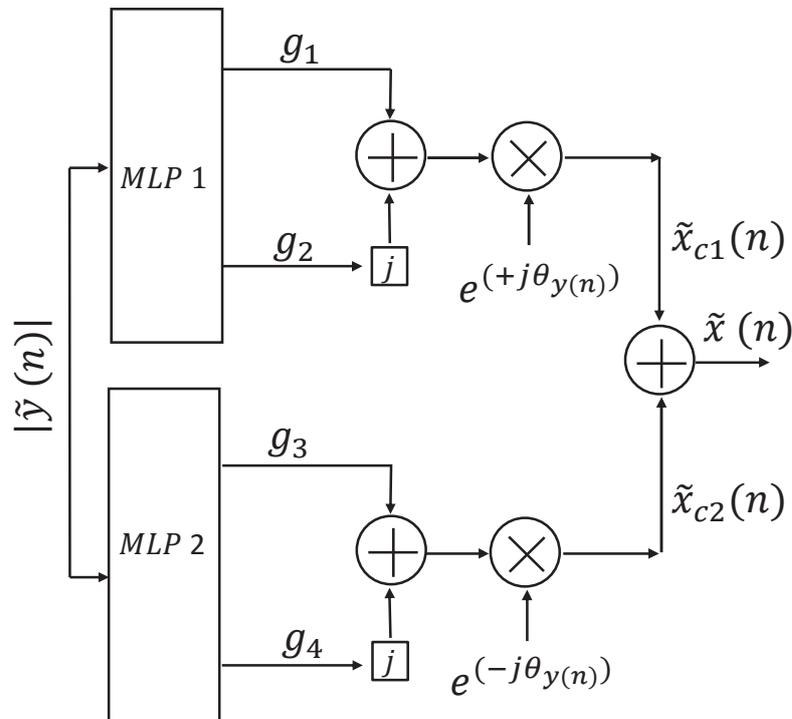
O modelo da FIGURA 45 consegue modelar todos os comportamentos relacionados às funções não lineares g_1 , g_2 , g_3 e g_4 de (40). Porém, também são geradas contribuições adicionais designadas como g_5 , g_6 , g_7 e g_8 , onde observa-se que as fases estão multiplicadas pelo inteiro 3. Essas contribuições são associadas a harmônicas da frequência f_c e por isso inúteis para a modelagem dos sinais medidos do PA, o que leva a crer que dificulta o processo de modelagem por parte da rede.

Os modelos discutidos nessa seção são modelos sem memória a fim de simplificar o entendimento dos conceitos sendo discutidos.

6.2 ESTUDO DE CASO COM MODELO CONJUNTO DE PA E DO MODULADOR I/Q BASEADO EM REDE TLP DE VALOR REAL E COM TREINAMENTO ITERATIVO

O modelo proposto baseado em rede TLP de valor real visa comparar diferentes técnicas de treinamento e é apresentado na FIGURA 46. Diferente dos modelos da literatura o modelo proposto utiliza uma técnica de treinamento iterativa. Duas redes TLP de valor real são empregadas e ambas possuem como sinal de entrada a amplitude do sinal de saída do PA, $|\tilde{y}(n)|$. Para facilitar o estudo o modelo proposto é um modelo que não inclui os efeitos de memória. A saída da rede MLP 1 são as componentes real (g_1) e imaginária (g_2) de um número complexo. Para se obter o sinal de saída completo do modelo, \tilde{x}_{c1} , devem ser somadas as saídas da rede multiplicando o fator j pela parte imaginária, $g_1 + jg_2$, e multiplicar por $e^{+j\theta_{y(n)}}$ para recuperar a informação da fase do sinal de entrada. A saída da rede MLP 2 são as componentes real (g_3) e imaginária (g_4) de um número complexo que posteriormente será multiplicado por $e^{-j\theta_{y(n)}}$ para recuperar a informação da fase do sinal de entrada gerando o sinal \tilde{x}_{c2} .

FIGURA 46 – MODELO PROPOSTO SEM MEMÓRIA COM DUAS REDES TLP DE VALOR REAL



FONTE: (FREIRE, MARCONDES e LIMA, 2018)

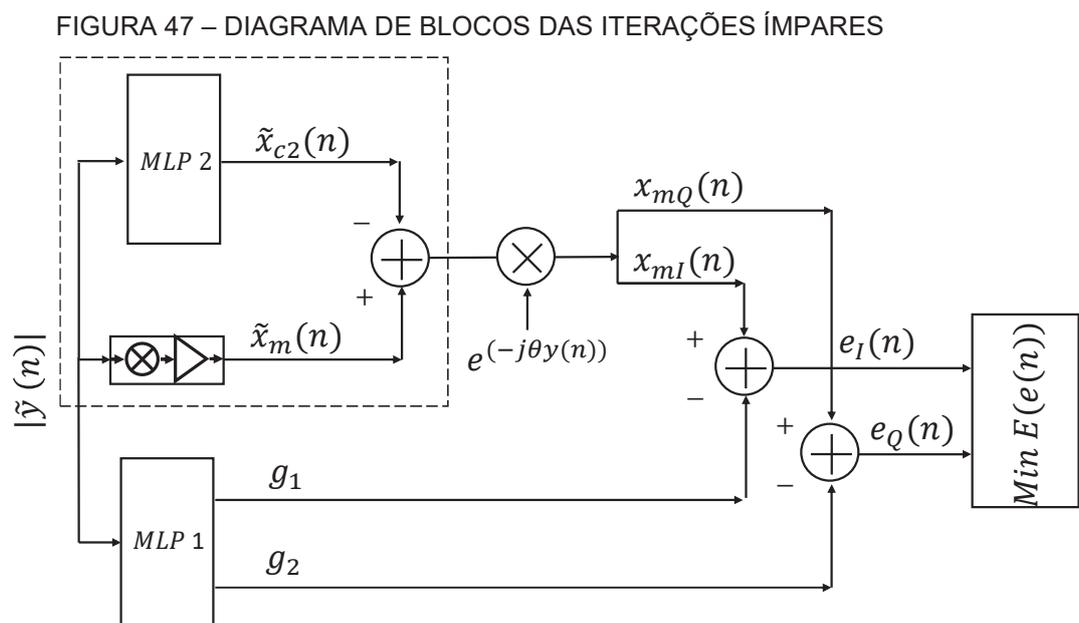
Não se tem conhecimento direto dos sinais \tilde{x}_{c1} e \tilde{x}_{c2} então foi adotada uma abordagem de treinamento iterativo. Uma informação que se tem conhecimento é que existe uma relação entre esses dois termos, onde o segundo termo possui uma componente que é o conjugado da componente do primeiro termo e essa informação será utilizada para se realizar o treinamento iterativo. Cada rede é treinada separadamente, porém o sinal desejado de cada rede depende da resposta da outra rede como será explicado adiante.

Nas iterações ímpares, descrita na FIGURA 47, a rede MLP 1 é treinada. O sinal desejado da rede MLP 1 provém da subtração da saída da rede MLP 2, \tilde{x}_{c2} , do sinal medido do PA, designado como \tilde{x}_m :

$$\tilde{x}_{c1} = \tilde{x}_m - \tilde{x}_{c2}. \quad (43)$$

É retirada a fase de entrada desse sinal utilizando o termo $e^{-j\theta y(n)}$. Esse sinal então é decomposto em suas componentes em fase, $x_{mI}(n)$, e em quadratura, $x_{mQ}(n)$, para então ser comparado com os sinais de saída da rede MLP 1, g_1 e g_2 , gerando os sinais de erro, $e_I(n)$ e $e_Q(n)$.

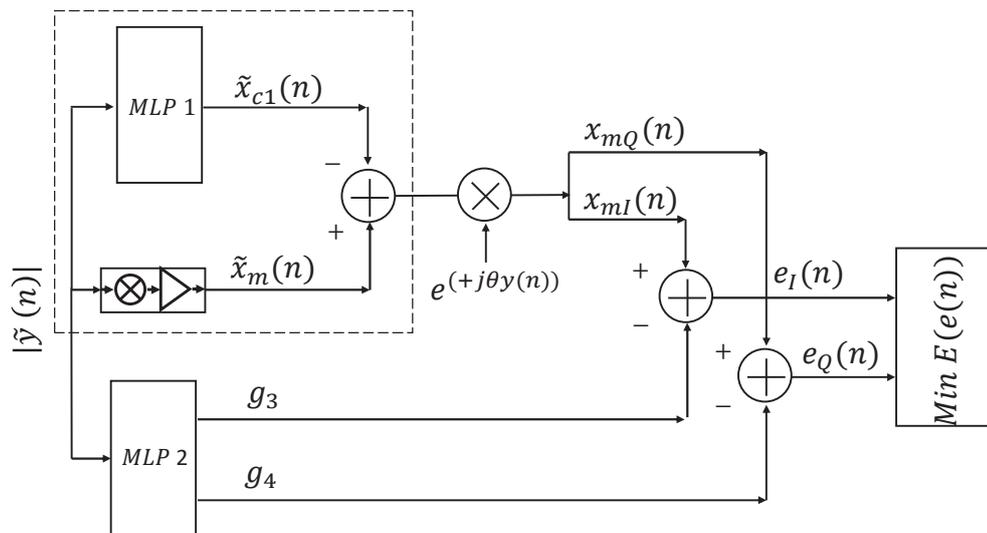
Na primeira iteração, quando a rede MLP 1 será treinada pela primeira vez, ainda não se têm o sinal \tilde{x}_{c2} , pois a rede MLP 2 nunca foi treinada, dessa forma esse sinal é igual a zero.



