



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola d'Enginyeria de Telecomunicació
i Aeroespacial de Castelldefels

TRABAJO DE FIN DE GRADO

TÍTULO DEL TFG: Linealización de un Amplificador Balanceado con Modulación de Carga Mediante un Predistorsionador Digital Basado en Redes Neuronales para Comunicaciones en Vehículos Aéreos no Tripulados

TITULACIÓN: Doble grado en Ingeniería de Sistemas Aeroespaciales e Ingeniería de Sistemas de Telecomunicación

AUTOR: Estefania Guillena Berenguer

DIRECTOR: Pere Gilabert

FECHA: 25 de septiembre de 2020

Título: Linealización de un Amplificador Balanceado con Modulación de Carga Mediante un Predistorsionador Digital Basado en Redes Neuronales para Comunicaciones en Vehículos Aéreos no Tripulados

Autor: Estefania Guillena Berenguer

Director: Pere Gilabert

Fecha: 25 de septiembre de 2020

Resumen

Este proyecto final de carrera se centra en el análisis y linealización de un amplificador balanceado con modulación de carga (LMBA), diseñado para operar con señales de gran ancho de banda sin degradar la eficiencia del amplificador.

Este tipo de amplificadores puede ser utilizado en vehículos aéreos no tripulados que requieran de una alta tasa de transmisión de datos, donde es crítico mantener unos niveles altos de eficiencia para minimizar el consumo y maximizar la autonomía de dichos vehículos no tripulados. También se pueden utilizar en estaciones bases para ser utilizados con el nuevo estándar 5G.

En este proyecto se analizan los parámetros que intervienen en el LMBA (el desfase entre las dos señales que controlan el amplificador y la relación entre estas dos amplitudes), así como sus efectos sobre la linealidad y la eficiencia. Además, se proponen una serie de pasos a seguir para su correcta configuración y obtener así el máximo rendimiento de este amplificador.

Finalmente, se realiza la linealización del amplificador con un predistorsionador digital basado en una red neuronal artificial formada de varias capas ocultas, y se comparan los resultados obtenidos con un DPD basado en un modelo de comportamiento del tipo GMP (Generalized Memory Polynomial) para diferentes señales LTE.

Title : Load Modulated Balanced Amplifier Linearization Using a Digital Predistorter Based on Neural Network for Unmanned Aerial Vehicle Communications

Author: Estefania Guillena Berenguer

Advisor: Pere Gilabert

Date: September 25, 2020

Overview

This final degree project focuses on the analysis and linearization of a load modulated balanced amplifier (LMBA), designed to operate with wide bandwidth signals without degrading the amplifier's efficiency.

This type of amplifier can be used in unmanned aerial vehicles that require a high data transmission rate, where it is critical to maintain high levels of efficiency to minimize consumption and maximize flight autonomy, as well as in base stations to be used with the new 5G standard.

In this project the parameters involved in the LMBA (the phase shift between the two signals that control the amplifier and the relationship between these two amplitudes) and its effects on the linearity and efficiency are analyzed. In addition, a number of steps are proposed to follow for proper settings and get the most out of this amplifier.

Finally, the linearization of the amplifier is performed with a digital predistorter based on an artificial neural network made up of several hidden layers, and these results are compared to the results obtained with the GMP (Generalized Memory Polynomial) method for different LTE signals.

ÍNDICE GENERAL

Agradecimientos	1
Abstract	3
CAPÍTULO 1. Introducción	5
1.1. Vehículos aéreos no tripulados	5
1.2. Tecnología 5G	7
CAPÍTULO 2. Amplificadores de potencia	11
2.1. Introducción	11
2.2. Parámetros de análisis de un amplificador de potencia	11
2.2.1. Back-off	11
2.2.2. Adjacent Channel Power Ratio (ACPR)	12
2.2.3. Error cuadrático medio normalizado (NMSE)	12
2.2.4. Error Vector Magnitude (EVM)	13
2.2.5. Eficiencia energética	13
2.2.6. Peak-to-Average Power Ratio (PAPR)	14
2.3. Distorsión no lineal del amplificador de potencia	15
2.3.1. Distorsión AM-AM	15
2.3.2. Distorsión AM-PM	17
2.3.3. Efecto de memoria	18
2.4. Clases de amplificadores de potencia	19
2.5. Linealidad versus eficiencia	20
2.6. Técnicas para optimizar la eficiencia	21
2.6.1. Técnicas de alimentación dinámica	21
2.6.2. Técnicas de modulación de carga dinámica	22
2.7. Técnicas de linealización	24
2.7.1. Feedback	24
2.7.2. Feedforward	25
2.7.3. Predistorsión	25

CAPÍTULO 3. Predistorsión	27
3.1. Introducción	27
3.2. Predistorsión analógica	27
3.3. Predistorsión digital	28
3.3.1. Predistorsión no adaptativa	29
3.3.2. Predistorsión adaptativa	29
3.4. Modelos de predistorsión digital	31
3.4.1. Series de Volterra	31
3.4.2. Memory Polynomial	32
3.4.3. Generalized Memory Polynomial	32
3.4.4. GMP basado en B-splines	33
3.4.5. Redes neuronales	35
3.5. Estimación de los coeficientes del predistorsionador	35
3.5.1. Método de mínimos cuadrados	35
3.5.2. Implementación de DPD adaptativo directo	36
CAPÍTULO 4. Redes neuronales artificiales	39
4.1. Introducción	39
4.1.1. Principio de funcionamiento de las redes neuronales artificiales	39
4.2. Problemas de regresión y de clasificación	40
4.3. Redes neuronales supervisadas y no supervisadas	40
4.4. Estructura de una red neuronal artificial	41
4.4.1. Función de entrada	42
4.4.2. Función de activación	42
4.4.3. Función de salida	44
4.4.4. Número de neuronas y de capas ocultas	44
4.4.5. Función de entrenamiento	45
4.4.6. Datos de entrenamiento, validación y test	50
4.4.7. Tipos de redes neuronales artificiales	50
4.5. Red neuronal como predistorsionador digital	53
4.5.1. Topología polar y cartesiana	53
4.5.2. Variables de entrada a la red neuronal	54
4.5.3. Red neuronal artificial adaptativa	55

CAPÍTULO 5. Configuración de parámetros de LMBA para máxima eficiencia	57
5.1. Parámetros que intervienen en las prestaciones del LMBA	57
5.2. Setup experimental	59
5.3. Desfase entre señales del LMBA	60
5.3.1. Efecto del desfase sobre la eficiencia	60
5.3.2. Efecto del desfase sobre la linealidad	60
5.3.3. Variación del desfase óptimo con la frecuencia	62
5.3.4. Predicción del desfase óptimo para NMSE	62
5.3.5. Efecto del retardo entre señales sobre el desfase óptimo	63
5.3.6. Predicción del desfase óptimo aplicando retardo en banda base	66
5.3.7. Efecto del desfase entre señales sobre la linealización del amplificador mediante DPD	66
5.3.8. Validación con señales diferentes	68
5.4. Efecto de la relación entre amplitudes sobre el desfase óptimo	72
5.5. Efecto de la tensión de alimentación del amplificador auxiliar sobre la linealidad y la eficiencia	76
5.6. Criterios de selección de los parámetros del LMBA	77
CAPÍTULO 6. Implementación de ANN como DPD sobre LMBA	79
6.1. Introducción	79
6.2. Redes neuronales artificiales en Matlab	80
6.3. Selección de la red neuronal artificial	80
6.3.1. Selección de las variables de entrada	81
6.3.2. Selección de hiperparámetros de la red	83
6.3.3. Funciones de aprendizaje, entrenamiento y activación	85
6.3.4. Selección de la topología de la red neuronal	85
6.3.5. Entrenamiento de la red neuronal	87
6.4. Resultados obtenidos con la red neuronal	88
6.4.1. Esquema del código utilizado	88
6.4.2. Red neuronal adaptativa versus no adaptativa	92
6.4.3. Aprendizaje directo versus aprendizaje indirecto	92
6.4.4. Comparativa de predistorsión realizada con ANN y GMP	94
6.4.5. Parámetros para optimizar la eficiencia	99
6.5. Validación en entorno dinámico	99

Conclusiones	103
Bibliografía	105
APÉNDICE A. Código de Matlab utilizado	109

ÍNDICE DE FIGURAS

1.1 Vehículos aéreos no tripulados.	5
1.2 Esquema de cobertura móvil mediante estaciones base situadas en drones. . .	6
2.1 ACPR en el espectro de la señal.	12
2.2 Representación del vector error de magnitud.	13
2.3 Esquema de la potencia de alimentación en un PA.	14
2.4 Curva característica AM-AM de un PA ideal.	15
2.5 Curva característica AM-AM de un PA real.	16
2.6 Representación del ensanchamiento espectral debido a la distorsión no lineal. .	17
2.7 Representación del efecto Warping en una modulación 16QAM.	17
2.8 Curva característica AM-PM.	18
2.9 Representación del efecto de memoria en un PA.	19
2.10 Curva AM-AM con efecto de memoria	19
2.11 Linealidad versus eficiencia.	21
2.12 Diagrama de bloques del seguimiento de envolvente.	21
2.13 Diagrama de bloques de EER.	22
2.14 Diagrama de bloques de LINC.	23
2.15 Diagrama de bloques de amplificador Doherty.	23
2.16 Diagrama de bloques de LMBA.	24
2.17 Esquema de linealización feedback.	24
2.18 Esquema de linealización feedforward.	25
3.1 Esquema de predistorsión.	27
3.2 Predistorsionador analógico CRLHTLs.	28
3.3 Predistorsionador analógico basado en transistores.	28
3.4 Arquitectura DPD en bucle cerrado.	29
3.5 DPD adaptativo con aprendizaje indirecto.	30
3.6 DPD adaptativo con aprendizaje directo.	30
3.7 Curva de Bézier cuadrática.	33
3.8 Ejemplo de curvas B-splines de orden 4.	34
3.9 Esquema de la implementación de DPD directo.	37
4.1 Esquema de una neurona del cerebro humano.	39
4.2 Esquema simplificado de una neurona de una red neuronal artificial.	40
4.3 Esquema de una red neuronal artificial de una sola capa.	41
4.4 Diagrama de bloques de una neurona.	41
4.5 Función de activación log-sigmoidea.	42
4.6 Función de activación tan-sigmoidea.	43
4.7 Función de activación lineal (purelin).	43
4.8 Función de activación ReLU.	44
4.9 Esquema de minimización del error.	46
4.10 Esquema de una red neuronal artificial prealimentada.	51
4.11 Esquema de una red neuronal artificial en cascada hacia delante.	52
4.12 Esquema de una red neuronal artificial recurrente.	52

4.13	Esquema de una red neuronal artificial convolucional.	53
4.14	Topologías convencionales de redes neuronales para DPD.	54
4.15	Posibles esquemas de variables de entrada de la red neuronal artificial.	55
5.1	Diagrama de bloques del LMBA utilizado en este proyecto.	57
5.2	Diagrama de bloques del LMBA con implementación de DPD.	58
5.3	Setup experimental empleado para realizar las pruebas sobre el LMBA.	59
5.4	LMBA utilizado para realizar la parte experimental.	59
5.5	Valores de eficiencia medios obtenidos en las campañas de medidas.	60
5.6	Valores de NMSE medios obtenidos en las campañas de medidas.	61
5.7	Valores de peor ACPR medios obtenidos en las campañas de medidas.	61
5.8	Desfases para optimizar NMSE, ACPR y eficiencia.	62
5.9	Ajuste de las curvas de regresión sobre el desfase para optimizar NMSE.	62
5.10	Valores de NMSE y ACPR con las diferentes regresiones encontradas.	63
5.11	Valores de NMSE en función del retardo en banda base para diferentes frecuencias.	64
5.12	Valores de ACPR del canal superior e inferior en función del retardo en banda base para la frecuencia de 2 GHz y 2.4 GHz.	65
5.13	Comparativa del desfase óptimo en términos de NMSE con y sin retardo en banda base.	65
5.14	Comparativa del desfase óptimo en términos de ACPR para los dos canales adyacentes con y sin retardo en banda base	65
5.15	Ajuste de las diferentes curvas de regresión a los desfases óptimos obtenidos en la campaña de medida con retardo en banda base.	66
5.16	Valores de NMSE obtenidos con las diferentes regresiones aplicando un retardo de -2 muestras en banda base.	67
5.17	Valores de ACPR obtenidos con las diferentes regresiones aplicando un retardo de -2 muestras en banda base.	67
5.18	Linealización mediante DPD con el mejor y el peor desfase para la linealidad.	68
5.19	Espectro de las señales utilizadas para la validación de la predicción del desfase óptimo.	69
5.20	Valores de ACPR, NMSE y eficiencia obtenidos con las campañas de medidas de la señal de 20 MHz y la señal de 60 MHz.	70
5.21	Validación de la predicción de desfase óptimo para la señal LTE de 60 MHz	70
5.22	Valores de ACPR, NMSE y eficiencia obtenidos con las campañas de medidas de las señales de 20 MHz y 200 MHz	71
5.23	Validación de la predicción de desfase óptimo para la señal LTE de 200 MHz.	72
5.24	Valores de Eficiencia y linealidad en función de p para una señal a 2 GHz.	73
5.25	Efecto de la relación entre amplitudes sobre el desfase óptimo y la linealidad a 2 GHz.	74
5.26	Efecto de la relación entre amplitudes sobre la eficiencia a 2 GHz.	75
5.27	Valores de linealidad máximos, mínimos y con el desfase óptimo para $p = 3$ a diferentes frecuencias.	76
5.28	Valores de linealidad y eficiencia para una señal LTE de 20 MHz de ancho de banda a una frecuencia de 2 GHz en función de la tensión de alimentación.	76
6.1	Esquema del RF WebLab de la Universidad de Chalmers	79
6.2	Esquema de red neuronal artificial en la App de Matlab.	80

6.3	Curva AM-AM sin tener en cuenta los efectos de memoria del PA.	81
6.4	Curva AM-AM con la predicción obtenida mediante la red neuronal seleccionada.	83
6.5	Esquema de las variables de entrada y salida seleccionadas para la red neuronal.	84
6.6	Valores de linealidad para cada red neuronal testada en función de las iteraciones.	85
6.7	Topologías de red neuronal artificial testadas con el modelado de un amplificador de potencia.	86
6.8	Esquema de entrenamiento y comportamiento de la red neuronal artificial utilizada como DPD.	87
6.9	Formación del histograma bidimensional de una señal 16-QAM sin efecto de memoria.	90
6.10	Reducción de los datos a partir del histograma de malla lineal.	91
6.11	Esquema de las muestras seleccionadas con este método.	91
6.12	Comparativa en la implementación de una red neuronal con esquema adaptativo o no adaptativo sobre una señal LTE de 200 MHz de ancho de banda.	93
6.13	Comparativa en la implementación de una red neuronal con aprendizaje directo e indirecto sobre una señal LTE de 200 MHz de ancho de banda.	94
6.14	Comparativa de curva AM-AM y espectro frecuencial sobre la señal LTE 20 MHz con GMP y ANN.	96
6.15	Comparativa de curva AM-AM y espectro frecuencial sobre la señal LTE 60 MHz con GMP y ANN.	97
6.16	Comparativa de curva AM-AM y espectro frecuencial sobre la señal LTE 200 MHz con GMP y ANN.	98
6.17	Comparativa de rendimiento de la señal LTE de 200 MHz con diferentes valores de PAPR máximo y ganancia de predistorsión.	100
6.18	Resultados obtenidos con las diferentes configuraciones de señal.	102

ÍNDICE DE TABLAS

1.1	Clasificación de los vehículos no tripulados según la EASA.	6
1.2	Características principales de cada generación de telefonía móvil.	8
1.3	Valor EVM máximo permitido en función de la modulación empleada.	9
2.1	Características principales de las diferentes clases de amplificadores de potencia.	20
5.1	Rendimiento del LMBA después de aplicar DPD cuando se selecciona el mejor y el peor desfase entre señales.	68
6.1	Resultados de NMSE en el modelado del amplificador de potencia.	82
6.2	Resultados de NMSE en el modelado del amplificador de potencia teniendo en cuenta el valor absoluto de la señal.	82
6.3	Resultados de NMSE en el modelado del amplificador de potencia teniendo en cuenta el valor absoluto de la señal y sus potencias.	83
6.4	Resultados de NMSE en el modelado del amplificador de potencia en función de las capas ocultas de la red neuronal artificial.	84
6.5	Resultados de NMSE en el modelado del amplificador de potencia en función de la topología de red neuronal empleada.	86
6.6	Configuración de la señal y valores de los parámetros utilizados en la fase experimental.	88
6.7	Configuración del modelo GMP empleado en la linearización del LMBA.	95
6.8	Valores de linealidad, potencia, eficiencia y EVM obtenidos con DPD basado en ANN y GMP para la señal LTE de 20 MHz.	95
6.9	Valores de linealidad, potencia, eficiencia y EVM obtenidos con DPD basado en ANN GMP para la señal LTE de 60 MHz.	97
6.10	Valores de linealidad, potencia, eficiencia y EVM obtenidos con DPD basado en ANN y GMP para la señal LTE de 200 MHz.	99
6.11	Configuración empleada para cada una de las señales en el entorno dinámico.	101
6.12	Resultados obtenidos en el entorno dinámico.	101

AGRADECIMIENTOS

A continuación, no me gustaría dejar de reconocer en las siguientes líneas la labor de apoyo incondicional de todas aquellas personas que han sido un pilar esencial en la elaboración de este TFG: Pere Gilabert, director de este trabajo, por haber visto en mí a la figura más apta para desarrollar esta línea de investigación en redes neuronales, y por su gran labor docente; a mis padres, que siempre han sido una presencia positiva en mi vida, apoyándome en todos mis proyectos, siendo un sostén básico para poder seguir creciendo personal y profesionalmente; a mi hermana, por todo el cariño y refuerzo positivo que me ha dado para continuar con mis estudios; y, finalmente, a mi marido, por todo su amor, afecto y gran paciencia, que me ha servido para ser constante en mi esfuerzo para finalizar esta doble titulación.

ABSTRACT

La evolución tecnológica de los últimos años ha situado a los vehículos aéreos no tripulados (comúnmente conocidos como drones) como unos de los dispositivos ampliamente utilizados para un abanico cada vez más extenso de aplicaciones. Algunas de estas aplicaciones conllevan la transmisión de una gran cantidad de datos, como puede ser la transmisión de video en tiempo real, o su utilización como estaciones base de telefonía móvil para dar cobertura en caso de catástrofes.

Una de las principales características de los drones es su autonomía. Contar con dispositivos energéticamente eficientes es vital para este tipo de vehículos aéreos. Dentro de un transmisor de RF, el amplificador de potencia (PA) es el dispositivo de mayor consumo energético. Por esta razón, resulta de gran importancia considerar la implementación de un PA eficiente.

El amplificador balanceado con modulación de carga (LMBA) se diseña para poder amplificar señales de gran ancho de banda sin sacrificar la eficiencia. Este tipo de amplificadores son una buena opción en transmisores montados sobre drones que dan servicio en aplicaciones que requieren de una elevada tasa de transmisión de datos, además de poder ser utilizados en la novedosa tecnología móvil 5G. Pero esta implementación requerirá utilizar algún método de linealización que permita obtener un buen rendimiento del dispositivo debido a las características no lineales inherentes del LMBA.

En este trabajo se lleva a cabo un detallado análisis de los parámetros que intervienen en este amplificador y cómo realizar una correcta configuración del mismo. Finalmente, se implementa la linealización del LMBA, por medio de un predistorsionador digital (DPD) formado por una red neuronal artificial.

CAPÍTULO 1. INTRODUCCIÓN

1.1. Vehículos aéreos no tripulados

Los vehículos aéreos no tripulados (UAS, del inglés Unmanned Aircraft Systems), comúnmente conocidos como drones, son sistemas aéreos que realizan operaciones sin piloto a bordo ni tripulación. Estas aeronaves no tripuladas pueden ser controladas desde tierra (RPAS, del inglés Remotely Piloted Aircraft System) o ser programadas para volar de forma autónoma. Los drones, en la actualidad, son empleados en un amplio abanico de aplicaciones. Estos vehículos aéreos van desde simples juguetes, con un sistema de comunicaciones sencillo, que apenas transmite datos; hasta armas militares, con capacidad para transmitir información de telemetría, órdenes y control, imágenes captadas en infrarrojos o radar, streaming de video y audio en tiempo real, etc. Por lo tanto, estas aeronaves presentan múltiples formas y diseños, así como diferentes mecanismos de propulsión para adaptarse a las necesidades de cada aplicación (1.1).



(a)



(b)

Figura 1.1: Vehículos aéreos no tripulados.

Este tipo de vehículo aéreo, al igual que otras muchas aplicaciones, surgió en el ámbito militar, donde los RPAS han alcanzado un grado de madurez importante. En el ejército estadounidense, por ejemplo, los RPAS conforman un tercio de su flota total, encargándose de todas las misiones de inteligencia, vigilancia y reconocimiento; pero, hoy en día, los drones ya son ampliamente desarrollados en el ámbito civil, donde se han creado multitud de dispositivos diferentes para ser utilizados con distintos propósitos. Algunas de estas aplicaciones son: la toma de imágenes en zonas no accesibles, búsqueda y rescate de personas en catástrofes, el uso de drones en la agricultura, o drones para el envío y entrega de paquetería, etc. En el ámbito de las comunicaciones se propone su utilización para dar cobertura wifi o de datos móviles en zonas remotas, o en zonas en donde ha habido una catástrofe y no es posible recibir cobertura por otro medio (donde el dron actuaría como estación base de telefonía), etc.

Existen multitud de formas de clasificar a los drones: según su forma, su mecanismo de propulsión o según su aplicación, entre otros. Según el reglamento de EASA (European Union Aviation Safety Agency) [1], los drones se clasificaron en 5 clases, C0-C4. Cada clase marca una serie de limitaciones en: la masa máxima de despegue (incluida la carga útil), la altura de vuelo desde el punto de despegue o la velocidad máxima horizontal a alcanzar. Las características de estas 5 clases se muestran en la tabla 1.1.

Como se ha comentado previamente, estos dispositivos no disponen de piloto a bordo,

Tabla 1.1: Clasificación de los vehículos no tripulados según la EASA.

Clase	Carga máxima al despegue	Altura máxima	Velocidad máxima	Tensión de alimentación máxima
C0	250g	19 m/s	120m	24 V (DC)
C1	900g	19 m/s	120m	24 V (DC)
C2	4kg	-	120m	48 V (DC)
C3	25kg	-	120m	48 V (DC)
C4	25kg	-	-	-

sino que, por lo general, son controlados remotamente. Para ello, para transmitir y recibir los datos de control de la aeronave, se utilizan sistemas de comunicaciones inalámbricas. Por lo general, estas comunicaciones de control no requieren de un gran ancho de banda y suelen emplear las bandas de radio industriales, científicas y médicas (ISM) de baja frecuencia.

Una opción para transmitir los comandos de control de la aeronave no tripulada es utilizar transceptores de circuitos integrados, que permiten una baja tasa de transferencia de datos, con un bajo consumo de potencia y un alto alcance, pero suficiente para este tipo de aplicaciones. Muchos de estos transceptores funcionan a la frecuencia de 2.4 GHz y presentan tasas de transmisión desde los 200 kbps hasta los 2 Mbps aproximadamente.

Este tipo de transceptores son útiles para la transmisión de comandos de control, pero el ancho de banda que permiten es limitado y no es suficiente para aplicaciones que requieren, por ejemplo: transmisión de video en tiempo real, de datos LiDAR (Laser Imaging Detection and Ranging) o el uso mismo de estos drones como estaciones base de datos móviles, entre otros.

En el caso de utilizar drones como pequeñas estaciones móviles (1.2), el sistema de comunicaciones tiene que ser capaz de transmitir grandes cantidades de datos, ocupando grandes anchos de banda, a una alta potencia, y poder proporcionar así el servicio demandado por los diferentes usuarios. Debido a las limitaciones de espacio y peso en el dron, quedará muy restringido el suministro de energía disponible para el transceptor de RF. En este caso, resultará crucial disponer en el aparato de una estación base inalámbrica de alta eficiencia.

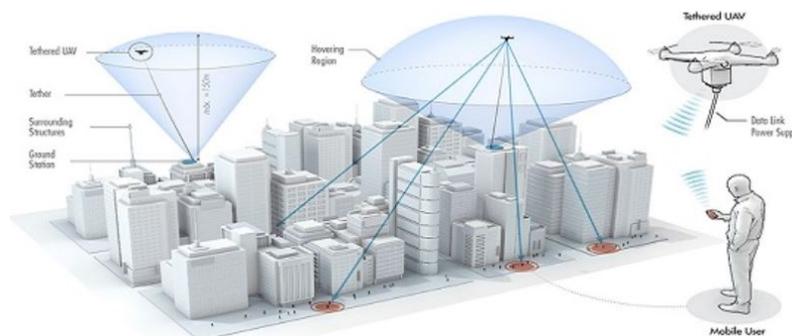


Figura 1.2: Esquema de cobertura móvil mediante estaciones base situadas en drones.

Debido a la clasificación de la EASA, mostrada en la tabla 1.1, y a las necesidades de ali-

mentación de los transmisores de RF, sólo es posible construir un sistema de RF de mayor potencia en drones a partir de la clase C2 (debido a la tensión máxima de alimentación disponible).

Los drones de las clases C2 y C3 ya cubren la gran mayoría de aplicaciones, permitiendo montar cámaras de video y LiDAR como carga útil. Por lo general, estas aplicaciones requieren de enlaces de comunicaciones ascendente y descendentes, que permitan transmisión de video; por lo tanto, requieren de un buen ancho de banda que a cierta potencia permitan cubrir un buen alcance. En el caso de los drones de clase C4, además, se pueden encontrar más alejados del centro de control, ya que pueden alcanzar altitudes mayores y requerir mayor potencia de transmisión. Para las aplicaciones que precisan de mayor ancho de banda y mayor potencia de transmisión, es crucial optimizar la eficiencia del amplificador de potencia montado en el sistema de comunicaciones del dron. Optimizar la eficiencia, en términos de consumo energético, permitirá mejorar la autonomía del dron, o aumentar la potencia de transmisión para cubrir mayor distancia. Además, en la actualidad, es posible utilizar el protocolo de comunicaciones móviles LTE para controlar y transmitir señales de video y audio desde un dron, que las retransmitirá usando las estaciones base de las redes de telefonía móvil dentro de su alcance.

En este trabajo se presenta la linealización, mediante redes neuronales, de un amplificador de potencia balanceado con modulación de carga, que permite trabajar con anchos de banda grandes sin degradar la eficiencia. Este tipo de amplificador podría ser utilizado en sistemas de comunicaciones de drones que requieran de grandes anchos de banda, o que funcionen con el estándar LTE. En concreto, se ha trabajado a nivel de laboratorio, para demostrar su validez, con señales LTE de anchos de banda de 20, 60 y 200 MHz.

1.2. Tecnología 5G

Debido al aumento de la demanda por parte de los usuarios de las prestaciones en las comunicaciones móviles, la industria celular se ha movilizado en busca de nuevas soluciones que permitan dar una mejor experiencia al usuario en términos de calidad y velocidad de transmisión. Inicialmente, la telefonía celular estaba pensada para transmisiones de voz analógicas. Posteriormente, debido a la aparición de la tecnología digital y a la necesidad de dar cobertura a más canales de voz surgió la segunda generación de telefonía (2G) en los años noventa. La aparición de la tecnología digital dio pie a la transferencia de datos con mensajes de texto en la telefonía celular pero los estándares de 2G no permitían una velocidad de transmisión suficiente para la demanda de datos generada, por esta razón se desarrolló la tercera generación. Con esta generación se aumentó la capacidad y la velocidad de la transferencia de datos, pero, nuevamente, la necesidad por parte de los usuarios de utilizar aplicaciones que requieren una mayor tasa de transferencia como la transmisión de video de alta calidad o tiempo real ha llevado a la necesidad de desarrollar la cuarta generación que permita la transmisión de datos de alta velocidad.

Desde hace algunos años han surgido una serie de servicios que no pueden ser cubiertos con la tecnología celular desarrollada hasta el momento (el 4G). Estas nuevas aplicaciones van desde la realidad virtual que requiere un gran ancho de banda, al internet de las cosas que conectará a la red millones de nuevos dispositivos o las comunicaciones entre vehículos que requieren de una baja latencia. Por esta razón se ha desarrollado la

quinta generación de telefonía móvil para mejorar las tecnologías desarrolladas hasta el momento y dar respuesta a las nuevas necesidades que han surgido con esta serie de nuevos servicios.

La red 5G está pensada para ser una red multiservicio de baja latencia con un aumento considerable de la tasa de transmisión comparada con sus antecesores y con amplia cobertura. Para poder aumentar la capacidad en estas redes 5G se utilizan técnicas como MIMO masivo, antenas activas, agregación de portadoras, modulaciones de alto orden, densificación con small cells y la utilización de un amplio espectro radioeléctrico que va desde 1 GHz hasta los 100 GHz. En la tabla 1.2 se muestra la evolución en la telefonía móvil a través de una comparativa entre las diferentes generaciones desarrolladas [3].

Tabla 1.2: Características principales de cada generación de telefonía móvil.

Gen.	Estándar	Modulación	Frecuencia máx.	Ancho de banda	Bit rate máx.	PAPR máx.
2G	GSM/EDGE	TDMA/FDMA , GMSK , 8PSK	1.9 GHz	0.2 MHz	1 Mbit/s	3.3 dB
3G	UMTS	W-CDMA, QPSK/64QAM	3 GHz	5 MHz	100 Mbit/s	10.6 dB
4G	LTE	OFDMA , MIMO 4x4 , 64QAM	5.9 GHz	20 MHz	300 Mbit/s	12 dB
4G	LTE-Advanced	OFDMA/SC-FDMA, MIMO 8x8, CA , 64QAM	5.9 GHz	20 MHz (100 con CA)	Mayor a 300 Mbit/s	12 dB
4G	WiMAX	OFDMA, MIMO 2x2 , QPSK/64QAM	5.8 GHz	20 MHz	100 Mbit/s	12 dB
5G	5G	OFDMA, mas-sive MIMO, ad-vanced CA	40 GHz	Mayor a 100 MHz	Mayor a 1 Gbit/s	Mayor a 12 dB

Tal y como se muestra en esta tabla, con el surgimiento del estándar LTE comienza a utilizarse la modulación ortogonal frequency division multiplexing (OFDM) que permite una utilización más eficiente del espectro frecuencial además de ser escalable y fácil de implementar por medio de la transformada rápida de Fourier. Esta modulación también se utiliza en 5G, pero con variantes como el “Filter Bank Multi-Carrier” (FBMC) donde estos filtros permiten reducir los lóbulos laterales de cada subportadora OFDM. Sin embargo, el precio a pagar por usar este tipo de modulaciones es un importante aumento del Peak-to-Average Power Ratio (PAPR) que requiere del uso de amplificadores de potencia muy lineales en amplitud y en fase para cumplir con los requerimientos marcados por los estándares. En la tabla, también se observa como la tendencia a lo largo de los años consiste en crear estándares que permitan un incremento en el ancho de banda de la señal para aumentar así la tasa de transferencia de datos y utilizar frecuencias centrales cada vez más elevadas. Por lo tanto, es de esperar que los siguientes estándares que aparezcan seguirán en la misma línea.

Según la normativa especificada para 5G por el organismo ITU-R [2], el valor límite de ACPR para para este estándar se establece en -45 dB. Por lo tanto, las señales transmitidas deben mantener un valor de ACPR en todos los canales adyacentes inferior a los -45 dB. En cuanto al valor de EVM máximo permitido, este depende de la modulación empleada. En la tabla 1.3 los valores máximos permitidos que especifica la normativa para 5G.

Tabla 1.3: Valor EVM máximo permitido en función de la modulación empleada.

Modulación	EVM máxima(%)
QPSK	17.5
16QAM	12.5
64QAM	8
256QAM	3.5

La linealización implementada en este trabajo sobre el LMBA se ha realizado para cumplir con las especificaciones de ACPR y EVM mostradas previamente para el estándar 5G, con la idea de poder incorporar este tipo de amplificadores en las comunicaciones móviles de esta quinta generación.

CAPÍTULO 2. AMPLIFICADORES DE POTENCIA

2.1. Introducción

El amplificador de potencia (PA, del inglés Power Amplifier) es un componente de vital importancia en las comunicaciones inalámbricas. Gracias al PA la señal en el transmisor toma un nivel de potencia lo suficientemente alto como para permitir su transmisión por el medio aire y llegar al receptor con un nivel de potencia superior a la sensibilidad del mismo.

Para poder llevar a cabo su labor el PA requiere de un alto consumo de energía. Dentro de la cadena de transmisión este componente consume entre el 50% y el 80% de la energía suministrada. Por ello, es necesario utilizar topologías de PA que busquen maximizar la eficiencia y obtener el máximo de potencia a la salida a partir de la energía aportada.

Pero el consumo de energía no es la única problemática de este dispositivo; el PA es la mayor fuente de no linealidades dentro del transmisor y, a mayor eficiencia, mayor distorsión no lineal. Por ello, se requieren técnicas de linealización que permitan reducir esta distorsión no deseada.

2.2. Parámetros de análisis de un amplificador de potencia

2.2.1. Back-off

A la hora de trabajar con amplificadores de potencia es importante conocer el concepto de punto de operación o back-off. Existen dos puntos de operación: el de entrada, input back-off (IBO); y el de salida, output back-off (OBO). Estos dos puntos de operación quedan definidos con las siguientes ecuaciones:

$$IBO = 10 \log \frac{P_{0in}}{P_{in}} \quad (2.1)$$

$$OBO = 10 \log \frac{P_{0out}}{P_{out}} \quad (2.2)$$

Donde P_{in} y P_{out} son las potencias medias de entrada y de salida del amplificador de potencia, y P_{0in} y P_{0out} son las potencias de saturación a la entrada y a la salida de éste. Estas potencias de saturación corresponden al punto de compresión de 1 dB. Por lo tanto, un valor de back-off elevado significará que la potencia media de la señal se encontrará alejada del punto de saturación del amplificador.

Si se trabaja con niveles elevados de back-off se disminuye el efecto de la no linealidad del amplificador, al trabajar con potencias medias alejadas de la saturación del dispositivo, pero se consiguen niveles de eficiencia bajos. Para aumentar la eficiencia, el amplificador debe trabajar con niveles de back-off bajos, aunque esto implique una mayor distorsión.

2.2.2. Adjacent Channel Power Ratio (ACPR)

Este parámetro mide la relación entre la potencia que se introduce en el canal adyacente, debido a un recrecimiento espectral, y la potencia de la señal en su banda correspondiente. Esta relación de potencias marca la distorsión fuera de la banda provocada por el PA. Los dos anchos de banda utilizados para determinar la potencia no tienen por qué ser iguales, y un valor menor de ACPR significa un menor recrecimiento espectral, así como una menor interferencia en el canal adyacente. Podemos definir el ACPR con la siguiente ecuación:

$$ACPR = 10 \log \frac{P_{outBand}}{P_{inBand}} \quad (2.3)$$

En la figura 2.1 se muestra el concepto de ACPR en el espectro frecuencial de una señal. A partir de la imagen se puede determinar que el ACPR tiene un valor aproximado de -45 dB.

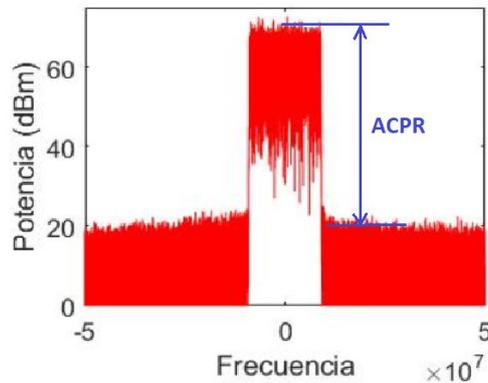


Figura 2.1: ACPR en el espectro de la señal.

2.2.3. Error cuadrático medio normalizado (NMSE)

Este parámetro no es específico de los amplificadores de potencia, ni de las señales de radio frecuencia, pero permite establecer una relación entre el comportamiento real del amplificador y un modelo matemático establecido.

Este parámetro determina la desviación producida entre el valor real y el valor estimado. En este caso, estos valores hacen referencia al valor que toma la señal a la entrada del amplificador y el valor que toma ésta a la salida. El NMSE se define a partir de la siguiente expresión:

$$NMSE = \frac{1}{N} \sum_i \frac{(P_i - M_i)^2}{\overline{PM}} \quad (2.4)$$

$$\overline{P} = \frac{1}{N} \sum_i P_i \quad (2.5)$$

$$\bar{M} = \frac{1}{N} \sum_i M_i \quad (2.6)$$

Donde P_i son los valores reales, M_i son los valores estimados y N es la cantidad de muestras empleadas. Cuanto menor sea el valor de NMSE, mejor será la estimación del modelo bajo estudio.

2.2.4. Error Vector Magnitude (EVM)

El EVM permite cuantificar la distorsión provocada dentro de la banda de la señal. Esta magnitud tiene en cuenta la distorsión provocada, tanto en la componente en fase, como en la componente en cuadratura provocada por el PA. Este parámetro consiste en calcular el vector error producido entre el símbolo de modulación I-Q recibido y el que deberíamos recibir si no existiese distorsión. Este valor se expresa en porcentaje al compararlo con el módulo del símbolo de referencia y viene descrito por la siguiente función:

$$EVM = \frac{1}{N} \sqrt{\frac{S_{err}}{S_{ref}}} [\%] \quad (2.7)$$

Donde $S_{err} = \frac{1}{N} \sum_i^N (\Delta I^2 + \Delta Q^2)$, $S_{ref} = \frac{1}{N} \sum_i^N (I_{ref}^2 + Q_{ref}^2)$ y N es el número de muestras utilizadas.

En la figura 2.2 se muestra de forma gráfica el concepto del vector error.

