Research Repository UCD

Title	Mixture of Experts Approach for Behavioral Modeling of RF Power Amplifiers
Authors(s)	Brihuega, Alberto, Abdelaziz, Mahmoud, Anttila, Lauri, Li, Yue, Zhu, Anding, Valkama, Mikko
Publication date	2021-01-20
Publication information	Brihuega, Alberto, Mahmoud Abdelaziz, Lauri Anttila, Yue Li, Anding Zhu, and Mikko Valkama. "Mixture of Experts Approach for Behavioral Modeling of RF Power Amplifiers." IEEE, 2021.
Conference details	The 2021 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR 2021), Virtual Conference, 17-20 January 2021
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/12389
Publisher's statement	© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/PAWR51852.2021.9375546

Downloaded 2023-04-07T10:52:29Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

Mixture of Experts Approach for Behavioral Modeling of RF Power Amplifiers

Alberto Brihuega¹, Mahmoud Abdelaziz², Lauri Anttila¹, Yue Li³, Anding Zhu³, and Mikko Valkama¹

¹Department of Electrical Engineering, Tampere University, Tampere, Finland

²Zewail City of Science and Technology, Giza, Egypt

³School of Electrical and Electronic Engineering, University College Dublin, Dublin 4, Ireland

Email: alberto.brihuegagarcia@tuni.fi

Abstract—Piece-wise (PW) behavioral models are commonly adopted when modeling RF power amplifiers (PAs) that exhibit strong amplitude dependent nonlinear distortion characteristics. In this paper, we consider a new PW model for PAs based on the mixture of experts (ME) approach. We first introduce the ME framework theory while also extend it such that it can be applied to model complex baseband signals and nonlinearities. Then, we show how the ME model allows overcoming some of the intrinsic shortcomings that existing PW behavioral models commonly exhibit, which translates into improved modeling accuracy. The proposed solution is validated with extensive RF measurements on a load modulated balanced PA at 2.1 GHz, excited by a 320 MHz OFDM signal, and is benchmarked against several state-of-the-art PW models.

Index Terms—5G new radio, behavioral modeling, power amplifiers, nonlinear distortion, piece-wise polynomials.

I. INTRODUCTION

Over the years, multiple power amplifier (PA) technologies have been developed with the goal of delivering enhanced power efficiency (PE) at different power back-off levels and over wide bandwidths. Good examples are the Doherty PA (DPA) and the load modulated balanced (LMBA) PA, which leverage the concept of load modulation that allows the PE to be optimized dynamically at a specific power back-off, by tuning the load impedance. Due to the operation principle of DPAs and LMBA PAs, their nonlinear distortion characteristics become amplitude dependent. This makes their modeling through classical global polynomials (GPs) very challenging. To tackle this issue, different piece-wise (PW) models have been proposed [1]–[4]. PW models utilize separate submodels that operate over a specific subregion of the overall amplitude response of the PA. Thus, they are capable of conveniently modeling such distinct amplitude dependent behaviour. A vector-switched (VS) model was proposed in [2], and is based on hard partitions of the PA input signal space, which define the range of operation of each of the submodels. Zhu *et al.* proposed in [3] a decomposed piece-wise (DPW) Volterra model, where each signal is decomposed into several sub-signals that are then processed by the different submodels before the sample is reconstructed. A PW behavioral model based on a vector rotation decomposition of the canonical PW linear (CPWL) basis functions (BFs) was proposed in [4], and was shown to require less amount of coefficients when modeling systems with non Volterra-like behavior, e.g.,

as that exhibited by DPA or LMBA PAs. Despite these PW models provide significantly better modeling accuracy than GPs, they have some inherent limitations. For instance, the VS model does not impose any continuity constraint between the submodels, which may limit its performance [1], [4]. The DVR model in [4] considers an approximation of the original CPWL so that the model is linear in parameters, which may limit its performance. Additionally, in general, memory modeling capabilities may be compromised in PW models as the different submodels operate independently, while memory effects may involve samples belonging to different subregions.

In this paper, we propose a new PW behavioral model for PAs based on the mixture of experts (ME) framework [5], [6]. The ME model is a probabilistic framework that allows to combine multiple regression functions, the so-called experts, and make them cooperate with a gating function to solve a regression problem. One distinct feature of the ME model compared to other PW models is the fact that ME utilizes soft partitions of the data. This implies that the submodels work across overlapping regions. This is a very important feature, as it avoids potential non-smooth transitions between submodels, and facilitates the modeling of memory effects between regions. The ME model theory and its extension to the complex baseband case are introduced in Section II, while RF measurement results are provided in Section III.