FONTE: (FREIRE, MARCONDES e LIMA, 2018)

Na FIGURA 48 é mostrado o processo de treinamento nas iterações pares. O processo é o mesmo que para as iterações ímpares, mas para treinar a rede MLP 2. Para se retirar a fase do sinal a ser modelado se utiliza o termo $e^{(+j\theta y(n))}$. Ao se fazer isso é forçado que a saída de uma rede seja o conjugado da outra rede. As redes são treinadas até o erro alcançar um patamar em que não há grandes variações nos valores.

FIGURA 48 – DIAGRAMA DE BLOCOS DAS ITERAÇÕES PARES



FONTE: (FREIRE, MARCONDES e LIMA, 2018)

O sistema a ser modelado foi simulado e foram definidos modelos de PA e modulador. Foi utilizado um sinal LTE OFDMA como o sinal de envoltória de entrada com 10 MHz de largura de banda. Esse sinal é amostrado em uma frequência de 61,44 MHz. O PA sendo utilizado para teste é modelado por polinômios de 5° ordem seguido por um filtro centrado em 1,2288 GHz cujos coeficientes encontram-se em (TEMPORAL, 2019). O modulador I/Q utilizado para teste tem um desbalanceamento de 1,2 na amplitude e $\pi/3$ na fase, além de uma frequência da portadora de 1,2288 GHz. Foram utilizados valores elevados de desequilíbrios para permitir uma maior diferenciação no tratamento destas imperfeições pelas diferentes abordagens. Para se realizar a medida do sinal de envoltória é usado um modulador ideal usando um filtro passa baixas com 122,88 MHz.

O modelo proposto será validado e comparado com o modelo da literatura da FIGURA 45. Na FIGURA 49 e na FIGURA 50 são mostrados os valores de NMSE e ACEPR em função do número de parâmetros da rede, respectivamente. O modelo

proposto obteve melhorias na ordem de 11 dB em NMSE e 16 dB em ACEPR para números de parâmetros similares em comparação com o modelo usual. Na Tabela 2 se encontram os valores de referência da comparação.

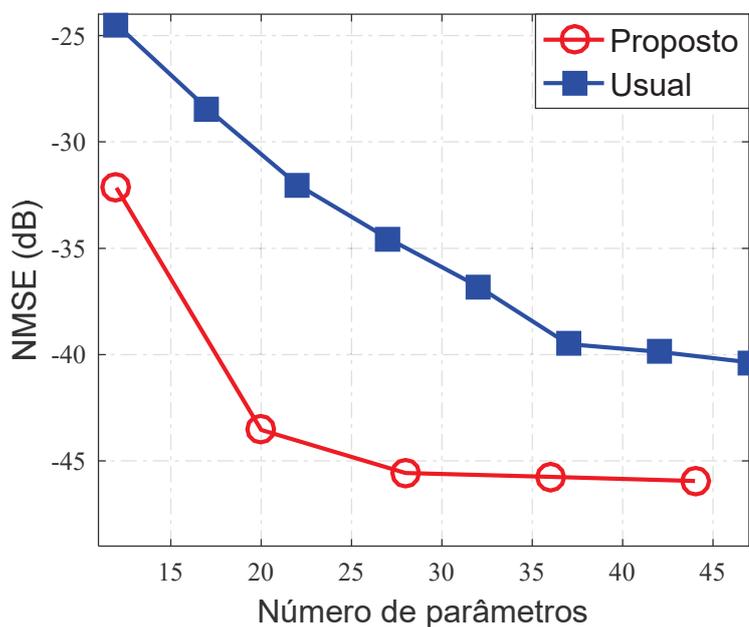
Tabela 2 – Valores comparativos de NMSE e ACEPR

	NMSE (dB)	ACEPR (dB)	número de parâmetros
Usual	-34,5	-52	27
Proposto	-45,5	-68	28

FONTE: O autor (2021)

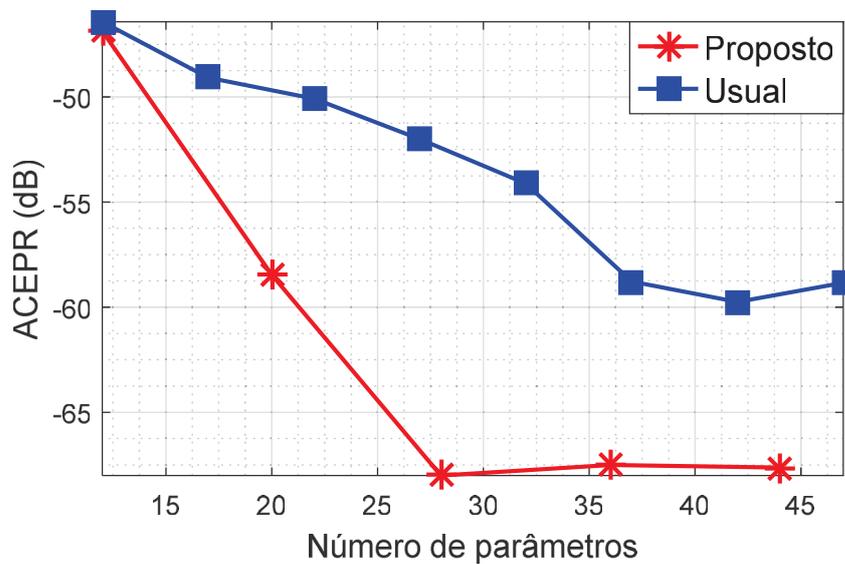
A FIGURA 51 mostra os valores dos PSDs dos sinais de erro dos modelos comprovando maior acurácia de modelagem em todo o espectro de frequência.

FIGURA 49 – VALORES DE NMSE



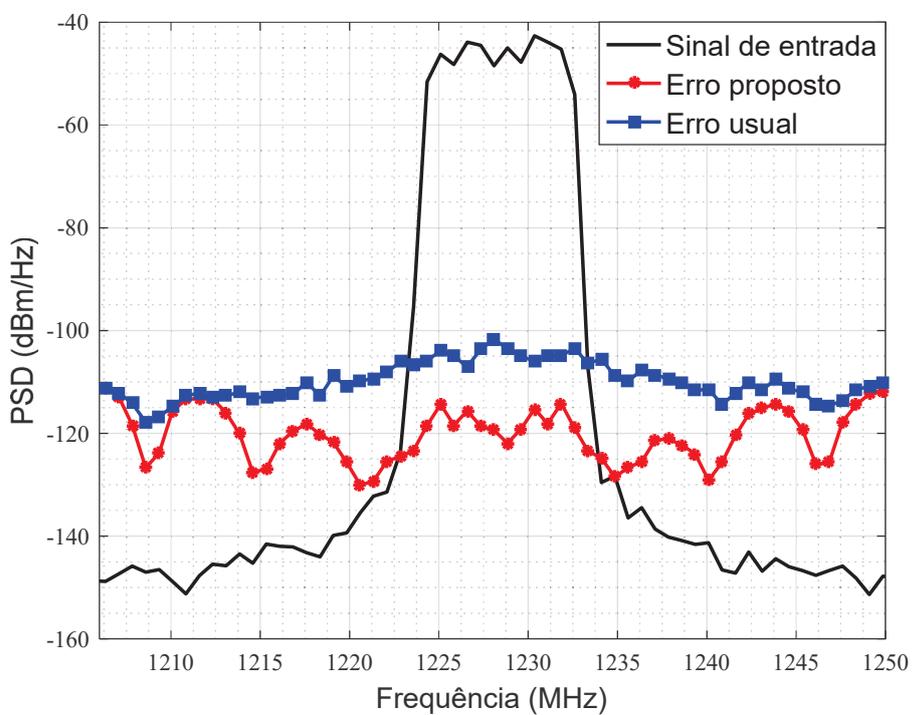
FONTE: (FREIRE, MARCONDES e LIMA, 2018)

FIGURA 50 – VALORES DE ACEPR



FONTE: (FREIRE, MARCONDES e LIMA, 2018)

FIGURA 51 – PSD DOS SINAIS DE ERRO



FONTE: (FREIRE, MARCONDES e LIMA, 2018)

LEGENDA: O modelo proposto apresenta 5 neurônios e 44 parâmetros e o modelo usual apresenta 9 neurônios e 52 parâmetros.