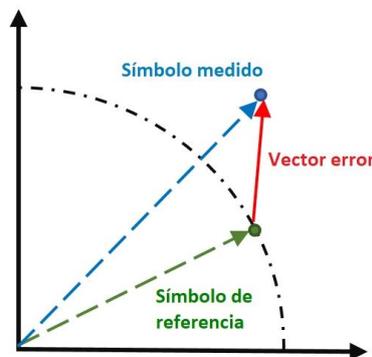


Figura 2.2: Representación del vector error de magnitud.

2.2.5. Eficiencia energética

La eficiencia energética en un PA mide la relación entre la energía de la señal a la salida del amplificador y la energía con la que se ha alimentado el dispositivo para realizar la amplificación de la señal. Por lo que podemos determinar que esta eficiencia marca la capacidad del PA de convertir potencia continua (potencia con la que se alimenta al PA) en potencia de radio frecuencia (potencia con la que sale la señal del PA). En la figura 2.3 se muestran las diferentes potencias que intervienen en el PA.

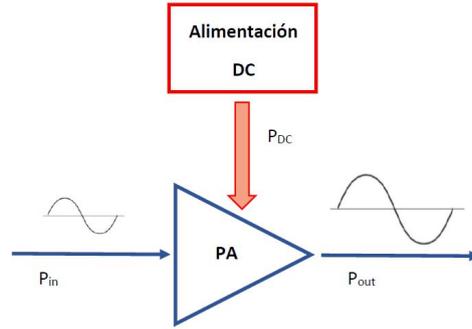


Figura 2.3: Esquema de la potencia de alimentación en un PA.

Matemáticamente, esta eficiencia se expresa como:

$$\eta = \frac{P_{out}}{P_{DC}} [\%] \quad (2.8)$$

También se puede calcular la eficiencia de potencia agregada (PAE), en la cual se tiene en cuenta la potencia con la que llega la señal al PA (P_{in}), de tal manera que sólo se tiene en cuenta la variación de potencia generada por la amplificación. Esta eficiencia queda definida con la siguiente expresión:

$$PAE = \frac{(P_{out} - P_{in})}{P_{DC}} [\%] \quad (2.9)$$

Si la señal que se transmite consiste en una señal con envolvente variable la eficiencia del amplificador también será variable, mostrando mayor eficiencia cuando la amplitud de la señal es mayor. Al ser esta eficiencia dependiente del tiempo, es posible calcular la potencia media como la relación entre las potencias medias de salida y de alimentación.

$$\eta_{AVG} = \frac{P_{outAVG}}{P_{dcAVG}} [\%] \quad (2.10)$$

2.2.6. Peack-to-Average Power Rario (PAPR)

El PAPR mide la relación entre la potencia de pico y su valor RMS, o valor cuadrático medio de la potencia de señal al cuadrado. El PAPR es equivalente al cuadrado del factor de cresta, el cual indica el número de picos que presenta una señal. Una señal con PAPR alto significa que es una señal que presenta numerosos picos de potencia y, por esta razón, es una señal más compleja de tratar. Sin embargo, una señal con una modulación eficiente, que permite transmitir numerosos bits en un solo símbolo, presenta un PAPR alto. Por lo tanto, para aumentar la velocidad de transmisión, es necesario trabajar con señales eficientes en cuanto a modulación, pues presentan un PAPR elevado.

Para medir el PAPR de una señal se utiliza la siguiente expresión:

$$PAPR = 10 \log \frac{|x|_{peak}^2}{x_{rms}^2} [dB] \quad (2.11)$$

2.3. Distorsión no lineal del amplificador de potencia

2.3.1. Distorsión AM-AM

La característica AM-AM representa la relación de amplitud entre la señal modulada a la entrada del PA y la señal modulada a la salida [4]. Cuando esta relación es lineal, no existe distorsión AM-AM.

Idealmente, la característica AM-AM de un amplificador de potencia es lineal y queda descrita con la siguiente ecuación:

$$V_{out}(t) = G * V_{in}(t) \quad (2.12)$$

Según la ecuación 2.12, a la salida del amplificador la tensión de la señal presenta un nivel G veces mayor que a la entrada, siendo G la ganancia escalar de voltaje del amplificador. Idealmente, este comportamiento lineal se lleva a cabo hasta cierta tensión de entrada, a partir de la cual el amplificador ya no puede realizar su función y se halla en régimen de saturación. En la figura 2.4 se muestra la curva AM-AM de un amplificador de potencia ideal.

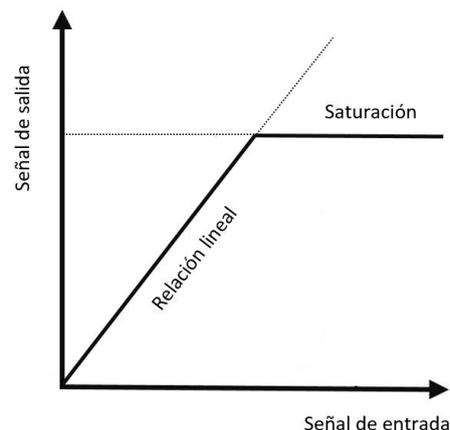


Figura 2.4: Curva característica AM-AM de un PA ideal.

Si analizamos este comportamiento lineal en el dominio de la frecuencia, en el caso de tener a la entrada una senoide de frecuencia f_0 , a la salida tendremos la misma senoide de frecuencia f_0 con un nivel de potencia mayor. Con este comportamiento ideal, la señal a la salida del amplificador sólo estará compuesta por los componentes frecuenciales que aparecen en la señal a la entrada.

Pero el comportamiento del PA no es ideal y, cuando la señal se acerca al límite de potencia, la salida del amplificador sufre una compresión, dejando de existir esa relación lineal entre las tensiones de entrada y salida. La figura 2.5 muestra la curva AM-AM real de un amplificador de potencia.

En un PA real, la transición entre la zona lineal y la saturación no se hace de forma abrupta, sino que hay una zona de transición que provoca una distorsión de amplitud en la señal a la salida del amplificador. A mayor potencia de entrada, mayor será el efecto no lineal que sufrirá la señal.

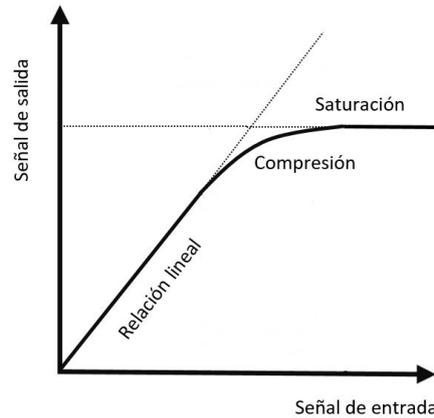


Figura 2.5: Curva característica AM-AM de un PA real.

Debido a esta zona de transición, no podemos aproximar el comportamiento del amplificador como un comportamiento lineal, pero sí lo podemos aproximar por un comportamiento polinómico. El comportamiento del PA real se puede expresar con la siguiente ecuación:

$$V_{out}(t) \simeq \sum_{k=1}^{\infty} a_k V_{in}^k(t) \quad (2.13)$$

La ecuación 2.13 muestra que el voltaje de la señal a la salida, $V_{out}(t)$, está formada por una serie de términos proporcionales al voltaje de la señal a la entrada, $V_{in}(t)$. El primer término de esta serie ($k = 1$) corresponde al término deseado a la salida (término lineal). El resto de términos de la serie corresponden a componentes frecuenciales no deseados y que no forman parte de la señal de entrada.

Para ser más específicos, los componentes de la señal de salida con k par corresponden a la distorsión armónica (HD), y describen los múltiplos enteros de las componentes frecuenciales de la señal de entrada. Si la señal de entrada es un tono de frecuencia f_0 , a la salida obtendremos una señal con frecuencia f_0 , más una serie de tonos a $2f_0, 3f_0, 4f_0$, etc., que van disminuyendo su amplitud debido a la distorsión armónica generada por el PA.

Los términos de la serie que corresponden a k impar introducen una distorsión de intermodulación (IMD). Esta distorsión aparece cuando la señal de entrada está formada por 2 o más componentes frecuenciales. Por ejemplo, si la señal de entrada está compuesta por dos tonos a f_1 y f_2 , la distorsión de intermodulación de tercer orden ($k = 3$) introduce nuevas componentes frecuenciales a $2f_1 - f_2, 2f_2 - f_1, 2f_1 + f_2$ y $2f_2 + f_1$.

La distorsión armónica puede ser filtrada, pero la distorsión de intermodulación introduce componentes frecuenciales cercanos a la banda de la señal y no pueden ser filtrados. Esta distorsión introduce un ensanchamiento del espectro de la señal y produce distorsión fuera de banda. Mantener el ACPR por debajo de un cierto valor permite tener bajo control la distorsión debida a la intermodulación. Un ejemplo de ensanchamiento del espectro se muestra en la figura 2.6.

El efecto de la distorsión no lineal también se aprecia en la constelación de la modulación recibida. Esta distorsión en señales digitales provoca que los símbolos de mayor potencia

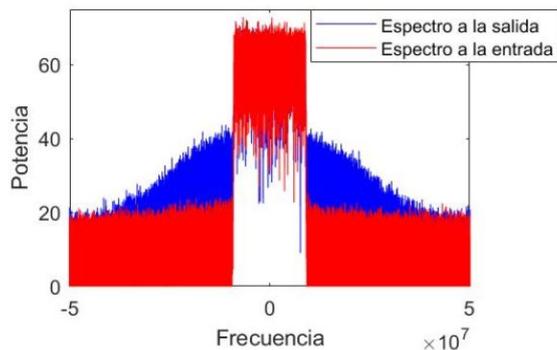


Figura 2.6: Representación del ensanchamiento espectral debido a la distorsión no lineal.

se transmiten a menor potencia debido a la compresión sufrida en el PA. Este hecho puede provocar una distorsión en los límites de decisión de la constelación y producir un aumento en la tasa de error en el bit (BER). Este efecto es conocido con el nombre de warping y se muestra en la figura 2.7.

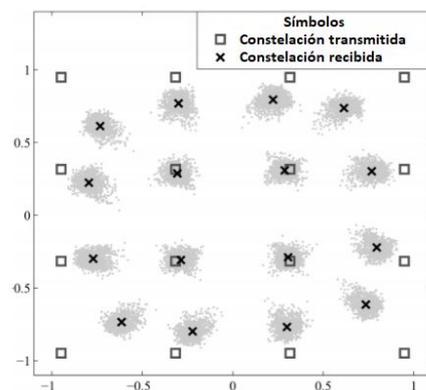


Figura 2.7: Representación del efecto Warping en una modulación 16QAM.

2.3.2. Distorsión AM-PM

La característica AM-PM define la desviación indeseada producida en la fase de la señal a la salida del PA debido a la distorsión no lineal que introduce el amplificador en la amplitud de la señal. Supongamos que a la entrada tenemos una señal portadora, $P_{in}(t)$, y una señal moduladora definida como:

$$M(t) = A_M \cos \omega_M(t) \quad (2.14)$$

Idealmente, la fase de la señal a la salida sería:

$$\phi(P_{in}(t)) = K_\phi \quad (2.15)$$

Donde K_ϕ es una constante. Esto quiere decir que la fase se mantiene igual, aunque la amplitud haya variado. Pero, en un PA real, la fase a la salida deja de ser constante y

queda determinada por la siguiente expresión:

$$\phi(P_{in}(t)) = K_{\phi} \cos [\omega_C t + A_M \cos (\omega_M t)] = K_{\phi} \sum_{n=-\infty}^{\infty} J_n(A_M) \cos [(\omega_C + n\omega_M)t] \quad (2.16)$$

Donde J_n es la función de Bessel de orden n y ω_C es la frecuencia de la portadora. A partir de la ecuación 2.16, el espectro de la señal a la salida de una portadora modulada en fase con una señal moduladora sinusoidal da como resultado la aparición de productos de intermodulación.

En la figura 2.8 se muestra un ejemplo de curva AM-PM. En este tipo de gráfica se puede apreciar la diferencia de fase entre la entrada y la salida. Esta diferencia de fase es dependiente de la amplitud de la señal de entrada.

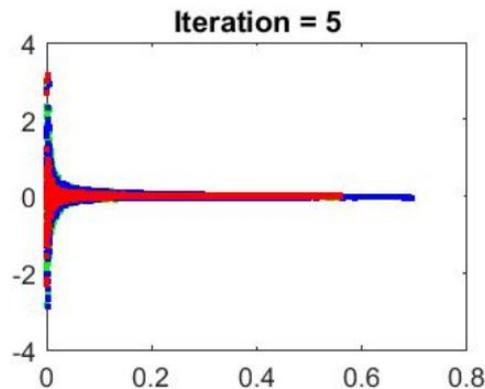


Figura 2.8: Curva característica AM-PM.

2.3.3. Efecto de memoria

Además de la distorsión AM-AM, AM-PM y del ensanchamiento del espectro, también hay que considerar que el amplificador introduce un efecto de memoria. Esto quiere decir que la señal a la salida del amplificador está condicionada por muestras pasadas de la señal (existe una dependencia temporal), por lo que, para un mismo valor de entrada, no se va a obtener siempre el mismo valor a la salida, sino que este valor dependerá de las muestras anteriores (2.9).

Estos efectos de memoria se producen debido a dos fuentes distintas: por un lado, tenemos los efectos de memoria eléctricos vinculados a la circuitería del amplificador, porque una señal de entrada de envolvente no constante provoca variaciones en la impedancia del componente activo; y, por otro lado, tenemos los efectos de memoria térmicos, que se producen por una variación no instantánea en la temperatura de manera del elemento activo.

El efecto de memoria puede causar una variación de la potencia de los productos de intermodulación en función del ancho de banda de la señal. Por este mismo motivo también puede verse rota la simetría entre los productos de intermodulación superior e inferior.

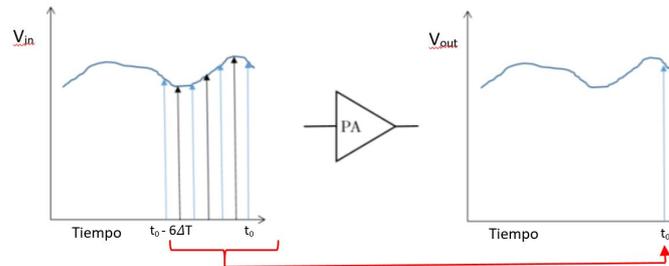


Figura 2.9: Representación del efecto de memoria en un PA.

Como consecuencia, podemos considerar que el efecto de memoria es una distorsión dependiente del ancho de banda.

En la curva AM-AM, si existe efecto de memoria, se produce un aumento en el grosor debido a la posibilidad de tener más de una salida diferente para una misma entrada. En la figura 2.10 se muestra la curva AM-AM con efecto de memoria.

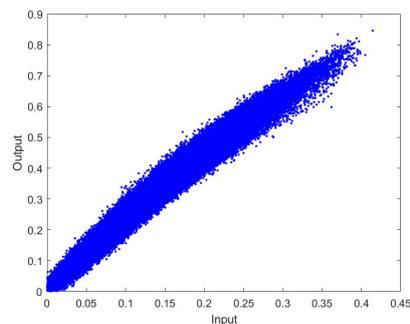


Figura 2.10: Curva AM-AM con efecto de memoria

2.4. Clases de amplificadores de potencia

Los amplificadores de potencia para radiofrecuencia se pueden dividir en dos grandes categorías: la primera de estas categorías consiste en aquellos amplificadores que preservan la forma de la onda a la salida del amplificador, conocidos como amplificadores lineales; mientras que la segunda categoría, o amplificadores no lineales, está formada por aquellos amplificadores que no preservan la forma de la onda. Estas dos grandes categorías quedan divididas, a su vez, en diferentes clases. Estas clases dependen del ángulo de conducción de la corriente de drenaje del transistor. Cada una de ellas presenta una topología, una linealidad y una eficiencia distintas.

Dentro de los amplificadores lineales encontramos las clases A, B y AB; mientras que en el grupo de amplificadores no lineales encontramos las clases C, D, E y F. En la tabla 2.1 se muestran las principales características de las distintas clases [4].

Tabla 2.1: Características principales de las diferentes clases de amplificadores de potencia.

Clases PA	Descripción	Ángulo de conducción	Linealidad	Eficiencia máxima
Clase A	Trabaja en zona lineal de su curva característica.	2π	Máxima linealidad.	50 %
Clase B	Dos transistores, cada uno amplifica la mitad del periodo.	π	Linealidad alta pero distorsión cruzada.	78.5 %
Clase AB	Presenta menor distorsión cruzada que la clase B.	$\pi < AB < 2\pi$	Menor a la clase A pero mejor que clase B.	$50\% < AB < 78,5\%$
Clase C	Mejora la eficiencia gracias a conducir corriente durante menos de medio ciclo.	$C < \pi$	Baja	100 %
Clases D, E y F	Funcionan como interruptores, por lo que genera una importante distorsión en la forma de la onda.	Conmutación	Gran distorsión.	100 %

2.5. Linealidad versus eficiencia

A partir de la tabla 2.1 se observa cómo, para un PA, la eficiencia y la linealidad son parámetros opuestos. Si queremos un dispositivo lineal se debe sacrificar eficiencia con valores de back-off elevados y, si lo que se buscan son niveles altos de eficiencia, se tendrá que lidiar con efectos no lineales importantes. Además, en las señales que presentan un PAPR alto, para evitar la distorsión no lineal, se requiere aumentar el back-off, degradando consecuentemente todavía más la eficiencia. Debido a que los amplificadores de potencia son dispositivos de alto consumo es importante maximizar la eficiencia y, por lo tanto, aumentar el back-off no es una opción. Por esta razón se utilizan técnicas que permiten aumentar la eficiencia, aunque ello conlleve una mayor distorsión no lineal. Esta distorsión se solventa por medio de otro conjunto de técnicas de linealización. En las figuras 2.11 se muestra cómo una alta linealidad conlleva una baja eficiencia y como un alto PAPR provoca mayor distorsión si no se aumenta el nivel de back-off.

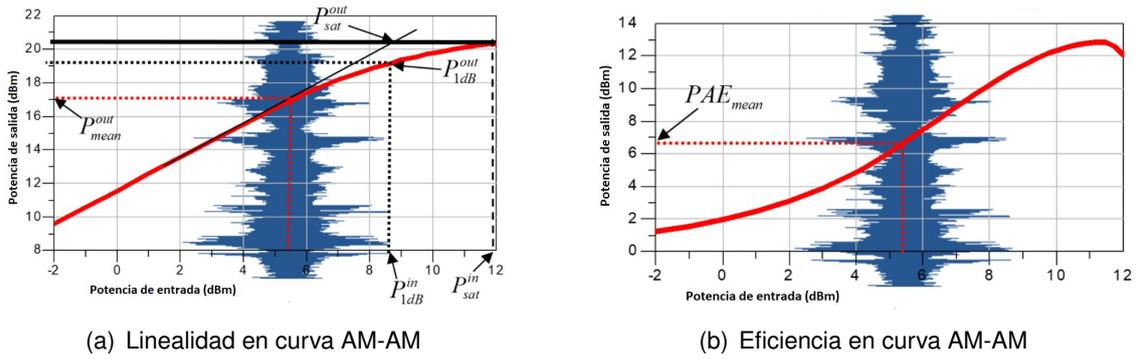


Figura 2.11: Linealidad versus eficiencia.

2.6. Técnicas para optimizar la eficiencia

2.6.1. Técnicas de alimentación dinámica

2.6.1.1. Envelope tracking

Esta técnica de seguimiento de envolvente permite adaptar la tensión de la fuente de alimentación a la envolvente de la señal [5]. Si la señal presenta una amplitud baja, la fuente de alimentación se adapta bajando su nivel de tensión, aumentando la alimentación cuando la envolvente toma niveles altos de amplitud. Esta técnica, que adapta la alimentación a la envolvente, permite aumentar la eficiencia, sobre todo, en señales que presentan niveles de PAPR elevados.

El ahorro de energía que se produce con esta técnica se puede perder si el detector de envolvente y el modulador de la tensión de alimentación son ineficientes. Además, hay que tener en cuenta que la variación de la tensión de alimentación introduce no linealidades extra, que deben ser tratadas con alguna técnica de linealización. La figura 2.12 muestra el esquema de bloques de esta técnica.

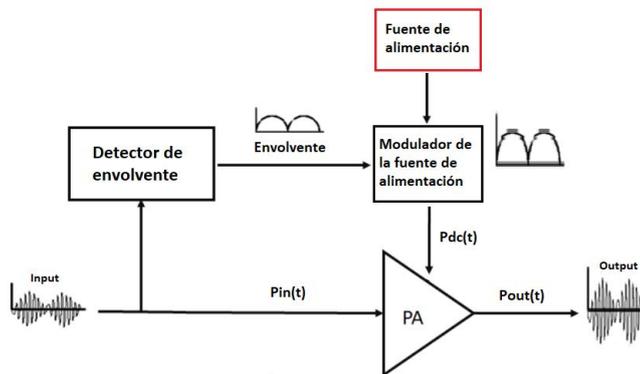


Figura 2.12: Diagrama de bloques del seguimiento de envolvente.

2.6.1.2. Envelope Elimination and Restoration (EER)

Esta técnica consiste en separar la señal de entrada en dos componentes: la primera de ellas es la envolvente de la señal, que se encargará de modular la tensión de alimentación como en la técnica anterior; la segunda componente corresponde a la portadora de la señal modulada en fase. Esta última componente se amplifica mediante un PA de alta eficiencia alimentado con la tensión de alimentación modulada. Amplificar una señal de envolvente constante nos permite utilizar un amplificador no lineal de alta eficiencia. Sin embargo, esta técnica no es útil para señales con ancho de banda grande. En la figura 2.13 se muestra el diagrama de bloques de esta técnica.

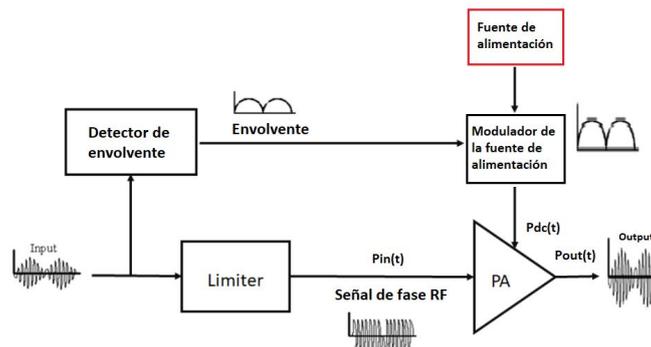


Figura 2.13: Diagrama de bloques de EER.

2.6.2. Técnicas de modulación de carga dinámica

2.6.2.1. Linear Amplification using Nonlinear Components (LINC)

Esta topología fue propuesta por D. Cox en [6], y consiste en dividir la señal en dos señales de RF con envolvente constante y modulación de fase con cambios de fase opuestos. Las dos componentes de la señal se amplifican independientemente con dos amplificadores separados. Como las dos componentes son de envolvente constante, se pueden amplificar cerca de la zona de saturación del amplificador sin sufrir una importante distorsión. Posteriormente, se combinan las dos señales para obtener la señal de la entrada amplificada. El combinador implementado en esta técnica consiste en un combinador aislado que permite obtener una señal amplificada lineal, aunque produce una disipación de energía extra en el dispositivo. El diagrama de bloques simplificado de esta técnica se muestra en la figura 2.14.

2.6.2.2. Amplificador Doherty

Esta arquitectura mejora significativamente la eficiencia combinando dos PA para obtener una salida con diferentes niveles de potencia. Este tipo de arquitectura está formada por un PA principal, un PA auxiliar, y por dos inversores de impedancia de $\lambda/4$ para balancear el desfase entre los dos caminos. Si el nivel de potencia de la señal es bajo (niveles de back-off altos), sólo se encuentra activo el PA principal; sin embargo, si la señal presenta niveles de potencia altos (niveles de back-off bajos), entra en funcionamiento el PA auxiliar,

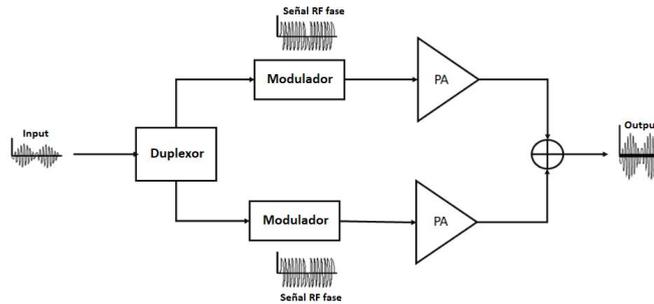


Figura 2.14: Diagrama de bloques de LINC.

obteniendo la potencia de la señal a la salida a partir de la amplificación producida por los dos PAs. Generalmente, el amplificador principal es de clase B o AB, mientras que el auxiliar es una clase C.

Este tipo de arquitectura que se muestra en la figura 2.15 nos permite obtener eficiencias altas, pero requiere de dos PA en lugar de uno y, además, el ancho de banda de la señal está limitado por las líneas de transmisión de $\lambda/4$ utilizadas, que dependen de la frecuencia de la señal. Los amplificadores Doherty introducen una distorsión no lineal, que principalmente se debe a la compresión de ganancia que aparece justo antes de que el amplificador auxiliar comience a conducir [7].

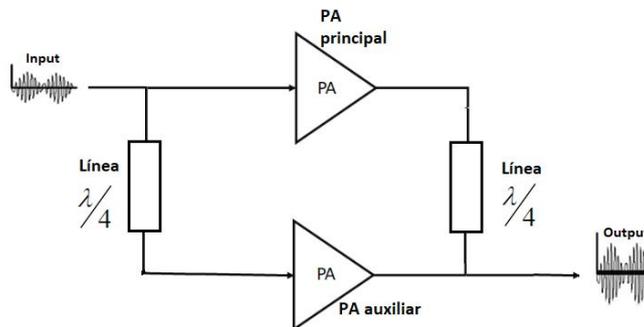


Figura 2.15: Diagrama de bloques de amplificador Doherty.

2.6.2.3. Load Modulated Balanced Amplifier

Este tipo de amplificador se propone en [7], y se basa en un PA balanceado (BPA) donde se inyecta una potencia de señal de control (CSP), por medio del puerto aislado de un acoplador de 90° , para modular la carga. Una característica importante del LMBA es que el CSP siempre contribuye a la potencia de la señal a la salida, independientemente de la modulación de carga; a diferencia de los amplificadores Doherty, donde la contribución del amplificador auxiliar y la modulación de carga están relacionadas. Otra característica importante es que este tipo de amplificadores permite trabajar con señales con un ancho de banda importante. La limitación en el ancho de banda viene determinada por el ancho de banda del acoplador utilizado, que suele ser bastante mayor que el ancho de banda que soportan los amplificadores Doherty. En la figura 2.16 se muestra la arquitectura LMBA. En el capítulo 5 [8] de este trabajo se presenta más en profundidad este amplificador.

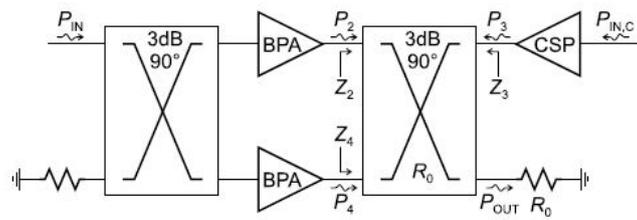


Figura 2.16: Diagrama de bloques de LMBA.

2.7. Técnicas de linealización

2.7.1. Feedback

Esta técnica permite una buena reducción de los productos de intermodulación, pero, a cambio, requiere sacrificar ganancia del amplificador para obtener esta linealidad. En la figura 2.17 se muestra el esquema de la configuración directa de esta técnica.

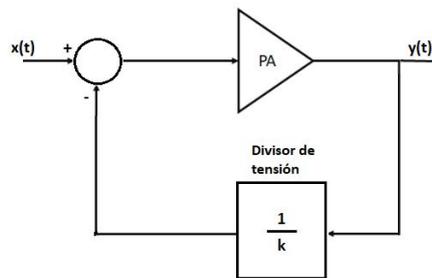


Figura 2.17: Esquema de linealización feedback.

La señal de entrada del PA se compara con la señal a la salida atenuada. Esta atenuación se realiza por medio de un divisor de tensión. La señal diferencia se amplifica un valor G para obtener la siguiente relación a la salida:

$$y(t) = \frac{kG}{(k + G)}x(t) \quad (2.17)$$

Donde k es la atenuación que se aplica por medio del divisor de tensión, y G es la ganancia del amplificador. Si consideramos que G es mucho mayor que k , la relación entrada salida se puede reescribir de la siguiente forma:

$$y(t) = kx(t) \quad (2.18)$$

Tal y como se muestra en la ecuación 2.18, la relación entre la entrada y la salida es lineal, pero la ganancia obtenida es bastante menor que la ganancia del amplificador. Esta técnica genera mejor linealidad que la técnica feedforward sin la necesidad de utilizar dos amplificadores, pero solo puede operar con anchos de banda pequeños, debido al retardo producido por el bucle cerrado (Δt_s). Además de la configuración direct feedback, existen otras configuraciones como polar feedback, cartesian feedback y envelope feedback.

2.7.2. Feedforward

En un linealizador feedforward podemos diferenciar dos circuitos: el primero se trata de un circuito cancelador de la señal; y el segundo, de un circuito cancelador del error. Con el circuito que cancela la señal se obtiene la señal de error, que consiste en los productos de intermodulación que se han generado en el PA. Para obtener la señal error se requiere comparar, mediante un combinador de 180° , la señal a la salida del amplificador (debidamente atenuada) con la señal retardada de la entrada. Posteriormente, en el circuito cancelador del error, la señal de error se amplifica y se combina en contrafase con la señal a la salida del amplificador (previamente retardada). Con este procedimiento se eliminan los productos de intermodulación. En la figura 2.18 se muestra el esquema de este linealizador.

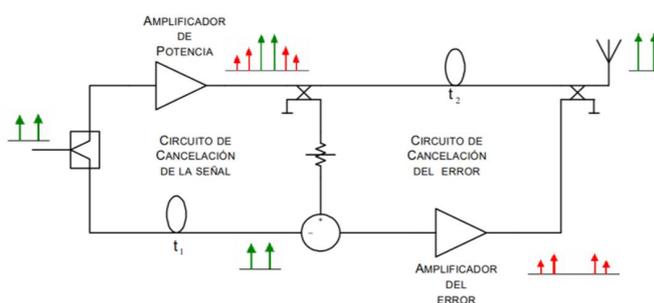


Figura 2.18: Esquema de linealización feedforward.

A partir del esquema, si tenemos como señal de entrada dos tonos, primero, en la etapa de cancelación de señal se divide la señal en dos ramas: la primera rama pasa por el amplificador de potencia, y en ella aparecen los productos de intermodulación debido a los efectos no lineales del amplificador; en la segunda rama sólo se aplica un retardo igual al tiempo requerido para que la primera rama pase por el amplificador. Posteriormente, en la segunda etapa, la etapa de cancelación de error, se realiza la diferencia entre las dos ramas mencionadas, obteniendo sólo los productos de intermodulación. En esta segunda etapa, el circuito continúa con dos ramas: la primera rama es la señal amplificada, por lo que el espectro de la señal está afectado por los productos de intermodulación; y, la segunda rama, sólo está compuesta por estos productos de intermodulación. A continuación, la segunda rama amplifica la señal, y en la primera rama se introduce un retardo para compensar el efecto de este amplificador. Finalmente, las señales de las dos ramas se suman en contrafase, eliminando los productos de intermodulación. Así logramos tener la señal de entrada amplificada sin distorsión. El concepto de esta técnica es sencillo, pero presenta varias desventajas: primero, la eficiencia es baja debido a la necesidad de operar con dos PAs; y, segundo, se pierde linealidad debido al desajuste de fases en los diferentes componentes. Por esta razón es importante un buen ajuste de los retardos aplicados, para compensar los efectos de las etapas de amplificación.

2.7.3. Predistorsión

Esta técnica de linealización consiste en aplicar una modificación no lineal a la señal de entrada, antes de que la señal pase por el PA. La relación entrada-salida que se aplica

con el predistorsionador es la función inversa del modelo matemático del comportamiento entrada-salida que presenta el PA. De esta manera, la relación entrada-salida del conjunto predistorsionador y PA será lineal. Esta técnica se desarrolla en mayor detalle en el capítulo 3 de este trabajo.

CAPÍTULO 3. PREDISTORSIÓN

3.1. Introducción

La predistorsión fue propuesta por Adel A. M. Saleh y J. Salz en 1983 [9]. Consiste en evitar la distorsión no lineal aplicando la función inversa del comportamiento del PA a la señal de entrada u . En la figura 3.1 se muestra el esquema básico del concepto de predistorsión.

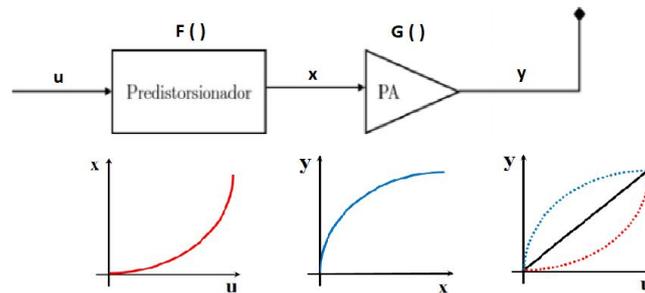


Figura 3.1: Esquema de predistorsión.

A partir del esquema de la figura 3.1, la señal a la salida y , se verá afectada por las funciones de transferencia del predistorsionador y del amplificador, donde estos dos componentes se encuentran colocados en cascada. La relación entrada-salida del conjunto predistorsionador-amplificador puede ser descrita con la siguiente expresión:

$$y = G(x) = F(G(u)) \quad (3.1)$$

Esta técnica se puede aplicar sobre la señal de entrada en banda base (BB), en frecuencia intermedia (IF) o en radio frecuencia (RF). Sin embargo, es aconsejable realizar la predistorsión en BB o IF por ser más robusto; no depende de la banda frecuencial de operación y porque los componentes requeridos, como los convertidores analógico-digital, son más económicos si trabajan a frecuencias más bajas. A pesar de ello, como contrapartida, realizar la predistorsión en BB o IF implica que el convertidor a RF introducirá una distorsión adicional. Además, la predistorsión se puede aplicar tanto sobre una señal analógica (APD) como sobre una señal digital (DPD). El esquema puede ser: de bucle cerrado, que corresponde a un modelo adaptativo, o de bucle abierto, modelo no adaptativo.

3.2. Predistorsión analógica

La predistorsión analógica (APD) utiliza un circuito analógico para predistorsionar la señal de entrada. Este circuito genera el comportamiento inverso del PA, distorsionando la señal antes de que llegue al amplificador.

Una de las ventajas de la APD es que la expansión del ancho de banda se produce en el

dominio analógico, permitiendo una relajación en los requerimientos de velocidad del reloj del convertidor digital-analógico para generar la señal analógica a transmitir.

A lo largo de los años se han propuesto diversas soluciones para generar la función inversa por medio de circuitos analógicos. La empresa Scintera, por ejemplo, comercializa una solución APD basada en polinomios de la serie Volterra, implementada con multiplicadores analógicos. Otras soluciones se basan en las características no lineales de los diodos. También se propone la utilización de composite right-left-handed transmission lines (CRLHTLs) [10], que permiten variar sus características no lineales en función de la potencia de entrada o del voltaje de polarización. El esquema de este APD se muestra en la figura 3.2 [12].

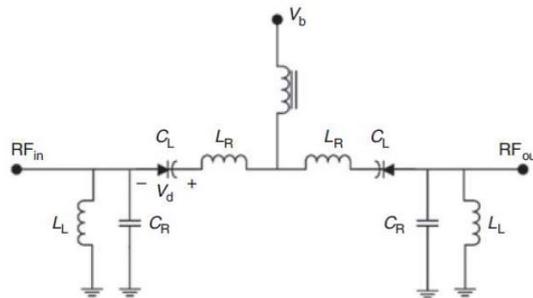


Figura 3.2: Predistorsionador analógico CRLHTLs.

Finalmente, otra topología muy utilizada consiste en utilizar transistores, tal y como se presenta en [11] y se muestra en la figura 3.3 [11].

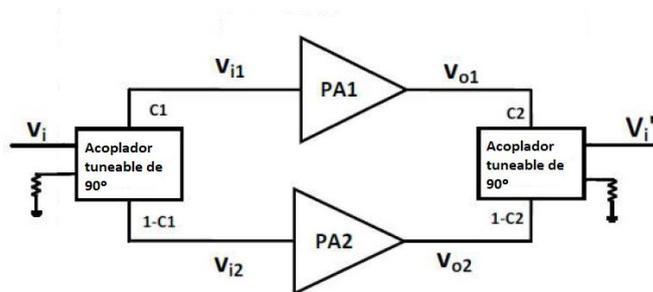


Figura 3.3: Predistorsionador analógico basado en transistores.

3.3. Predistorsión digital

La predistorsión digital (DPD) se lleva a cabo sobre la señal digital compleja en banda base a transmitir, que se genera en el procesador digital de señal, y consiste en aplicar una transformación matemática a la señal.

El DPD se describe mediante modelos de comportamiento, o blackbox, que expresan matemáticamente la función de transferencia inversa no lineal del PA. A diferencia de un modelo físico, donde es necesario conocer la electrónica con la que está diseñado el PA, para implementar DPD sólo es necesario disponer de muestras de entrada y salida de

PA. El rendimiento de un tipo de DPD frente a otro dependerá del modelo matemático que se utilice para obtener el comportamiento del PA, y del procedimiento que se utilice para calcular los coeficientes del predistorsionador.

La implementación de esta técnica de linealización se puede realizar mediante una topología de bucle cerrado, donde se van modificando los coeficientes a lo largo del tiempo, o mediante una topología de bucle abierto, donde no hay retroalimentación y los coeficientes son fijos.