II. MIXTURE OF EXPERTS FOR PA MODELING

Let $x(n)$ and $y(n)$ denote the I/Q samples of the PA input and output signals, respectively. The output signal can be estimated or approximated as

$$\hat{y} = \Psi(\mathbf{x})\alpha, \quad (1)$$

where $\hat{\mathbf{y}} = [\hat{y}(1), \hat{y}(2), \dots, \hat{y}(N)]^T$, with $\hat{y}(n)$ being an estimate of $y(n)$, and $\mathbf{x} = [x(1), x(2), \dots, x(N)]^T$. $\Psi(\mathbf{x}) \in \mathbb{C}^{N \times B}$ is the matrix containing the regressors of the model, B is the total number of regressors, and $\alpha \in \mathbb{C}^{B \times 1}$ are the model coefficients. Assuming N statistically independent data points and I experts, the ME model can be formulated as the following decomposition of the input/output data [5], [6]

$$P(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^N \sum_{i=1}^I P(\mathbf{z}_i^{(n)} = 1 | \mathbf{x}^{(n)}, \mathbf{v}_i) P(\mathbf{y}^{(n)} | \mathbf{x}^{(n)}, \mathbf{w}_i) \quad (2)$$

Fig. 1: Block diagram of the mixture of experts principle.

where $\mathbf{z}$ is a set of hidden/latent variables, $P(\mathbf{z}_i^{(n)} = 1|\mathbf{x}^{(n)}, \mathbf{v}_i)$ is the gating function of parameters $\mathbf{v}_i$, and measures the probability of the i th expert given the input. $P(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}, \mathbf{w}_i)$ is the probability of the i th expert, with parameters $\mathbf{w}_i$, of generating $\mathbf{y}^{(n)}$. The superscript (n) refers to the n th element of the corresponding vector.

The gating function is defined as a mixture model as follows

$$P(\mathbf{x}|\pi, \mathbf{v}) = \sum_{i=1}^I \pi_i P(\mathbf{x}|\mathbf{v}_i), \quad (3)$$

where π_i are the so-called mixing probabilities, and sum up to one. By invoking Bayes' rule and the total probability theorem, the gating functions $P(\mathbf{z}_i^{(n)} = 1|\mathbf{x}^{(n)}, \mathbf{v}_i)$ can be expressed as

$$P(\mathbf{z}_i^{(n)} = 1|\mathbf{x}^{(n)}, \mathbf{v}_i) = \frac{a_i P(\mathbf{x}^{(n)}|\mathbf{z}_i^{(n)} = 1, \mathbf{v}_i)}{\sum_{j=1}^I a_j P(\mathbf{x}^{(n)}|\mathbf{v}_j)} \quad (4)$$

where $a_i = P(\mathbf{z}_i^{(n)} = 1)$ is the prior probability of the i th expert. $P(\mathbf{x}^{(n)}|\mathbf{v}_i)$ is considered to be a density of the exponentially family, and specifically a Gaussian density, which allows obtaining $\mathbf{v}_i$ in closed form. As the nonlinearities act on the envelope of the signal, the gates are assumed to make soft partitions based on the amplitude of the input signal, denoted as $\mathbf{A}^{(n)} = |\mathbf{x}^{(n)}|$, similar to other PW models [1], [2]. Hence, in the following, $P(\mathbf{z}_i^{(n)} = 1|\mathbf{x}^{(n)}, \mathbf{v}_i)$ will be expressed as a function of $\mathbf{A}^{(n)}$ rather than of $\mathbf{x}^{(n)}$.

As for the experts, they are also commonly chosen from the exponential family so that their parameters can also be obtained in closed form. In this work, it is considered that the experts are Gaussian distributed, i.e., $P(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}, \mathbf{w}_i) = \mathcal{N}(\mathbf{y}^{(n)}|\hat{\mathbf{y}}_i^{(n)}, \sigma_{e_i}^2)$, where $\mathbf{w}_i = \{\hat{\mathbf{y}}_i^{(n)}, \sigma_{e_i}^2\}$, $\hat{\mathbf{y}}_i^{(n)} = \Psi(\mathbf{x})\alpha_i$ is the mean, and $\sigma_{e_i}^2$ is the variance. As $\mathbf{y}^{(n)}$ is complex-valued, the probability density function reads

$$P(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}, \mathbf{w}_i) = \frac{1}{\pi \sigma_{e_i}^2} \exp\left(-\frac{|\mathbf{y}^{(n)} - \hat{\mathbf{y}}_i^{(n)}|^2}{\sigma_{e_i}^2}\right) \quad (5)$$

It is important to note that for a perfect sample estimate, i.e., $\hat{\mathbf{y}}^{(n)} = \mathbf{y}^{(n)}$, $P(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}, \mathbf{w}_i)$ reaches its maximum value.