6.3 ESTUDO DE CASO DE 15 MODELOS CONJUNTO DE PA E DO MODULADOR I/Q BASEADO EM REDES TLP DE VALOR REAL E COMPLEXO

Nessa seção serão avaliados diversos modelos sem memória com diferentes arquiteturas e sinais de entrada e saída. O objetivo é avaliar como a rede TLP influencia a acurácia e complexidade computacional do modelo conjunto PA e modulador.

Serão avaliados modelos baseados em redes TLP de valores reais com diferentes decomposições dos sinais banda base de valor complexo de entrada do PA. Em (FREIRE, MARCONDES e LIMA, 2018), (FREIRE, FRANÇA e LIMA, 2014), (FREIRE, FRANÇA e LIMA, 2015) e (LIMA, CUNHA e PEDRO, 2011), estão descritos os modelos base para os modelos analisados nesse estudo.

Todos os modelos analisados são resumidos na Tabela 6. A classificação dos modelos é feita em base nos sinais de entrada e saída da rede TLP. Os modelos com TLP de valores reais vão dos modelos *A* até *C*. Os modelos com TLP de valor complexo vão de *D* até *F*. Existe uma subclassificação que indica diferentes quantidade de redes TLP no modelo ou diferentes arranjos do sinal de saída das redes.

Por exemplo, os modelos *C* têm como sinais desejados os componentes em fase e em quadratura do sinal de entrada medido do PA, $\tilde{x}(n) = x_1^I + jx_1^Q$ e $\tilde{x}^*(n) = x_2^I - jx_2^Q$. Esses sinais são multiplicados por $e^{j\theta(n)}$ e $e^{-j\theta(n)}$ respectivamente para recuperar a informação da fase. Porém, o modelo C_1 utiliza uma rede com 4 saídas: $x_1^I, x_1^Q, x_2^I, x_2^Q$. Já o modelo C_4 utiliza duas redes, onde cada rede tem duas saídas cada, a rede 1 gera as componentes em fase, x_1^I e x_2^I , e a rede 2 gera as componentes em quadratura, x_1^Q, x_2^Q .

A Tabela 3 e a Tabela 4 mostram os sinais de entrada e saída possíveis utilizados nos modelos, respectivamente. A designação do sinal serve para facilitar a visualização. Todos os modelos seguem as premissas necessárias para um modelo gerar apenas contribuições dentro da banda.

Tabela 3 – Definição dos sinais de entrada usados nos modelos analisados

Designação das entradas	Valores de entrada
I_1	$\tilde{y}(n)$
I_2	$ \tilde{y}(n) $
I_3	y_I, y_Q

FONTE: O autor (2021)

Tabela 4 – Definição dos sinais de saída usados nos modelos analisados

Designação da saída	Valores de saída
O_1	\tilde{x}
O_2	$x_I + jx_Q$
O_3	$\tilde{x}_1 e^{j\theta(n)} + \tilde{x}_2 e^{-j\theta(n)}$
O_4	$\tilde{x}_1 \cos(\theta(n)) + j\tilde{x}_2 \sin(\theta(n))$
O_5	$(x_1^I + jx_1^Q) e^{j\theta(n)} + (x_2^I + jx_2^Q) e^{-j\theta(n)}$
O_6	$(x_1^I + jx_1^Q) \cos(\theta(n)) + j(x_2^I + jx_2^Q) \sin(\theta(n))$

FONTE: O autor (2021)

Em modelos que se utilizam da estratégia de retirar a informação da fase do sinal de entrada é necessário reaver essa informação no sinal de saída da rede TLP. É proposto nesse estudo uma nova forma de recuperar a informação da fase. É comumente utilizado os termos $e^{j\theta(n)}$ e $e^{-j\theta(n)}$ multiplicado aos sinais de saída para reaver a informação da fase, porém de maneira equivalente podem ser usados os termos $\cos(\theta(n))$ e $\sin(\theta(n))$. Para se comprovar essa equivalência tem-se que:

$$e^{j\theta(n)} = \cos(\theta(n)) + j\sin(\theta(n)). \quad (44)$$

Então agora a resposta a ser modelada seria:

$$\tilde{x}(n) = (a_3 \tilde{y}(n) + a_4 \tilde{y}(n)) \cos(\theta(n)) + j(a_3 \tilde{y}(n) - a_4 \tilde{y}(n)) \sin(\theta(n)). \quad (45)$$

O sistema a ser modelado foi descrito na Seção 6.2. Na análise dos modelos será usada a quantidade de FLOPs para avaliar sua complexidade. O número de FLOPs é obtido somando a quantidade de multiplicações e somas. O número de multiplicações pode ser representado como uma função do número de entradas (n_e), saídas (n_s), neurônios (n) e quantidade de redes (r):

$$\text{multiplicações} = r * (n_e * n + n_s * n). \quad (46)$$

Em (46) cada rede do modelo possui quantidades iguais de entradas, saídas e neurônios.

É importante ressaltar as diferenças entre multiplicação de valores complexos e multiplicação de valores reais. A última é muito menos complexo requerendo apenas 1 FLOP, enquanto a primeira requer 6 FLOPs. Na Tabela 5 estão os fatores de conversão de operações com números complexos e reais para FLOP.

Tabela 5 – Conversão de operações para FLOPs

Operação	Número de FLOPs
adição real	1
multiplicação real	1
adição complexa	2
multiplicação complexa	6

FONTE: O autor (2021)

Na Tabela 6 são resumidos os 15 modelos em estudo que serão validados e analisados. Os vários modelos foram simulados até sua convergência (valores que não mudam de forma significativa com o aumento do número de neurônios da rede TLP), o que se busca são modelos que convirjam mais rapidamente a valores de NMSE e ACEPR.

Em sequência serão apresentados os gráficos dos resultados de NMSE por número de FLOPs dos modelos propostos por tipo de sinais de entrada e saída a serem modelados. Também são apresentados para cada grupo de modelos os diagramas com os respectivos sinais de entrada e saída. Os grupos de modelos são classificados por letras e são eles: *B, C, E* e *F*. Os grupos *B* e *C* apresentam modelos de valores reais, já os grupos *E* e *F* apresentam modelos de valores complexos. Os grupos de modelos *A* e *D* são os modelos de valores reais e valores complexos existentes na literatura respectivamente e não são mostrados nesse momento.

Tabela 6 – Resumo dos modelos sendo analisados.

Modelo	I	O	FLOPs	NMSE (dB)
A_1	3	2	32	-31,82
A_2^*	3	2	36	-36,79
B_1	2	5	30	-38,28
B_2^{**}	2	5	32	-38,47
B_3^*	2	5	36	-38,72
B_4^*	2	5	36	-39,33
C_1	2	6	30	-39,33
C_2^{**}	2	6	32	-35,87
C_3^*	2	6	36	-35,27
C_4^*	2	6	36	-37,79
D	1	1	32	-27,98
E_1	1	3	48	-32,86
E_2^*	1	3	32	-30,53
F_1	1	4	48	-36,16
F_2^*	1	4	32	-30,41

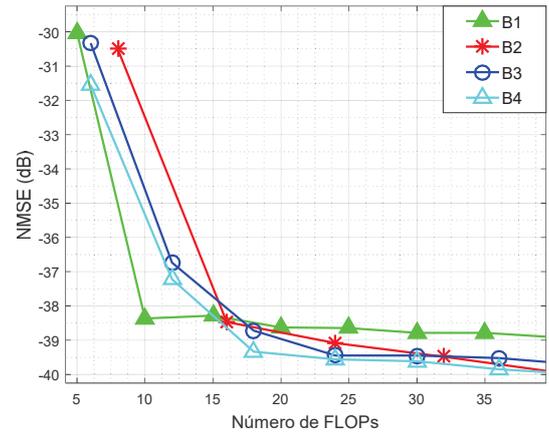
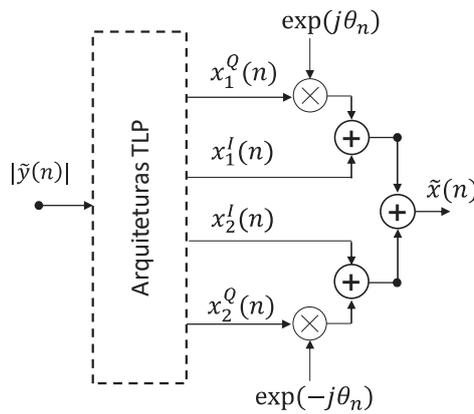
FONTE: O autor (2021)

Nota: *modelo que usa 2 redes TLP. ** modelo que usa 4 redes TLP

Os modelos 1, 2, 3 e 4 dos conjuntos B e C seguem características similares de arquitetura. A principal diferença entre os dois grupos é a forma de recuperar a informação da fase no sinal de saída da rede. Os modelos B_1 e C_1 possuem 1 rede e quatro saídas. Os modelos B_2 e C_2 possuem 4 redes e uma saída cada rede. Os modelos B_3 e C_3 possuem 2 redes e duas saídas cada rede. Os modelos B_4 e C_4 possuem 2 redes e duas saídas cada rede, porém o que difere os modelos 3 e 4 são os sinais a serem modelado por cada rede.