3.3.1. Predistorsión no adaptativa

La predistorsión no adaptativa corresponde a la implementación de DPD en bucle abierto. No existe una retroalimentación entre la entrada y la salida y los coeficientes del DPD se mantienen fijos. Este tipo de DPD no se puede adaptar a un PA que modifica su comportamiento en función del tiempo.

3.3.2. Predistorsión adaptativa

La predistorsión adaptativa corresponde a la implementación de DPD en bucle cerrado. Existe una retroalimentación que permite modificar los coeficientes del DPD y, por lo tanto, el DPD se adapta a un PA con respuesta variante en el tiempo. Esta respuesta con dependencia temporal se puede deber: a variaciones térmicas, al envejecimiento del equipo o a la implementación de controles automáticos de ganancia. Por otro lado, la implementación del método adaptativo conlleva un aumento en el coste del dispositivo y en el consumo de energía. En la figura 3.4 se muestra la arquitectura de predistorsión adaptativa.

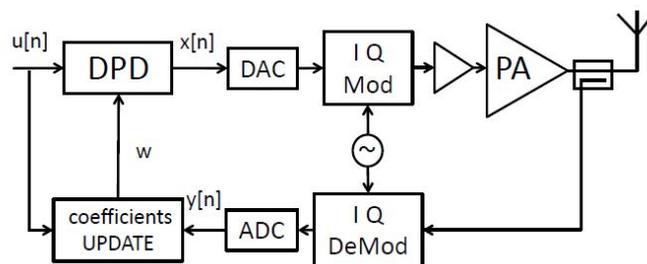


Figura 3.4: Arquitectura DPD en bucle cerrado.

La implementación de la predistorsión adaptativa se puede realizar de dos formas: mediante aprendizaje directo o aprendizaje indirecto. La diferencia entre una y la otra reside en el método implementado para actualizar los coeficientes necesarios para aplicar la predistorsión.

3.3.2.1. Aprendizaje indirecto

EL esquema de este tipo de aprendizaje se muestra en la figura 3.5. En este método, primero se estiman los coeficientes del filtro Post-inverse y, posteriormente, estos coeficientes se copian en el predistorsionador. El filtro Post-inverse recibe a la entrada la señal

obtenida a la salida del amplificador de potencia ($y[n]$), y regula sus coeficientes por medio del error generado entre la señal de entrada al amplificador ($x[n]$) y la señal a la salida del filtro Post-inverse.

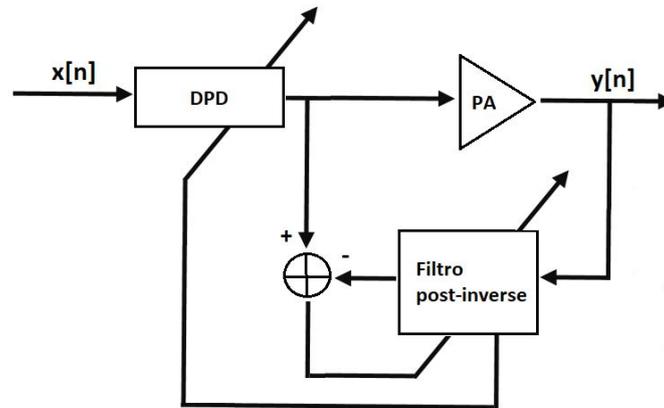


Figura 3.5: DPD adaptativo con aprendizaje indirecto.

Este método presenta una serie de desventajas: en primer lugar, la señal empleada en el filtro adaptativo $y[n]$ es una señal ruidosa, esto hace que el filtro converja en valores sesgados; además, teóricamente, los filtros no lineales no pueden ser permutados, por lo que colocar una copia de este filtro delante del amplificador de potencia no asegura una buena reducción o eliminación de la distorsión no lineal.

3.3.2.2. Aprendizaje directo

En este método de aprendizaje, los coeficientes del predistorsionador se actualizan directamente a partir de la entrada del predistorsionador y un error de referencia. Este método es recursivo y los coeficientes se basan en valores anteriores de los mismos en base a un criterio establecido, como puede ser el error cuadrático medio. Con el método directo se evita la utilización de una señal afectada por ruido, como es la señal $y[n]$. Ésta es la principal ventaja que presenta este método frente al indirecto. En la figura 3.6 se muestra el esquema de un predistorsionador adaptativo con método de aprendizaje directo donde $d[n]$ es la señal que se desea obtener a la salida del amplificador.

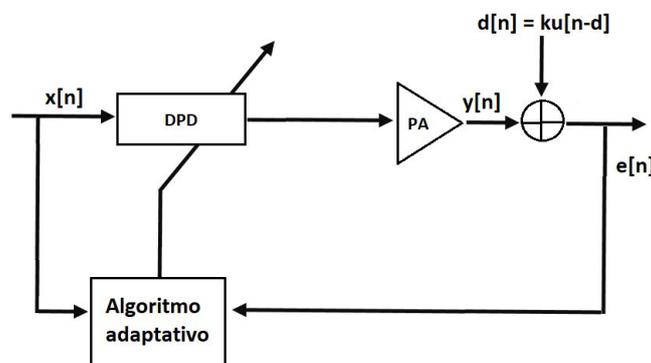


Figura 3.6: DPD adaptativo con aprendizaje directo.

3.4. Modelos de predistorsión digital

Para poder construir la función del predistorsionador es necesario identificar el comportamiento no lineal del PA y el efecto de memoria por medio de un modelo matemático o modelo DPD.

El modelo de DPD es una parte básica de todo el algoritmo DPD. Existen una gran variedad de modelos, pero básicamente pueden ser divididos en dos grandes grupos: modelos de comportamiento sin memoria y con memoria.

Los modelos sin memoria consisten en un mapeo no lineal entre la señal de entrada y la de salida. Los modelos sin memoria se pueden entender como un sistema *Single Input Single Output* (SISO), ya que la señal de salida depende únicamente de la señal de entrada actual.

Los modelos con efecto de memoria son los que mejor se ajustan al comportamiento del PA. En estos modelos se tiene en cuenta que la salida del amplificador no sólo depende de la señal de entrada en ese instante, sino que también depende del historial reciente de la señal. Este efecto de memoria se hace más relevante a medida que la señal aumenta su ancho de banda. Un modelo con efecto de memoria consiste en el mapeo no lineal de varias señales de entrada. Los modelos con memoria se pueden entender como un sistema *Multiple Input Single Output* (MISO). A continuación, se presentan los modelos de DPD más utilizados.

3.4.1. Series de Volterra

Estas series permiten modelar el comportamiento no lineal de manera similar a las series de Taylor. La diferencia entre ambas radica en que las series de Taylor permiten aproximar la respuesta de un sistema no lineal, cuando esta respuesta sólo depende del valor de la entrada en un instante determinado; sin embargo, en las series de Volterra, el valor a la salida del sistema no lineal depende de la entrada en toda su extensión temporal. Esta característica de las series de Volterra permite tener en cuenta el efecto de memoria a la hora de modelar el comportamiento del amplificador de potencia.

El modelo obtenido al aplicar series de Volterra está formado por una serie de potencias que tiene en cuenta los efectos de memoria, por lo que la salida del sistema se corresponde con una suma de series polinómicas de la entrada en el instante actual, y de retardos de la misma.

$$v_0(t) = \sum_k \sum_n n a_{k,n} v_i^n(t - k\Delta) \quad (3.2)$$

Donde $a_{k,n}$ son los coeficientes a determinar de las series de Volterra. Estas series se han aplicado en diferentes ámbitos, como pueden ser: la medicina, la neurociencia, la biología o la electrónica. En electrónica se han implementado para modelar la distorsión producida en amplificadores y mezcladores de radiofrecuencia. A continuación, se desarrolla la serie de Volterra de tercer orden de un sistema causal en el tiempo discreto:

$$y_{SV}(n) = \sum_{k=0}^{N-1} h_k^{(1)} x(n-k) + \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} h_{k,l,m}^{(3)} x(n-k)x(n-l)x^*(n-m) \quad (3.3)$$

Donde N es la profundidad de retardo, x e y son la envolvente compleja de la señal a la entrada y a la salida del sistema, y $h_k^{(1)}$ y $h_k^{(3)}$ son los kernels de la serie de Volterra de orden 1 y 3. Conocer estos kernels permite caracterizar completamente el sistema al igual que un sistema LTI queda caracterizado al completo por su respuesta al impulso. El kernel h_0 describe el comportamiento del sistema si a la entrada tenemos una señal continua. El kernel h_1 describe el comportamiento lineal del sistema y los kernels superiores se encargan de caracterizar el comportamiento no lineal.

Alguna de las desventajas de utilizar este modelo es que conlleva un elevado coste computacional, tiene una convergencia lenta y requiere del cálculo de un gran número de parámetros, muchos de los cuales son innecesarios. Por ello, existen diferentes grupos de investigación que trabajan en la generación de modelos basados únicamente en los coeficientes que modelan la no linealidad, omitiendo los coeficientes de menor influencia en estos efectos. Por lo general, para encontrar el valor de los coeficientes de las series de Volterra se utiliza la técnica de mínimos cuadrados.

3.4.2. Memory Polynomial

Este modelo se caracteriza por estar formado por distintas etapas que introducen retardos y funciones estáticas no lineales. Sólo considera los términos de la diagonal de la matriz de kernels de la serie general de Volterra; de esta manera se reduce considerablemente el número de parámetros a tener en cuenta. Podemos describir este modelo como un modelo que introduce pares de muestras retardadas hasta una determinada profundidad de retardo (valor de M), y cada retardo se considera hasta un determinado orden K ; con ello se consigue describir los efectos de memoria y las no linealidades del PA. La relación de entrada y salida de este modelo queda descrita por la siguiente expresión:

$$y_{MP}(n) = \sum_{m=0}^M \sum_{k=0}^K a_{m,k} x(n - \tau_m) |x(n - \tau_m)|^k \quad (3.4)$$

Donde $a_{m,k}$ son los coeficientes a determinar de este modelo. El retardo queda definido como $\tau_m = m * \tau_0$, donde τ_0 es el periodo de muestreo de la señal. La problemática de este modelo es que no puede describir los efectos de memoria que presentan los amplificadores a largo plazo debido a que esto significa un importante aumento en el número de coeficientes a considerar.

3.4.3. Generalized Memory Polynomial

Este modelo está basado en el Memory Polynomial, pero tiene en cuenta múltiples retrasos con diferencias temporales, tanto positivas como negativas, de manera que aparecen términos cruzados. Este modelo queda descrito con la siguiente expresión:

$$\begin{aligned}
y_{GMP}(n) = & \sum_{k=0}^{K_a-1} \sum_{l=0}^{L_a-1} a_{k,l} x(n-l) |x(n-l)|^k \\
& + \sum_{k=1}^{K_b} \sum_{l=0}^{L_b-1} \sum_{m=1}^{M_b} b_{k,l,m} x(n-1) |x(n-l-m)|^k \\
& + \sum_{k=1}^{K_c} \sum_{l=0}^{L_c-1} \sum_{m=1}^{M_c} c_{k,l,m} x(n-1) |x(n-l+m)|^k \quad (3.5)
\end{aligned}$$

Donde K_a y L_a son los coeficientes sin retrasos, ni positivos ni negativos, en la parte del modelo que corresponde al modelo Memory Polynomial. Los valores K_b , L_b y M_b son los coeficientes de la señal y su envolvente, retrasados positivamente; y K_c , L_c y M_c son los coeficientes de la señal y su envolvente, retrasados negativamente. Todos estos coeficientes tienen una dependencia lineal.

Una ventaja de este modelo es que sus coeficientes, al igual que sucede con el modelo Memory Polynomial, al aparecer con una relación lineal, se calculan de manera simple con algoritmos como el de mínimos cuadrados. La simplicidad del cálculo de estos coeficientes, debido a su dependencia lineal, hace que el algoritmo sea estable y que no presente una gran carga computacional.

3.4.4. GMP basado en B-splines

Una B-spline es una generalización de una curva de Bézier. Estas curvas, también conocidas como trazados vectoriales, permiten unir dos puntos o nodos por medio de curvas de diferente orden, donde la forma de la curva viene marcada por una serie de puntos invisibles denominados puntos de control. Un ejemplo de estas curvas se muestra en la figura 3.7. Las curvas de Bézier se utilizan ampliamente en los gráficos generados por ordenador para modelado de curvas suaves.

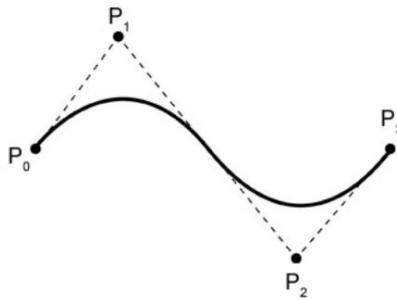


Figura 3.7: Curva de Bézier cuadrática.

El método basado en B-splines como DPD, consiste en aproximar la curva característica AM-AM del amplificador por medio de funciones de Bézier a trozos de orden n [13]. Para construir la matriz basada en las B-splines es necesario definir el orden de las curvas, la cantidad de nodos y los intervalos escogidos. Las B-splines se encuentran controladas por una serie de nodos, $T = [t_0, t_1, \dots, t_m]$. Estas B-splines se definen con la siguiente expresión:

$$\varphi(t) = \sum_{i=0}^{m+1-n} \alpha_i \beta_{i,n}(t) \quad (3.6)$$

donde α_i define los puntos de control, t son los valores de la secuencia de entrada y $\beta_{i,n}$ es la i -ésima función de base local de grado n . El número de puntos de control siempre es igual a la longitud del vector de nodos menos el orden de las curvas B-splines utilizadas ($PC = m + 1 - n$). Para construir los valores de $\beta_{i,n}$ se utiliza el algoritmo de Cox-De Boor. Para calcular los valores de la B-spline de orden n se necesita calcular previamente los valores de las B-splines de orden $n - 1$. Por esta razón, primero siempre se tiene que calcular el valor de las B-splines de orden 0. Estas funciones sólo toman el valor 0 o 1, de modo que: $\beta_{i,0}(t) = 1$, si $t_i \leq t < t_{i+1}$; y $\beta_{i,0}(t) = 0$, en el resto de casos. Una vez calculados los valores de $\beta_{i,0}(t)$ se pueden calcular los valores de grado 1 y sucesivos empleando la siguiente expresión:

$$\beta_{i,n}(t) = \frac{t - t_i}{t_{i+n} - t_i} \beta_{i,n-1}(t) + \frac{t_{i+n+1} - t}{t_{i+n+1} - t_{i+1}} \beta_{i+1,n-1}(t) \quad (3.7)$$

En la figura 3.8 se muestran las curvas B-splines de orden 4 con el siguiente vector de nodos $T = [0, 0, 0, 0, 0, 33, 0, 67, 1, 1, 1, 1]$. Este vector siempre se completa colocando el primer y último valor tantas veces como el grado de las funciones empleadas. En este caso, como el grado es 4, el valor 0 y 1 aparecen 4 veces en el vector de nodos. En este caso el número de puntos de control y de curvas B-splines de orden 4 son 6, que corresponden a las 6 curvas que se muestran en la figura.

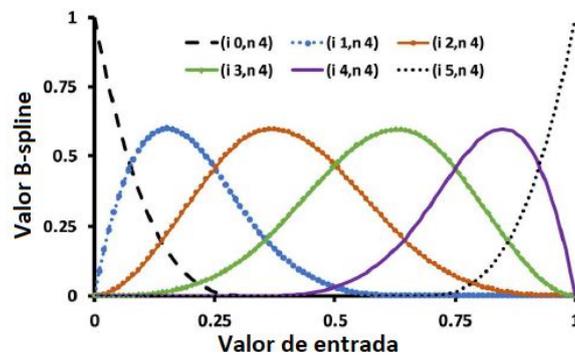


Figura 3.8: Ejemplo de curvas B-splines de orden 4.

Si expresamos este método siguiendo el esquema GMP, el modelo de B-Splines queda descrito con la siguiente ecuación:

$$\begin{aligned}
y_{BS}(n) = & \sum_{l=0}^{L_a-1} x(n-l) \varphi_l^a(|x(n-l)|) \\
& + \sum_{l=0}^{L_b-1} \sum_{m=1}^{M_b} x(n-l) \varphi_{l,m}^b(|x(n-l-m)|) \\
& + \sum_{l=0}^{L_c-1} \sum_{m=1}^{M_c} x(n-l) \varphi_{l,m}^c(|x(n-l+m)|) \quad (3.8)
\end{aligned}$$

donde $\varphi(\cdot)$ son las B-splines para una entrada determinada, y $L_{a,b,c}$ y $M_{b,c}$ son la profundidad de memoria y productos de memoria cruzados utilizados. En este caso, los valores de la secuencia de entrada t se corresponden con $|x(n+l \pm m)|$. Para implementar este método como predistorsionador se crea la matriz X con los valores de las B-splines para los diferentes valores de entrada y sus retardos. Una vez generada la matriz se aplica el método de mínimos cuadrados, que permite encontrar el valor de los coeficientes del modelo.

3.4.5. Redes neuronales

Las redes neuronales se pueden implementar para resolver una infinidad de problemas diferentes, y la linealización de un PA no es una excepción. Las redes neuronales nos permiten realizar el modelado del comportamiento del PA por medio de un aprendizaje previo con datos de entrada y salida del PA. Las redes neuronales están formadas por diferentes capas y cada capa está compuesta por neuronas. El aprendizaje se utiliza para identificar los coeficientes a aplicar en cada neurona. Estos coeficientes se seleccionan buscando minimizar el error cuadrático medio entre la predicción con la red neuronal y el valor real a la salida del PA. El funcionamiento de las redes neuronales y su implementación como DPD se desarrolla en el capítulo 4 de este trabajo.

3.5. Estimación de los coeficientes del predistorsionador

3.5.1. Método de mínimos cuadrados

Este método se utiliza para encontrar el valor óptimo de los coeficientes del modelo de comportamiento del PA. Los mínimos cuadrados se utilizan junto a modelos de DPD como la serie de Volterra, MP, GMP y B-splines.

El vector estimado de salida de un modelo de comportamiento (\hat{y}) puede describirse como el producto entre la matriz de datos de entrada, formada a partir de un modelo determinado como MP o GMP, y un vector que representa los parámetros o coeficientes del modelo.

$$\hat{y} = X\omega \quad (3.9)$$

Si tomamos como ejemplo el modelo MP, la matriz X está formada por los términos con memoria y sus potencias de la señal de entrada, y ω es el vector de parámetros del

modelo $(a_{p,m})$. La estimación del error se define como la diferencia entre la salida del PA y la salida estimada por el modelo.

$$e = y - \hat{y} = y - X\omega \quad (3.10)$$

El objetivo de este método es encontrar los valores de ω que minimizan este error. Tomando la norma l_2 al cuadrado del error, el problema de los mínimos cuadrados se define como sigue:

$$\min_{\omega} \|e\|_2^2 = \min_{\omega} \|y - X\omega\|_2^2 \quad (3.11)$$

a partir de la derivada de la función de coste ($J(\omega) = \|e\|_2^2$), e igualándola a 0 para encontrar un mínimo. La solución de los mínimos cuadrados se muestra a continuación:

$$\omega = (X^H X)^{-1} X^H y \quad (3.12)$$

donde H denota la conjugada traspuesta de la matriz X . Para obtener los coeficientes del modelado del comportamiento del PA, sólo es necesario resolver la ecuación 3.12.

3.5.2. Implementación de DPD adaptativo directo

Para aplicar los mínimos cuadrados en el DPD adaptativo con aprendizaje directo se calcula el error entre la señal antes del DPD ($u[n]$) y la señal después del PA ($y[n]$), las dos normalizadas para eliminar la ganancia lineal.

$$e[n] = y[n] - u[n] \quad (3.13)$$

Denominamos U a la matriz que contiene las funciones bases de un modelo determinado y ω el vector de parámetros. La matriz U es una matriz $L \times M$, conteniendo L muestras y M funciones base. El vector de parámetros es de longitud M y el vector de salida del DPD, x , es de longitud L . Este vector de salida se puede expresar como:

$$x = u - U\omega \quad (3.14)$$

Los coeficientes ω se pueden obtener por medio del descenso del gradiente de forma iterativa:

$$\omega_{j+1} = \omega_j + \mu_j \Delta\omega \quad (3.15)$$

donde j marca la iteración y μ es la ratio de aprendizaje, que puede ser constante o puede variar con cada iteración. El incremento del coeficiente $\Delta\omega$ se calcula por medio de la solución de mínimos cuadrados con al siguiente expresión:

$$\Delta\omega = (U^H U)^{-1} U^H e \quad (3.16)$$

En la figura 3.9 se muestra el esquema de DPD directo.

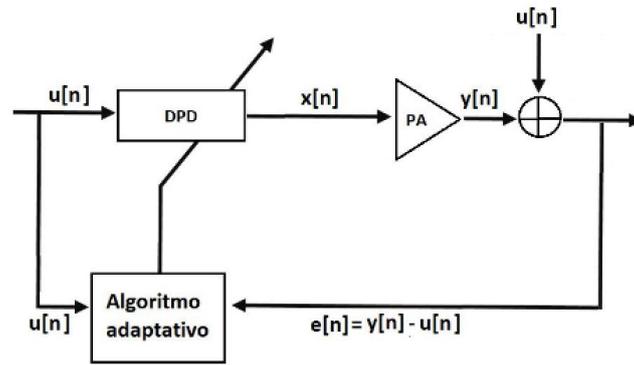


Figura 3.9: Esquema de la implementación de DPD directo.

CAPÍTULO 4. REDES NEURONALES ARTIFICIALES

4.1. Introducción

Las redes neuronales son un tipo de algoritmo surgido a finales de los años 40, y que se basaba en investigaciones de plasticidad neuronal llevadas a cabo por neurocientíficos y psicólogos de la época. Estos algoritmos tratan de imitar al cerebro para resolver diferentes problemas. Este tipo de algoritmo se utilizó para distintas aplicaciones en los años 80 y 90, pero cayó en desuso debido a la falta de recursos, ya que estas redes son computacionalmente muy costosas. En la actualidad, gracias a la evolución tecnológica y a la aparición de las redes neuronales convolucionales y recurrentes, se ha podido retomar este tipo de algoritmo con éxito. Esta técnica se ha convertido en una técnica vanguardista que se puede utilizar en una amplia variedad de aplicaciones.

4.1.1. Principio de funcionamiento de las redes neuronales artificiales

Aunque el cerebro humano puede aprender, en principio, diferentes cosas que puedan parecer completamente distintas como: determinar qué objeto tenemos delante por medio de la vista (procesar imágenes), reconocer diferentes olores o reconocer palabras por medio de la escucha; las redes neuronales artificiales tratan de imitar al cerebro, y se basan en la hipótesis de que el cerebro realiza todos estos aprendizajes distintos con un único algoritmo.

En la figura 4.1 aparece el esquema simplificado de una neurona del cerebro humano. Las dendritas de la neurona se encargan de recibir los diferentes inputs de información y de comunicarse con el núcleo de la célula nerviosa, que es donde tienen lugar los diferentes cálculos y procesamiento de la información. Finalmente, el axón transportará a la salida la información obtenida en la neurona, por medio de sus terminaciones, a las dendritas de otras neuronas.

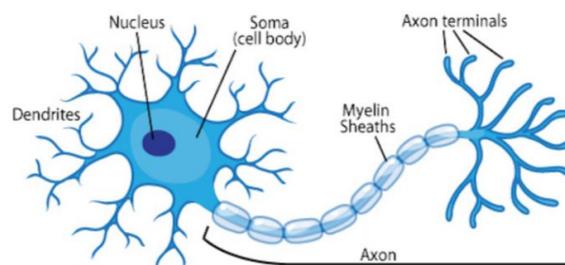


Figura 4.1: Esquema de una neurona del cerebro humano.

Las redes neuronales siguen la misma filosofía. Estas redes están formadas por diferentes neuronas a las que llegan una serie de entradas. Las neuronas realizan unos determinados cálculos con estas entradas y se obtiene una salida que se transmite como entrada

a otras neuronas, si es necesario. En la figura 4.2 aparece el esquema de una neurona dentro de una red neuronal artificial.

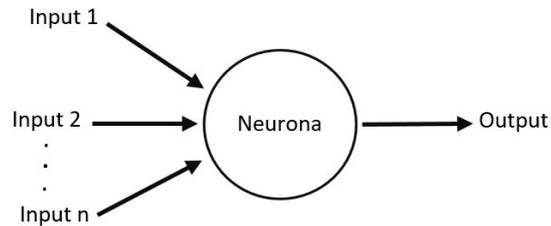


Figura 4.2: Esquema simplificado de una neurona de una red neuronal artificial.

Aunque una neurona es un elemento pequeño y simple, al combinarse miles o millones de estas neuronas se pueden resolver problemas muy complejos. Las redes neuronales artificiales presentan una serie de características típicas del cerebro. Estas redes son capaces de aprender a partir de la experiencia, son capaces de generalizar ofreciendo respuestas válidas a entradas que han sufrido alguna alteración, y son capaces de abstraer, es decir, obtener y separar las cualidades de un objeto.

4.2. Problemas de regresión y de clasificación

Las redes neuronales artificiales se pueden implementar para resolver dos tipos de problemas: los problemas de regresión, donde la salida es un valor continuo; o problemas de clasificación, donde la salida es un valor binario, 0 o 1.

En los problemas de regresión se consigue adaptar la relación de entradas y salida por medio de una curva no lineal, y en los problemas de clasificación se trata de determinar si el sistema cumple o no con una característica determinada. El caso de aplicar una red neuronal como predistorsionador corresponde a un problema de regresión donde, dadas una serie de entradas, se obtiene un valor continuo a la salida.

4.3. Redes neuronales supervisadas y no supervisadas

Las redes neuronales artificiales se pueden clasificar en: redes neuronales supervisadas y no supervisadas, en función de cómo se realiza el entrenamiento de la red.

En el aprendizaje supervisado el entrenamiento se lleva a cabo mediante conjuntos de datos de entrada y salida. A partir de estos conjuntos de entrenamiento el programa es capaz de obtener una salida válida a una entrada que no forma parte de las muestras de entrenamiento.

En el aprendizaje no supervisado, el entrenamiento se lleva a cabo sólo con datos de la entrada (no se facilitan los datos de salida correspondientes). La red aprende a partir de las experiencias de los patrones de entrenamiento anteriores.

En este trabajo, para implementar el DPD por medio de redes neuronales, se utiliza un modelo de aprendizaje supervisado, donde se entrena a la red neuronal artificial por medio de patrones de entrada y salida del LMBA.

4.4. Estructura de una red neuronal artificial

En la figura 4.3 se muestra el esquema de red neuronal artificial de una sola capa, que tomaremos como punto de partida para analizar el funcionamiento básico de las redes neuronales artificiales.

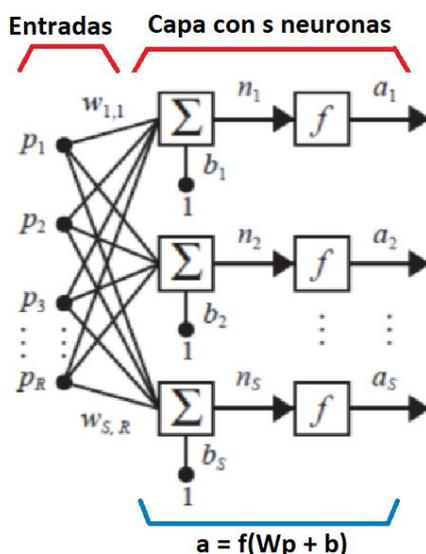


Figura 4.3: Esquema de una red neuronal artificial de una sola capa.

Esta red neuronal está alimentada por una serie de variables de entrada; estas variables de entrada forman el vector p :

$$p = [p_1, p_2, p_3, \dots, p_R] \tag{4.1}$$

Si nos centramos en cada una de las neuronas que forman la red, se observa que la neurona la podemos dividir en una serie de funciones que se aplican para obtener la salida deseada. En la figura 4.4 [14] se muestra el esquema de una neurona artificial y sus funciones.

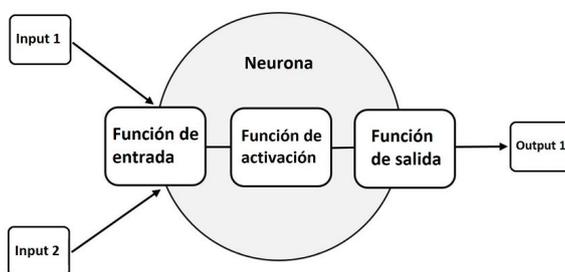


Figura 4.4: Diagrama de bloques de una neurona.

4.4.1. Función de entrada

La función de entrada determina la relación entre las diferentes variables de entrada. Las variables de entrada se multiplican con su peso correspondiente al llegar a la neurona y se relacionan entre ellas de alguna manera. Esta relación puede ser: una suma, una multiplicación, el valor máximo, el valor mínimo, etc. En nuestra red neuronal, la función de entrada consiste en la suma de las diferentes variables de entradas ponderadas por sus pesos, tal y como se muestra en la ecuación 4.2.

$$f(x) = b + \sum_{i=1}^N \omega_i x_i \quad (4.2)$$

4.4.2. Función de activación

Las neuronas del cerebro tienen un estado de activación, es decir, pueden estar activas (excitadas) o no activas (no excitadas). Lo mismo sucede en las neuronas de una red neuronal artificial. Estas neuronas tienen diferentes estados de activación que toman distintos valores dentro de un conjunto determinado. Este conjunto de valores normalmente se encuentra comprendido entre 0 y 1, o entre -1 y 1. Para obtener el valor del estado de activación de una neurona se utiliza la función de activación. Existen diferentes funciones de activación. A continuación, se comentan las más utilizadas.

4.4.2.1. Función de activación log-sigmoidea

Las neuronas pueden tomar valores entre $-\infty$ y $+\infty$. Esta función reescala estos valores para que estén comprendidos entre 0 y 1 siguiendo la ecuación 4.3. Esta función de activación se muestra en la figura 4.5.

$$\text{logsig}(x) = \frac{1}{1 + e^{-x}} \quad (4.3)$$

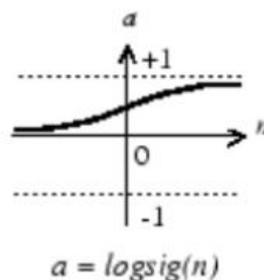


Figura 4.5: Función de activación log-sigmoidea.

4.4.2.2. Función de activación tan-sigmoidea

Una vez aplicada esta función, los valores de las neuronas se encuentran comprendidos entre -1 y 1 siguiendo la ecuación 4.4

$$tansig(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (4.4)$$

Esta función de activación es equivalente a la función tangente hiperbólica, pero la tangente sigmoidea se ejecuta a mayor velocidad, y la variación entre los valores obtenidos con las dos funciones es mínimo. Esta función de activación se muestra en la figura 4.6.

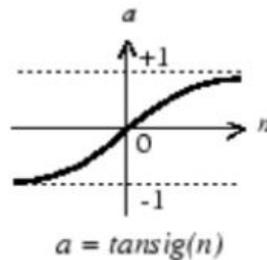


Figura 4.6: Función de activación tan-sigmoidea.

Esta función de activación es la función utilizada en la red neuronal artificial implementada en este proyecto.

4.4.2.3. Función de activación lineal

Esta función de activación se conoce en Matlab como *purelin*, y con ella se obtiene una salida lineal, donde $purelin(x) = x$. Este tipo de función de activación se suele implementar en la capa de salida cuando el problema a resolver es una regresión. Esta función se muestra en la figura 4.7.

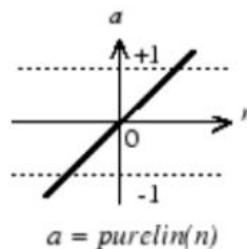


Figura 4.7: Función de activación lineal (*purelin*).

4.4.2.4. Función de activación ReLU

Durante mucho tiempo se ha utilizado la función de activación sigmoidea para las redes neuronales, pero, cuando la red incrementa su número de capas ocultas, esta función hace que la red sea muy difícil de entrenar debido al problema de desaparición de gradiente. En la actualidad, estas redes, en vez de utilizar como función de activación la función sigmoidea, utilizan la función ReLU para solventar este inconveniente. Los valores a la salida de esta función están definidos entre 0 e ∞ . En la figura 4.8 se muestra la función de activación ReLU.

$$ReLU(x) = \max(0, x) \quad (4.5)$$

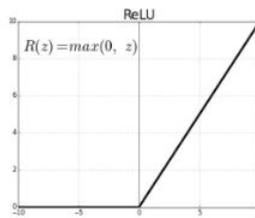


Figura 4.8: Función de activación ReLU.

4.4.3. Función de salida

La función de salida es la función que determina el valor a la salida de la neurona y, por ende, el valor que se transmite a las distintas neuronas de la capa siguiente o, si ya es la última capa, como salida de la red neuronal. Al igual que los valores de entrada no pueden tomar cualquier valor, sino que están comprendidos entre 0 y 1 ó -1 y 1, los valores a la salida de la función de salida también se encuentran comprendidos entre esos valores.

Las dos funciones de salida más comúnmente utilizadas son: la función identidad, que consiste en que el valor a la salida de la función de activación es el valor a la salida de la neurona; y la función binaria, donde el valor de salida es 1 cuando el valor después de la función de activación está por encima de un valor umbral, o la salida toma valor 0 en caso contrario.

4.4.4. Número de neuronas y de capas ocultas

Las redes neuronales están formadas por capas y en cada capa encontramos una cierta cantidad de neuronas. La red neuronal más simple consta de una capa de entrada, una capa oculta y una capa de salida. A esta red se la conoce como red de dos capas debido a que, en el cómputo de capas de una red neuronal, sólo se tienen en cuenta las que tienen verdaderamente entradas; y la capa de entrada realmente no tiene, sino que está constituida propiamente por estas mismas entradas. La capa de entrada está formada por tantas neuronas como variables de entrada. Las capas ocultas tienen tantas neuronas como se quieran asignar, y la capa de salida contiene un número de neuronas igual a las variables de salida que presenta el problema a resolver.

Una red sin capas ocultas hace que las variables de entrada sean independientes entre sí; no existe una relación entre las contribuciones de cada entrada en el valor de salida de la red. La mayoría de problemas de la vida real no funcionan así, y las variables de entrada tienen una importante dependencia entre sí. Tener capas ocultas en la red neuronal nos permite caracterizar esas relaciones sutiles entre las variables de entrada. Por otro lado, disponer de más neuronas en una capa oculta nos permite disponer de mayor capacidad de cálculo.

Por lo tanto, dos hiperparámetros que se deben definir para generar la arquitectura de la red neuronal artificial, y que determinarán el rendimiento de la misma, son las capas ocultas y la cantidad de neuronas que forman parte de cada capa.

La selección de estos dos hiperparámetros se puede realizar de manera empírica entrenando diferentes redes neuronales con distinto número de capas ocultas y distinta cantidad de neuronas, y se comparan los resultados para seleccionar la red que menor error genera para resolver un problema concreto. Este método es el que se ha empleado en este proyecto. El principal inconveniente que presenta es que no se obtiene como resultado la arquitectura de red óptima, pero se pueden obtener buenos resultados empleando poco tiempo de búsqueda.

Otra opción consiste en emplear algún algoritmo de optimización, como pueden ser los propios los algoritmos genéticos. Estos algoritmos son mecanismos de búsqueda estocásticos iterativos basados en la teoría de la selección natural darwiniana. Después de una evolución generacional en el algoritmo se mantienen los mejores genes. En este caso los genes son hiperparámetros de la red neuronal. Este método se utiliza en [15] para encontrar los hiperparámetros óptimos de una red neuronal, utilizada como DPD, para linealizar un PA.

4.4.5. Función de entrenamiento

El proceso de aprendizaje de una red neuronal artificial se conoce como entrenamiento y existen diversas funciones para llevarlo a cabo. En esta sección se explicarán diferentes funciones de entrenamiento, empezando por la más sencilla, el descenso del gradiente, y finalizando con la implementada en nuestra red neuronal, el algoritmo de Levenberg-Marquardt.

Antes de analizar cada función de entrenamiento por separado, empezamos por la formulación del problema de aprendizaje. Este problema, en redes neuronales, se formula para minimizar la función error asociada, que denotaremos como f .

Esta función a minimizar generalmente se encuentra formada por dos términos: el primer término se conoce como término de error, y evalúa cómo de bien se ajustan, a los valores reales proporcionados para el entrenamiento, aquellos valores obtenidos a la salida de la red; y el segundo término se conoce como término de regularización, y se encarga de evitar un sobreajuste a los datos proporcionados.

El valor de la función error depende de los pesos sinápticos que se aplican entre neuronas y los bias de cada neurona de la red neuronal. Estos parámetros se pueden escribir como un vector de parámetros, que llamaremos ω . Como la función error depende de estos parámetros podemos escribir la función como $f(\omega)$. El objetivo de la función de entrenamiento de la red neuronal es encontrar los valores de ω , que corresponden a un mínimo local de la función $f(\omega)$.

Por lo general, la función a minimizar no es una función lineal, por lo que los algoritmos para encontrar estos mínimos no son sencillos ni exactos. Estos algoritmos realizan una búsqueda de los parámetros de forma iterativa para aproximarse a un valor mínimo en el error. En el esquema de la figura 4.9 se representa esta idea para el caso de una sola dimensión (un solo parámetro de peso). Inicialmente, por lo general, el vector de parámetros toma valores al azar, pero, seguidamente se genera un nuevo vector de parámetros

para reducir el valor de la función error. Este proceso se repite hasta obtener un valor de la función error por debajo de un umbral determinado previamente, o hasta cumplir alguna condición específica que detenga este proceso.

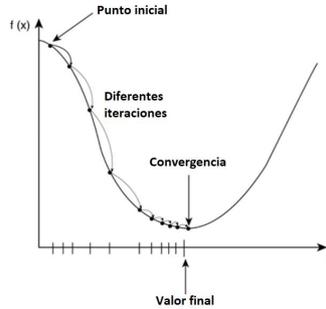


Figura 4.9: Esquema de minimización del error.