Na FIGURA 52 são mostrados os resultados do grupo de modelos B . O modelo que convergiu mais rapidamente foi o modelo B_1 , porém o valor de NMSE por número de FLOPs alcançou valores próximos de -38 dB enquanto os outros modelos obtiveram resultados próximos de -39 dB. Lembrando que a classificação de um modelo comparativamente a outros modelos depende das necessidades da aplicação. O guia dessa classificação são os valores de acurácia e quantidade de FLOPs. Nesse caso, o modelo B_1 para o valor específico de 10 FLOPs obteve o melhor resultado dentre os modelos. Os modelos B_2 , B_3 e B_4 obtiveram valores de convergência aproximados, porém o modelo B_4 chegou mais rapidamente no valor de convergência, ou seja, para um mesmo valor de NMSE apresentou menores valores de número de FLOPs.

FIGURA 52 – VALORES DE NMSE GRUPO B

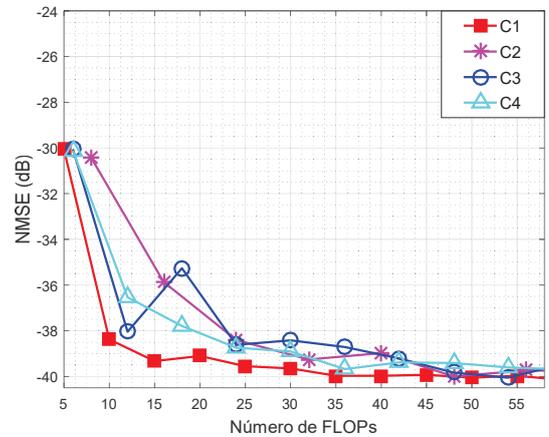
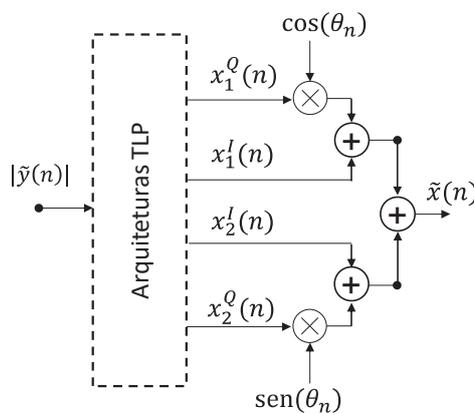


Modelos (B)

FONTE: O autor (2021)

Na FIGURA 53 são apresentados os resultados do grupo de modelos *C*. O modelo que convergiu mais rapidamente e que também apresentou o menor valor de NMSE por quantidade de FLOPs foi o modelo *C*₁.

FIGURA 53 - VALORES DE NMSE GRUPO C

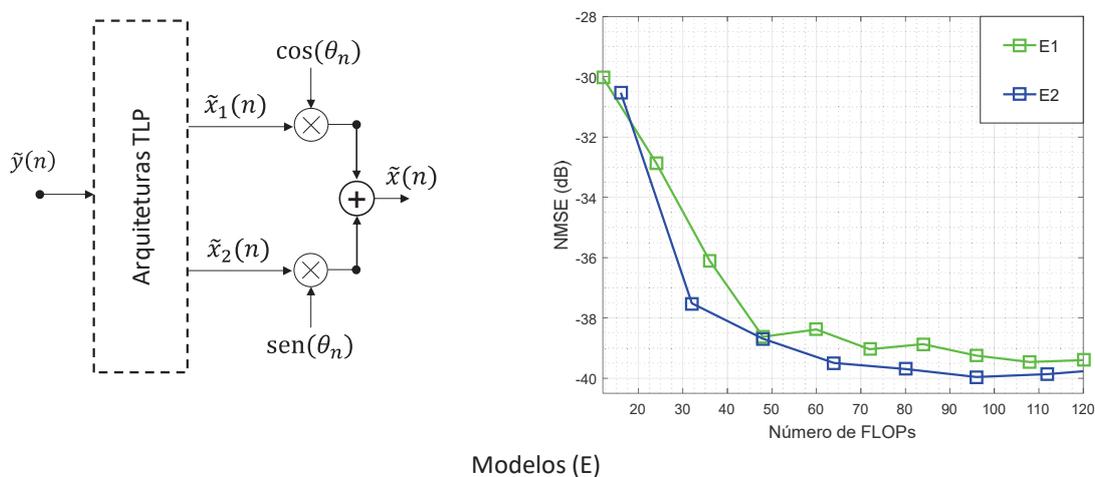


Modelos (C)

FONTE: O autor (2021)

Os grupos de modelos *E* e *F* utilizam valores complexos. A diferença entre os dois grupos está na forma de reaver a informação da fase no sinal de saída da rede. Os modelos *E*₁ e *F*₁ utilizam uma rede e duas saídas. Os modelos *E*₂ e *F*₂ utilizam duas redes e cada rede possui uma saída. Pode-se observar na FIGURA 54 e na FIGURA 55 que os modelos com duas redes convergiram para valores menores de NMSE e convergiram mais rapidamente a esses valores.

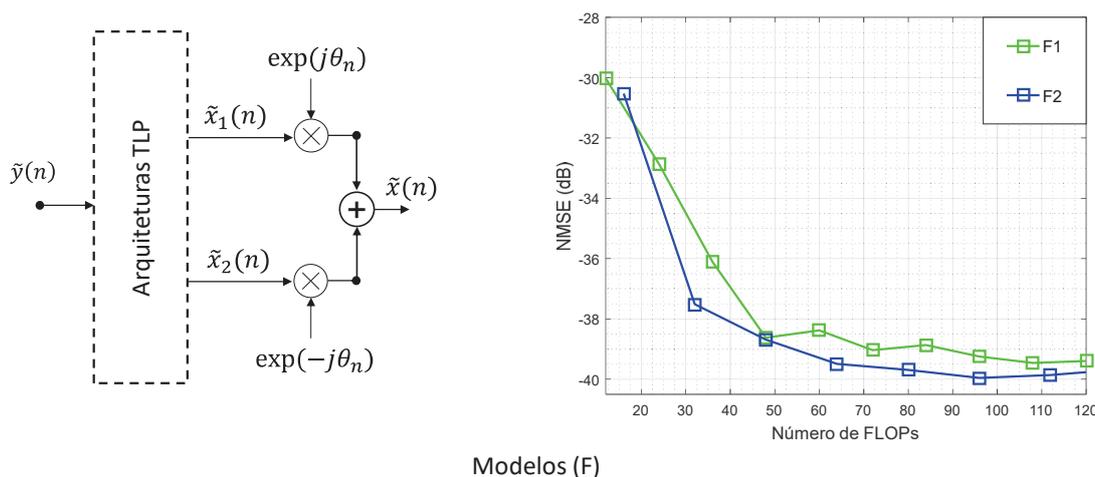
FIGURA 54 - VALORES DE NMSE GRUPO E



Modelos (E)

FONTE: O autor (2021)

FIGURA 55 - VALORES DE NMSE GRUPO F



Modelos (F)

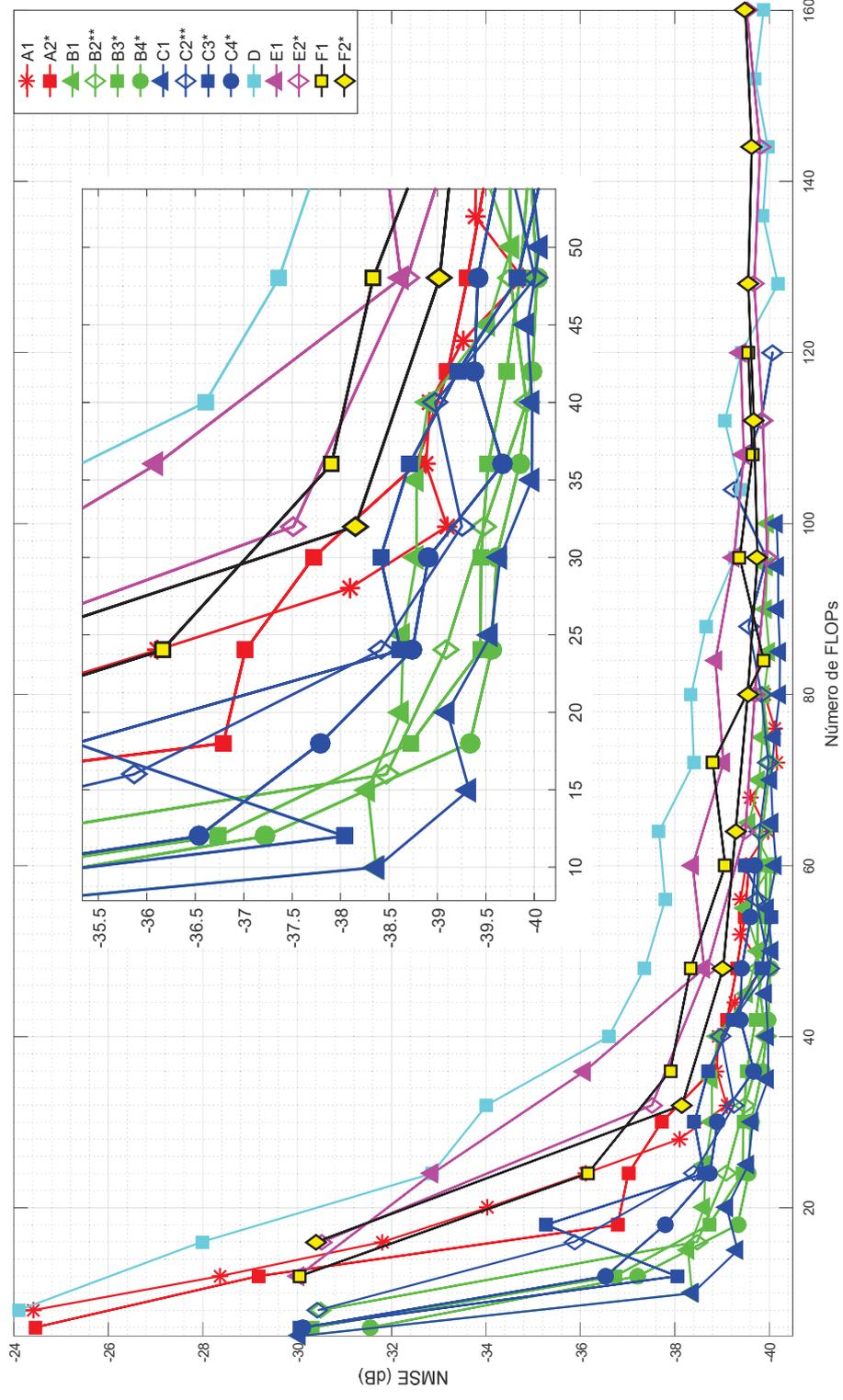
FONTE: O autor (2021)

A fim de apresentar uma visualização geral dos modelos a FIGURA 56 mostra o NMSE dos modelos A_1 até F_2 em função dos FLOPs. Os modelos que não obedecem às restrições do comportamento passa-banda do PA claramente obtiveram pior desempenho, são os modelos de valor real A_1 e A_2 e o modelo de valor complexo D . No gráfico, com o auxílio da imagem em aproximação, pode-se observar que os modelos B_2 , B_4 e C_1 apresentam valores próximos de NMSE e boa estabilidade de treinamento não tendo pontos com variações acentuadas. Porém, o modelo C_1 converge mais rapidamente alcançando valores menores de NMSE para valores próximos de FLOPs.

Os modelos B_1 e C_1 possuem a mesma estrutura. A diferença entre os dois modelos está na forma como é recuperada a informação da fase. O modelo proposto C_1 obtém resultados superiores, como pode ser visto na Tabela 6, demonstrando que os sinais na forma $\cos(\theta(n))$ e $\sin(\theta(n))$ geraram melhor acurácia do modelo com mesma quantidade de parâmetros. Os modelos com 1 ou duas redes de valor real obtiveram melhor resposta que os modelos com quatro redes, B_2 e C_2 .

A FIGURA 57 apresenta o NMSE dos modelos em termos do número de parâmetros. Os modelos baseados em TLP de número complexo apresentaram desempenho comparável com o melhor modelo baseado em TLP de números reais. Fica claro as qualidades de modelagem das redes de valor complexo, porém sua aplicação prática é dificultada pela grande quantidade de FLOPs que é necessária, mais especificamente a quantidade de FLOPs para realizar multiplicações complexas, já que cada operação de multiplicação complexa se converte em 6 FLOPs.

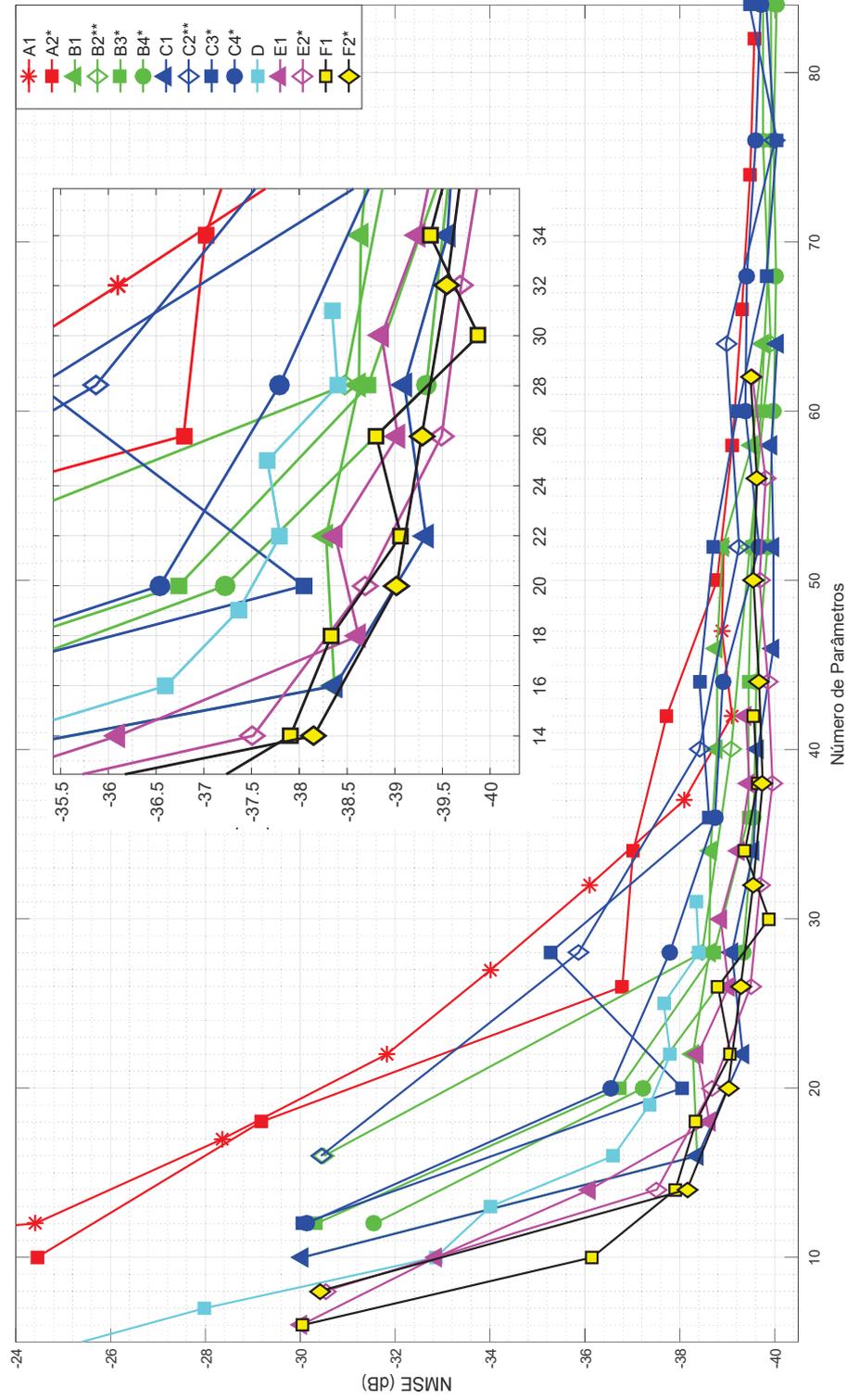
FIGURA 56 – VALORES DE NMSE



FONTE: (FREIRE, MARCONDES e LIMA, 2018)

LEGENDA: NMSE dos 15 modelos em estudo em função do número de FLOPs e zoom dos valores de NMSE para menores números de FLOPs.

FIGURA 57 – VALORES DE NMSE



FONTE: (FREIRE, MARCONDES e LIMA, 2018)

LEGENDA: NMSE dos 15 modelos em estudo em função do número de parâmetros e zoom dos valores de NMSE para menores números de parâmetros.

7 CONSIDERAÇÕES FINAIS

Para sistemas DPD é necessário um modelo do inverso do comportamento do PA sendo linearizado. Esses modelos devem ter uma acurácia suficiente para linearizar o PA e apresentar baixa complexidade computacional. Ainda não se têm normas e padrões que definam os valores das métricas utilizadas para mensurar a acurácia de um modelo, dessa forma deve ser avaliado caso a caso o quão acurado deve ser um modelo. Isso se deve principalmente pelo fato do comportamento dos PAs variarem de acordo com a tecnologia do PA, com a modulação do sinal que processam, com o ambiente de operação dentre outros fatores. Por exemplo, PAs de tecnologias distintas podem ser mais ou menos sensíveis à memória de longo termo, sendo necessário adaptar os modelos para cada aplicação. A complexidade computacional usualmente empregada é a quantidade de parâmetros que uma rede neural utiliza. Porém, pode-se utilizar os números de FLOPs necessários para implementar tal modelo. Neste trabalho algumas análises efetuadas não dão enfoque à complexidade de implementação dos modelos, analisando apenas as interações das redes TLP e dos sinais processados por ela sendo então utilizada a métrica de parâmetros de rede. A complexidade de operação também é analisada para certos modelos.