Si la función $f(\omega)$ es derivable se pueden calcular la primera y segunda derivada y generar, así, el vector gradiente formado por las derivadas parciales de ésta (4.6), así como la matriz Hessiana, que se compone de las derivadas parciales del vector gradiente (4.7).

$$\nabla f(\omega) = \frac{\partial f}{\partial \omega_i}, i = 1, \dots, n \quad (4.6)$$

$$H_{ij}f(\omega) = \frac{\partial^2 f}{\partial \omega_i \partial \omega_j}, i, j = 1, \dots, n \quad (4.7)$$

En las ecuaciones 4.8 y 4.9 se muestra la forma vectorial y matricial del gradiente y de la matriz Hessiana de $f(\omega)$. Para encontrar un mínimo en la función $f(\omega)$ se tiene que cumplir que el vector gradiente sea un vector nulo y que la matriz Hessiana esté formada por valores positivos.

$$\nabla f(\omega) = \begin{bmatrix} \frac{\partial f}{\partial \omega_1} \\ \frac{\partial f}{\partial \omega_2} \\ \dots \\ \frac{\partial f}{\partial \omega_n} \end{bmatrix} \quad (4.8)$$

$$\nabla^2 f(\omega) = \begin{bmatrix} \frac{\partial^2 f}{\partial \omega_1^2} & \frac{\partial^2 f}{\partial \omega_1 \partial \omega_2} & \dots & \frac{\partial^2 f}{\partial \omega_1 \partial \omega_n} \\ \frac{\partial^2 f}{\partial \omega_2 \partial \omega_1} & \frac{\partial^2 f}{\partial \omega_2^2} & \dots & \frac{\partial^2 f}{\partial \omega_2 \partial \omega_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial \omega_3 \partial \omega_1} & \frac{\partial^2 f}{\partial \omega_3 \partial \omega_2} & \dots & \frac{\partial^2 f}{\partial \omega_3 \partial \omega_n} \end{bmatrix} \quad (4.9)$$

Aunque por lo general el vector de parámetros de la red neuronal es multidimensional, muchos de los algoritmos de entrenamiento utilizados en redes neuronales se basan en una optimización unidimensional. Primero se busca una dirección de entrenamiento, d en la cual se espera un cambio del error máximo, y seguidamente se utiliza un factor de entrenamiento v que indica cómo de grande será el salto en la dirección establecida ($f(\omega + vd)$).

4.4.5.1. Descenso del gradiente

Este algoritmo es el más simple. Sólo utiliza el vector gradiente, es decir, únicamente emplea la primera derivada de la función error, por lo que se conoce como método de primer orden.

En este método, para obtener los valores de los parámetros de la siguiente iteración (ω_{i+1}), se trasladan los valores de la iteración previa (ω_i) en la dirección de entrenamiento d . Esta dirección es la opuesta al gradiente de la función error de la iteración i ($d = -g_i$).

$$\omega_{i+1} = \omega_i - gv \quad (4.10)$$

El factor de entrenamiento, v , puede mantenerse fijo a lo largo del proceso de entrenamiento, o puede calcularse para cada iteración mediante procesos de optimización.

Este algoritmo es muy sencillo pero, si la función error es algo compleja, suele requerir de muchas iteraciones para converger y encontrar los valores óptimos del vector de parámetros.

4.4.5.2. Descenso del gradiente con momento

Este método, además de tener en cuenta el valor de la primera derivada de la iteración actual, tiene en cuenta los valores de la iteración anterior. El gradiente actual se multiplicará por la tasa de aprendizaje (v), y el valor de la actualización anterior por una constante conocida como coeficiente del momento (γ).

$$\omega_{i+1} = \omega_i - gv + \gamma \Delta \omega_{i-1} \quad (4.11)$$

Con este método se obtiene una media móvil del gradiente ponderada exponencialmente, de tal forma que se avanza más rápido cuando se mueve en la dirección correcta, provocando una convergencia más rápida.

4.4.5.3. Método de Newton

Este método es de segundo orden, ya que utiliza las derivadas segundas de la función error. El objetivo del método de Newton es encontrar las mejores direcciones de disminución del error por medio de las derivadas segundas. Para cumplir con este objetivo se utiliza el desarrollo de Taylor de segundo orden de la función $f(\omega)$, tal y como se muestra en la ecuación [4.12](#)

$$f(\omega) = f_0 + g_0(\omega - \omega_0) + \frac{1}{2}(\omega - \omega_0)^2 H_0 \quad (4.12)$$

Donde H_0 es el Hessiano de $f(\omega)$ en este punto. Teniendo en cuenta que el gradiente debe ser cero, para encontrar un mínimo llegamos a la siguiente expresión para el gradiente:

$$g(\omega) = g_0 + H_0(\omega - \omega_0) = 0 \quad (4.13)$$

Finalmente, a partir del vector de parámetros ω_0 , este método permite generar los parámetros de las iteraciones sucesivas con la expresión:

$$\omega_{i+1} = \omega_i - H_i^{-1} g_i \quad (4.14)$$

El problema es que con esta ecuación nos podemos encontrar que el Hessiano es negativo y, en vez de llevarnos a un mínimo, nos lleva a un máximo de la función error. Por esta razón, la ecuación del método de Newton se suele utilizar modificada, introduciéndose una tasa de entrenamiento:

$$\omega_{i+1} = \omega_i - (H_i^{-1} g_i) \nu \quad (4.15)$$

Con este método se obtiene una dirección y una velocidad de entrenamiento. Normalmente este método requiere de menos iteraciones que el descenso del gradiente, pero el cálculo del Hessiano y la inversa conlleva un incremento en el coste computacional del algoritmo.

4.4.5.4. Gradiente conjugado

Este método surge con la idea de acelerar el tiempo de convergencia del método del descenso del gradiente y de reducir el coste computacional del método de Newton.

En este método, la búsqueda del mínimo en la función error se realiza por medio de direcciones conjugadas que permiten una convergencia más rápida. Dos direcciones (u y v) son conjugadas con respecto a una matriz (en este caso H) si $u^T H v = 0$.

En este método las diferentes direcciones se forman por medio de la siguiente expresión:

$$d_{i+1} = g_{i+1} + d_i \gamma_i \quad (4.16)$$

Donde γ se conoce como parámetro conjugado. La dirección de entrenamiento se reestablece al valor de $-g$ para evitar una acumulación del error. Una vez disponemos del valor de la dirección el vector de parámetros se calcula de la forma habitual:

$$\omega_{i+1} = \omega_i + d_i \nu \quad (4.17)$$

4.4.5.5. Método Cuasi-Newton

El problema del método de Newton es que es computacionalmente muy costoso, por tener que calcular la matriz Hessiana y su inversa. Los métodos Cuasi-Newton son métodos en

los que se evita tener que calcular estas dos matrices. En lugar de realizar estos cálculos, se construye directamente una aproximación de la matriz inversa Hessiana a partir de las derivadas primeras de la función error. Considerando que esta matriz aproximada de la matriz inversa, en la iteración i , la denominamos G_i , el cálculo del vector de parámetros se realiza con la siguiente expresión:

$$\omega_{i+1} = \omega_i - (G_i g_i) v \quad (4.18)$$

Este método es altamente utilizado, debido a que presenta una convergencia rápida y requiere menos coste computacional que el método de Newton.

4.4.5.6. Algoritmo de Levenberg-Marquardt

El algoritmo de Levenberg-Marquardt, o método de mínimos cuadrados amortiguado, funciona con funciones de error cuadráticas y no requiere el cálculo de la matriz Hessiana exacta, sino que la calcula de forma aproximada por medio del gradiente y de la matriz Jacobiana.

La función a minimizar es la función error cuadrático, que se calcula con la siguiente expresión:

$$f(\omega) = \sum_{i=0}^m e_i^2(\omega) \quad (4.19)$$

donde m es el número de datos proporcionados para el entrenamiento. La matriz Jacobiana de la función error se calcula a partir de las derivadas parciales de los errores con respecto a los diferentes parámetros del vector ω . Esta matriz es de dimensión $M \times N$. Tiene tantas columnas como parámetros en el vector ω y tantas filas como datos de entrenamiento.

$$Jf(\omega) = \begin{bmatrix} \frac{\partial f_1}{\partial \omega_1} & \frac{\partial f_1}{\partial \omega_2} & \dots & \frac{\partial f_1}{\partial \omega_n} \\ \frac{\partial f_2}{\partial \omega_1} & \frac{\partial f_2}{\partial \omega_2} & \dots & \frac{\partial f_2}{\partial \omega_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_m}{\partial \omega_1} & \frac{\partial f_m}{\partial \omega_2} & \dots & \frac{\partial f_m}{\partial \omega_n} \end{bmatrix} \quad (4.20)$$

El cálculo del vector gradiente de la función error ($\nabla f(\omega)$), denominado previamente como g_i , se realiza a partir del Jacobiano con la ecuación 4.21, donde e es el vector error que tiene dimensión $M \times 1$. La aproximación de la matriz Hessiana (H) también se calcula a partir del Jacobiano con la ecuación 4.22.

$$\nabla f(\omega) = 2J^T e \quad (4.21)$$

$$H \approx 2J^T J + \lambda I \quad (4.22)$$

Donde I es la matriz identidad y λ es el factor de amortiguamiento que se encarga de que el Hessiano sea positivo. Finalmente, el vector de parámetros de la función error se calcula con la siguiente ecuación:

$$\omega_{i+1} = \omega_i - (J_i^T J_i + \lambda_i I)^{-1} (2J_i^T e_i) \quad (4.23)$$

El algoritmo de Levenberg-Marquardt se puede entender como una combinación del método de Newton y del algoritmo de descenso del gradiente. A partir de la ecuación 4.22 se observa que, cuando λ es cero, este método se convierte en el método de Newton con la matriz Hessiana aproximada y, cuando λ es muy grande, este método se convierte en el descenso del gradiente.

Este algoritmo para entrenar redes neuronales es muy rápido, pero sólo se puede utilizar cuando se consideran funciones cuadráticas como función de error. Además, este método no es compatible con términos de regularización y, si la red neuronal es muy compleja, o hay muchos datos de entrenamiento, la matriz Jacobiana es enorme y requiere de mucha memoria.

4.4.6. Datos de entrenamiento, validación y test

A la hora de construir el modelo matemático, que representa la red neuronal, se utilizan tres conjuntos de datos diferentes:

Primero, se utiliza un conjunto de datos denominados datos de entrenamiento. Este primer conjunto consiste en ejemplos utilizados para ajustar los coeficientes de los pesos entre las neuronas de la red neuronal artificial.

Seguidamente, el modelo ajustado se utiliza para predecir las respuestas de las observaciones en un segundo conjunto de datos llamado el conjunto de datos de validación. Este conjunto de datos se utiliza para detectar sobreajuste y aplicar regularización para corregirlo.

Finalmente, los datos de test se utilizan para obtener una evaluación imparcial del modelo y realizar un último ajuste en los valores de los coeficientes de la red.

4.4.7. Tipos de redes neuronales artificiales

4.4.7.1. Red feedforward

La red neuronal feedforward, o red prealimentada, es un tipo de red neuronal donde no aparecen bucles o retroalimentaciones. Este tipo de red neuronal es la más sencilla y fue la primera en aparecer. En este tipo de redes la información sólo se desplaza en una única dirección, hacia adelante. La información pasa de los nodos o neuronas de entrada a las neuronas de las capas ocultas y, de éstas últimas, a las neuronas o nodos de la capa de salida. Esto quiere decir que cada neurona de una capa tiene conexiones dirigidas a las neuronas de la capa siguiente. En la figura 4.10 se muestra un esquema básico con dos capas ocultas.

Estas redes neuronales utilizan alguna técnica de entrenamiento que permite obtener los valores óptimos de los coeficientes o pesos de la red. Una de las técnicas más utilizadas

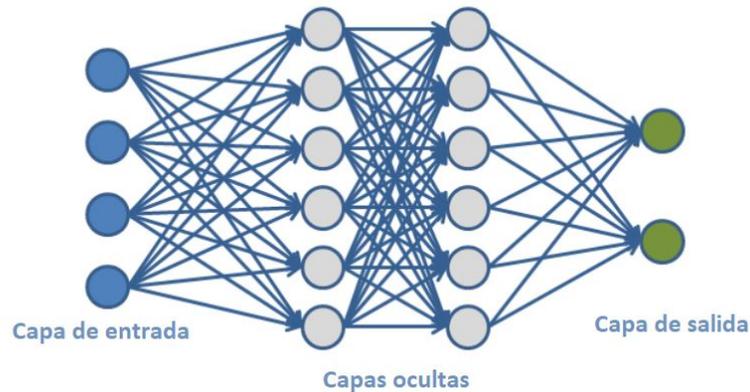


Figura 4.10: Esquema de una red neuronal artificial prealimentada.

es la propagación hacia atrás o *backpropagation*. Con esta técnica, los valores obtenidos a la salida de la red se comparan con los valores esperados para calcular el error, generado por medio de alguna función error. Este dato se propaga hacia atrás, en la red, llegando a todas las neuronas de las capas ocultas. Sin embargo, las neuronas de las diferentes capas ocultas sólo reciben una fracción de la señal error total, basándose en la contribución relativa aportada por esa neurona a la salida generada. Este proceso, en el cual cada neurona sólo recibe una señal error que describe su contribución al error general, hace que las neuronas de las capas intermedias se organicen para aprender a reconocer distintas características del espacio total de datos de entrada.

4.4.7.2. Red en cascada hacia delante

Este tipo de red neuronal se diferencia de la red *feedforward* en que cada capa de la red no sólo recibe los inputs de la capa inmediatamente anterior, sino que también recibe como entrada los datos de todas las capas previas. Los datos de entrada que recibe una neurona de una capa determinada será la suma de todas las neuronas de todas las capas anteriores de la red, teniendo en cuenta la capa de entrada. Este tipo de red conlleva un aumento en la complejidad, debido a que aumentan considerablemente la cantidad de pesos sinápticos a calcular. En la figura 4.11 se muestra el esquema de este tipo de redes neuronales con dos capas ocultas.

4.4.7.3. Red recurrente

Las redes recurrentes se caracterizan por tener bucles de retroalimentación. Esta retroalimentación hace que la información se mantenga durante algunas épocas de entrenamiento. Mientras que una red *feedforward* actúa prácticamente de forma instantánea, una red recurrente posee memoria y, por lo tanto, genera sistemas dinámicos donde su respuesta depende del estado anterior de la red.

Este tipo de redes son muy eficientes para tratar secuencias temporales, pudiendo tratar secuencias muy largas; además, recuerda como entrada los valores obtenidos previamente a la salida. Sin embargo, este tipo de redes son muy costosas en tiempo de entrenamiento y en requerimiento de memoria RAM. Por esta razón, en muchos casos se

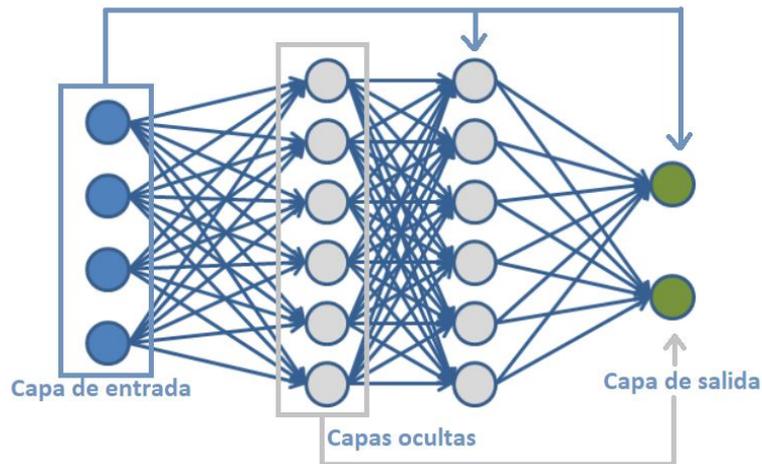


Figura 4.11: Esquema de una red neuronal artificial en cascada hacia delante.

utilizan redes feedforward que presentan tantas capas ocultas como pasos temporales a tener en cuenta. En la figura 4.12 se muestra el esquema de una red neuronal recurrente con retroalimentación de los valores de salida a la entrada de la red neuronal artificial. Este tipo de red neuronal se propone para modelar el comportamiento de un PA en [16].

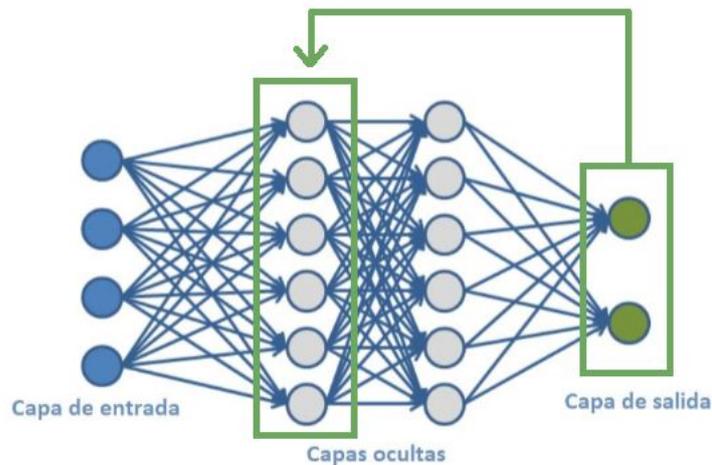


Figura 4.12: Esquema de una red neuronal artificial recurrente.

4.4.7.4. Red convolucional

Este tipo de redes neuronales son muy efectivas para tareas de clasificación y segmentación de imágenes, pero también se puede adaptar para otro tipo de aplicaciones para ser empleada como DPD.

Las redes convolucionales procesan sus capas imitando al córtex visual del ojo humano para identificar objetos, obteniendo así visión artificial.

Primero de todo es necesario organizar los datos de entrada en forma de matriz. Este tipo de redes se caracterizan por aplicar una serie de filtros realizando el producto escalar

entre valores cercanos dentro de la matriz de entrada y unas matrices llamadas kernels. Estos kernels son las neuronas de las redes convolucionales y recorren todas las matrices de entrada (de izquierda-derecha, de arriba-abajo), y genera una nueva matriz de salida que, en definitiva, será la nueva capa de neuronas ocultas. Se obtienen tantas matrices de salida como número de kernels utilizados. Durante el entrenamiento de estas redes los coeficientes que se calculan son los valores de los kernels, es decir, de los filtros que se aplican a la matriz de entrada.

Las redes convolucionales se estructuran encadenando capas convolucionales con capas de reducción. En las capas de reducción se disminuye la cantidad de parámetros seleccionando los más representativos. Finalmente, se colocan capas de red neuronal feedforward para tratar cada píxel con una neurona separada, en vez de utilizar filtros (kernels). En la figura 4.13 se muestra el esquema de red neuronal convolucional aplicada a la linealización de un PA, tal y como se propone en [17].

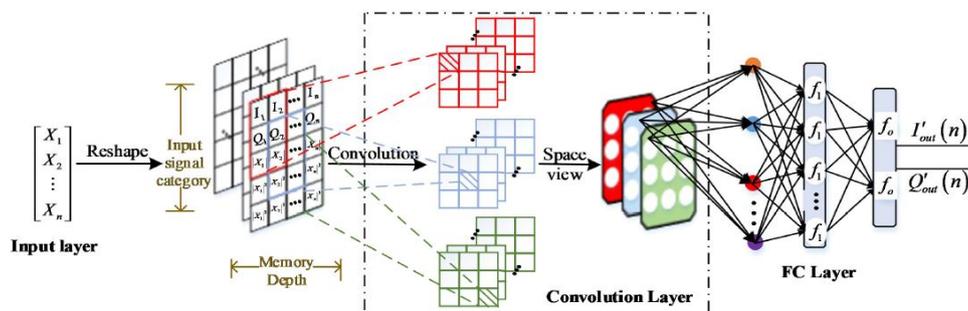


Figura 4.13: Esquema de una red neuronal artificial convolucional.

4.5. Red neuronal como predistorsionador digital

4.5.1. Topología polar y cartesiana

Para poder utilizar una red neuronal como DPD, primero de todo, es necesario seleccionar la topología que seguirá la red neuronal. La señal a linealizar es una señal compleja que tiene parte en fase y parte en cuadratura. Sin embargo, las neuronas de la red neuronal sólo aceptan valores reales.

Una primera opción consiste en obtener como salida la amplitud y el desfase predistorsionados de la señal, pero para ello es necesario utilizar dos redes neuronales diferentes que generalmente no convergen a la vez y, además, pueden llevar a problemas de sobreentrenamiento o subentrenamiento. La segunda opción consiste en obtener a la salida los valores de parte real e imaginaria (fase y cuadratura) de la señal predistorsionada, y de esta manera se pueda utilizar una sola red neuronal que genere dos salidas. Estos dos esquemas posibles se muestran en la figura 4.14. Estas dos opciones se presentan en [18].

Para linealizar el LMBA se va a utilizar una topología cartesiana, así sólo será necesario utilizar una red neuronal que presente a su salida la componente en fase y la componente en cuadratura de la señal predistorsionada. Pero la entrada no sólo estará formada por

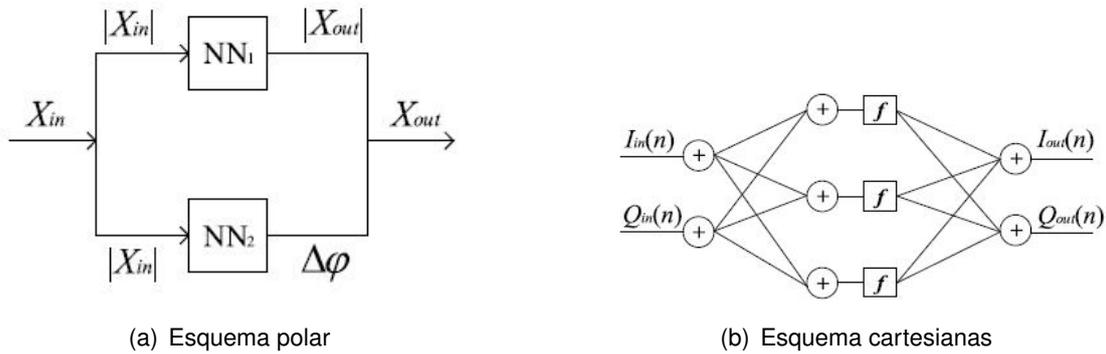
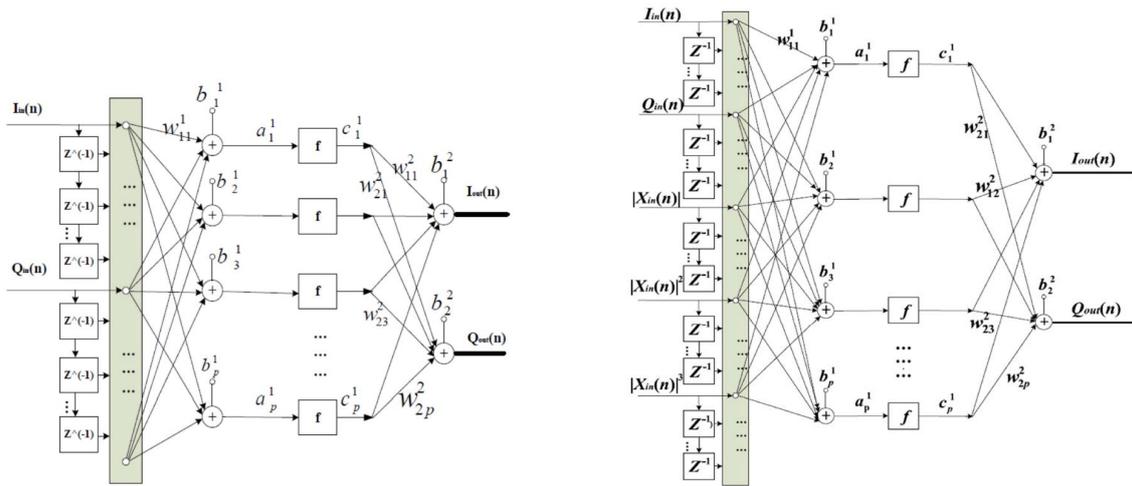


Figura 4.14: Topologías convencionales de redes neuronales para DPD.

la componente en fase y cuadratura de los datos actuales, sino que también se considerarán muestras pasadas de la señal para compensar los posibles efectos de memoria del amplificador.

4.5.2. Variables de entrada a la red neuronal

Para conseguir unos buenos resultados de linealidad, al aplicar un predistorsionador basado en redes neuronales artificiales, es de vital importancia seleccionar unas correctas variables de entrada. Si sólo consideramos como variables de entrada la componente en fase y la componente en cuadratura de la señal actual, no tendremos en cuenta los efectos de memoria del PA, es decir, la dependencia en el comportamiento del PA de muestras anteriores de la señal. Por esta razón, es una buena opción no considerar sólo la componente en fase y cuadratura del momento actual, sino también tener en cuenta muestras pasadas [19]. Además de estas variables de entrada, también se pueden considerar el valor absoluto de la señal con sus retardos y elevados a potencias. Esta selección de variables de entrada se propone en [18]. En la figura 4.15 aparecen los esquemas con las variables de entrada posibles.



(a) Esquema de red neuronal con parte en fase y parte en cuadratura como datos de entrada

(b) Esquema de red neuronal con parte en fase y parte en cuadratura más valor absoluto y sus potencias como datos de entrada

Figura 4.15: Posibles esquemas de variables de entrada de la red neuronal artificial.

4.5.3. Red neuronal artificial adaptativa

Para aplicar la red neuronal como predistorsionador, primero se entrena esta red con datos de la señal a la entrada y a la salida del PA. Una vez entrenada la red ya se puede aplicar como DPD, pero si el comportamiento del PA sufre modificaciones a lo largo del tiempo, es conveniente aplicar una topología adaptativa, donde se vaya reentrenando la red con una serie de datos de la señal recogidos a lo largo del tiempo. Esta implementación adaptativa se puede realizar mediante aprendizaje directo o aprendizaje indirecto.

CAPÍTULO 5. CONFIGURACIÓN DE PARÁMETROS DE LMBA PARA MÁXIMA EFICIENCIA

5.1. Parámetros que intervienen en las prestaciones del LMBA

El LMBA se propuso como una nueva técnica que permite mantener la eficiencia del amplificador (no aumentar el nivel de back off) en aplicaciones que requieren de anchos de banda importantes [7]. Este diseño se basa en un amplificador balanceado (BPA) y en la inyección de una señal de control (CSP) en el puerto aislado de un acoplador de 90° de salida, para modular la carga, tal y como se muestra en la figura 5.1. Una característica distintiva del LMBA es que, en principio, la potencia del CSP siempre se recupera íntegramente a la salida del LMBA; esto quiere decir que la señal auxiliar siempre contribuye a la potencia total a la salida, independientemente de la modulación de carga impuesta. Esta situación es muy diferente a lo que sucede en un amplificador Doherty, donde sólo se recupera toda la potencia de la señal auxiliar cuando las fases de las dos señales (la auxiliar y la principal) se encuentran alineadas. Esto significa que, en un LMBA, la modulación de carga y la recuperación de la potencia de la señal auxiliar son independientes, mientras que en un amplificador Doherty están relacionadas.

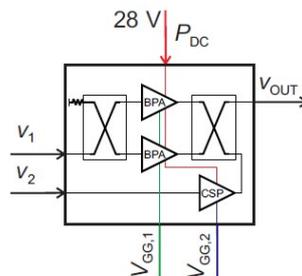


Figura 5.1: Diagrama de bloques del LMBA utilizado en este proyecto.

Otra característica clave del LMBA, que mejora las prestaciones del amplificador Doherty, es el ancho de banda de RF del amplificador. En el caso del LMBA este parámetro viene marcado por el ancho de banda del acoplador de 90° , mientras que, en los amplificadores Doherty, su ancho de banda viene limitado por el del combinador utilizado. Debido a que, por lo general, el ancho de banda permitido en un acoplador es mayor que el permitido en un combinador, el LMBA puede manejar señales de mayor ancho de banda. Sin embargo, una desventaja que presenta el LMBA frente al amplificador Doherty es que el LMBA es intrínsecamente no lineal, por lo que es necesario implementar alguna técnica de linealización. En este proyecto se utiliza como técnica de linealización un predistorsionador basado en redes neuronales artificiales.

Volviendo a la estructura del LMBA que se muestra en la figura 5.1, la señal v_1 controla el par de amplificadores balanceados (BPA). Estos amplificadores se basan en dos transistores CGH40025F de Wolfspeed. Estos amplificadores son de clase AB, con $V_{GG,1}$ a

-2.8 V, correspondiente a 80 mA de corriente de drenaje inactiva. Por otro lado, la señal v_2 controla al amplificador de potencia de la señal de control (CSP), que también se basa en un transistor CGH40025F; pero este amplificador es de clase C, con $V_{GG,2}$ a -5.3V.

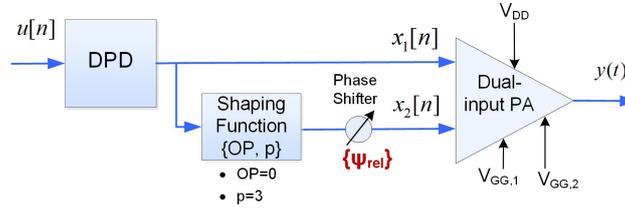


Figura 5.2: Diagrama de bloques del LMBA con implementación de DPD.

Si ahora nos basamos en el esquema mostrado en la figura 5.2, la señal compleja BPA la definimos como $x_1[n]$, mientras que la señal compleja CSP la definimos como $x_2[n]$. La señal $x_1[n]$ es directamente la señal después de aplicar predistorsión ($x_1[n] = x[n]$), y la señal $x_2[n]$ se genera mediante una función de conformación que introduce dos grados de libertad y la aplicación de un desfase con respecto a la señal $x_1[n]$. Para ser más específicos podemos definir $x_2[n]$ como:

$$x_2[n] = x_{sf}[n]e^{j\Psi_{rel}} \quad (5.1)$$

Donde Ψ_{rel} es el desfase entre las dos señales y $x_{sf}[n]$ se define con la siguiente ecuación:

$$x_{sf}[n] = A_s[n]k_0e^{i\phi_x} \quad (5.2)$$

Donde ϕ_x es la fase de la señal principal ($\phi_x = \text{phase}(|x[n]|)$) y $A_s[n]$ y k_0 se definen con las siguientes ecuaciones:

$$k_0 = \frac{\max(|x[n]|)}{\max(|A_s[n]|)} \quad (5.3)$$

$$A_s[n] = ((x_{min})^6 + (|x[n]|)^6)^{\frac{1}{p}} \quad (5.4)$$

Finalmente, x_{min} se define como:

$$x_{min} = \max(|x[n]|)OP \quad (5.5)$$

A partir de las ecuaciones que relacionan las dos señales, $x_1[n]$ y $x_2[n]$, se observa cómo existen 3 grados de libertad. Por un lado, tenemos: el parámetro p , que establece la relación entre los módulos de las amplitudes de las señales; otro parámetro que podemos configurar es el porcentaje de offset, (OP), que también influye sobre la relación de amplitudes; y, por último, tenemos el parámetro Ψ_{rel} , que determina el desfase entre las dos señales. Este capítulo del trabajo se centra principalmente en el efecto del parámetro Ψ_{rel} y el parámetro p sobre el rendimiento del LMBA, prestando mayor atención al desfase entre las señales.

5.2. Setup experimental

Este proyecto se realiza en dos partes experimentales, para las que se va a utilizar el mismo setup y el mismo dispositivo LMBA. La primera parte consiste en analizar el impacto en términos de rendimiento de los parámetros de libre configuración del LMBA, y la segunda parte consiste en, una vez establecidos los parámetros óptimos del LMBA, linealizar el amplificador por medio de predistorsión basada en redes neuronales artificiales. En la figura 5.3 se muestra el equipamiento de laboratorio utilizado. El banco de pruebas se controla por medio de Matlab, de los instrumentos de interfaz de generación de onda (Keysight AWG M8190A) y de la interfaz de adquisición (Keysight DSO 90404A). La señal generada digitalmente se convierte de banda base (procesada con un reloj de 614,4 MSa/s) a 1,8-2,5 GHz de frecuencia RF, y se realiza la conversión de digital a analógico de la señal RF (con una frecuencia de reloj de 7,9872 GHz y 14 bits) para alimentar el LMBA utilizado en este proyecto. La señal de salida del PA se muestrea en RF con un DSO (90404A de Keysight) a 20 GSa/s con resolución de 8 bits. Finalmente, la señal digital se baja a banda base, se alinea en el tiempo y se realiza el procesado DPD.

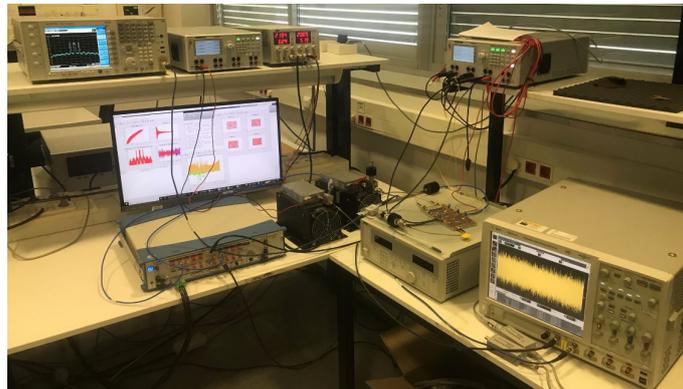


Figura 5.3: Setup experimental empleado para realizar las pruebas sobre el LMBA.

Las dos señales de RF generadas son las entradas al LMBA, mientras que la salida se analiza para recoger los datos de linealidad, eficiencia y potencia. El LMBA utilizado y facilitado por el Prof. Roberto Quaglia de la Universidad de Cardiff [7] se muestra en la figura 5.4.

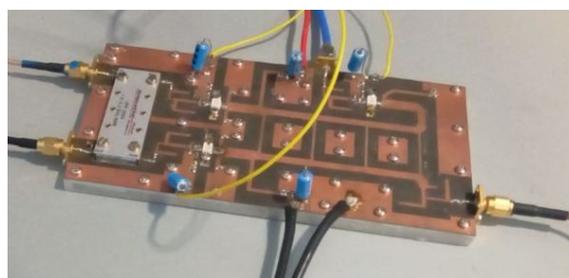


Figura 5.4: LMBA utilizado para realizar la parte experimental.

5.3. Desfase entre señales del LMBA

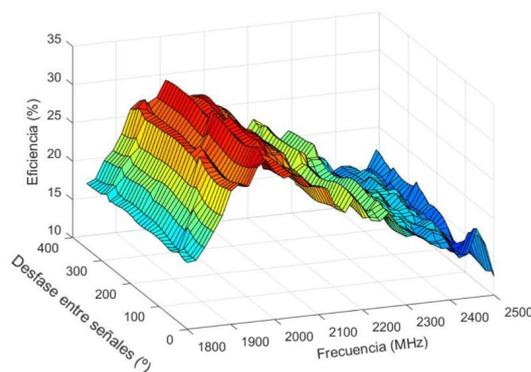
En esta sección se estudia el efecto del desfase entre señales sobre la linealidad y la eficiencia del LMBA. Para ello fijamos el resto de parámetros que intervienen en la configuración de este amplificador. El parámetro p , que relaciona las amplitudes entre las dos señales, lo mantenemos a 3, y el porcentaje de offset de la señal auxiliar (OP) lo fijamos a 0.

Una vez fijados estos parámetros, realizamos dos campañas de medidas con señales LTE de 20 MHz de ancho de banda, en un rango de frecuencias entre 1.8 GHz y 2.5 GHz, y con una resolución de 20 MHz. En cada frecuencia portadora se han tomado medidas de ACPR, NMSE y eficiencia; variando el desfase entre señales, entre 0° y 350° , y con una resolución de 10° .

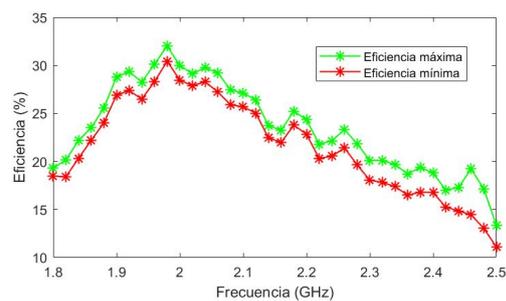
5.3.1. Efecto del desfase sobre la eficiencia

En la figura 5.5(a) se muestran los valores de eficiencia obtenidos para cada desfase y cada frecuencia. Estos valores corresponden a la media de las dos campañas de medidas realizadas. A partir de esta gráfica se observa cómo la eficiencia se ve afectada por la frecuencia, obteniendo los mejores valores de eficiencia alrededor de los 2 GHz de frecuencia portadora. Sin embargo, el desfase no provoca una variación sustancial sobre la eficiencia.

Si analizamos el valor máximo y el valor mínimo obtenido en cada frecuencia, comprobamos como, en el peor caso, la variación de eficiencia entre el peor y mejor caso alcanza el 5 %, pero, en la gran mayoría de frecuencias analizadas, esta variación ronda entre el 1 y 2 %. Estos resultados se observan claramente en la figura 5.5(b).



(a) Eficiencia en función de la frecuencia y el desfase



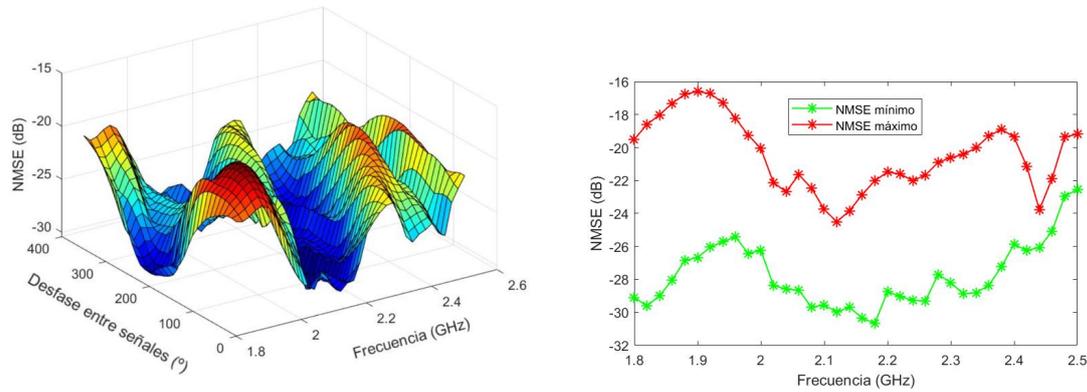
(b) Eficiencia máxima y mínima en función de la frecuencia

Figura 5.5: Valores de eficiencia medios obtenidos en las campañas de medidas.