Esse trabalho analisou as características do sinal processado pelo PA e suas interações com as funções de ativação das redes TLP. Na literatura foi demonstrado que os modelos de valor real com as funções de ativação genéricas normalmente utilizadas para as várias aplicações na literatura geravam contribuições fora da banda de operação de um PA. Isso ocorre porque as funções de ativação apresentam um comportamento não linear que modifica a informação de fase do sinal sendo processado por ela. Existem até agora duas formas de mitigar esse problema: modificar as entradas de tal forma que não possuam a informação da frequência da portadora; ou utilizar uma função de ativação que não deforme a informação da fase do sinal de entrada.

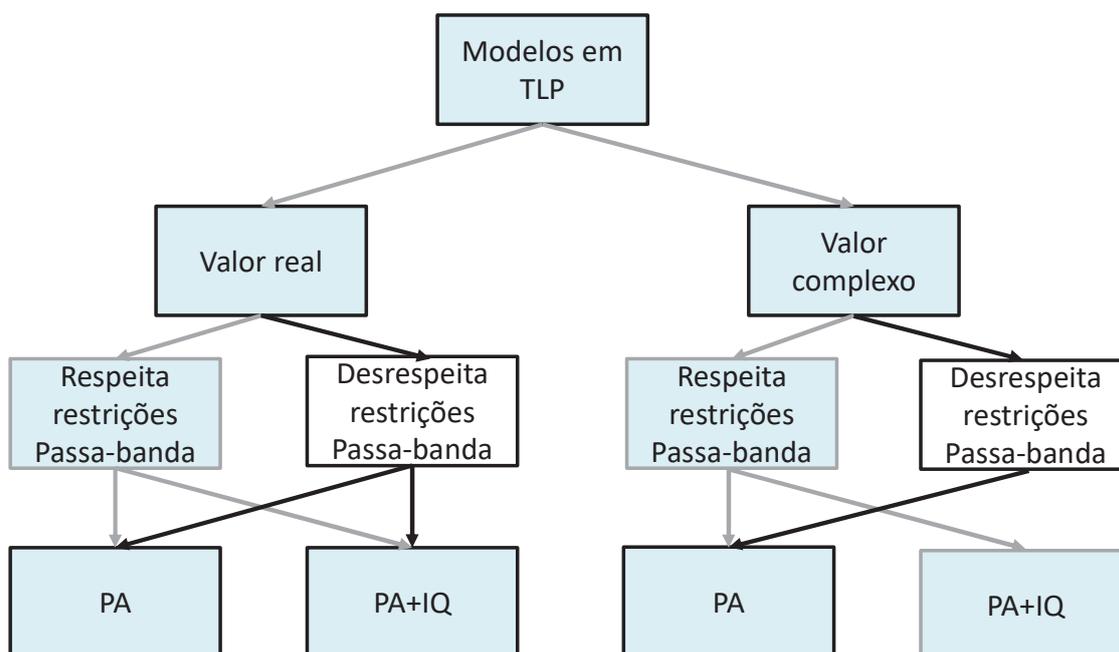
As contribuições desse trabalho foram analisar os modelos baseados em redes TLP de valor complexo para verificar se suas funções de ativação também geravam contribuições fora da banda. Esse fato foi comprovado e mesmo as funções que não geravam contribuições espúrias não eram capazes de mapear todas as características de um PA. Dessa forma foram propostos dois modelos de PA

baseados em redes neurais complexas que mitigavam essa questão. Um modelo modificava os sinais de entrada para não apresentar a informação da frequência da portadora e outro modelo que utilizava uma nova função de ativação proposta pela autora. Os modelos apresentaram melhor acurácia com quantidades aproximadas de valores de parâmetros de valores complexos.

Outra contribuição foi analisar os modelos do modulador em quadratura e do PA de forma conjunta. Os modelos da literatura utilizavam as mesmas funções de ativação dos modelos de PA de valor real. Dessa forma fica claro a não adequação desses modelos às restrições do PA. Além disso, esses modelos envolvem modelar um sinal e seu conjugado, e esses sinais têm natureza muito distintas não sendo possível mapear esses dois sinais em um mesmo espaço amostral. Em um estudo de caso foi proposto um modelo que possuía duas redes TLP e que modificava as entradas das redes para não possuir a informação da frequência da portadora. Por um processo de treinamento iterativo cada rede deve mapear separadamente as características mencionadas. Ainda, a fim de avaliar as características ideais para modelagem do sistema conjunto de modulador e PA foram estudados quinze modelos com arquiteturas diferentes utilizando redes TLP de valor real e de valor complexo. O modelo que obteve uma convergência mais rápida, exigindo assim menos números de parâmetros para valores semelhantes de NMSE utilizou uma rede TLP de valor real, onde as entradas respeitavam as restrições impostas pelo sinal passa banda do PA. Uma característica desses modelos que retiram a informação da fase do sinal de entrada é que essa informação precisa ser reinserida no sinal de saída da rede. O diferencial desse último modelo foi reinserir essa informação da fase através de sinais senos e cossenos e não na forma exponencial, demonstrando que a forma de apresentar os sinais à rede influencia na facilidade de modelagem dos sinais e por conseguinte na relação entre acurácia e número de parâmetros. Os modelos de valor complexo apresentam características matemáticas interessantes para modelagem de sistemas com sinais de natureza senoidais, porém, ainda que apresentem bons resultados de acurácia quando comparados com os números de parâmetros das redes, ainda não são adequados para implementação apresentando valores de FLOPs elevados.

A FIGURA 58 apresenta um resumo dos modelos apresentados neste trabalho e auxiliam na visualização das contribuições deste trabalho.

FIGURA 58 – DIAGRAMA DE CONTRIBUIÇÕES DO TRABALHO



FONTE: O autor (2021)

LEGENDA: O diagrama representa os modelos da literatura e propostos neste trabalho e separa os modelos baseado em características abrangentes. As características de modelos em cor cinza e azul têm ambas as características inovadoras. As características com preto e azul têm a característica anterior como inovação. E os modelos com preto e branco são os modelos existentes na literatura. As setas cinza indicam o caminho das contribuições do trabalho, onde não há seta é que não existe tal modelo.

Esse trabalho apresentou embasamentos teóricos que demonstram que se levar em consideração as características individuais das redes neurais e do sistema ao qual ela deve modelar, a rede tem mais facilidade de mapear essas características e convergir. Os próximos trabalhos devem levar em consideração essas teorias e adaptar os modelos para serem menos complexos para implementação. Recentemente novos recursos de treinamentos de redes neurais estão sendo estudados que melhoram o desempenho das redes neurais, como o algoritmo de otimização Adam (JARAUT, ABDELHAFIZ, *et al.*, 2021) e outras topologias de redes que possuem características interessantes como a rede neural convolucional, que não muda a estrutura da rede ao se mudar a dimensão de seu sinal de entrada e possui uma característica de compartilhamento de pesos (HU, LIU, *et al.*, 2021). Essas novas abordagens podem simplificar a implementação das redes neurais e se beneficiar das sugestões propostas neste trabalho. Além disso, existem vários trabalhos que focam

na simplificação da implementação de sistemas que lidam com números complexos, alguns exemplos são: (PAVULURI, PRASAD e RAMBADU, 2013), (VARGHESE, PRADEEP, *et al.*, 2015) e (TOMASKAR e KHANDALE, 2014).

REFERÊNCIAS

BASSEY, J.; QIAN, L.; LI, X. A survey of complex-valued neural network. **arXiv:2101.12249**, p. 1-15, 2021.

BENEDETTO, S.; BIGLIERI, E. Modeling and performance evaluation of nonlinear satellite links - a Volterra series approach. **IEEE Trans. Aerosp. Electron. Syst.**, v. 15, n. 4, p. 494-507, 1979.

BENEDETTO, S.; BIGLIERI, E. **Principles of Digital Transmission: with wireless applications**. 2. ed. New York: Kluwer Academic/Plenum Publishers, 1999.

BENVENUTO, N. et al. **Non linear satellite radio links equalized using blind neural networks**. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Toronto, USA: [s.n.]. 1991. p. 14-17.

BENVENUTO, N.; PIAZZA, F. On the complex backpropagation algorithm. **IEEE Transactions on Signal Processing**, 40, n. 4, 1992. p. 967-969.

BENVENUTO, N.; PIAZZA, F.; UNCINI, A. **A neural network approach to data predistortion with memory in digital radio systems**. Proceedings IEEE International Conference on Communication. Geneva, Switzerland: [s.n.]. 1993. p. 232-236.

BOYD, S.; CHUA, L. Fading memory and the problem of approximating nonlinear operators with Volterra series. **IEEE Trans. Circuits Syst.**, CAS-32, 1985. p. 1150-1161.

CHANG, K. **RF and Microwave Wireless Systems**. Texas: JOHN WILEY & SONS, INC., 2000.

CHEN, M. et al. Machine Learning for Wireless Networks with Artificial Intelligence: A Tutorial on Neural Networks, oct. 2017. Disponível em: <<https://arxiv.org/pdf/1710.02913.pdf>>. Acesso em: 2018.

CHEN, S. et al. Adaptive equalization of finite non linear channels using multilayer perceptrons. **Signal Processing**, v. 20, p. 107-119, 1990.

CHURCHLAND, P. S.; SEJNOWSKI, T. J. **The Computational Brain**. Cambridge: MIT press, 1992.

CISCO. **Cisco Visual Networking Index: Forecast and Methodology, 2016–2021**. [S.l.]. 2017.

DESGREYS, et al. **Wideband Power Amplifier Predistortion: Trends, challenges and solutions**. 2017 IEEE 12th International Conference on ASIC (ASICON). Guiyang: [s.n.]. 2017.

DUKE, U. Accurate neural network computer vision without the 'black box'. **ScienceDaily**, 15 Dez. 2020. Disponível em: <www.sciencedaily.com/releases/2020/12/201215140827.htm>.

EDWARDS, H. M. **Linear Algebra**. New York: Birkhäuser, 1995.

ERICSSON. 5G Energy Performance, 2015. Disponível em: <http://www.5gamericas.org/files/3114/3898/7332/Ericsson_wp-5g-energy-performance.pdf>.

ERICSSON. **Mobile network traffic growth remains steady**. [S.l.]. 2021.

FAULKNER, ; MATTSSON, T. Spectral sensitivity of power amplifiers to quadrature modulator misalignment. **IEEE Trans. Veh. Technol.**, v. 41, n. 4, p. 516-525, 1992.

FREIRE, L. B. C. et al. **Behavioral modeling of power amplifiers using perceptrons with complex activation functions**. Proceedings of the 35th South Symposium on Microelectronics. São Miguel das Missões: [s.n.]. 2020. p. 1-4.

FREIRE, L. B. C. et al. Behavioral Modeling of a multimode power amplifier using perceptrons with AM-AM and AM-PM complex-valued activation function. **Journal of Integrated circuits and systems**, 16, 2021. p. 1-11.

FREIRE, L. B. C.; FRANÇA, C. D.; LIMA, E. G. D. Low-pass equivalent behavioral modeling of RF power amplifiers using two independent real-valued neural networks. **PIER C**, v. 52, p. 125-133, 2014.

FREIRE, L. B. C.; FRANÇA, C. D.; LIMA, E. G. D. A modified real-valued feed forward neural network low-pass equivalent behavioral model for RF power amplifiers. **Progress In Eletromagnetics Research C**, v. 57, p. 43-52, 2015.

FREIRE, L. B. C.; LIMA, E. G. D. **Complex-valued three-layer perceptron for the behavioral modeling of RF power amplifiers**. Proceedings of the 32nd South Symposium on Microelectronics. Rio Grande, Rio Grande do Sul: [s.n.]. 2017. p. 1-4.

FREIRE, L. B. C.; MARCONDES, B. T.; LIMA, E. G. D. **A Three-layer Perceptron Tailored for the Joint Modeling of I/Q Modulator Mismatches and Power Amplifier Distortions**. South Symposium on Microelectronics (SIM). Curitiba: [s.n.]. 2018.

FREIRE, L. B. C.; MARCONDES, B.; LIMA, E. G. D. **Study of joint modeling of I/Q modulator impairments and power amplifier distortions based on three-layer perceptrons**. Proceedings of the 34th South Symposium on Microelectronics. Pelotas: [s.n.]. 2019. p. 131-134.

FU, K.; LAW, C. L.; THEIN, T. T. Novel neural network model of power amplifier plus IQ imbalances. **Progress In Elec. Research B**, v. 46, p. 177-192, 2013.

GANGWAR, A.; BHARDWAJ,. An Overview: Peak to Average Power Ratio in OFDM system & its Effect. **International Journal of Communication and Computer Technologies**, v. 1, n. 2, p. 22-25, 2012.

GEORGIU, G. M.; KOUTSOUGERAS, C. Complex domain backpropagation. **IEEE Transaction on Circuits and Systems II: Analog and Digital Signal Processing**, 39, 1992. p. 330-334.

GHANNOUCHI, F. M.; YOUNES, M.; RAWAT, M. Distortion and mismatches mitigation and compensation of single- and multi-band wireless transmitters. **IET Microw. Antennas Propag.**, v. 7, n. 7, p. 518-534, 2013.

GHANNOUCHI, F. M.; YOUNES, M.; RAWAT, M. Distortion and mismatches mitigation and compensation of single- and multi-band wireless transmitters, 7, n. 7, 2013. p. 518-534.

GUARIENTO,. Laboratório de Limnologia/UFRJ, 14 ago. 2014. Disponível em: <<https://limnonews.wordpress.com/2014/08/14/o-que-sao-graus-de-liberdade/>>. Acesso em: jun. 2018.

HAMMI, O. Augmented twin-nonlinear two-box behavioral models for multicarrier LTE power amplifier. **The scientific world journal**, v. 14, 2014.

HAYKIN, S. **Redes Neurais: princípios e práticas** 2º edição. Ontário, Canadá: Bookman, 2001.

HAYKIN, S. **Neural networks and learning machines**. Ontario: Prentice Hall, 2009.

HIROSE, A. Nature of complex number and complex-valued neural networks. **Frontiers of Electrical and Eletronic Engineering in China**, 6, 2011. p. 171-180.

HIROSE, A. **Studies in Computacional Intelligence: Complex-Valued Neural Networks**. 2. ed. [S.l.]: Springer, v. 400, 2012.

HIROSE, A.; YOSHIDA, S. Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. **IEEE Trans. Neural Netw. Learn Syst.**, 23, 2012. p. 541-551.

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences**, v. 79, p. 2554-2558, 1982.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Netw.**, v. 2, p. 359-366, 1989.

HU, X. et al. Convolutional neural network for behavioral modeling and predistortion of wideband power amplifiers, 2021.

HUANG, Z. et al. **Forward behavioral modeling of concurrnt dual-band power amplifiers using extended real valued time delay neural networks**. International Conference on Microwave and Milimeter Wave Technology (ICMMT). Shenzhen, China: [s.n.]. 2012. p. 1-4.

IBNKAHLA, M. et al. Neural networks for modeling nonlinear memoryless communcation channels. **IEEE Transactions on Communications**, 45, n. 7, 1997. p. 768-771.

ISAKSSON, M.; WISSEL, D.; RONNOW, D. A comparative analysis of behavioral models for RF power amplifiers. **IEEE Trans. Microw. Theory Tech.**, v. 54, n. 1, p. 348-359, 2006.

JARAUT, P. et al. Augmented convolutional neural network for behavioral modeling and digital predistortion of concurrent multiband power amplifiers. **IEEE transactions on microwave theory and techniques**, 2021. p. 1-15.

KIM, J. et al. Analysis of adaptive digital feedback linearization techniques. **Circuits and Systems I: Regular Papers, IEEE Transactions on**, v. 57, n. 2, p. 345-354, 2010.

KRIESEL, D. A Brief Introduction to Neural Network, Mai. 2005. Disponivel em: <http://www.dkriesel.com/en/science/neural_networks>. Acesso em: Mai 2018.

KRÖSE, B.; VAN DER SMAGT, P. **An Introduction to Neural Network**. Amsterdam: The University of Amsterdam, 1996.

KURENKOV, A. A 'Brief' History of Neural Nets and Deep Learning, 2015. Disponivel em: <<http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning-part-4/>>. Acesso em: 2018.

LANDIN, P. **Linearization of High Efficiency Transmitters for Wireless Communications**. KTH Royal Institute of Technology. 2012.

LEE, Y.-S. et al. **A wideband multi-branch analog predistorter with memory-effect compensation for multi-carrier wcdma repeater systems**. in International Microwave Symposium Digest (IEEE MTT-S). Boston: [s.n.]. 2009.

LIMA, E. G.; CUNHA, T. R.; PEDRO, J. C. A Physically Meaningful Neural Network Behavioral Model for Wireless Transmitters Exhibiting PM-AM/PM-PM Distortions. **IEEE Trans. Microw. Theory Tech.**, v. 59, n. 12, p. 3512-3521, 2011.

LIU, T.; BOUMAIZA, S.; GHANNOUCHI, F. M. Dynamic behavioral modeling of 3G power amplifier using real valued time-delay neural networks. **IEEE transactions on microwave Theory and Techniques**, 52, n. 3, 2004. p. 1025-1033.

LUONGYINH, D.; KWON, Y. **Behavioral modeling of power amplifiers using fully recurrent neural networks**. Digest 2005 IEEE MTT-S Internamntional Microwave Symposium (IMS). Long Beach, Califórnia: [s.n.]. 2005. p. 1979-1982.

M. GHANNOUCHI, F.; HAMMI, O.; HELAOUI, M. **Behavioral Modeling and Predistortion of Wideband Wireless Transmitters**. [S.l.]: Wiley, 2015.