5.3.2. Efecto del desfase sobre la linealidad

En la figura 5.6(a) se muestran los valores de NMSE obtenidos en cada frecuencia y cada desfase de los analizados. En el caso del NMSE se observa una importante dependencia con el desfase entre señales. Además, también existe dependencia con la frecuencia.

En la figura 5.6(b) se muestra el mejor NMSE (NMSE mínimo) y el peor valor de NMSE (NMSE máximo) a cada frecuencia. En esta gráfica se aprecia cómo el desfase tiene un efecto directo sobre la linealidad y cómo, el escoger el mejor desfase o el peor desfase, puede derivar en una diferencia de hasta 10 dB a determinadas frecuencias (frecuencias entre 1.8 GHz y 2 GHz). A partir de los 2.4 GHz la linealidad empeora, y la diferencia entre el mejor y peor valor de NMSE disminuye mostrando menor efecto del desfase.

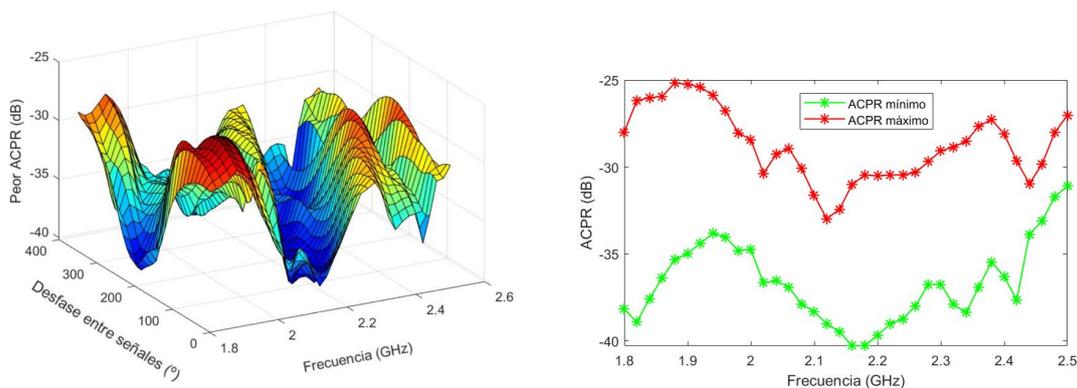


(a) NMSE en función de la frecuencia y el desfase (b) NMSE máximo y mínimo en función de la frecuencia

Figura 5.6: Valores de NMSE medios obtenidos en las campañas de medidas.

A continuación, se analizan los valores de ACPR. Se considera el peor caso de ACPR para cada frecuencia y cada desfase (el valor de ACPR mas alto). En la figura 5.7(a) se muestran estos valores. A partir de esta gráfica se observa cómo existe una dependencia importante del ACPR con el desfase y con la frecuencia. De hecho, el comportamiento es muy similar al del NMSE. Las dos gráficas presentan una tendencia muy semejante.

En el caso del ACPR, la diferencia entre utilizar el peor desfase y utilizar el mejor puede alcanzar los 10 dB, a determinadas frecuencias cercanas a los 2 GHz. Las curvas de mejor y peor ACPR, en función de la frecuencia, siguen la misma tendencia que el NMSE. Estos resultados se muestran en la figura 5.7(b).



(a) Peor caso de ACPR en función del desfase y la frecuencia (b) Valor máximo y mínimo de ACPR en función de la frecuencia

Figura 5.7: Valores de peor ACPR medios obtenidos en las campañas de medidas.

5.3.3. Variación del desfase óptimo con la frecuencia

En la figura 5.8 se muestran los desfases entre señales que optimizan el NMSE, el ACPR y la eficiencia, es decir, el valor de desfase entre señales que proporciona el mejor valor de cada uno de estos 3 parámetros. Estos valores se han obtenido a partir de la media entre las dos campañas de medidas tomadas. El desfase para optimizar el NMSE y el ACPR siguen una tendencia similar y prácticamente lineal para la mayor parte del rango de frecuencias utilizadas. Esta tendencia se pierde a frecuencias altas y bajas. Sin embargo, el desfase para optimizar la eficiencia no sigue el mismo patrón.

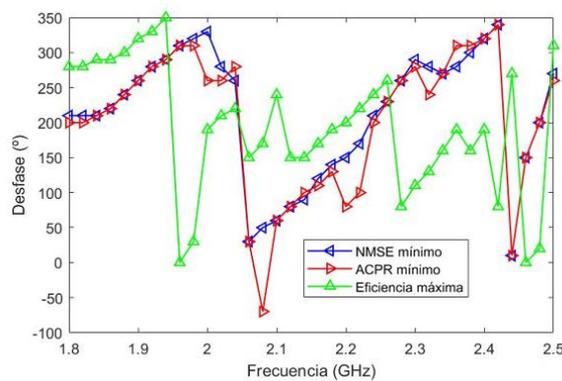


Figura 5.8: Desfases para optimizar NMSE, ACPR y eficiencia.

Teniendo en cuenta que la eficiencia no se ve muy afectada por el desfase, se decide modelar y predecir el desfase óptimo para la linealidad.

5.3.4. Predicción del desfase óptimo para NMSE

Se pretende predecir el desfase óptimo en términos de NMSE para una frecuencia determinada. Para ello se utilizan 3 regresiones polinómicas (de grados 1,3 y 5) y una regresión lineal a trozos. La función a trozos se divide en 3 rangos: 1.8 GHz-1.9 GHz, 1.9 GHz-2.42 GHz y 2.42 GHz-2.5 GHz. En la figura 5.9 se muestra cómo se ajustan estas curvas al desfase óptimo encontrado por medio de las campañas de medidas.

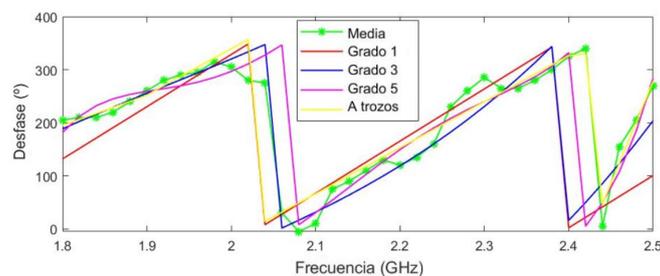
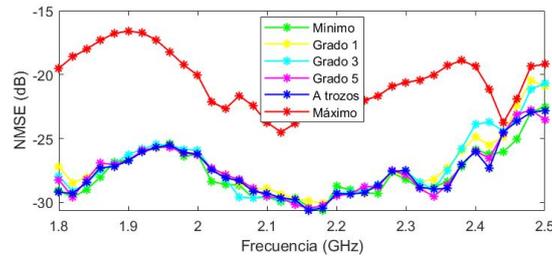


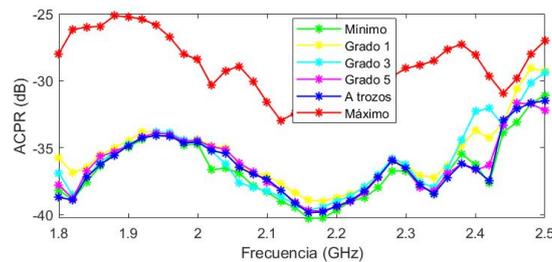
Figura 5.9: Ajuste de las curvas de regresión sobre el desfase para optimizar NMSE.

Con estas curvas de ajuste se obtienen los valores de NMSE y ACPR que se muestran en la figura 5.10. En esta gráfica se comparan estos valores con los valores máximos y mínimos encontrados en las campañas de medidas. Todas las curvas utilizadas muestran

unos valores de linealidad muy próximos a los mínimos para casi todo el rango de frecuencias (de 1.8 GHz a 2.4 GHz). A frecuencias mayores los valores obtenidos dejan de aproximarse a los óptimos. La regresión de grado 5 y la función a trozos son las curvas que mejor se ajustan, considerando también las frecuencias comprendidas entre los 2.4 GHz y 2.5 GHz. Además, como era de esperar, la utilización de regresiones para encontrar el desfase óptimo, en términos de NMSE, también nos permite obtener valores de ACPR muy próximos a los óptimos.



(a) Comparativa de valores de NMSE con regresiones y campaña de medidas



(b) Comparativa de valores de ACPR con regresiones y campaña de medidas

Figura 5.10: Valores de NMSE y ACPR con las diferentes regresiones encontradas.

5.3.5. Efecto del retardo entre señales sobre el desfase óptimo

En este subapartado se analiza la relación del desfase óptimo en función de la frecuencia con un retardo entre señales. Se pretende comprobar si el desfase requerido entre las dos señales se debe simplemente al hecho de compensar un retardo entre las dos señales. El retardo de la portadora se puede expresar con la siguiente ecuación:

$$\tau = \frac{\Psi_{rel}(rad)}{2\pi f} \quad (5.6)$$

A partir de esta ecuación podemos observar cómo un retardo constante para diferentes frecuencias se corresponde a un desfase lineal. Si despejamos el desfase de la ecuación obtenemos que la pendiente de la recta es $2\pi\tau$, y se deduce que a un mayor retardo encontramos una mayor pendiente. Los resultados mostrados previamente, donde se apreciaba una cierta tendencia lineal entre la frecuencia y el desfase óptimo, se ajustan a una posible hipótesis de retardo entre señales. Por esta razón realizamos un análisis del retardo.

A la hora de aplicar un retardo diferenciamos entre: retardar la señal auxiliar respecto a la principal en banda base, y retardarla en radiofrecuencia (una vez la señal se encuentra trasladada en frecuencia). Primero se toman medidas de linealidad aplicando un retardo de -25 a 25 muestras en banda base, a las frecuencias de 1.8 GHz, 1.9 GHz, 2 GHz, 2.2 GHz y 2.4 GHz, y fijando el desfase al valor óptimo encontrado para cada frecuencia. Los valores de NMSE obtenidos se muestran en la figura 5.11.

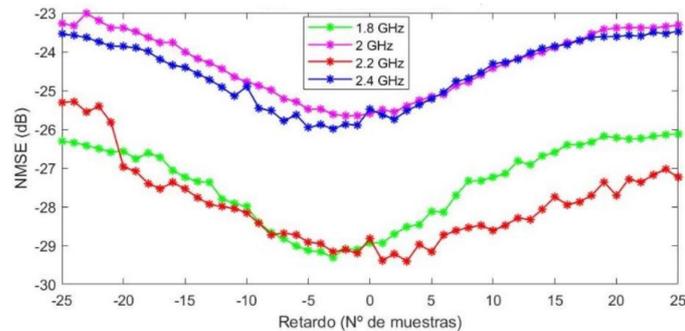


Figura 5.11: Valores de NMSE en función del retardo en banda base para diferentes frecuencias.

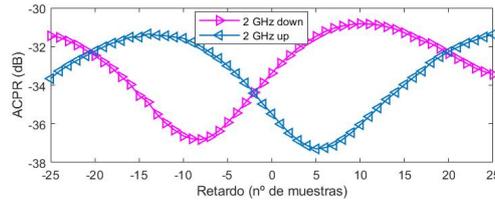
Las conclusiones que podemos obtener a partir de la gráfica son: que con un retardo nulo no obtenemos el mejor valor de NMSE, aunque este valor no difiere en más de 0.5 dB en ninguna de las 4 frecuencias observadas; además, también se puede concluir que el retardo que optimiza el NMSE tiene dependencia con la frecuencia, en donde para cada frecuencia analizada el retardo óptimo no es el mismo (existe una pequeña variación); y, por último, que en la frecuencia de 1.8 GHz la diferencia entre el valor más alto y más bajo de NMSE, en función del retardo, es de 3 dB, mientras que, al analizar el desfase, la diferencia entre el mejor y peor caso a esa frecuencia alcanza los 9 dB. A partir de este resultado podemos intuir que el retardo en banda base no es el causante de estas variaciones importantes de NMSE, entre el desfase óptimo y el peor desfase. Aun así, es importante considerar el ajuste de este retardo para obtener unos resultados mejores.

En la figura 5.12 se muestra el comportamiento del ACPR, en los dos canales adyacentes, en función del retardo para dos frecuencias diferentes (2 GHz y 2.4 GHz). En esta gráfica se aprecia cómo el retardo tiene un efecto importante en el ACPR, provocando un desbalance entre los dos canales. En algunos casos de retardo, la diferencia de ACPR entre los dos canales puede alcanzar los 3 dB. Este desequilibrio hace referencia a los efectos de memoria. Ajustando el retardo en banda base podemos compensar los dos canales. El retardo óptimo con este criterio es, aproximadamente, de -2 muestras.

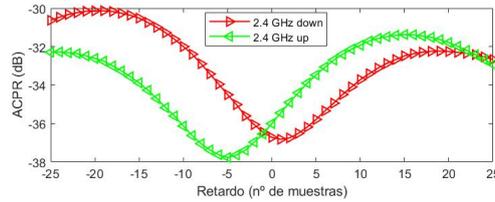
Para comprobar el efecto del retardo en banda base, en el desfase óptimo, realizamos una campaña de medidas en el rango de frecuencias analizado, con los desfases de 0° a 350° y con un retardo de -2 muestras en banda base (retardo que consideramos óptimo para equilibrar el ACPR en los dos canales).

Si se compara el desfase para optimizar NMSE sin retardo y el encontrado con retardo, se observa que son muy semejantes, coincidiendo en gran parte de las frecuencias analizadas (figura 5.13).

Si se analizan los resultados obtenidos (5.14), en el caso de no aplicar retardo, existe diferencia entre los desfases que optimizan el ACPR si se analizan el canal alto y el canal



(a) ACPR para la frecuencia de 2 GHz



(b) ACPR para la frecuencia de 2.4 GHz

Figura 5.12: Valores de ACPR del canal superior e inferior en función del retardo en banda base para la frecuencia de 2 GHz y 2.4 GHz.

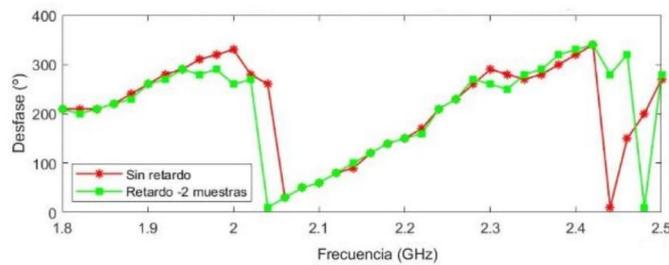


Figura 5.13: Comparativa del desfase óptimo en términos de NMSE con y sin retardo en banda base.

bajo. Sin embargo, al aplicar el retardo, encontramos que los dos canales se optimizan con el mismo desfase. Con este retardo óptimo se mantiene la tendencia lineal con la frecuencia y se optimiza, a la vez, el ACPR de los dos canales adyacentes. Esta tendencia se mantiene hasta los 2.4 GHz. A partir de esta frecuencia el desfase óptimo no sigue un patrón claro.

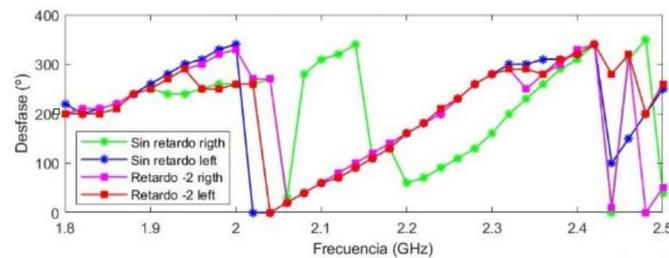


Figura 5.14: Comparativa del desfase óptimo en términos de ACPR para los dos canales adyacentes con y sin retardo en banda base

5.3.6. Predicción del desfase óptimo aplicando retardo en banda base

Con el desfase en banda base de -2 muestras, que consideramos óptimo, realizamos rectas de regresión de grado 1, 3, 5; y una función a trozos, tal y como se hizo en el caso de no aplicar delay. La regresión a trozos se compone de: una primera regresión de grado 1, para los desfases óptimos a las frecuencias entre 1.8 y 2.02 GHz; la segunda regresión de grado 1, para las frecuencias entre 2.04 y 2.42 GHz; y, finalmente, una tercera regresión de grado 1, para las frecuencias entre 2.44 y 2.5 GHz. En la figura 5.15 se muestran los desfases obtenidos en función de la frecuencia, con las distintas regresiones, y cómo se ajustan a la campaña de medidas tomadas con el retardo de -2 muestras en banda base. A partir de 2.4 GHz ninguna curva se ajusta a los resultados obtenidos.

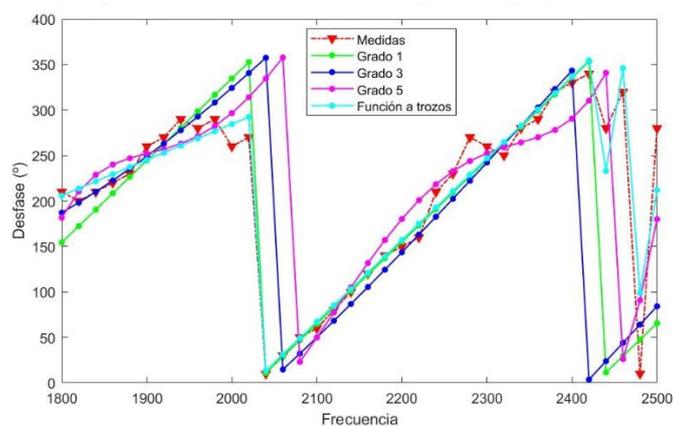


Figura 5.15: Ajuste de las diferentes curvas de regresión a los desfases óptimos obtenidos en la campaña de medida con retardo en banda base.

Aplicando estas regresiones se obtienen los valores de NMSE y ACPR que se muestran en la figura 5.16 y la figura 5.17. A partir de estas imágenes se aprecia que, si sólo consideramos las frecuencias comprendidas entre 1.8 y 2.4 GHz, la curva de regresión que mejores resultados de linealidad nos ofrece es la regresión de grado 1. Se considera que esta regresión es la mejor, porque ofrece valores de linealidad bastante mejores que el resto de regresiones, dentro del rango de frecuencias entre 2 GHz y 2.1 GHz. Esta regresión es la que utilizaremos para validar la predicción del desfase óptimo para otras señales y para aplicar predistorsión digital.

5.3.7. Efecto del desfase entre señales sobre la linealización del amplificador mediante DPD

Previamente se ha demostrado el importante efecto que ejerce el desfase entre señales sobre el ACPR y el NMSE, donde se pueden alcanzar unos 10 dB de diferencia entre utilizar el desfase óptimo y el peor desfase. Sin embargo, no se ha mostrado si existe algún tipo de repercusión en la linealización del LMBA mediante DPD cuando se selecciona un mal desfase entre señales. Para comprobar el efecto del desfase sobre la linealización se implementa DPD sobre una señal de 20 MHz de ancho de banda, a 2 GHz de frecuencia

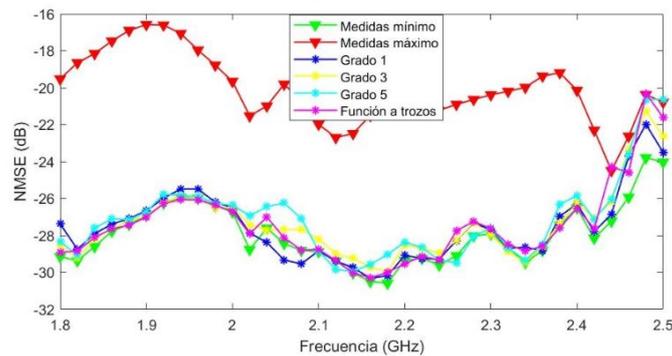


Figura 5.16: Valores de NMSE obtenidos con las diferentes regresiones aplicando un retardo de -2 muestras en banda base.

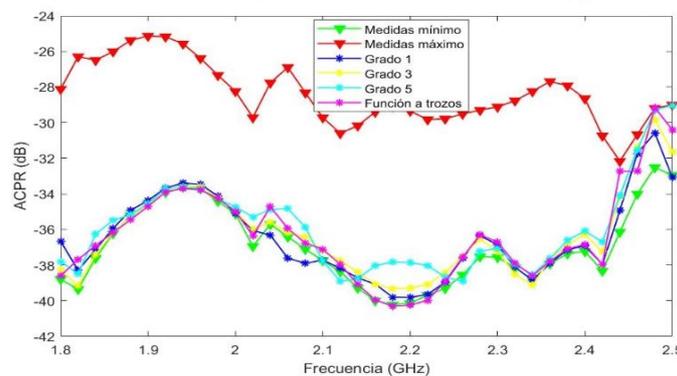
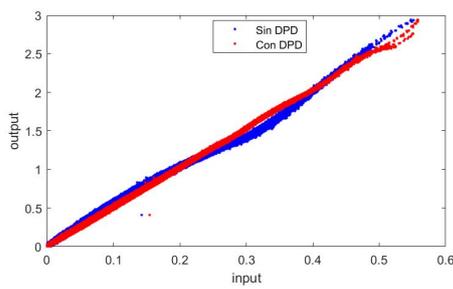


Figura 5.17: Valores de ACPR obtenidos con las diferentes regresiones aplicando un retardo de -2 muestras en banda base.

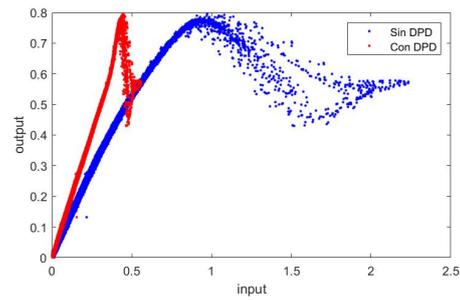
central con el desfase óptimo, en términos de linealidad (260°) y el peor desfase (100°). La predistorsión se realiza en ambos casos por medio de un modelo GMP basado en B-Splines.

En las figuras 5.18(a) y 5.18(b) se muestran las curvas AM-AM de la señal, con y sin DPD, cuando utilizamos el desfase óptimo (260°) y cuando usamos el peor desfase (100°). En estas gráficas se aprecia cómo no es posible linealizar el amplificador cuando seleccionamos el peor desfase, mientras que la linealización sí se realiza cuando del desfase es de 260° . Si analizamos los espectros de la señal a la salida del amplificador (figuras 5.18(c) y 5.18(d)) claramente se observa cómo el espectro de la señal correspondiente al desfase de 100° sufre un importante ensanchamiento, que no se puede mitigar mediante DPD; sin embargo, cuando aplicamos un desfase óptimo esto no sucede y, por lo tanto, se podrá cumplir con el requerimiento de -45 dB de ACPR al aplicar la predistorsión.

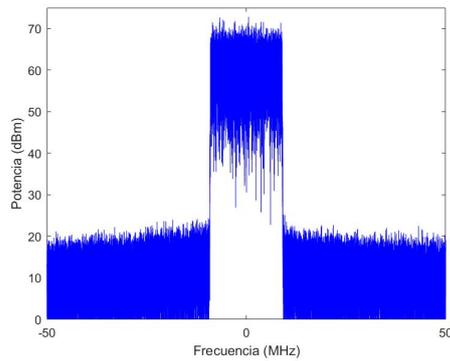
Los resultados de estas mediciones se recogen en la tabla 5.1. En esta tabla se comprueba cómo no es posible alcanzar los niveles de linealidad requeridos por el sistema si se escoge un mal desfase. Pero no sólo los valores de linealidad se ven afectados, sino también los valores de eficiencia y de potencia a la salida del amplificador son inferiores si se hace una mala selección de este parámetro.



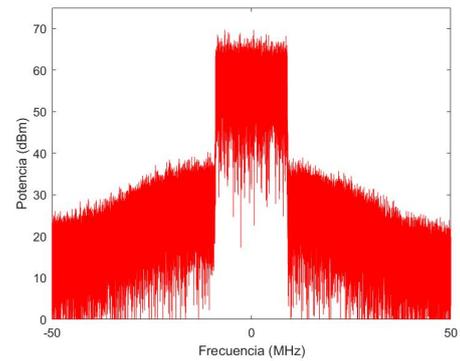
(a) Curva AM-AM cuando linealizamos con desfase óptimo (260°)



(b) Curva AM-AM cuando linealizamos con el peor desfase (100°)



(c) Espectro de la señal linealizada con el mejor desfase (260°)



(d) Espectro de la señal linealizada con el peor desfase (100°)

Figura 5.18: Linealización mediante DPD con el mejor y el peor desfase para la linealidad.

Tabla 5.1: Rendimiento del LMBA después de aplicar DPD cuando se selecciona el mejor y el peor desfase entre señales.

Desfase	NMSE	ACPR	Eficiencia	Potencia media
Óptimo (260°)	-35.0 dB	-47.5 dB	31.2%	36.6 dBm
Peor desfase (100°)	-24.1 dB	-30.3 dB	22.1 %	32.9 dBm

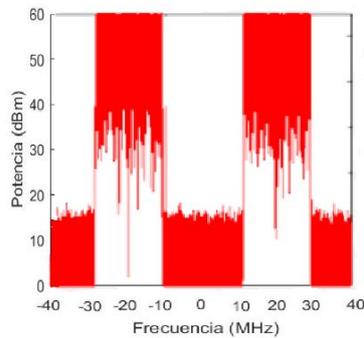
Con estos datos recogidos se comprueba la importancia de una buena selección del desfase entre señales, y cómo este parámetro se vuelve crucial para poder linealizar el LMBA.

5.3.8. Validación con señales diferentes

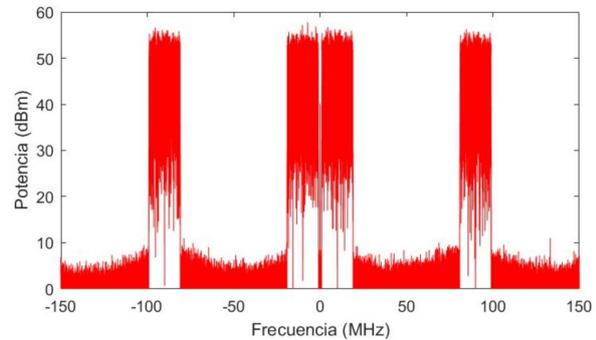
Se ha demostrado que es posible predecir el desfase óptimo entre señales, en términos de linealidad, para una señal LTE de 20 MHz. Ahora se pretende extrapolar esta predicción a otras señales de ancho de banda distinto. Para ello se analizan los resultados obtenidos, en términos de linealidad, con la regresión de grado 1 y la regresión a trozos, sobre una señal LTE de 60 MHz y otra señal LTE de 200 MHz.

Se realiza una campaña de medidas para obtener los valores de eficiencia y linealidad

para una señal LTE de 60 MHz, formada por 2 canales de 20 MHz y un canal vacío entre medio (5.19(a)); y para una señal LTE de 200 MHz, formada por canales de 20 MHz, siguiendo la estructura de señal de la figura 5.19(b).



(a) Espectro frecuencial de la señal LTE de 60 MHz de ancho de banda



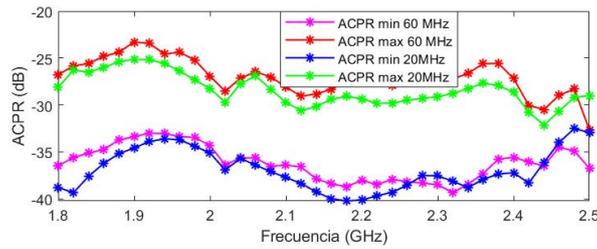
(b) Espectro frecuencial de la señal LTE de 200 MHz de ancho de banda

Figura 5.19: Espectro de las señales utilizadas para la validación de la predicción del desfase óptimo.

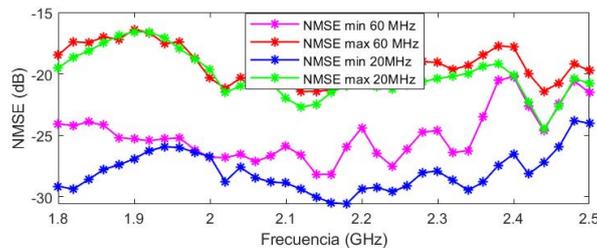
Primero, si se comparan los valores máximos y mínimos de eficiencia y linealidad encontrados en las campañas de medidas, para las señales de 20 MHz y 60 MHz, se observa que tanto el ACPR óptimo como la eficiencia máxima son bastante parecidos para los dos anchos de banda; sin embargo, el NMSE mínimo es mejor para la señal de 20 MHz, alcanzando una diferencia de hasta 5 dB a determinadas frecuencias. Estos resultados aparecen en la figura 5.20.

Ahora se aplican dos de las regresiones encontradas para la señal de 20 MHz, la regresión de grado 1 y la función a trozos. Con estas dos regresiones se realiza la predicción del desfase entre señales para cada frecuencia, y se toman las medidas de linealidad y eficiencia. En la figura 5.21(a) se muestra el desfase óptimo, en términos de ACPR y NMSE, y el desfase con las dos regresiones. Según esta gráfica parece que la función a trozos se adapta mejor que la regresión de grado 1, sobre todo con respecto al ACPR. Sin embargo, ninguna de las dos ofrece unos buenos resultados de desfase a partir de 2.4 GHz.

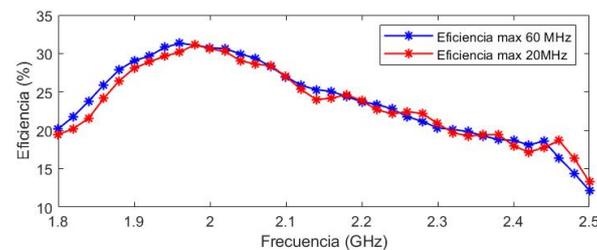
En la figura 5.21(b) aparecen los valores de NMSE y ACPR obtenidos para las diferentes frecuencias, utilizando la predicción del desfase óptimo por medio de las regresiones encontradas. También aparecen los valores máximos y mínimos recogidos durante la campaña de medidas para la señal de 60 MHz. Los valores de linealidad se aproximan a los valores óptimos para las dos regresiones, sólo si se consideran aquellas frecuencias inferiores a los 2.4 GHz. A partir de esta frecuencia los valores dejan de ser óptimos y se aproximan a los valores máximos.



(a) Comparativa de ACPR entre las señales de 20 MHz y 60 MHz

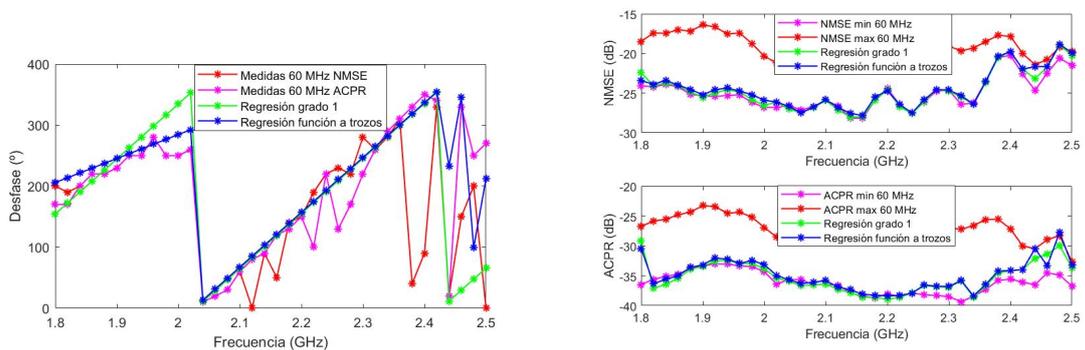


(b) Comparativa de NMSE entre las señales de 20 MHz y 60 MHz



(c) Comparativa de eficiencia entre las señales de 20 MHz y 60 MHz

Figura 5.20: Valores de ACPR, NMSE y eficiencia obtenidos con las campañas de medidas de la señal de 20 MHz y la señal de 60 MHz.

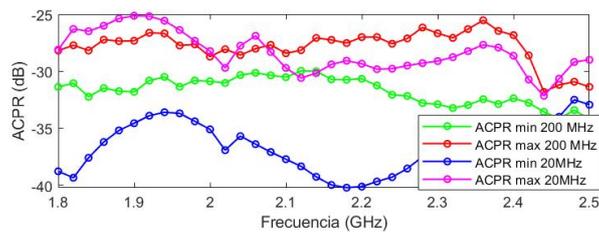


(a) Desfase óptimo y regresiones para señal de 60 MHz (b) Valores de ACPR y NMSE con las regresiones

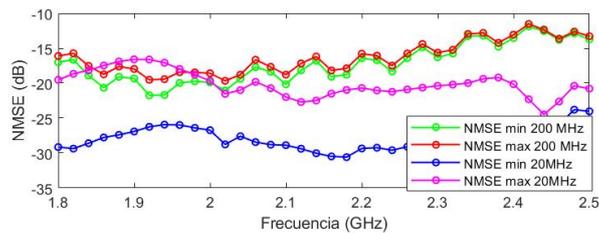
Figura 5.21: Validación de la predicción de desfase óptimo para la señal LTE de 60 MHz

Exactamente igual que se ha procedido con la señal LTE de 60 MHz, se comparan los valores máximos y mínimos de linealidad y eficiencia encontrados para las señales de 20 MHz y 200 MHz. Los resultados obtenidos se muestran en la figura 5.22. Tanto en el

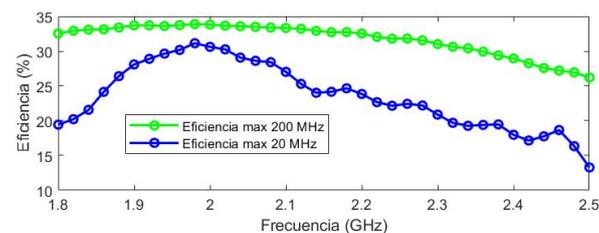
ACPR como en el NMSE los valores máximos y mínimos encontrados para la señal de 200 MHz se encuentran mucho más próximos que en el caso de la señal de 20 MHz. No existe tanta diferencia entre el mejor y el peor caso. De hecho, el NMSE apenas muestra variación con el desfase. El ACPR se encuentra más afectado por este parámetro, pero, como mucho, la diferencia es de 7 dB entre el mejor y el peor valor para frecuencias cercanas a los 2.34 GHz. Además, es importante comentar que los valores de linealidad óptimos se encuentran bastante alejados de los encontrados para la señal de 20 MHz. En cuanto a la eficiencia, la señal de 200 MHz muestra mejores valores máximos que la señal de 20 MHz, y esta eficiencia máxima muestra un comportamiento más plano con la frecuencia que la señal de 20 MHz.



(a) Comparativa de ACPR entre las señales de 20 MHz y 200 MHz



(b) Comparativa de NMSE entre las señales de 20 MHz y 200 MHz

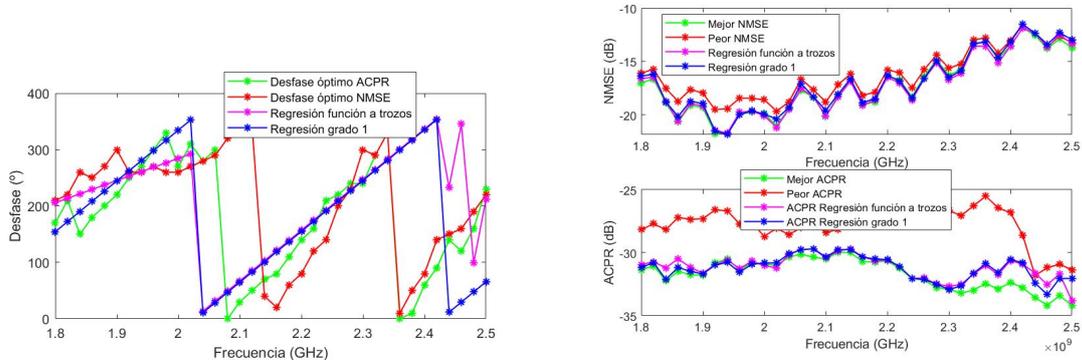


(c) Comparativa de eficiencia entre las señales de 20 MHz y 200 MHz

Figura 5.22: Valores de ACPR, NMSE y eficiencia obtenidos con las campañas de medidas de las señales de 20 MHz y 200 MHz

En la figura 5.23(a) se muestra cómo se ajustan las dos regresiones a los desfases óptimos, en términos de linealidad de la señal de 200 MHz; mientras que, en la figura 5.23(b), se muestran los valores de ACPR y NMSE con los desfases obtenidos mediante la utilización de las dos regresiones, y la comparativa con los valores máximos y mínimos registrados con la campaña de medidas. Mediante la gráfica de NMSE (5.23(b)) no se pueden sacar conclusiones claras, porque los valores máximos y mínimos son muy cercanos. Sin embargo, si nos fijamos en los valores de ACPR, se observa cómo las dos regresiones se ajustan perfectamente a los valores óptimos encontrados para un rango entre 1.8 GHz

y 2.3 GHz. En el rango comprendido entre 2.3 GHz y 2.4 GHz, los valores de ACPR no coinciden con los óptimos, pero se encuentran alejados de los peores valores. A partir de 2.4 GHz los valores de ACPR se encuentran cercanos a los valores máximos, tal y como se ve en las anteriores pruebas realizadas con señales de diferente ancho de banda.



(a) Desfase óptimo y desfase con regresiones para la señal de 200 MHz (b) Valores de ACPR y NMSE obtenidos con las regresiones

Figura 5.23: Validación de la predicción de desfase óptimo para la señal LTE de 200 MHz.

A partir de las medidas tomadas para las señales de 60 MHz y 200 MHz, se puede considerar que las regresiones encontradas para la señal de 20 MHz son válidas para señales con un ancho de banda mayor, en el rango comprendido entre 1.8 GHz y 2.4 GHz de frecuencia central. Más allá de esta frecuencia los resultados no son válidos, pero, en general, el amplificador no presenta buenas prestaciones a estas frecuencias; la eficiencia se ve muy degradada, llegando a mostrar valores inferiores al 15 % a 2.5 GHz, para las señales de 20 MHz y 60 MHz. Este LMBA muestra su mejor punto de trabajo a frecuencias cercanas a los 2 GHz, que es a la frecuencia que muestra un máximo de eficiencia.

5.4. Efecto de la relación entre amplitudes sobre el desfase óptimo

También es necesario realizar un análisis de cómo afecta la relación entre amplitudes a partir del parámetro p , tanto en la eficiencia como en la linealidad; y comprobar si este parámetro modifica el valor del desfase óptimo en términos de linealidad ya que, en el análisis del desfase entre señales, se ha establecido un valor de p fijo, considerando que la relación entre amplitudes es independiente del desfase entre señales.

Tal y como se ha comentado previamente, el parámetro p se encarga de establecer la relación entre los valores absolutos de la amplitud de las dos señales del LMBA. Esta relación se muestra en la siguiente ecuación:

$$A_s[n] = ((x_{min})^6 + (|x[n]|)^6)^{\frac{1}{p}} \tag{5.7}$$

donde $A_s[n]$ es el valor absoluto de la señal $x_2[n]$. Como el valor de porcentaje de offset lo hemos fijado a 0, x_{min} queda definido como:

$$x_{min} = \max(|x[n]|)OP = 0 \quad (5.8)$$

Y, por lo tanto, podemos reescribir la ecuación que determina la relación entre amplitudes como:

$$A_s[n] = (|x[n]|)^{\frac{6}{p}} \quad (5.9)$$

A partir de la ecuación 5.9, si p toma un valor 3, por ejemplo, el valor absoluto de la amplitud de la señal auxiliar ($x_2[n]$) será el cuadrado del valor absoluto de la amplitud de la señal principal ($x_1[n]$). Si ahora p toma el valor de 6, los dos valores absolutos de la amplitud serán iguales.

Para el análisis del efecto de p , sobre la linealidad y la eficiencia, se realiza una campaña de medidas a 2 GHz con el desfase óptimo entre señales encontrado previamente (260°) para distintos valores de p . Los valores de p utilizados van entre 1 y 10, con una resolución de 0.5.

La figura 5.24 muestra los valores de eficiencia, NMSE y ACPR en función del parámetro p . A partir de la imagen se extrae que la eficiencia empieza a decaer para valores de p mayores a 4. Este parámetro muestra un importante efecto sobre la eficiencia, mayor que el desfase entre señales, alcanzando una diferencia del 7% entre $p = 4$ y $p = 10$. Sin embargo, el parámetro p no presenta un importante impacto en el NMSE y en el ACPR como el desfase entre señales, aunque sí que muestra un mínimo en la linealidad para $p = 5$.

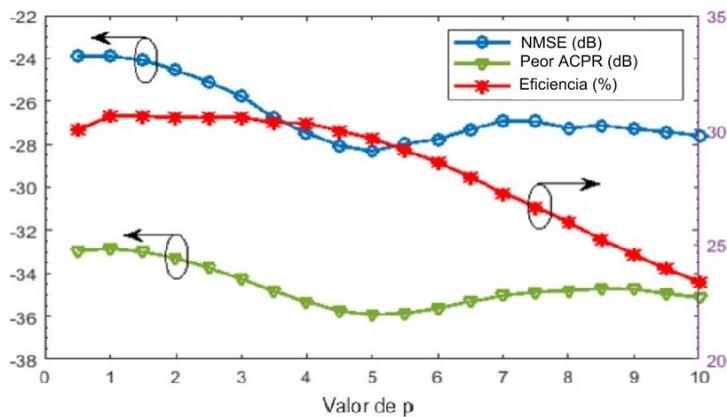


Figura 5.24: Valores de Eficiencia y linealidad en función de p para una señal a 2 GHz.

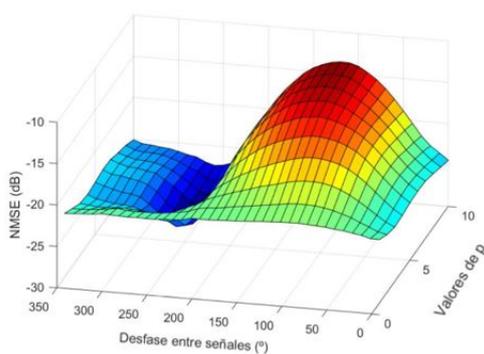
El análisis anterior del efecto de p sobre la linealidad y la eficiencia se ha realizado considerando que el desfase óptimo no varía con la relación entre amplitudes (parámetro p), pero no se ha demostrado la independencia de estos dos parámetros. Para comprobar si el desfase óptimo sufre alguna modificación al modificar el valor de p , se realiza una nueva campaña de medidas para valores de p entre 1 y 10, en una señal a 2 GHz. Para cada p se recogen los valores de NMSE, ACPR, eficiencia y potencia media, para desfases entre 0 y 350 grados, con una resolución de 10 grados.

En las figuras 5.25(b) y 5.25(a) se muestran los valores de linealidad en función del desfase entre señales y el valor de p (relación entre amplitudes). Tanto el ACPR como el NMSE

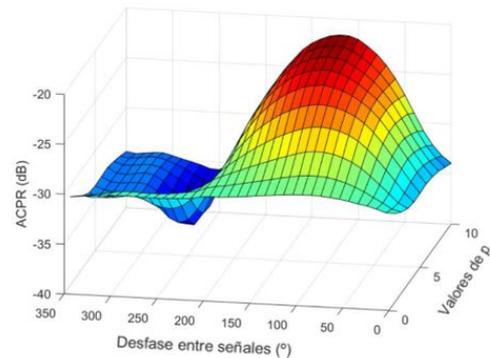
siguen una tendencia prácticamente idéntica, y se observa cómo el valor de desfase óptimo y el peor desfase apenas sufren modificaciones en función del valor de p . A partir de estas gráficas también se aprecia cómo, a valores pequeños de p , la diferencia entre el peor y mejor valor de linealidad es insignificante; mientras que, a medida que aumentamos p , esta diferencia se incrementa.

En la figura 5.25(c) se muestran el desfase óptimo y el peor desfase entre señales, en términos de linealidad. El desfase óptimo para NMSE y ACPR son prácticamente el mismo, y este valor se mantiene constante a partir de $p = 4$ (260°).

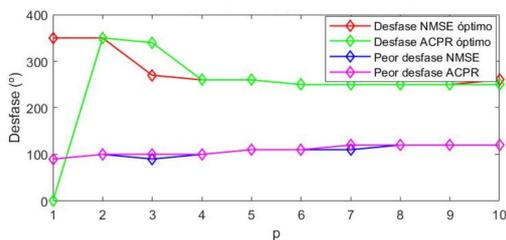
En la figura 5.25(d) se muestran los valores de ACPR y NMSE máximos, mínimos y los obtenidos si dejamos el desfase fijo a 260° (valor óptimo encontrado para $p = 3$). En la gráfica se observa cómo, a valores pequeños de p , igual que se observa en las figuras de superficie, la diferencia entre el mejor y peor caso es muy pequeño (valores de p en los cuales el desfase, en función de p , no se mantiene constante). Para p igual a 1 apenas hay diferencia, y para p igual a 3 la variación de ACPR y NMSE ronda los 5 dB; mientras que, si aumentamos p (por ejemplo, para p igual a 6) esta diferencia puede alcanzar los 15 dB. Además, los valores de NMSE y ACPR obtenidos a 260° de desfase se aproximan a los valores óptimos para todo el rango de p analizado, por lo que es posible considerar que los parámetros de desfase y p son independientes a la frecuencia de 2 GHz.



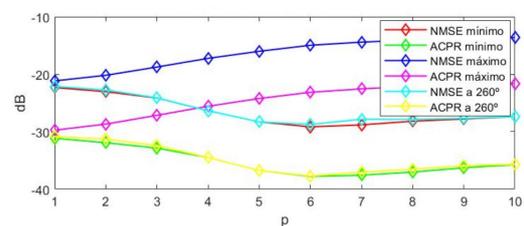
(a) Valores de NMSE en función del desfase y del valor de p



(b) Valores de ACPR en función del desfase y del valor de p



(c) Desfase óptimo y peor desfase para la linealidad en función de p

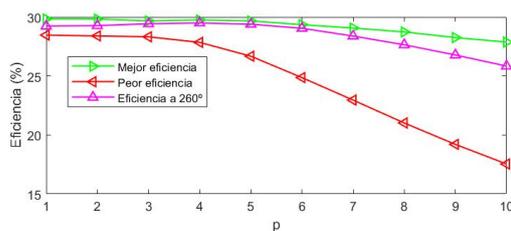


(d) Valores máximos y mínimos de ACPR, NMSE y con el desfase de 260°

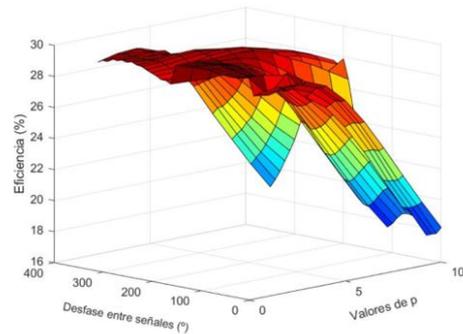
Figura 5.25: Efecto de la relación entre amplitudes sobre el desfase óptimo y la linealidad a 2 GHz.

Con los valores de eficiencia encontrados con el desfase de 260° nos aproximamos a los valores óptimos de eficiencia encontrados para todo el rango de p , tal y como se muestra en la figura 5.26(a). Al analizar estos resultados se comprueba que es cierto que la efi-

ciencia no se ve fuertemente afectada por el desfase, aunque esto sólo se cumple para valores de p pequeños. A medida que aumenta p , la diferencia entre la mejor eficiencia (mejor desfase) y la peor eficiencia (peor desfase) aumenta, alcanzando valores de hasta un 10% para p igual a 10. Al igual que pasaba en el caso de la linealidad, el valor de desfase para optimizar la eficiencia se mantiene prácticamente constante a medida que se modifica el valor de p (5.26(b)).



(a) Valores de eficiencia máximos, mínimos y a 260° de desfase en función de p



(b) Valores de eficiencia en función del desfase y del valor de p

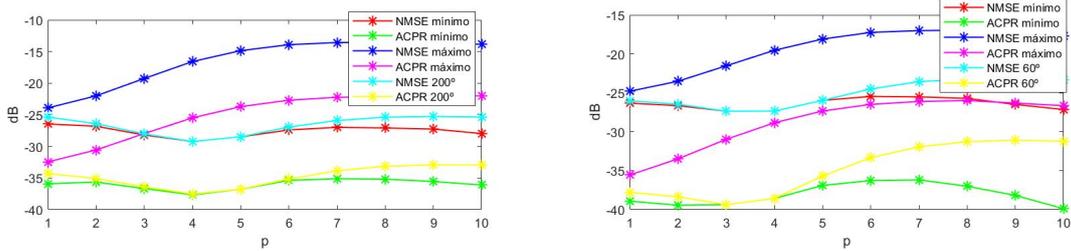
Figura 5.26: Efecto de la relación entre amplitudes sobre la eficiencia a 2 GHz.

Aunque se puede considerar que la relación entre amplitudes y el desfase entre señales son independientes a 2 GHz, esto no quiere decir que esto sea cierto en todo el rango de frecuencias analizado. Por ello se realizan medidas del efecto de p sobre la linealidad a 1.8 GHz y 2.2 GHz de frecuencia central.

En la figura 5.27(a) se muestran los valores de NMSE y ACPR mejores y peores, y los valores de linealidad cuando el desfase es el óptimo encontrado para esta frecuencia, con el valor de p fijado a 3 (200°). Con los resultados obtenidos, el desfase de 200° se ajusta al valor óptimo hasta $p = 7$; para valores de p superiores, el valor de NMSE y ACPR encontrados se acerca mucho más al valor óptimo que al peor valor, aunque se aleja cada vez más de los valores mínimos.

En la figura 5.27(b) se muestran los valores de NMSE y ACPR mejores y peores, y los valores de linealidad cuando el desfase es el óptimo encontrado para esta frecuencia con el valor de p fijado a 3 (60°). En este caso, el valor obtenido a 60° se ajusta al valor óptimo, pero sólo hasta $p = 7$; para valores de p superiores, los valores de NMSE y ACPR encontrados a 60° se alejan considerablemente del valor óptimo, alcanzando una diferencia de 9 dB en el peor caso de ACPR, entre el valor óptimo y el encontrado a 60°, para un valor de p de 10.

Estos resultados muestran que la independencia entre el desfase y p depende de la frecuencia. Para poder predecir un valor óptimo de desfase en términos de linealidad, sin depender del valor de p , se hace necesario reducir el rango de p utilizado. Como a partir de $p = 5$ se muestra una pendiente de descenso en la eficiencia, sólo se consideran valores de p entre 3 y 5. De esta manera, es posible considerar cierta independencia entre los parámetros, y reducir así el rango de búsqueda del valor óptimo de p .



(a) Valores de linealidad máximos, mínimos y con el desfase óptimo para $p = 3$ a 1.8 GHz (b) Valores de linealidad máximos, mínimos y con el desfase óptimo para $p = 3$ a 2.2 GHz

Figura 5.27: Valores de linealidad máximos, mínimos y con el desfase óptimo para $p = 3$ a diferentes frecuencias.

5.5. Efecto de la tensión de alimentación del amplificador auxiliar sobre la linealidad y la eficiencia

Otro parámetro a tener en cuenta, que modifica las prestaciones del LMBA, es la tensión de alimentación V_{gs} del amplificador de la señal auxiliar (CSP). Para comprobar el efecto de V_{gs} se realizan medidas de linealidad y eficiencia, para una señal LTE de 20 MHz de ancho de banda a una frecuencia de 2 GHz, variando este parámetro. El rango de valores de V_{gs} considerado es entre 3.5 Voltios y 5.5 Voltios, con una resolución de 0.1 Voltios. Los resultados de estas mediciones se muestran en la figura 5.28. Los valores de linealidad (ACPR y NMSE) varían aproximadamente 3 dB entre el mejor y peor caso, para el rango de V_{gs} considerado. Los mejores valores de linealidad se muestran a tensiones pequeñas, que es cuando el amplificador se comporta como un amplificador de clase AB. A medida que aumenta la tensión, el amplificador pasa a comportarse como un amplificador de clase C y pierde linealidad. La eficiencia muestra una variación del 0.5 %, entre el mejor y peor caso, mostrando sus valores máximos en torno a los 4-4.5 Voltios de tensión alimentación del CSP.

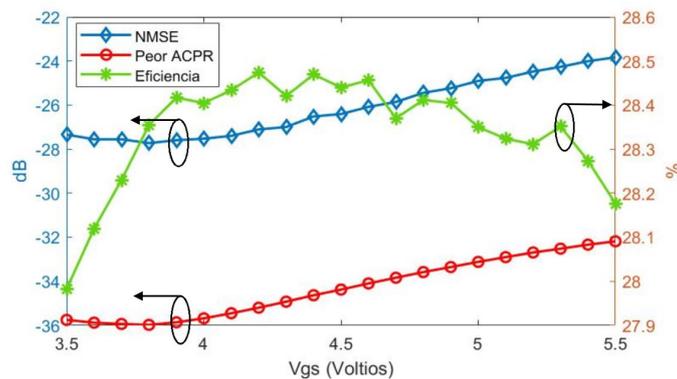


Figura 5.28: Valores de linealidad y eficiencia para una señal LTE de 20 MHz de ancho de banda a una frecuencia de 2 GHz en función de la tensión de alimentación.

Los valores de linealidad y eficiencia en función de la tensión de alimentación pueden verse afectados por el ancho de banda de la señal y la frecuencia, por lo que el valor óptimo de V_{gs} puede verse modificado dependiendo de la señal transmitida. Sin embargo, se

puede considerar que este valor óptimo se encontrará en un rango pequeño comprendido entre 3.5 Voltios y 5 Voltios.

5.6. Criterios de selección de los parámetros del LMBA

Una vez analizados los parámetros que intervienen en las prestaciones del LMBA, se sugieren una serie de criterios o pasos a seguir para realizar una correcta configuración de este amplificador.

1. Realizar un barrido de retardos, en banda base, que permita encontrar el retardo óptimo para balancear los valores de ACPR de los distintos canales adyacentes.
2. Una vez establecido este retardo, realizar una campaña de medidas para obtener los desfases que optimizan la linealidad, y generar una recta de regresión de grado 1, que permita predecir el desfase óptimo para diferentes señales a diferentes frecuencias.
3. Con el retardo en banda base y la predicción de desfase óptimo, lo ideal es realizar un ajuste fino que permita encontrar el valor de p óptimo dentro del rango $p = 3$ y $p = 5$, y el valor de V_{gs} óptimo en el rango $V_{gs} = 3,5$ y $V_{gs} = 5$, para cada ancho de banda y frecuencia central de la señal.

Con estos pasos se logra una configuración óptima del LMBA, que permitirá obtener las mejores prestaciones de este amplificador después de implementar algún tipo de linealización por medio de DPD.

CAPÍTULO 6. IMPLEMENTACIÓN DE ANN COMO DPD SOBRE LMBA

6.1. Introducción

Esta sección corresponde a la parte final experimental, en la que se selecciona una red neuronal artificial que se entrena y se aplica como predistorsionador, para conseguir una respuesta lineal del LMBA que permita cumplir con los requerimientos de 5G. Además, los valores de linealidad y eficiencia obtenidos con la red neuronal se van a comparar con los valores obtenidos por medio de otro método de DPD, como es el GMP, para una señal LTE con tres anchos de banda diferentes (20 MHz, 60 MHz y 200 MHz).

La red neuronal artificial implementada se ha generado en Matlab, y para seleccionar su topología se han utilizado las muestras de entrada y salida de una señal LTE de 5 MHz de ancho de banda de un amplificador de potencia disponible en el laboratorio virtual de la Universidad de Chalmers [20].

El setup de medición de este laboratorio virtual se basa en un chasis PXI (PXIe-1082) con un PC host integrado de National Instruments. El chasis viene equipado con un transceptor de señal vectorial (PXIe- 5646R VST) con ancho de banda instantáneo de 200 MHz. La señal generada desde el transmisor (frecuencia central 2,14 GHz) alimenta a un amplificador lineal antes del GaN PA del setup (CGH40006-TB, placa de prueba para el transistor CGH40006P). Se coloca un atenuador de RF de 30 dB en la salida del PA desde donde se conecta la señal de salida al receptor VST. Se utiliza un PC integrado en el chasis PXI para controlar los instrumentos y para descargar y cargar archivos de datos a petición de los usuarios. La fuente de alimentación (PXI-4130) suministra energía al PA y mide el consumo de corriente del amplificador de potencia. En la figura 6.1 se muestra el esquema de este setup experimental.

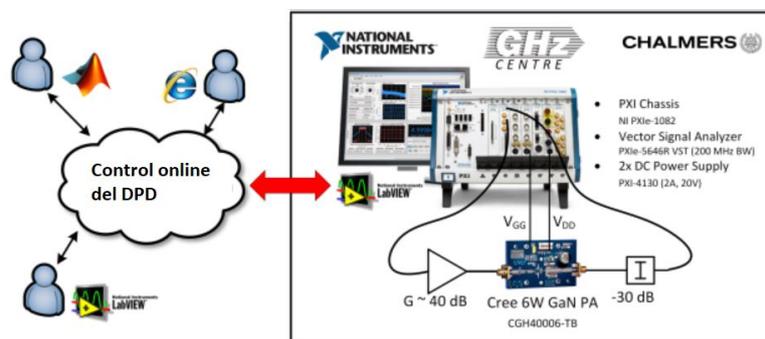


Figura 6.1: Esquema del RF WebLab de la Universidad de Chalmers

Con los datos de este amplificador se han seleccionado las variables de entrada, los hiperparámetros de la red (número de capas ocultas y neuronas por capa, función de activación y entrenamiento) y el tipo de red neuronal artificial. La selección de cada uno de estos parámetros se ha realizado en base al NMSE obtenido al modelar el comportamiento del PA con las diferentes redes neuronales analizadas. Una vez seleccionada la

mejor opción de red neuronal, ésta se ha entrenado con los datos del LMBA del setup del laboratorio.

6.2. Redes neuronales artificiales en Matlab

Matlab es un lenguaje de alto nivel pensado para realizar cálculos complejos de forma automatizada. En Matlab el cálculo, la visualización y la programación se integran en un entorno sencillo de utilizar. Su nombre, Matlab, significa laboratorio de matrices, por lo que este lenguaje de programación está pensado para operar con vectores y matrices de manera fácil, permitiendo resolver este tipo de problemas en un tiempo mucho menor que otros lenguajes de programación.

Matlab se utiliza para diferentes propósitos como: cálculos numéricos, desarrollo de algoritmos, modelado de prototipos, análisis de datos y obtención de gráficos y desarrollo de aplicaciones. Una ventaja fundamental de Matlab es el uso de las "toolboxes". Estas "toolboxes" son grupos de comandos en Matlab que extienden el ambiente de Matlab para solventar problemas específicos en diferentes áreas de la ciencia y la tecnología. Algunas de las "toolboxes" más importantes son para el procesamiento digital de señal, sistemas de control y, el que se va a utilizar en este trabajo, para machine learning y redes neuronales.

La "toolbox" de redes neuronales permite crearlas de manera sencilla, y de acuerdo a tus necesidades, por medio de una interfaz gráfica fácil de utilizar. Esta "toolbox" permite seleccionar el tipo de red neuronal entre feedforward, red recurrente y red en cascada. También permite determinar si el problema a resolver es de clasificación o regresión, y seleccionar que cantidad de muestras que se utilizarán para el entrenamiento, para la validación y para el test. Finalmente, permite escoger los hiperparámetros de la red (capas ocultas, neuronas por capa, épocas máximas de entrenamiento, variables de entrada y de salida) y las diferentes funciones que forman parte de la red, como la función de activación y entrenamiento. Matlab permite realizar todos estos cambios por medio de unas pocas líneas de código. En la figura 6.2 se muestra el tipo de esquema que presenta la "toolbox" de redes neuronales.

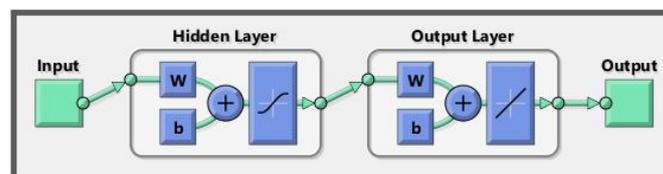


Figura 6.2: Esquema de red neuronal artificial en la App de Matlab.

6.3. Selección de la red neuronal artificial

A continuación, se seleccionan todas las características de la red neuronal generada por medio de la "toolbox" de Matlab. Como se ha comentado previamente, en el capítulo 4 se implementará una topología de red en la cual se separa la parte real (o parte en fase) de la parte imaginaria (o parte en cuadratura) de la señal, tanto a la entrada como a la

salida de la red. De esta forma, se deja de tener valores complejos y la red sólo trabaja con valores reales.

Todos los parámetros referentes a la red neuronal podrían ser escogidos por medio de algún procedimiento de optimización, como puede ser la aplicación de un algoritmo genético. Sin embargo, en este trabajo se han seleccionado mediante un método de prueba y comparación, donde la red neuronal finalmente escogida no es una red optimizada para nuestro problema (aunque sí es una red que ofrece buenos resultados).

6.3.1. Selección de las variables de entrada

Existen varias posibilidades de variables de entrada a considerar. Por un lado, se puede tener en cuenta la parte en fase y la parte en cuadratura de la señal en un instante de tiempo determinado, pero si sólo se consideran esas variables no se estarán teniendo en cuenta los efectos de memoria del amplificador. Tal y como se muestra en la figura 6.3, únicamente seleccionando la muestra actual de la señal como variable de entrada, no es suficiente para modelar el comportamiento del PA.

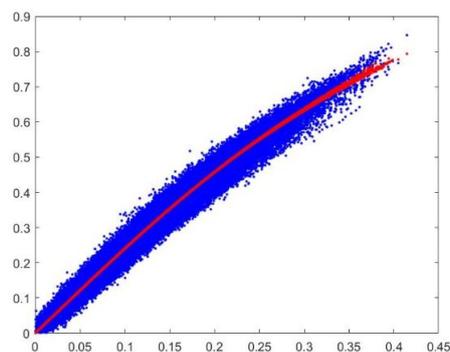


Figura 6.3: Curva AM-AM sin tener en cuenta los efectos de memoria del PA.

Otra opción consiste en considerar las componentes en fase y cuadratura pasadas de la señal, además de las componentes de la muestra actual. Este conjunto de variables de entrada se expresa como se muestra a continuación:

$$x[n] = [x_I[n], x_I[n-1], \dots, x_I[n-m], x_Q[n], x_Q[n-1], \dots, x_Q[n-m]] \quad (6.1)$$

Donde m determina la profundidad de retardo empleada. A partir de los valores de NMSE obtenidos con las diferentes variables de entrada, se selecciona el retardo óptimo para considerar en el modelado del comportamiento del PA. Estos datos se muestran en la tabla 6.1, y se decide tener en cuenta las componentes en fase y cuadratura hasta un retardo de 7 muestras. Estas medidas se han tomado utilizando una red neuronal de una única capa oculta formada por 20 neuronas.

También es posible obtener una mejoría en el modelado del comportamiento del PA si se consideran términos a la entrada que determinan la envolvente de la señal, es decir, el módulo y retardos del mismo. El valor de profundidad del retardo, considerado para las componentes en fase y cuadratura, es el óptimo encontrado previamente ($m = 7$). En este caso, las variables de entrada a la red neuronal seguirán la siguiente expresión:

Tabla 6.1: Resultados de NMSE en el modelado del amplificador de potencia.

Cantidad de retardos fase y cuadratura	NMSE
m = 2	-27.1 dB
m = 4	-27.4 dB
m = 6	-28.2 dB
m = 7	-28.4 dB
m = 8	-28.4 dB

$$x[n] = [x_I[n], x_I[n-1], \dots, x_I[n-m], x_Q[n], x_Q[n-1], \dots, x_Q[n-m], |x[n]|, |x[n-1]|, \dots, |x[n-p]|] \quad (6.2)$$

donde p denota la profundidad de retardo del valor absoluto utilizado. Volvemos a utilizar la red neuronal de una capa oculta, y 20 neuronas para tomar mediciones de NMSE y seleccionar la profundidad de retardo del valor absoluto de la señal. Los resultados obtenidos se muestran en la tabla 6.2.

Tabla 6.2: Resultados de NMSE en el modelado del amplificador de potencia teniendo en cuenta el valor absoluto de la señal.

Cantidad de retardos valor absoluto	NMSE
p = 3	-31.8 dB
p = 5	-32.6 dB
p = 7	-32.7 dB
p = 9	-32.7 dB

Aunque la diferencia entre utilizar $p = 5$ y $p = 7$ es mínima, se decide seleccionar $p = 7$, así que se utiliza la misma profundidad de retardo, tanto en las componentes en fase y cuadratura de la señal como en valor absoluto. Si se comparan los valores obtenidos cuando sólo se utilizan las componentes de fase y cuadratura, o cuando se consideran también la envolvente de la señal, se observa una disminución importante en el NMSE, de aproximadamente 4 dB, al utilizar el valor absoluto de la señal.

Por último, además del valor absoluto, también podemos tener en cuenta potencias de este valor. La siguiente expresión contempla estas potencias como variables de entrada:

$$x[n] = [x_I[n], x_I[n-1], \dots, x_I[n-m], x_Q[n], x_Q[n-1], \dots, x_Q[n-m], |x[n]|, |x[n-1]|, \dots, |x[n-p]|, |x[n]|^2, |x[n-1]|^2, \dots, |x[n-p]|^2, \dots, |x[n]|^s, |x[n-1]|^s, \dots, |x[n-p]|^s] \quad (6.3)$$

Se toman valores de NMSE con redes neuronales de una capa oculta y 20 neuronas con $m = 7$ y $p = 7$, para diferentes potencias del valor absoluto de la señal. En este caso, se observa una ligera mejoría, de unos 0.5 dB, entre usar las potencias o no usarlas. Estos resultados obtenidos aparecen en la tabla 6.3.

Tabla 6.3: Resultados de NMSE en el modelado del amplificador de potencia teniendo en cuenta el valor absoluto de la señal y sus potencias.

Potencias del valor absoluto	NMSE
s = 1	-32.7 dB
s = 2	-32.9 dB
s = 4	-33.1 dB

A partir de todas las medidas realizadas para diferentes variables de entrada, se decide considerar las componentes en fase y cuadratura y el valor absoluto de la señal hasta un retardo de 7 muestras, y el valor absoluto hasta la cuarta potencia. Esto quiere decir que se utilizarán 48 variables de entrada. En la figura 6.4 se muestra la curva AM-AM del amplificador y del modelo obtenido con esta red neuronal, que considera las 48 variables de entrada comentadas previamente. Al observar esta gráfica se aprecia cómo ahora sí se tiene en cuenta el efecto de memoria del amplificador al realizar el modelado.

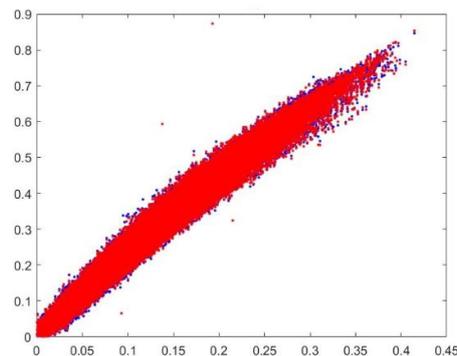


Figura 6.4: Curva AM-AM con la predicción obtenida mediante la red neuronal seleccionada.

En el esquema de la figura 6.5 se muestran las variables de entrada y de salida seleccionadas para la red neuronal que se empleará como predistorsionador.

6.3.2. Selección de hiperparámetros de la red

Los hiperparámetros de una red neuronal artificial están formados por: el número de capas ocultas empleadas, la cantidad de neuronas por capa, la ratio de aprendizaje utilizada y una serie de parámetros que determinan el rendimiento de la red. Los hiperparámetros no se utilizan para modelar los datos directamente, como podrían ser los pesos utilizados para establecer los valores que toma cada neurona, pero influyen en la capacidad y características de aprendizaje de la red neuronal artificial.

Primero se decide buscar la cantidad de capas ocultas que tendrá la red. Para tomar esta decisión se sigue analizando el valor de NMSE obtenido en el modelado del comporta-

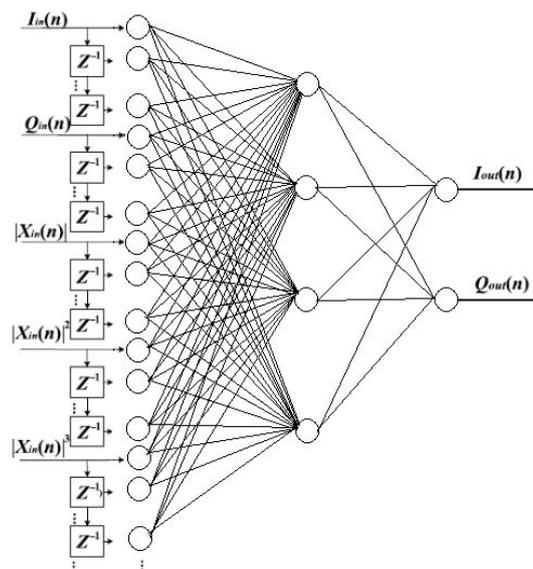


Figura 6.5: Esquema de las variables de entrada y salida seleccionadas para la red neuronal.

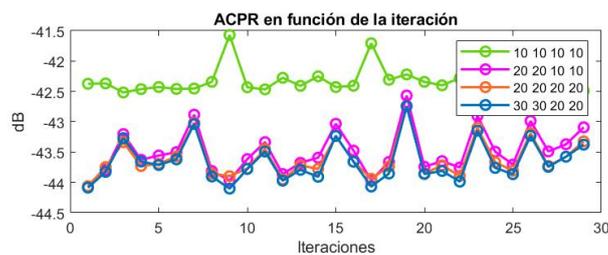
miento del amplificador de potencia del laboratorio online de la Universidad Tecnológica Chalmers. Se generan y se entrenan diversas redes neuronales, de hasta 5 capas ocultas, y se comparan los resultados de NMSE obtenidos en cada una de ellas. La tabla 6.4 recoge estos resultados, y en base a ellos se decide escoger una red neuronal de 4 capas ocultas. Es posible que con 1 o 2 capas ocultas hubiese sido suficiente para obtener buenos resultados, pero, en este caso, se escoge el mejor resultado obtenido, aunque la diferencia sea de apenas de 0.4 dB y conlleve un importante incremento de la complejidad de la red.

Tabla 6.4: Resultados de NMSE en el modelado del amplificador de potencia en función de las capas ocultas de la red neuronal artificial.

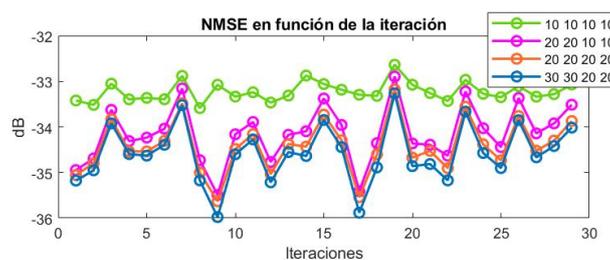
Estructura de la red neuronal	NMSE
1 capa oculta, 30 neuronas	-33.3 dB
2 capas ocultas, 20 y 10 neuronas	-33.3 dB
3 capas ocultas de 10 neuronas cada una	-33.5 dB
4 capas ocultas de 10 neuronas cada una	-33.7 dB
5 capas ocultas de 10 neuronas cada una	-33.6 dB

En cuanto a la cantidad de neuronas por capa se utilizan diferentes redes neuronales, y se comparan los resultados de éstas, en términos de NMSE y peor ACPR, a lo largo de diferentes iteraciones. Estas redes neuronales se utilizan como predistorsionador digital, con un esquema no adaptativo, en el PA de la Universidad Tecnológica Chalmers. En las figuras 6.6(a) y 6.7(c) se observan los resultados obtenidos. Tanto en el ACPR como en

el NMSE, la red neuronal que mejores resultados genera es la compuesta por dos capas ocultas de 30 neuronas, seguidas por dos capas ocultas de 20 neuronas. Sin embargo, la diferencia con las redes formadas por 20 neuronas en cada una de sus capas ocultas, y la formada por dos capas ocultas de 20 neuronas, y dos capas ocultas de 10 neuronas, es mínima. La red neuronal de 10 neuronas en todas sus capas ocultas se descarta por mostrar resultados considerablemente inferiores a las otras 3 redes testeadas. Finalmente, se decide utilizar la red neuronal compuesta por [20, 20, 10, 10] neuronas, por mostrar resultados que difieren de los mejores obtenidos en apenas 0.2 dB en el peor caso, y contar con menor complejidad computacional (requiere de menor cantidad de coeficientes) que la red compuesta por [30, 30, 20, 20] neuronas.



(a) Valores de ACPR según la cantidad de neuronas en cada capa oculta



(b) Valores de ACPR según la cantidad de neuronas en cada capa oculta

Figura 6.6: Valores de linealidad para cada red neuronal testada en función de las iteraciones.

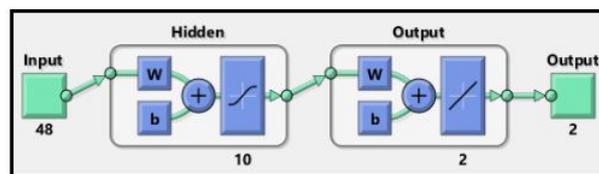
6.3.3. Funciones de aprendizaje, entrenamiento y activación

La red generada para linealizar el LMBA utiliza como función de aprendizaje el algoritmo de descenso del gradiente con momento. Con este algoritmo se tienen en cuenta los gradientes de iteraciones anteriores para calcular el error. Mientras que, como función de entrenamiento, la red utiliza el algoritmo de Levenberg-Marquardt. La función de activación empleada en cada neurona es la función tangente sigmoidea, excepto en la capa de salida, donde se utiliza una función de activación lineal denominada purelin en Matlab.

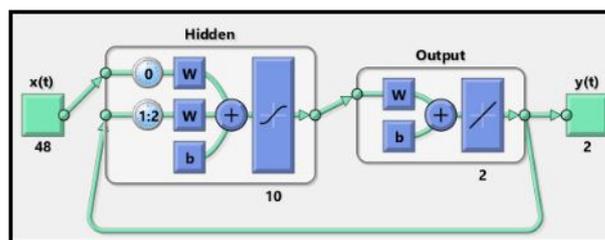
6.3.4. Selección de la topología de la red neuronal

Para la selección de la topología de red, se generan redes neuronales de entre las diferentes topologías con una sola capa oculta de 10 neuronas, para reducir el tiempo de entrenamiento de la red. Las redes testeadas son una red feedforward, una red recurrente

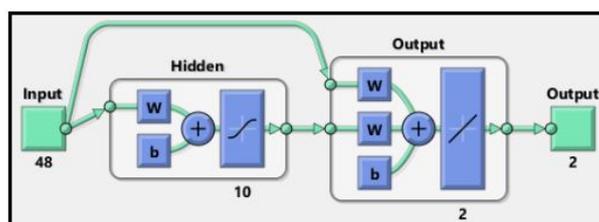
y una red en cascada. Las variables de entrada seleccionadas siguen siendo la parte real y parte imaginaria de la señal, el valor absoluto, hasta orden 4, y los retardos de estas componentes, hasta una profundidad de retardo de 7 muestras. En el caso de la red recurrente, previamente se selecciona el retardo óptimo de las muestras de salida para tener en cuenta a la entrada. Se escoge una red recurrente con 5 muestras de retardo de la señal de salida. Para comparar las redes se ha medido el NMSE obtenido en el modelado del amplificador de potencia del weblab de la Universidad Tecnológica Chalmers. En la figura 6.7 se muestran las 3 topologías empleadas. Los resultados obtenidos se muestran en la tabla 6.5. Además de estas tres topologías, se ha intentado implementar una red neuronal convolucional, pero los resultados obtenidos se encontraban muy alejados de los generados por las otras topologías y no se recogen en esta tabla.