MATHEWS, V.; SICURANZA, G. **Polynomial Signal Processing**. New York: Wiley, 2000.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of mathematical Biophysics**, v. 5, p. 115-133, 1943.

MINSKY, M.; PAPERT, S. **Perceptrons: An Introduction to Computational Geometry**. [S.l.]: The MIT Press, 1969.

MKADEM, F. et al. **Behavioral modeling and digital predistortion of power amplifiers with memory using two hidden layers artificial neural networks**. IEEE MTT-S Int. Microwave Symposium Digest. [S.l.]: [s.n.]. 2010. p. 656-659.

MORGAN, D. R. et al. A generalized memory polynomial model for digital predistortion of RF power amplifiers. **IEEE Trans. Signal Process.**, 54, 2006. p. 3852-3860.

MUHA, M. S. et al. **Validation of power amplifier nonlinear block models**. IEEE MTT-S International Microwave Symposium Digest. Anaheim, CA: [s.n.]. 1999. p. 759-762.

NASKAS, N.; PAPANANOS, Y. **Adaptive baseband predistorter for radio frequency power amplifiers based on multilayer perceptron**. Proceedings 9th International Conference on Electronics, Circuit and Systems. Dubrovnik, Croatia: [s.n.]. 2002. p. 1107-1110.

NEIL BRAITHWAITE, R. **Analog Linearization Techniques Suitable for RF Power Amplifiers Used in Integrated Transmitters**. Compound Semiconductor Integrated Circuit Symposium (CSICS), 2013 IEEE. Monterey: [s.n.]. 2013.

NUNES DA SILVA, I. et al. **Artificial Neural Networks: A Practical Course**. São Carlos, São Paulo: Springer, 2017.

P. STAPLETON, S.; FRASER, S.; BURNABY, B. C. Adaptive Feedforward Linearization For RF Power Amplifiers. **Microwave Journal**, 1 Oct. 1999.

PAVULURI, M. K.; PRASAD, T. S. R. K.; RAMBADU, C. Design and implementation of complex floating point processor using FPGA. **International Journal of VLSI design & Communication Systems**, 4, 2013. p. 53-61.

PEARSON, R. K.; POTTMAN, M. Gray-box identification of block oriented nonlinear models. **Journal of Process Control**, 10, n. 4, 2000. p. 301-315.

PEDRO, J. C.; CARVALHO, N. B.; LAVRADOR, P. M. **Modeling nonlinear behavior of band-pass memoryless and dynamic systems**. Microwave Symposium digest IEEE MTT-S International. [S.l.]: [s.n.]. 2003. p. 2133-2136.

PEDRO, J. C.; MAAS, S. A. A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches. **IEEE Trans. Microw. Theory Tech.**, v. 53, n. 4, p. 1150-1163, 2005.

RAWAT, M. et al. Joint mitigation of nonlinearity and modulator imperfections in dual-band concurrent transmitter using neural networks. **Electronics Letters**, v. 49, n. 4, p. 253 - 255, 2013.

RAWAT, M.; GHANNOUCHI, F. M. A Mutual Distortion and Mismatches Compensator for Wideband Direct-Conversion Transmitters Using Neural Networks. **IEEE Transactions on Broadcasting**, v. 58, n. 2, p. 168-177, 2012.

RAWAT, M.; RAWAT, K.; GHANNOUCHI, F. M. Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks. **IEEE Transactions on Microwave Theory and Techniques**, 58, n. 1, 2010. p. 95-104.

ROBLIN, P. et al. Frequency-Selective Predistortion Linearization of RF Power Amplifiers. **IEEE Trans. Microwave Theory**, v. 56, n. 1, p. 65-76, 2008.

RUDIN, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nat Mach Intell**, p. 206-215, 2019.

RUGH, W. J. **Nonlinear System Theory - The Volterra-Wiener Approach**. Baltimore: The Johns Hopkins Univ. Press, 1981.

SALEH, A. A. M. Frequency-independent and frequency-dependent Nonlinear Models of TWT amplifiers. **IEEE transactions on Signal Processing**, 29, n. 11, 1981. p. 1715-1720.

SAMEK, W.; STANCZAK, S.; WIEG, T. The Convergence of Machine Learning and Communications. **ITU Journal: ICT Discoveries**, v. 1, n. 13, p. 1-10, 2017.

SANTOS, E. L. et al. A fully integrated CMOS power amplifier with discrete gain control for efficiency enhancement. **Microelectronics Journal**, v. 70, p. 34-42, 2017.

SARROFF, A. M. **Complex neural networks for audio**. Hanover, New Hampshire. 2018.

SCHREURS, D.; ET AL. **RF power amplifier behavioral modeling**. [S.I.]: Cambridge University Press, 2008.

SCHUARTZ, L. et al. Reduced complexity polynomials with memory applied to the linearization of power amplifiers with real time discrete gain control. **Circuits Systems and Signal Processing**, 38, 2019. p. 3901-3930.

SHANNON, C. E. A Mathematical Theory of Communication. **Bell System Technical Journal**, v. 27, p. 379–423 and 623–656, 1948.

SOLTANI TEHRANI, A. **Behavioral modeling of wireless transmitters for distortion mitigation**. Gothenburg, Sweden: Chalmers University of Technology, 2012.

STALLINGS, W. **Wireless communications and networks second edition**. [S.I.]: Pearson, 2004.

STOICA, P.; MOSES, R. L. Spectral analysis of Signals. Upper Saddle River: Pearson Prentice Hall, 2005.

TAKAQI, H.; WALKE, B. H. **Spectrum Requirement Planning in Wireless Communications: Model and Methodology for IMT - Advanced**. New York: Wiley, 2008.

TEHRANI, A. S. et al. A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavior models. **IEEE transactions on microwave theory and techniques**, 58, 2010. p. 1510-1520.

TEMPORAL, B. M. **Aplicação de pré-distorsões em rede de transmissores sem fio passíveis de imprecisões no modulador em quadratura.** Curitiba, Paraná. 2019.

TOMASKAR, R. A.; KHANDALE, G. D. FPGA implementation of complex multiplier using Urdhva Tiryakbham Sutra of Vedic mathematics. **Int. Journal of Engineering Research and Applications**, 4, 2014. p. 1-5.

TRABELSI, C.; ET AL. Deep Complex Networks. **arXiv:1705.09792**, p. 1-19, 2018.

TSE, D.; VISWANATH, P. **Fundamentals of Wireless Communication.** [S.I.]: Cambridge University Press, 2005.

VAN VEEN, F. The Asimov Institute, Set. 2016. Disponível em: <<https://www.asimovinstitute.org/author/fjodorvanveen/>>. Acesso em: Mai. 2018.

VARGHESE, A. et al. **FPGA implementation of area-efficient IEEE 754 complex divider.** International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST). [S.I.]: [s.n.]. 2015. p. 1120-1126.

VUOLEVI, J. H. K. **Analysis, measurement and cancellation of the bandwidth and amplitude dependence of intermodulation distortion in RF power amplifiers.** Oulu. 2001.

VUOLEVI, J. H. K.; RAHKONEN, T. **Distortion in RF Power Amplifiers.** [S.I.]: Artech House, 2002.

WANG, S. et al. **Impact of the normalization gain of digital predistortion on linearization performance and power added efficiency enhancement.** Proceedings of the 47th European Microwave Conference (EuMC). [S.I.]: [s.n.]. 2017. p. 1050-1053.

WEISSTEIN, E. W. Affine Transformation, maio 2018. Disponível em: <<http://mathworld.wolfram.com/AffineTransformation.html>>. Acesso em: maio 2018.

WIDROW, B.; HOFF, M. E. **Adapting switching circuits (technical report 4).** [S.I.]. 1960.

WOOD, J. System-Level Design Considerations for digital pre-distortion wireless base station transmitters. **IEEE Transactions on microwave theory and techniques**, 65, 2017. p. 1880-1890.

XIN, H.; ET AL. Convolutional neural network for behavioral modeling and predistortion of wideband power amplifiers. **IEEE Transactions on Neural Networks and Learning System**, 2021.

YI, J. et al. Analog predistortion linearizer for high-power RF amplifiers. **Microwave Theory and Techniques , IEEE Transactions on**, v. 48, n. 12, p. 2709-2713, 2000.

ZHANG, Q. Adaptive equalization using the back propagation algorithm. **IEEE trans. On Circuits and Systems**, v. 37, p. 848-849, 1990.

ZHANG, Q. J.; GUPTA, K. C. **Neural Networks for RF and Microwave Design**. Norwood: Artech House, 2000.

ZHU, A.; BRAZIL, T. J. Behavioral model of RF power amplifiers based on pruned Volterra series. **IEEE Microw. and Wireless comp. letters**, 14, 2004. p. 563 - 565.

ZUCKER, R. **Hyperbolic Functions**. In **Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables**. Dover, New York: M. Abramowitz and I. A. Stegun, 1972. p. 83-86.