(a) Esquema de red neuronal artificial feedforward



(b) Esquema de red neuronal artificial recurrente



(c) Esquema de red neuronal artificial en cascada

Figura 6.7: Topologías de red neuronal artificial testeadas con el modelado de un amplificador de potencia.

Tabla 6.5: Resultados de NMSE en el modelado del amplificador de potencia en función de la topología de red neuronal empleada.

Topología de la red neuronal	NMSE
Red feedforward	-30.3 dB
Red recurrente	-29.7 dB
Red en cascada	-29.5 dB

En el modelado del amplificador de potencia se obtienen los mejores resultados con la

red feedforward que, además, es la topología más sencilla y con menor cantidad de coeficientes a ser estimados. Para predecir el comportamiento del amplificador de potencia no se requiere considerar las variables de entrada de una capa en todas las capas siguientes, ni es necesario tener en cuenta los valores obtenidos a la salida. Para resolver este problema basta con considerar muestras pasadas de los valores de entrada de la señal.

6.3.5. Entrenamiento de la red neuronal

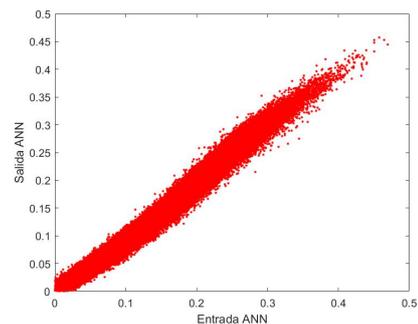
La red neuronal artificial finalmente seleccionada se trata de una red neuronal feedforward de 4 capas ocultas con 20, 20, 10 y 10 neuronas, respectivamente.

Para entrenar la red neuronal artificial se utilizan 307200 muestras complejas de valores de entrada y salida del LMBA de una señal LTE con diferentes anchos de banda, sin aplicar predistorsión. De todas estas muestras, el 70% se utilizan para el entrenamiento, un 15% se utiliza como test y el 15% restante se utiliza como muestras de validación de la red.

Considerando que, en el entrenamiento, las muestras recogidas a la entrada del LMBA son conocidas como $x[n]$, y las muestras a la salida $y[n]$, para realizar el modelado del comportamiento del LMBA se entrena la ANN considerando $x[n]$ como variables de entrada, y $y[n]$ como variables de salida. Sin embargo, para aplicar la red neuronal como predistorsionador, la función aplicada por la ANN debe ser la función inversa a la generada por el LMBA. Por esta razón, la red neuronal se entrena considerando las muestras $y[n]$, como las variables de entrada, y $x[n]$, como las variables de salida, justamente al revés que para establecer el modelado del comportamiento del amplificador. En la figura 6.8(b) se muestra la relación entrada-salida (curva AM-AM) de la señal LTE de 200 MHz de ancho de banda, a través de la red neuronal artificial entrenada, y en la figura 6.8(a) se muestra el esquema de entrenamiento.



(a) Esquema de entrenamiento de la red neuronal como DPD



(b) Curva AM-AM de señal a la entrada y salida de la red neuronal

Figura 6.8: Esquema de entrenamiento y comportamiento de la red neuronal artificial utilizada como DPD.

6.4. Resultados obtenidos con la red neuronal

En este apartado se realiza una explicación del código implementado para linealizar la señal del LMBA, y se selecciona el modo de implementación de la red neuronal como DPD (implementación adaptativa con aprendizaje directo o indirecto o no adaptativa) a partir de la comparativa de los resultados obtenidos. Finalmente, con la mejor configuración encontrada, se linealizan diferentes señales (con diferente ancho de banda), utilizando la red neuronal y el modelo GMP para determinar qué método genera los resultados.

6.4.1. Esquema del código utilizado

En la primera parte del código se determina la señal que se utilizará, es decir, su configuración, su ancho de banda y su frecuencia central. También se definen los parámetros que relacionan las dos señales (la señal principal y la señal auxiliar) que intervienen en este tipo de amplificadores. Estos parámetros son: el desfase entre señales (Ψ_{rel}), el parámetro p que determina la relación entre amplitudes y el porcentaje de offset (OP).

Además de estos valores, en esta sección del código, también se establece: el tipo de predistorsión a utilizar, si la implementación será por medio de aprendizaje directo o indirecto, la tensión de alimentación del amplificador auxiliar (CSP), la ganancia de predistorsión, la ganancia en banda base, el PAPR máximo permitido que, en caso de ser superado, se forzará su reducción, el número de iteraciones a realizar y el número de averaging que se utilizará para obtener los resultados de la linealización en cada iteración.

Para realizar las comparativas entre aprendizaje adaptativo y no adaptativo, aprendizaje directo e indirecto, y utilización de ANN o GMP como predistorsionador, se utilizan los valores de los parámetros presentados en la tabla 6.6

Tabla 6.6: Configuración de la señal y valores de los parámetros utilizados en la fase experimental.

Parámetro	Valor
Frecuencia central	2 GHz
Ancho de banda	20 MHz, 60 MHz y 200 MHz
Parámetro p	3
Porcentaje de Offset	0
Desfase entre señales	Según regresión de grado 1
Número de iteraciones	15
Número de averaging	Aumentando en 1 en cada iteración
Ganancia de predistorsión	1
Tensión de alimentación CSP	-5.4 V
Ganancia en banda base	6.5 dB (20 MHz), 9 dB (60 MHz), 10.6 dB (200 MHz)
Máximo PAPR permitido	15 dB
DPD utilizado	GMP y ANN
Aprendizaje DPD	Directo, indirecto

Una vez configurada la señal y todos los parámetros involucrados, se aplica una función de reducción del factor de cresta, donde se reducen los picos de la señal que superen el PAPR máximo establecido. Seguidamente se aplica la predistorsión seleccionada en banda base y se generan las dos señales, la principal y la auxiliar, utilizando los valores de p , Ψ_{rel} y OP seleccionados. Posteriormente, las señales se suben en banda a la frecuencia central establecida y se mandan al LMBA. Seguidamente se recoge la señal a la salida del amplificador, se aplica un alineamiento y se realiza la conversión a banda base. Una vez en banda base se obtienen los símbolos recibidos y el EVM. Por último, se calcula el error cometido para actualizar los coeficientes del GMP o para reentrenar la red neuronal, se calcula el NMSE y el ACPR en cada canal adyacente, según la configuración de señal LTE seleccionada, y se da paso a la siguiente iteración. Este código empleado se muestra en el anexo A.

6.4.1.1. Técnica de reducción del factor de cresta (CFR)

Hoy en día, debido al aumento de la velocidad de transmisión en las señales de telecomunicaciones, se utilizan señales moduladas en amplitud y fase con mayor cantidad de símbolos. Este tipo de modulación produce un importante incremento en el PAPR de la señal. Este aumento en el PAPR significa una mayor distorsión de la señal en el PA o trabajar con elevado back-off. Para mantener cierto compromiso entre linealidad y eficiencia es común suprimir los picos de la señal, sin comprometer la información transmitida. De esta manera se reduce el PAPR de la señal.

Existen diferentes técnicas para realizar la reducción del factor de cresta. En este caso se utiliza el método de cancelación de pico. Para aplicar esta técnica se define un valor de PAPR umbral (A), que será el PAPR máximo permitido. Primero se define la salida del clipper ($c[n]$) como: $c[n] = \frac{A}{|u[n]|}$ si $|u[n]| > A$, o como $c[n] = 1$ si $|u[n]| \leq A$. Por lo tanto, la señal, una vez aplicada la salida del clipper, se expresa como:

$$p[n] = u[n] - u[n]c[n] \quad (6.4)$$

Finalmente, para obtener la señal con el PAPR reducido se aplica un filtro pasa bajos:

$$z[n] = u[n] - \alpha[n]c[n]h[n] \quad (6.5)$$

$$\alpha[n] = \frac{\max(|p[n]|)}{\max(|p[n]| * h[n])} \quad (6.6)$$

donde $h[n]$ es la respuesta impulsional del filtro pasa bajos y $*$ es la operación convolución.

El PAPR máximo que se puede reducir con CFR depende de la técnica empleada, y siempre conlleva un precio a pagar. Los efectos nocivos que aparecen cuando se reduce el PAPR dependen, tanto de esta técnica usada como de la reducción de PAPR a realizar. Por lo tanto, a mayor reducción peor serán estos efectos secundarios.

Para implementar este método se requiere establecer un valor de PAPR máximo o PAPR umbral. Este valor es un parámetro que puede ser modificado para obtener mejores valores de ACPR y eficiencia.

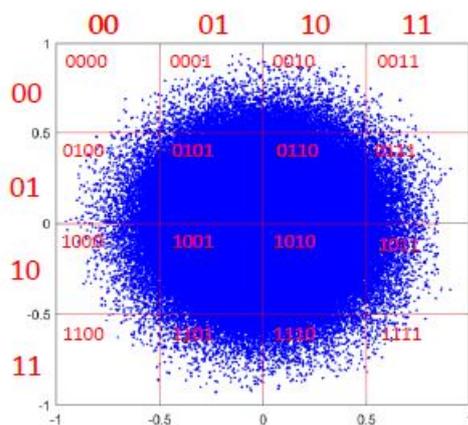
6.4.1.2. Selección de datos con función I-Q with memory Mesh-selecting

A la hora de implementar un esquema adaptativo de DPD, basado en redes neuronales, es necesario reentrenar la red por medio de una importante cantidad de datos significativos de entrada y salida del LMBA. En lugar de utilizar una serie de muestras consecutivas de la señal, se utiliza esta función para seleccionar muestras estadísticamente representativas del comportamiento del amplificador.

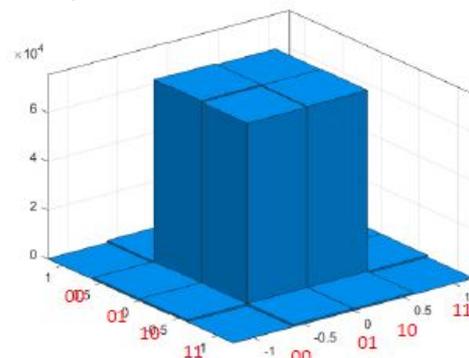
Por lo general, a la hora de determinar el modelo de un PA, ya sea con métodos como MP, GMP o redes neuronales, se dispone de mayor número de datos que de funciones base. En el caso de realizar este modelado por medio de redes neuronales, realizar una reducción adecuada del número de datos permite mantener las prestaciones de la red, disminuyendo el tiempo de entrenamiento y evitando el sobreajuste.

El método de selección de datos representativos utilizado en este trabajo se presenta en [21], y se basa en las características de la señal a transmitir. De este modo, el método de selección de datos es independiente del PA utilizado. Esta función está inspirada en un método más tradicional donde se genera un histograma de la curva AM-AM y se seleccionan las muestras manteniendo la proporcionalidad del histograma. En el caso de la función implementada, se evalúa cada valor complejo, del vector de datos de entrada, para construir un histograma multidimensional.

Este histograma multidimensional, o malla, se genera teniendo en cuenta las componentes en fase y cuadratura de la señal en el momento actual y las componentes pasadas. Por ejemplo, si no se consideran efectos de memoria, el histograma generado será bidimensional (un eje de componente en fase y el otro de componente en cuadratura). Si ahora se considera el retardo de una muestra, el histograma pasará a ser tetradimensional. En la figura 6.9 se muestra el ejemplo de histograma bidimensional sin considerar efecto de memoria.



(a) Constelación de la señal transmitida



(b) Histograma bidimensional de las componentes en fase y cuadratura de la señal

Figura 6.9: Formación del histograma bidimensional de una señal 16-QAM sin efecto de memoria.

Una vez construido el histograma multidimensional, cada histograma bidimensional I-Q se convierte en un histograma de malla lineal, donde se tiene en cuenta todos los histogramas bidimensionales para los diferentes retardos ($u[n]$, $u[n-1]$, $u[n-2]$, ..., $u[n-N]$). Seguida-

mente se realiza la reducción de muestras por un factor R, generando un nuevo histograma, pero manteniendo la misma proporcionalidad que el histograma lineal original. Esta reducción de muestras aparece en la figura 6.10.

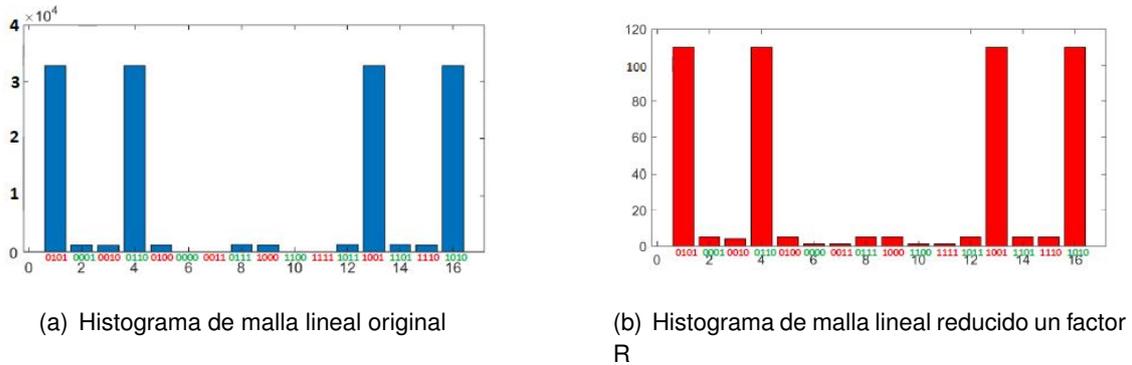


Figura 6.10: Reducción de los datos a partir del histograma de malla lineal.

Finalmente, una vez realizada la reducción de muestras, se pueden construir las matrices X_{cut} , U_{cut} y el vector y_{cut} correspondiente. En el caso de utilizar una red neuronal como DPD, las matrices X_{cut} y U_{cut} son matrices donde cada fila corresponde a una muestra de la señal y las columnas son las variables de entrada a la red neuronal. En la figura 6.11 aparecen las muestras seleccionadas en función del total de muestras y sobre la curva característica AM-AM.

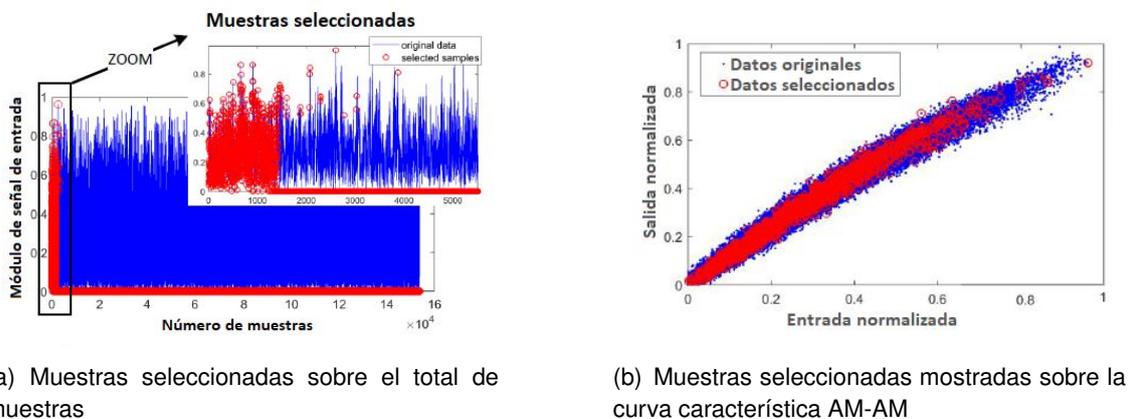


Figura 6.11: Esquema de las muestras seleccionadas con este método.

A la hora de aplicar esta función para seleccionar los datos que se utilizarán en el reentrenamiento de la red neuronal, el parámetro que puede ser modificado es el factor de reducción. Con este parámetro se determina la cantidad de datos representativos que se seleccionarán. Un valor elevado de factor de reducción conlleva seleccionar menor cantidad de datos para el reentrenamiento y, por lo tanto, este reentrenamiento requerirá de menor tiempo para llevarse a cabo. Es importante seleccionar un factor de reducción que nos permita obtener unos buenos resultados de linealidad al realizar una implementación adaptativa pero que, a su vez, sea rápido el proceso de reentrenamiento.

6.4.2. Red neuronal adaptativa versus no adaptativa

Experimentalmente se va a seleccionar entre un esquema no adaptativo o un esquema adaptativo. En el caso de la implementación no adaptativa, la red se entrena una sola vez, previamente a la transmisión de la señal. Este entrenamiento se realiza con muestras obtenidas del amplificador. Si, por el contrario, se utiliza un esquema adaptativo, la red neuronal se va reentrenando con un número reducido de muestras a lo largo de la transmisión. Si el comportamiento del LMBA no sufre modificaciones a lo largo del tiempo, sería viable utilizar un esquema no adaptativo, donde el valor de los coeficientes se calcula durante la fase de entrenamiento y una vez establecidos no son modificados. Sin embargo, si el comportamiento del LMBA sufre modificaciones debido, por ejemplo, a la variación en la temperatura de sus componentes, se requerirá implementar un esquema adaptativo, que recalculará el valor de los coeficientes de la red neuronal a lo largo del tiempo por medio nuevos entrenamientos. Implementar un esquema adaptativo aumenta la complejidad del predistorsionador digital.

Para realizar la comparativa entre los esquemas, se ha utilizado la señal de 200 MHz de ancho de banda. Se han tomado los valores de linealidad, eficiencia y potencia media a la salida durante 13 iteraciones, para un esquema adaptativo y un esquema no adaptativo. La red neuronal utilizada consiste en 4 capas ocultas de 20, 20, 10 y 10 neuronas, respectivamente, y una profundidad de retardo de 9 muestras para las variables de entrada. En cada una de las iteraciones se ha aumentado el averaging (Avg) en 1 para analizar a la misma vez el efecto de este parámetro.

En el esquema adaptativo se realiza un entrenamiento en cada iteración con todas las muestras obtenidas durante 100 épocas, como máximo.

En la figura 6.12(a) se observa el peor valor de ACPR obtenido en cada iteración. Con el esquema no adaptativo no se consigue alcanzar el valor de -45 dB requeridos para 5G, mientras que, con un esquema adaptativo, este valor se alcanza fácilmente a partir de un valor de $Avg = 4$. En cuanto al NMSE (6.12(b)), se obtienen mejores valores (hasta 7 dB) con el esquema adaptativo. La potencia media a la salida y la eficiencia (6.12(c), 6.12(d)) presentan prácticamente el mismo valor utilizando un esquema u otro.

A partir de estos resultados obtenidos, se concluye que es necesario implementar un esquema adaptativo que permita reentrenar la red neuronal, a lo largo del tiempo, para adaptarse a las modificaciones del comportamiento del LMBA.

6.4.3. Aprendizaje directo versus aprendizaje indirecto

Para seleccionar entre aplicar un aprendizaje directo o indirecto, se comparan los valores de linealidad y eficiencia obtenidos al aplicar la misma red neuronal, como DPD, sobre una señal LTE de 200 MHz. El entrenamiento en cada iteración se realiza con un factor de reducción de 5 (sólo se utiliza la quinta parte de las muestras obtenidas en cada iteración) y sólo se permiten 10 épocas, como máximo, en el entrenamiento de la red. De esta manera, se reduce el tiempo de entrenamiento. El número de averaging se mantiene fijo a 10.

En el caso de implementar un aprendizaje indirecto, la red se reentrena directamente con las muestras de entrada y salida del amplificador, obtenidos en la iteración previa.

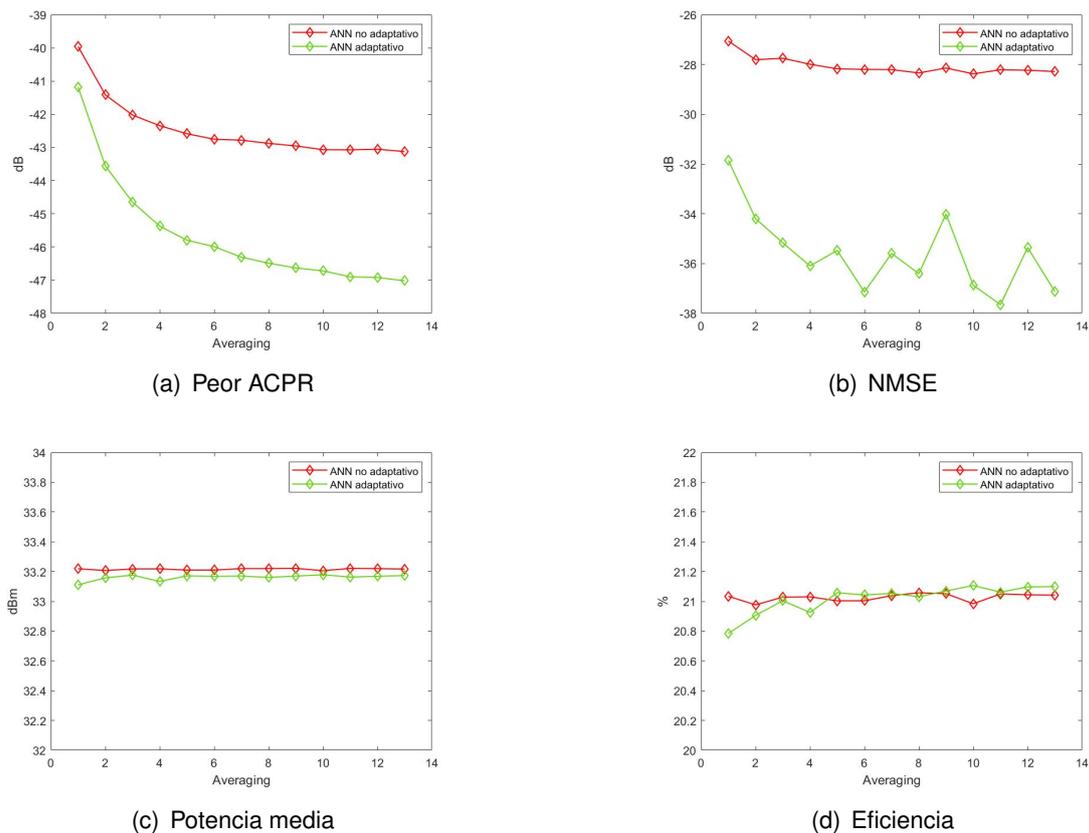


Figura 6.12: Comparativa en la implementación de una red neuronal con esquema adaptativo o no adaptativo sobre una señal LTE de 200 MHz de ancho de banda.

Es decir, se consideran los valores de $y[n]$ como variables de entrada de la red para su entrenamiento, y se consideran los valores de $x[n]$ como los valores a la salida de la red.

Si se implementa un aprendizaje directo, los valores de $y[n]$ no se utilizan directamente para el entrenamiento, sino que se calcula el error y se le aplica a la señal $x[n]$ por medio de un factor de aprendizaje (μ). En el aprendizaje directo la red neuronal se entrena con la señal $x[n] + \mu(u[n] - y[n])$, como entrada, y con $u[n]$, como salida. El valor de μ se establece a 0.9, debido a que es el valor con el que se obtienen mejores resultados de linealidad.

A partir de los resultados mostrados en las figuras 6.13, en las gráficas de linealidad (ACPR y NMSE) se muestra cómo los mejores resultados se obtienen con la retroalimentación del DPD basada en aprendizaje directo. La potencia media y la eficiencia no se ven afectadas prácticamente por el tipo de aprendizaje utilizado (sólo se aprecia una ligera mejoría en la eficiencia si utilizamos un esquema directo).

Con los resultados obtenidos en estos dos últimos apartados se decide utilizar la red neuronal con un esquema adaptativo basado en un aprendizaje directo.

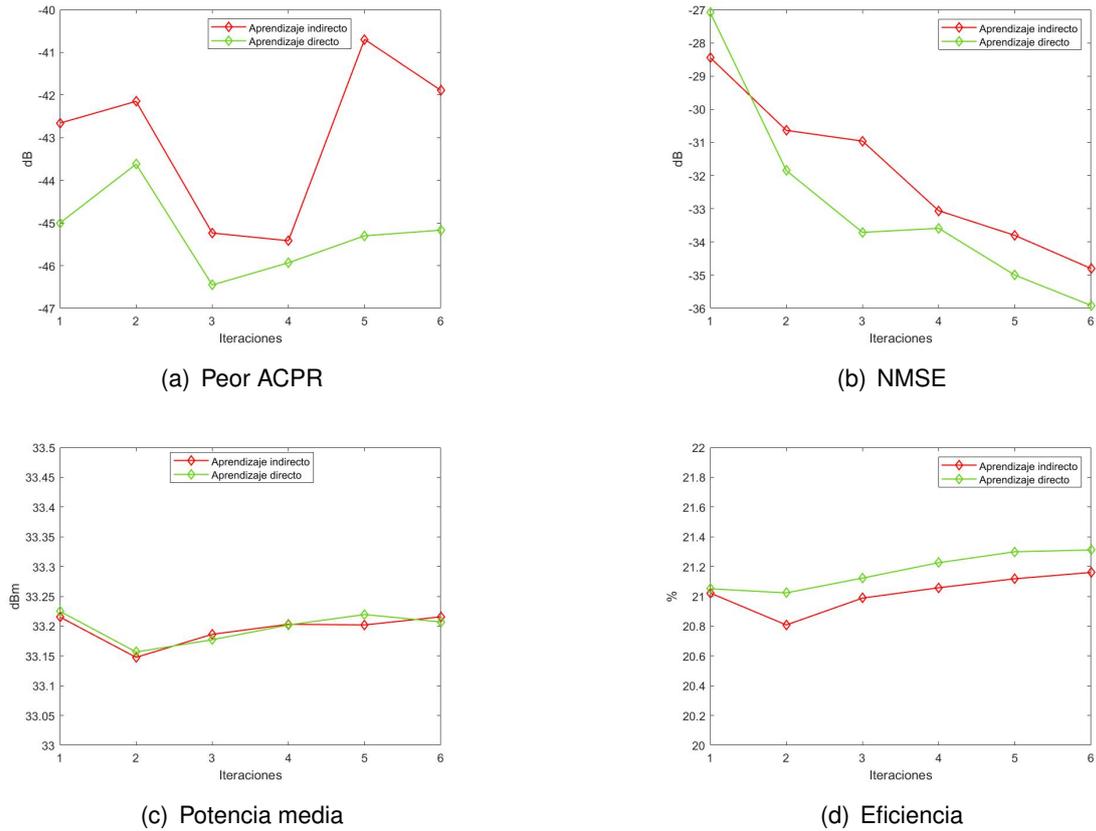


Figura 6.13: Comparativa en la implementación de una red neuronal con aprendizaje directo e indirecto sobre una señal LTE de 200 MHz de ancho de banda.

6.4.4. Comparativa de predistorsión realizada con ANN y GMP

A continuación se implementa el predistorsionador basado en la red neuronal seleccionada con esquema adaptativo de aprendizaje directo, limitando las épocas de entrenamiento máximas a 10 y utilizando un factor de reducción de 1. Este DPD se emplea para linealizar el LMBA, en el caso de utilizar 3 señales LTE con modulación 64QAM con distintos anchos de banda (20 MHz, 60 MHz y 200 MHz). Los resultados obtenidos con la ANN se comparan con los resultados generados mediante GMP. El modelo GMP se aplica con los valores de los parámetros mostrados en la tabla 6.7, y construyendo las funciones bases a partir de la siguiente ecuación:

$$\hat{y}[n] = \sum_{k=0}^{K_a} \sum_{l=0}^{L_a} a_{k,l} x(n - \tau_l^a) |x(n - \tau_l^a)|^k + \sum_{k=0}^{K_b} \sum_{l=1}^{L_b} \sum_{m=1}^{M_b} b_{k,l,m} x(n - \tau_l^b) |x(n - \tau_l^b - \tau_m^b)|^k + \sum_{k=0}^{K_c} \sum_{l=1}^{L_c} \sum_{m=1}^{M_c} c_{k,l,m} x(n - \tau_l^c) |x(n - \tau_l^c + \tau_m^c)|^k \quad (6.7)$$

Con la configuración de GMP mostrada en la tabla, la cantidad de coeficientes a determinar para el DPD es de 248, sin embargo, para implementar la red neuronal seleccionada

Tabla 6.7: Configuración del modelo GMP empleado en la linearización del LMBA.

Parámetro	Valor
k_a	7
k_b	5
k_c	5
τL_a	[0 : 1 : 9]
τL_b	[0 : 1 : 7]
τL_c	[0 : 1 : 7]
τM_b	[-1 : 1 : 1]
τM_c	[-1 : 1 : 1]

se requiere de la estimación de 1742 coeficientes durante el entrenamiento, un incremento bastante considerable con respecto al GMP.

6.4.4.1. Resultados obtenidos para la señal de 20 MHz

En el caso de la señal de 20 MHz se obtienen buenos resultados con los dos métodos de linealización. En términos de linealidad, la red neuronal presenta ligeramente mejores resultados que GMP, pero si nos centramos en la eficiencia, los mejores resultados los presenta el modelo GMP. Tanto con GMP como con la red neuronal se consiguen alcanzar los -45 dB de ACPR fácilmente (requerimiento 5G). Como la implementación del DPD basado en ANN conlleva un alto coste computacional, y el reentrenamiento con cada iteración requiere de varios minutos para llevarse a cabo, es conveniente utilizar, en el caso de la señal de 20 MHz, un modelo basado en GMP o semejante, en lugar de una red neuronal; ya que no existe justificación, en términos de rendimiento, para el aumento de complejidad y coste computacional que conlleva esta implementación. Las curvas AM-AM obtenidas con cada método, así como sus espectros, se muestran en las figuras 6.14; y los valores de linealidad, eficiencia, potencia y EVM recogidos al cabo de una serie de iteraciones se muestran en la tabla 6.8.

Tabla 6.8: Valores de linealidad, potencia, eficiencia y EVM obtenidos con DPD basado en ANN y GMP para la señal LTE de 20 MHz.

Tipo de DPD	Peor ACPR	NMSE	Potencia media	Eficiencia	EVM
ANN	-52.9 dB	-40.4 dB	35.3 dBm	26.8%	0.59%
GMP	-51.02 dB	-39.3 dB	35.8 dBm	28.6%	0.66%

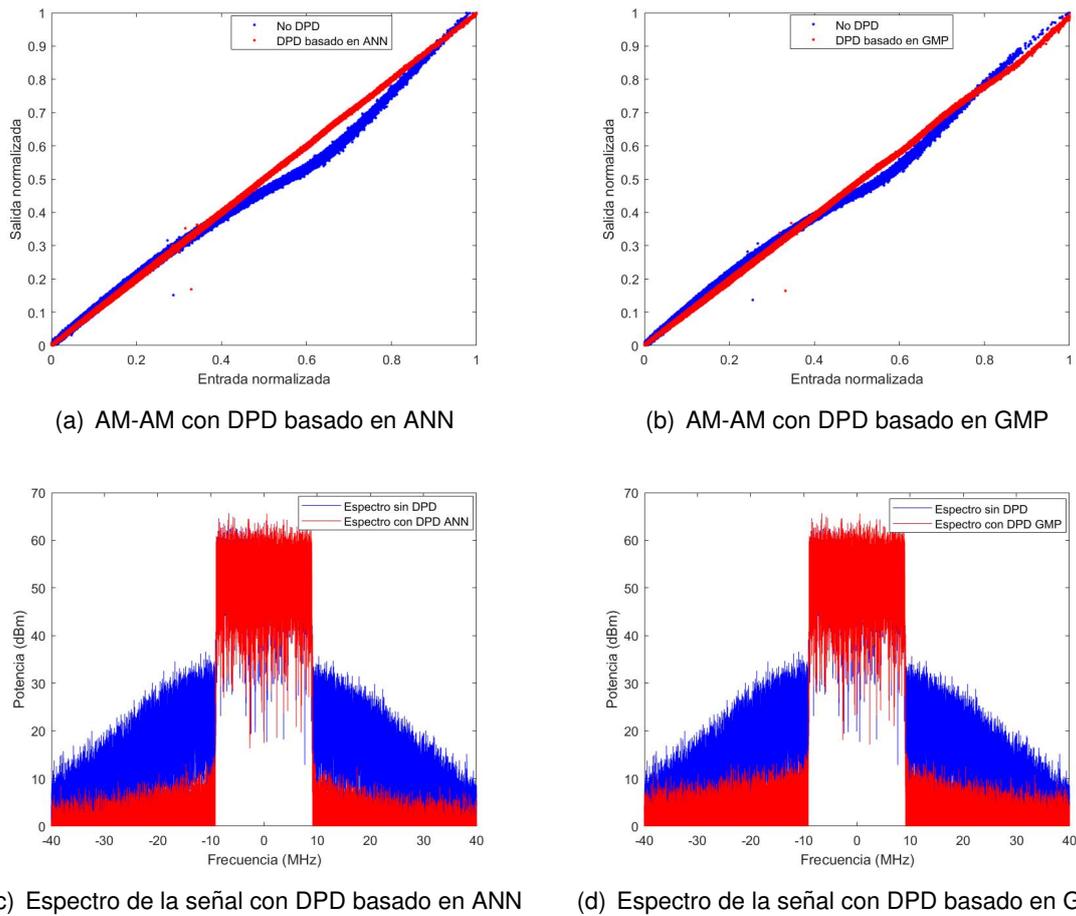


Figura 6.14: Comparativa de curva AM-AM y espectro frecuencial sobre la señal LTE 20 MHz con GMP y ANN.

6.4.4.2. Resultados obtenidos para la señal de 60 MHz

Una vez analizados los resultados para una señal de 20 MHz, realizamos la misma comparativa, pero a 60 MHz. Se utiliza la misma red neuronal, con los mismos parámetros de épocas máximas y factor de reducción, para la implementación adaptativa. El modelo GMP también se aplica con los valores mostrados previamente.

Las gráficas de las curvas AM-AM y del espectro de la señal, para los dos métodos de linealización, se muestran en la figura 6.15; y los valores de linealidad, eficiencia, potencia y EVM obtenidos al cabo de una serie de iteraciones aparecen en la tabla 6.9. A partir de las gráficas apenas se nota diferencia entre utilizar la red neuronal o el método GMP. Sin embargo, en los valores numéricos de la tabla, aunque con los dos métodos se cumplen los requerimientos de 5G, se obtiene un mejor rendimiento en el predistorsionador basado en GMP, tanto en linealidad como en eficiencia. Por lo tanto, al igual que sucedía con la señal de 20 MHz de ancho de banda, es preferible utilizar un DPD basado en GMP, en lugar de utilizar una red neuronal que aumentaría considerablemente la complejidad del problema sin obtener resultados mejores.

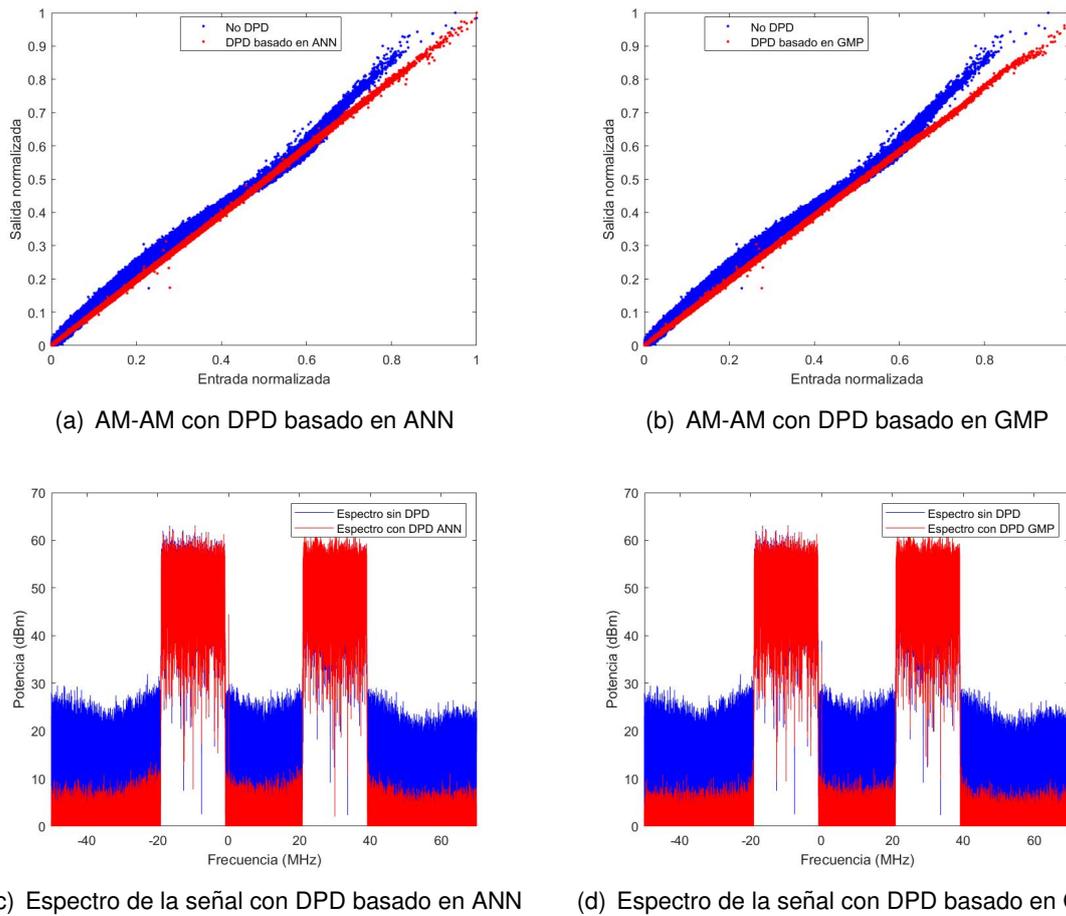


Figura 6.15: Comparativa de curva AM-AM y espectro frecuencial sobre la señal LTE 60 MHz con GMP y ANN.

Tabla 6.9: Valores de linealidad, potencia, eficiencia y EVM obtenidos con DPD basado en ANN GMP para la señal LTE de 60 MHz.

Tipo DPD	de	Peor ACPR	NMSE	Potencia media	Eficiencia	EVM
ANN		-49.9 dB	-39.2 dB	35.4 dBm	27.9%	0.59%
GMP		-51.0 dB	-39.6 dB	35.9 dBm	30.2%	0.59%

6.4.4.3. Resultados obtenidos para la señal de 200 MHz

Para concluir la comparativa entre los dos predistorsionadores, se pasa a linealizar el LMBA cuando se transmite una señal de 200 MHz de ancho de banda. En el caso del GMP, los resultados obtenidos no son nada buenos. El mejor valor de ACPR alcanzado es de -35 dB. Este valor dista en gran medida de los -45 dB de ACPR necesarios. Aunque se aumente el número de retardos y las potencias para la implementación de GMP, el valor de ACPR no presenta una modificación importante que permita situarse cerca de los -45 dB deseados.

En el caso de la red neuronal, con las variables de entrada establecidas previamente,

donde se tenía en cuenta hasta un retardo de 7 muestras, los resultados obtenidos son considerablemente mejores que en el caso de GMP. Sin embargo, el ACPR se mantiene sobre los -44 dB, sin alcanzar los -45 dB de requerimiento 5G. Por esta razón, se decide modificar la red neuronal y considerar como variables de entrada la señal hasta un retardo de 9 muestras (red neuronal de 1982 coeficientes). Con esta modificación se aumenta considerablemente los coeficientes a calcular, pero se alcanzan fácilmente los -45 dB de ACPR deseados. Para disminuir el tiempo de entrenamiento, que se ha visto afectado por la decisión de aumentar las variables de entrada, y por el ancho de banda de la señal, se reducen las épocas máximas y se aumenta el factor de reducción, obteniendo resultados aceptables con 5 épocas por iteración y un factor de reducción de 5.

En la figura 6.16 se muestran las curvas AM-AM y el espectro de la señal, cuando se utiliza GMP y cuando se utiliza la red neuronal con retardos de hasta 9 muestras. En la tabla 6.10 aparecen los valores de linealidad, eficiencia, potencia y EVM obtenidos con GMP y con las dos redes neuronales.

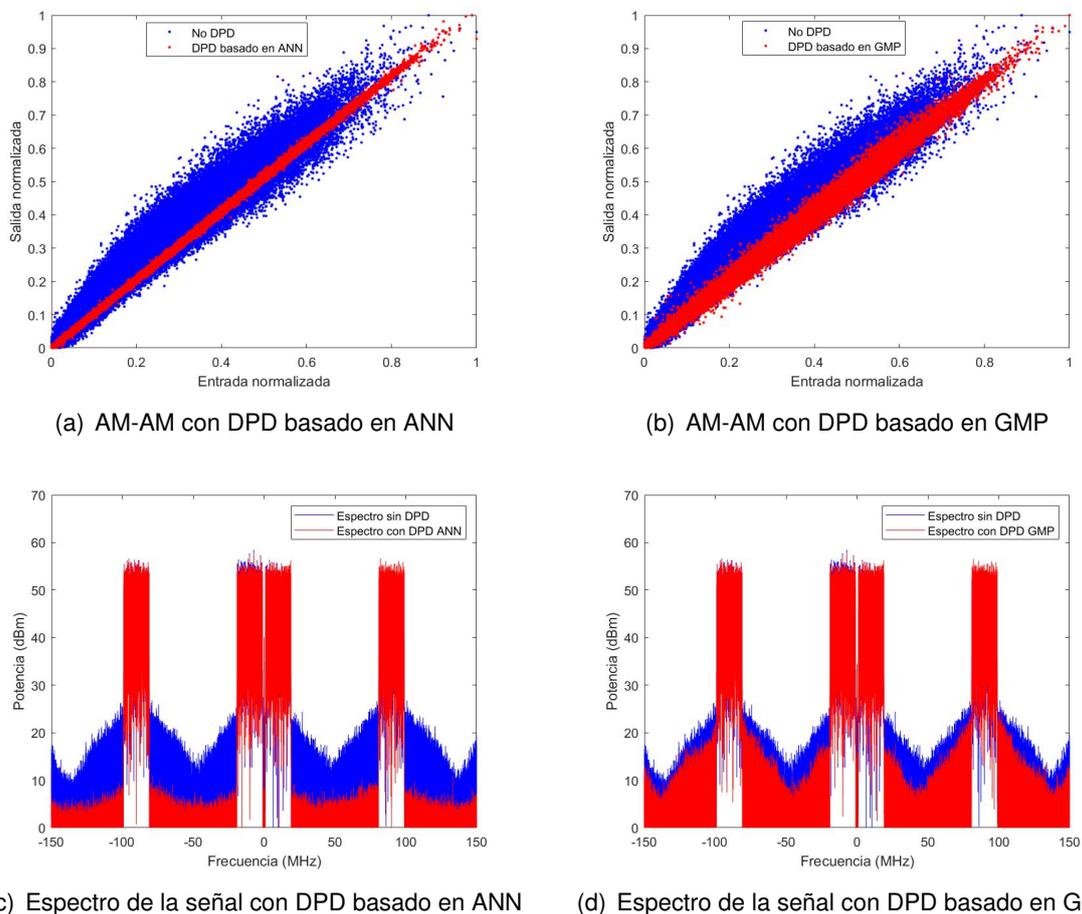


Figura 6.16: Comparativa de curva AM-AM y espectro frecuencial sobre la señal LTE 200 MHz con GMP y ANN.

A partir de los resultados obtenidos con los 3 anchos de banda distintos, podemos deducir que, para anchos de banda relativamente pequeños, no es necesario ni conveniente utilizar una red neuronal para linealizar el dispositivo; es mucho más eficiente utilizar otro método como GMP, que no requiere de tanto coste computacional. Sin embargo, cuando el ancho de banda de la señal se vuelve considerable y es complejo linealizar con

otros métodos, la utilización de una red neuronal, como DPD, nos proporciona resultados satisfactorios que permiten cumplir con los requerimientos deseados.

Tabla 6.10: Valores de linealidad, potencia, eficiencia y EVM obtenidos con DPD basado en ANN y GMP para la señal LTE de 200 MHz.

Tipo de DPD	Peor ACPR	NMSE	Potencia media	Eficiencia	EVM
ANN Retardo 7	-44.0 dB	-35.9 dB	33.2 dBm	21.3 %	0.69 %
ANN Retardo 9	-46.2 dB	-36.3 dB	33.3 dBm	21.4 %	0.89 %
GMP	-35.4 dB	-28.0 dB	33.7 dBm	22.7 %	2.08 %

6.4.5. Parámetros para optimizar la eficiencia

En las pruebas anteriores, el trabajo ha estado centrado en cumplir con el valor de ACPR requerido, pero no se ha tenido en cuenta la eficiencia obtenida. Una vez los valores de linealidad se encuentran por encima de los deseados, es posible modificar una serie de parámetros que permitirán aumentar la eficiencia y la potencia media a la salida. Se va a modificar el PAPR máximo permitido. Disminuir este parámetro permite mejorar el valor del ACPR, pagando el precio de empeorar el NMSE y EVM.

Esta mejora en el valor de ACPR posibilita modificar la ganancia de predistorsión que controla la potencia de salida. Disminuir el valor de ganancia de predistorsión genera una pendiente mayor en la linealidad, mejorando la eficiencia y la potencia, a la misma vez que se cumplen los requerimientos de ACPR.

En la figura 6.17 se muestran los resultados de ACPR, NMSE, potencia media y eficiencia, en el caso de mantener el valor de PAPR máximo en 15 dB y la ganancia de predistorsión a 1, y en los casos de disminuir el valor de PAPR a 9 dB y 8 dB, y la ganancia de predistorsión a 0.86 y 0.83, respectivamente. La configuración que mejor eficiencia y potencia media presenta corresponde a un valor de PAPR máximo de 8 dB y 0.83 de ganancia. Con esta configuración se gana aproximadamente un 4 % de eficiencia, y mantenemos por debajo de -45dB el peor caso de ACPR. Con esta configuración se sufre un importante empeoramiento el EVM, manteniéndose alrededor del 2,6 % en todas las iteraciones. El valor máximo permitido de EVM depende de la modulación utilizada. En este trabajo se considera que el valor máximo permitido es del 3 %.

6.5. Validación en entorno dinámico

La linealización del LMBA mostrada previamente se basa en la utilización de una señal de test, a una frecuencia central de 2 GHz, por lo que es necesario mostrar el comportamiento del DPD dentro de un entorno cambiante con una señal nueva que modifica su ancho de banda y su frecuencia.

Para cada ancho de banda la estrategia a seguir para linealizar el amplificador es distinta. Para las señales de 20 y 60 MHz se utiliza GMP, mientras que para 200 MHz se utiliza una

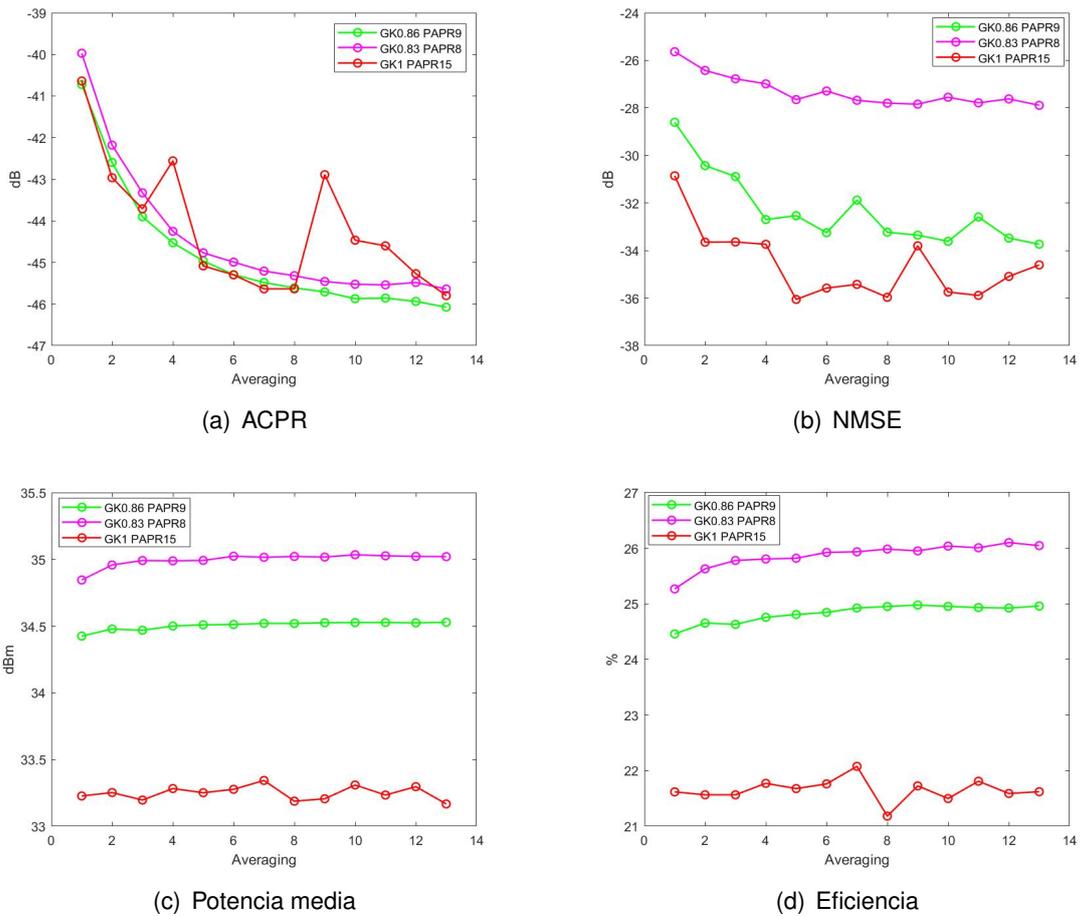


Figura 6.17: Comparativa de rendimiento de la señal LTE de 200 MHz con diferentes valores de PAPR máximo y ganancia de predistorsión.

red neuronal artificial previamente entrenada. Para cada frecuencia utilizada se establece el desfase entre señales del LMBA, a partir de la regresión de grado 1 encontrada en el capítulo 5 de este trabajo. Los valores de p (relación entre amplitudes) y OP (porcentaje de offset) se mantienen constantes a 3 y 0, respectivamente.

Se utilizan 5 configuraciones de señal diferentes. Las 3 primeras corresponden a una señal LTE de 200 MHz de ancho de banda, con frecuencias centrales de 1.8 GHz, 2 GHz y 2.2 GHz; seguidamente, se utiliza una señal LTE de 20 MHz de ancho de banda, a una frecuencia central de 1.9 GHz; y, finalmente, se emplea una señal LTE de 60 MHz de ancho de banda, con frecuencia central de 2.1 GHz.

Para cada una de las diferentes configuraciones se han establecido los mejores parámetros de: PAPR máximo, ganancia de predistorsión, factor de reducción y épocas máximas de entrenamiento (en caso de utilizar red neuronal); para obtener los mejores valores de eficiencia y potencia, cumpliendo con los requerimientos de 5G, y minimizando el tiempo de entrenamiento. Aun así, estos valores obtenidos en las mediciones pueden mejorarse si se optimizan los parámetros relacionados con el LMBA, que en este caso hemos mantenido constantes (relación entre amplitudes, porcentaje de offset y tensión de alimentación del amplificador auxiliar). Los valores de los parámetros utilizados en cada configuración de señal se muestran en la tabla 6.11.

Tabla 6.11: Configuración empleada para cada una de las señales en el entorno dinámico.

Configuración de la señal	DPD	PAPR máximo	Ganancia de predistorsión	Factor de reducción	Épocas máximas
$F_c = 2.2$ GHz, $B_w = 200$ MHz	ANN	9 dB	0.89	5	10
$F_c = 2$ GHz, $B_w = 200$ MHz	ANN	8 dB	1.1	5	15
$F_c = 1.8$ GHz, $B_w = 200$ MHz	ANN	8 dB	0.95	5	10
$F_c = 1.9$ GHz, $B_w = 20$ MHz	GMP	9 dB	0.77	-	-
$F_c = 2.1$ GHz, $B_w = 60$ MHz	GMP	12 dB	0.78	-	-

En la figura 6.18 aparecen los valores de peor ACPR, de NMSE, de potencia media y eficiencia, para cada una de las configuraciones de señal utilizada. Los mejores valores se obtienen para las señales de 20 MHz y 60 MHz. Los resultados obtenidos con la señal de 200 MHz son muy semejantes en las 3 frecuencias centrales analizadas. Los valores de ACPR encontrados en las señales de 200 MHz cumplen con los -45 dB de ACPR, pero sus valores de eficiencia y potencia se encuentran bastante alejados de los obtenidos en las señales de 20 MHz y 60 MHz. Sería necesario realizar una optimización de todos los parámetros que intervienen en el LMBA para mejorar los datos de eficiencia.

Tabla 6.12: Resultados obtenidos en el entorno dinámico.

Configuración de la señal	Peor ACPR	NMSE	Potencia media	Eficiencia	EVM
$F_c = 2.2$ GHz, $B_w = 200$ MHz	-46.1 dB	-32.9 dB	33.0 dBm	16.7%	1.2%
$F_c = 2$ GHz, $B_w = 200$ MHz	-46.1 dB	-34.0 dB	32.3 dBm	15.1%	1.1%
$F_c = 1.8$ GHz, $B_w = 200$ MHz	-46.1 dB	-27.2 dB	32.4 dBm	18.6%	2.7%
$F_c = 1.9$ GHz, $B_w = 20$ MHz	-48.0 dB	-37.7 dB	38.3 dBm	38.5%	0.8%
$F_c = 2.1$ GHz, $B_w = 60$ MHz	-51.2 dB	-38.3 dB	37.1 dBm	31.1%	0.7%

En la tabla 6.12 se muestran los valores de linealidad, eficiencia, potencia media y EVM obtenidos para cada configuración pasadas unas cuantas iteraciones. En todas las configuraciones se ha utilizado un averaging de 10 para realizar la linealización del LMBA. Los valores obtenidos por medio de la red neuronal se pueden mejorar si utilizamos más muestras para realizar el entrenamiento (menor valor del factor de reducción), permitiendo mayor número de épocas en cada entrenamiento. Sin embargo, aquí se ha realizado

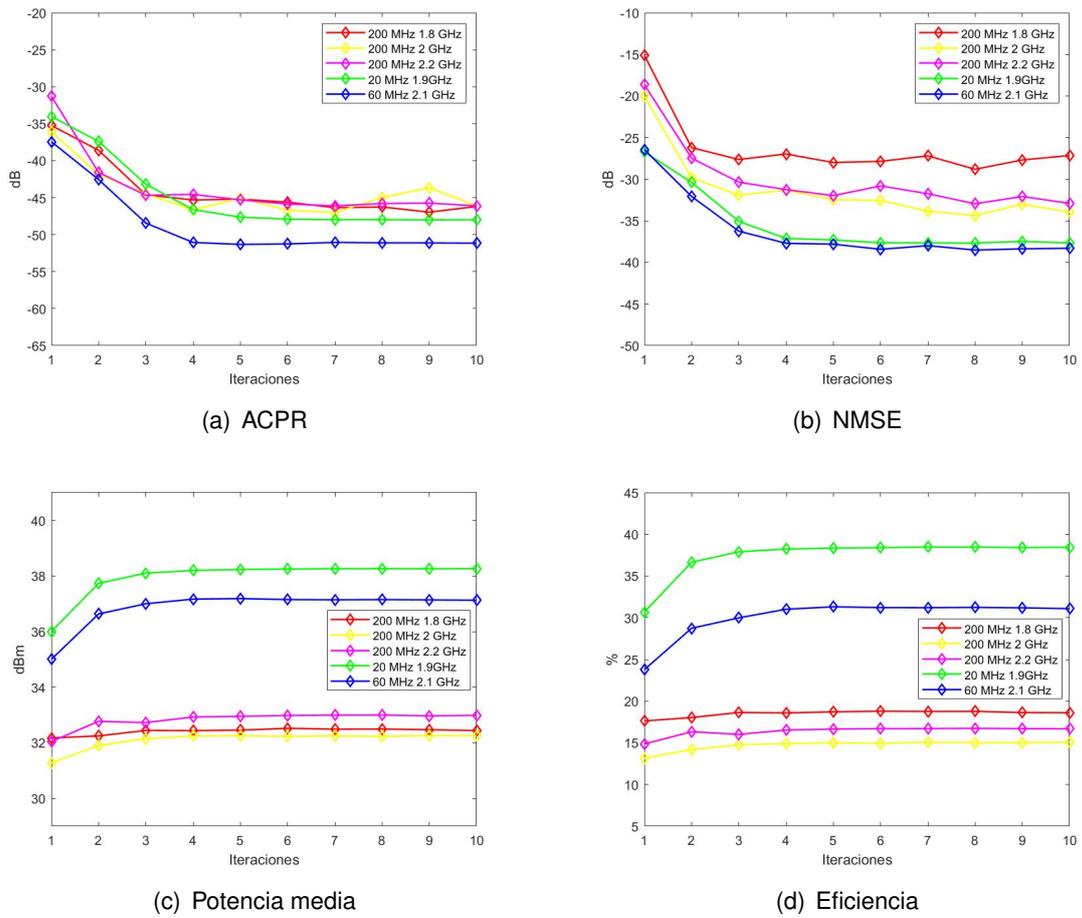


Figura 6.18: Resultados obtenidos con las diferentes configuraciones de señal.

una compensación entre los valores obtenidos de linealidad y eficiencia con el tiempo requerido de entrenamiento en cada iteración, intentando disminuir al máximo el tiempo requerido para hacer viable esta implementación.

CONCLUSIONES

El surgimiento de nuevas aplicaciones que requieren un incremento en la velocidad de transmisión generan la necesidad de utilizar señales con mayor ancho de banda. Para tratar con estas nuevas señales se requiere diseñar nuevos dispositivos de comunicaciones para adaptarse a las nuevas circunstancias. Uno de estos dispositivos que deben seguir evolucionando es el amplificador de potencia. Con la idea de manejar señales de mayor ancho de banda, sin dejar de lado la eficiencia, surge el LMBA. Este amplificador presenta una serie de parámetros que deben ser configurados correctamente para obtener buenos resultados de eficiencia y linealidad. Uno de estos parámetros es el desfase entre la señal principal y la auxiliar del amplificador. Este parámetro presenta un fuerte efecto sobre la linealidad y su valor óptimo es dependiente de la frecuencia. Esta dependencia entre el desfase y la frecuencia es prácticamente lineal y se puede predecir mediante una regresión de grado 1. Otro parámetro a tener en cuenta es la relación entre amplitudes; esta relación también afecta a la linealidad del dispositivo, aunque muestra una acción más pronunciada sobre la eficiencia. Estos dos parámetros, a valores de p (relación entre amplitudes) pequeños, se pueden considerar independientes entre ellos.

A partir de las mediciones realizadas en este trabajo, se propone una serie de criterios para realizar una correcta configuración del amplificador. Primero de todo, resulta conveniente buscar un retardo óptimo entre las señales en banda base que permita balancear el ACPR de los canales adyacentes. Seguidamente, se propone realizar una campaña de medidas para obtener la regresión lineal y, así, poder predecir el desfase óptimo. Finalmente, es posible realizar un ajuste fino de otros parámetros como el desfase entre amplitudes y la tensión de alimentación del amplificador auxiliar.

Como el LMBA es un amplificador que presenta una importante distorsión no lineal, aunque se realice una correcta configuración de sus parámetros, se requiere de la implementación de alguna técnica de linealización para cumplir con los requerimientos del estándar 5G. La utilización de un predistorsionador basado en GMP ofrece buenos valores de linealidad para señales con anchos de banda relativamente pequeños, sin embargo, a medida que el ancho de banda aumenta, el GMP pierde rendimiento y no es capaz de linealizar una señal de 200 MHz de ancho de banda. Para este tipo de señales, de mayor ancho de banda, se propone la utilización de un predistorsionador basado en redes neuronales. Se utiliza una red neuronal artificial de 5 capas (4 capas ocultas) implementada con un esquema adaptativo de aprendizaje directo para linealizar la señal de 200 MHz de ancho de banda. Con esta red se cumple con los requerimientos de ACPR y EVM.

Utilizar una red neuronal de varias capas ocultas conlleva un aumento considerable en la complejidad del DPD, debido al incremento en la cantidad de coeficientes a determinar. Por lo tanto, para señales de ancho de banda pequeños, donde es posible obtener buenos resultados de linealidad con otro método más sencillo como GMP, es preferible evitar la utilización de la red neuronal. Sin embargo, con el aumento del ancho de banda, es necesario considerar otras posibilidades de DPD, y las redes neuronales pueden ser unas buenas candidatas.

En este trabajo se han obtenido buenos resultados de linealidad para la señal de 200 MHz, por medio del DPD basado en redes neuronales, pero los resultados de eficiencia se encuentran por debajo de los valores esperados. Como futuro trabajo se pretende mejorar la eficiencia del LMBA, por medio de un ajuste fino de una serie de parámetros: la relación

entre amplitudes, el porcentaje de offset y la tensión de alimentación del amplificador auxiliar; y por medio de la optimización de la red neuronal empleada, utilizando algún algoritmo de optimización como el algoritmo genético.

BIBLIOGRAFÍA

- [1] “EASA: Commission Delegated Regulation (EU) 2019/945 of 12 March 2019 on unmanned aircraft systems and on third-country operators of unmanned aircraft systems”. *https://eur-lex.europa.eu/eli/reg_del/2019/945/oj*, (2019) 5
- [2] ETSI, “5G; NR; Base Station (BS) radio transmission and reception (3GPP TS 38.104 version 15.2.0 Release 15)”, (2018) 9
- [3] Chiara Ramella, Anna Piacibello, Roberto Quaglia, “High Efficiency Power Amplifiers for Modern Mobile Communications: The Load-Modulation Approach”. *Electronics*, **volum**(6), no. 96, (2017) 8
- [4] Peter B. Kenington, “High-Linearity RF Amplifier Design”. *Artech House microwave library*, (2000) 15, 19
- [5] Zhancang Wang, “Demystifying envelope tracking: Use for highefficiency power amplifiers for 4G and beyond”. *IEEE Microwave Magazine*, **volum**(16), no. 3, pp. 106–129, (2015) 21
- [6] D. Cox, “Linear amplification with nonlinear components”. *IEEE Transactions on Communications*. **volum**(22), 1942–1945. (1974) 22
- [7] R. Quaglia and S. Cripps, “A Load Modulated Balanced Amplifier for Telecom Applications”. *IEEE Trans. on Microwave Theory and Tech.* **volum**(66), 1328–1338. (2018) 23, 57, 59
- [8] Roberto Quaglia, Jeff Powell, “Analysis and Characterization of a Load Modulated Balanced Amplifier for Base-Station Applications”. *German Microwave Conference, GeMiC*, (2018) 23
- [9] A. A. M. Saleh and J. Salz, “Adaptive linearization of power amplifiers in digital radio systems”. *Bell System Technical Journal*. **volum**(62), no. 4, 1019–1033. (1983) 27
- [10] H. Deng, D. Zhou, D. Zhang, and Y. Zhang, “Analogue predistorter based on non-linear crlhtl”. *Electronics Letters*, **volum**(54), no. 11, 693–695. (2018) 28
- [11] Qi Cai, Wenquan Che, and Kaixue Ma, “A linear gan power amplifier using novel transistor based analog predistortion method”, 1–4. (2016) 28
- [12] H. Deng, D. Zhou, D. Zhang, Y. Zhang, “Analogue predistorter based on non-linear crlhtl”. *Electronics Letters*, **volum**(54), no. 11, pp. 693–695, (2018) 28
- [13] Kevin Chuang, “A Perspective on Linearization and Digital Pre- Distortion for Wireless Radio Systems”. *NanoSemi, Inc.*, (2019) 33
- [14] Damián Jorge Matich, “Redes Neuronales: Conceptos Básicos y Aplicaciones”. *Grupo de Investigación Aplicada a la Ingeniería Química (GIAIQ)*, (2001) 41
- [15] Siqi Wang, Morgan Roger, Julien Sarrazin, “Hyperparameter Optimization of Two-Hidden-Layer Neural Networks for Power Amplifiers Behavioral Modeling Using Genetic Algorithms”. *IEEE Microwave and wireless componets letters*, **volum**(29), no. 12, (2019) 45

- [16] Danh Luongvinh, Youngwoo Kwon, "Behavioral Modeling of Power Amplifiers Using Fully Recurrent Neural Networks". *School of Electrical Engineering, Seoul National University, Korea*, (2005) [52](#)
- [17] Xin Hu, Zhijun Liu, Xiaofei Yu, "Convolutional Neural Network for Behavioral Modeling and Predistortion of Wideband Power Amplifiers". *IEEE transactions on neural networks and learning system* [53](#)
- [18] Dongming Wang, Mohsin Aziz, Mohamed Helaoui, "Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters". *IEEE transactions on neural networks and learning system*, **volum**(30), no. 1, (2019) [53](#), [54](#)
- [19] Taijun Liu, Slim Boumaiza, "Dynamic Behavioral Modeling of 3G Power Amplifiers Using Real-Valued Time-Delay Neural Networks". *IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES*,, **volum**(52), no. 3, (2004) [54](#)
- [20] Chalmers University of Technology, "RF WebLab". *http : //dpdcompetition.com/rfweblab* [79](#)
- [21] Teng Wang, Pere L. Gilibert, "Mesh-Selecting for Computational Efficient PA Behavioral Modeling and DPD Linearization". *IEEE Microwave and Wireless Components Letters*, **volum**(30), no. 1, (2020) [90](#)

APÉNDICES

APÉNDICE A. CÓDIGO DE MATLAB UTILIZADO

En este anexo se muestra el código en Matlab utilizado para generar y entrenar la ANN empleada como DPD y para realizar la linealización del LMBA por medio del modelo GMP y de un modelo basado en redes neuronales.

Para crear la red neuronal y entrenarla para que funcione como DPD, se utilizan datos de la entrada ($x_{BB}[n]$) y datos de la salida del LMBA ($y_{BB}[n]$) de una señal LTE de 200 MHz de ancho de banda. La matriz de entrada a la red para su entrenamiento estará formada por los valores $y_{BB}[n]$ y los valores de salida son las componentes en fase y cuadratura de $x_{BB}[n]$.

```
1 % Cargar los datos a la entrada y a la salida del LMBA
2 xBB = load('xBB.mat');
3 yBB = load('yBB.mat');
4 % Separar fase y cuadratura para los datos de la salida
5 xBB = [real(xBB), imag(xBB)];
6 % Construir matriz con valores de entrada a la red
7 m = 9; p = 9; s = 4;
8 YBB = CreateInputMatrix(yBB, m, p, s);
9 % Generar la red neuronal artificial
10 net = fitnet([20,20,10,10], 'trainlm');
11 % Entrenar la red con los datos del LMBA
12 net = train(net, YBB, xBB);
13 save('net.mat');
```

En la primera parte del código se establece la configuración de la señal así como el método DPD empleado (GMP o ANN) y el tipo de aprendizaje (directo o indirecto).

```
1 MODE_SIGNAL='Single-Band';
2 Fc = 2.0e9;
3 PARAM.MOD.f_RF = Fc;
4 [PARAM]=LMBA_parameters(MODE_SIGNAL);
5 PARAM.Data_type='Test_signal';
6 CFR_type='PC_BB_preDPD';
7 PARAM.DPD.DPD_model='GMP';
8 PARAM.DPD.DPD_Adapt='Directo';
9 Aux_sign_mode='shaping';
10 N_iteations=15;
11 n_AVG=1;
12 % Ganancia de predistorsion
13 PARAM.DPD.GK=1.00;
14 Offset_Percentage=0.0;
15 p=3;
16 max_PAPR_BB=15;
17 Vgs_classC=5.4;
18 GainBB=6.5;
19 set_HMP_ch2(Vgs_classC);
20
21 % Red neuronal artificial como DPD
22 load('net.mat');
23 mu = 0.9
24 m = 9; p = 9; s = 4;
```

Seguidamente se establece el desfase entre señales a partir de la regresión de grado 1

encontrada para optimizar la linealidad.

```

1 %Se cargan los coeficientes obtenidos a partir de las medidas previas
2 Coeff = load('Coeficientes_grado1.mat');
3 Fase = Coeff(1)*Fc + Coeff(2);
4 while(Fase >= 360)
5     Fase = Fase - 360;
6 end

```

Una vez establecidos todos los parámetros involucrados, se crea un bucle hasta acabar las iteraciones donde se genera la señal, se aplica la predistorsión, se envía la señal, se recogen los datos del amplificador y se determinan los valores de NMSE, ACPR, eficiencia y EVM.

```

1 for iteration=1:N_iterations
2 % Signal generation
3     if (iteration==1) || (strcmp(PARAM.Data_type, 'New_data'))
4         [uBB,gBB,sBB,PARAM]= Signal_Generation(PARAM);
5         uBB=uBB/GainBB;
6         [RMS_BB,PEAK_BB,PAPR_BB]= Compute_Signal_Power(uBB); PAPR_BB_it(
            iteration)=PAPR_BB;
7     end
8 % Crest factor reduction
9     if (iteration==1) || (strcmp(PARAM.Data_type, 'New_data'))
10        if (strcmp(CFR_type, 'PC_BB_preDPD'))
11            [uBB_CFR,signalpeak_for_threshold]= PeakCancellation_target_PAPR(PARAM
                ,uBB,max_PAPR_BB,'LP');
12        else
13            uBB_CFR=uBB;
14        end
15    end
16    if (strcmp(PARAM.DPD.DPD_model, 'GMP'))
17        if (iteration==1)
18            w=0; U=0;
19        end
20        %Esta funcion aplica el DPD con modelo GMP
21        [xBB,U,w,PARAM]= Baseband_DPD(uBB_CFR,w,U,PARAM,iteration);
22    end
23    if (strcmp(PARAM.DPD.DPD_model, 'ANN'))
24        if (iteration > 1)
25            UBB_old = [real(uBB_old), imag(uBB_old)];
26            net = train(net,XBB_old,UBB_old);
27        end
28
29        U = CreateInputMatrix(uBB_CFR, m, p, s);
30        YBB = net(U);
31        yBB = complex(YBB(1,:),YBB(2,:));
32        yBB = yBB';
33    end
34
35    % Calculo del PAPR y de la potencia
36    [RMS_dpd,PEAK_dpd,PAPR_dpd]= Compute_Signal_Power(xBB);
37    PAPR_dpd_it(iteration)=PAPR_dpd;
38    if (max(abs(xBB))>1); warning('MAX_xBB_higher_than_1. '); end
39
40    %LMBA signals generation
41    switch(Aux_sign_mode)
42        case 'squared'

```

```

43     Amplitude=max(abs(xBB));
44     xBB_aux2=abs(xBB).^2/max(abs(xBB).^2)*Amplitude.*exp(1i*phase(xBB))
45 case 'linear'
46     xBB_aux2=xBB;
47 case 'shaping'
48     x_min=max(abs(xBB))*Offset_Percentage;
49     xBB_shape=(x_min^6+abs(xBB).^6).^(1/p);
50     Amplitude=max(abs(xBB));
51     xBB_aux2=xBB_shape/max(xBB.shape)*Amplitude.*exp(1i*angle(xBB));
52 end
53 %Se agrega el desfase obtenido con la regresion
54 xBB_aux=xBB_aux2*exp(+1i*pi/180*Fase);
55 % Aplicar retardo en banda base
56 xBB_aux=circshift(xBB_aux,-2);
57
58 %Up-conversion
59 [u_RF]=Upconversion_BB_to_RF(uBB,PARAM,0);
60 [x_RF1]=Upconversion_BB_to_RF(xBB,PARAM,0); x_RF=x_RF1;
61 [x_RF2]=Upconversion_BB_to_RF(xBB_aux,PARAM,0);
62
63 %Envio datos AWG M8190A
64 fprintf ('\nSENDING_data_to_Keysight_AWG_M8190A\n');
65 iqdownload((x_RF1+1i*x_RF2),PARAM.MOD.CLKRF, 'channelMapping', [1 0; 0 1], '
66     run',1);
67
68 %Recepcion datos de DSO90404A
69 fprintf ('\nGETING_data_from_Keysight_DSO90404A\n');
70 if (n_AVG>1)
71     y_RF_tot=zeros(PARAM.DEMOD.LSampling,1);
72     for i=1:n_AVG
73         [DIG]=get_agilentDSO9X(PARAM.DEMOD.CLKRF,PARAM.DEMOD.LSampling, '
74             READ', 'DSO90404A-4G',[1 0 0 0]);
75         ch1_0=DIG.ch1;
76         if (i>1);
77             [ch1_0, delay]=dpd_alignment(y_RF_tot, ch1_0, 'REAL', 'ECHO'); [
78                 ch1_0, ~]=dpd_alignmentfine(y_RF_tot, ch1_0, 20, 'MODULUS', '
79                 ECHO');
80         end;
81         y_RF_tot=y_RF_tot+ch1_0;
82     end
83     ch1=y_RF_tot/n_AVG;
84 else
85     [DIG]=get_agilentDSO9X(PARAM.DEMOD.CLKRF,PARAM.DEMOD.LSampling, 'READ', '
86     DSO90404A-4G',[1 0 0 0]);
87     ch1=DIG.ch1;
88 end
89
90 %Down-conversion de RF a BB
91 fprintf ('\n_Digital_DOWN-Conversion_RF_to_BB\n');
92 [yBB,y_RF,~,~]=Downconversion_RF_to_BB(ch1,u_RF,uBB_CFR,0,0,PARAM);
93
94 %De-channelization
95 fprintf ('\n_Doing_De-channelization_and_EVM_calculation\n');
96 [yBB_chan,y_simb,EVM,PARAM]=Dechannelization_signal_OFDM(yBB,gBB,sBB,PARAM)
97 ;
98
99 %Adaptacion DPD
100 fprintf ('\n_UPDATING_DPD_coefficients\n')

```

```

95  if strcmp(PARAM.DPD.DPD_model, 'OFF') || strcmp(PARAM.DPD.DPD_model, 'ANN')
96  else
97      fprintf ('\n_UPDATING_DPDCoefficients_Composite_Channels_\n');
98      [w,PARAM]=Adaptation_DPD(w,U,yBB,uBB,uBB.CFR,xBB,PARAM,iteration);
99  end
100
101  % Graficas de resultados
102  disp ('PLOTING_Results');
103  Plot_Results_LMBA;
104
105  % Seleccion de datos para el reentrenamiento de la red neuronal
106  if (strcmp(PARAM.DPD.DPD_model, 'ANN'))
107      x_relate=3;
108      train_mode='IQ_Memory_Mesh';
109      bits_num=3;
110      % Reduction factor
111      red_fact=10;
112      x_id = xBB + mu(uBB-yBB);
113      X_id = CreateInputMatrix(x_id, m, p, s);
114      [X_cut, x_cut, LUT_sel] = mesh_select_IQ_LUT(x_id, uBB.CFR, x_relate,
115          bits_num, red_fact, X_id, x_id, train_mode);
115      uBB_old=(nonzeros(LUT_sel.*uBB.CFR));
116      XBB_old = X_cut;
117      xBB_old = x_cut;
118  end
119 end

```