In order for the ME model to make a single prediction, the expectation of (2) is used, and reads [6]

$$\hat{\mathbf{y}}^{(n)} = \sum_{i=1}^I g_i(\mathbf{A}^{(n)}, \mathbf{v}) \hat{\mathbf{y}}_i^{(n)}, \quad (6)$$

which is a weighted sum of the outputs of the estimates of the individual experts, and where $g_i(\mathbf{A}^{(n)}, \mathbf{v}) = P(\mathbf{z}_i^{(n)} = 1|\mathbf{A}^{(n)}, \mathbf{v}_i)$.

To train the ME model, the expectation maximization (EM) algorithm is usually considered [7]. In order for the gate and expert parameters to be analytically solvable, the densities must belong to the family of the exponential densities, and additionally, instead of the likelihood function in (2), one should consider the joint density [6], which reads

$$P(\mathbf{x}, \mathbf{y}) = \prod_{n=1}^N \sum_{i=1}^I a_i P(\mathbf{A}^{(n)}|\mathbf{v}_i) P(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}, \mathbf{w}_i) \quad (7)$$

Then, the maximum-likelihood is calculated for $\ln P(\mathbf{x}, \mathbf{y}|\mathbf{v}, \mathbf{w})$ and is done by iterating the EM algorithm, which consists of the following two steps [6], [7].

1) *E-step*: In the k th iteration of the E-step, the expectation of the latent variables $h_i^k(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}) = \mathbb{E}\{P(\mathbf{z}_i^{(n)}|\mathbf{y}^{(n)}, \mathbf{x}^{(n)})\}$, which measures the relative probability of $x(n)$ belonging to expert i , commonly referred to as responsibility.

2) *M-step*: Compute the maximum likelihood parameters weighted by the responsibilities [5].

The overall ME principle is summarized in Fig. 1.

III. MEASUREMENT RESULTS

In this section, RF measurements on an in-house designed LMBA GaN PA are conducted in order to demonstrate the capabilities of the proposed ME model. The LMBA PA operates at 2.1 GHz carrier frequency with an average power of +37 dBm and 41% drain efficiency under the stimulus of a 320 MHz OFDM waveform composed of 16 20 MHz component

Fig. 2: (a) Hard and soft partitions provided by K -means and the ME model, respectively; (b) measured and modeled AM/AM and AM/PM responses; (c) measured and modeled spectra by different PW models.

carriers, further details on the PA design and its characteristics can be found in [8]. The sampling frequency of the signal is 1.2 Gsamples/s and its sample-level PAPR measured at 10^{-4} CCDF reads ca. 8 dB.

The ME model is assumed to utilize three experts, each of them utilizing generalized memory polynomial (GMP) [9] BFs with $P = 7$, $M = 7$ and $G = 4$. The VS and DPW reference models are considered to utilize three regions, given by K -means, and have the same parameterization as ME, while the GP model utilizes $P = 9$, $M = 12$ and $G = 5$. The DVR CPWL model utilizes fifteen uniformly spaced regions, and memory depth of $M_{CPWL} = 12$ [4]. The PA output data is recorded, taken to baseband and synchronized to the digital waveform. Then, the reference models are fitted by utilizing a least-squares approach, while the proposed ME model is fitted by iterating the EM algorithm until convergence.

The hard partitions provided by the K -means algorithm utilized e.g., in [2] as well as the soft partitions given by the ME model after the convergence of the EM algorithm are illustrated in Fig. 2(a). The vertical blue lines define the amplitude intervals over which the different submodels operate. On the other hand, the soft partitions should be interpreted as how much a given expert contributes – from zero to one – to generating a given output sample. In the context of hard partitions, these contributions are either one or zero. The different experts in the ME model can be interpreted as GP that learn to specialize.

The performance of the different models is compared in terms of the normalized-mean-square-error (NMSE) and the adjacent channel error power ratio (ACEPR), and are given in Table I along with the number of coefficients of each model. As it can be seen, the modeling accuracy of the GP falls significantly behind the accuracy of the PW models due to the strong amplitude dependent characteristics of the LMBA PA. The best modeling accuracy is provided by the proposed ME model, both for the NMSE and the ACEPR metrics, as a result of its improved memory-modelling capabilities.

IV. CONCLUSIONS

In this paper, we have proposed a new PW behavioral model for PAs based on the ME framework. The ME model utilizes

TABLE I: MODELING ACCURACY OF DIFFERENT CONSIDERED MODELS IN TERMS OF NMSE, ACEPR AND AMOUNT OF MODEL PARAMETERS

	No. parameters	NMSE (dB)	ACEPR (dB)
GP model	585	-30.43	-31.31
VS model	672	-32.24	-32.42
DPW model	672	-31.55	-33.43
DVR CPWL model	663	-32.37	-33.89
ME model	672	-32.97	-34.42

soft partitions of the data, which implies that the different submodels overlap with one another. This feature greatly contrasts with other PW models wherein the partitions are disjoint and the models operate and are learned independently. ME was shown to be capable of providing the best modeling accuracy amongst the tested models, which are very promising results. In future work, we will consider the ME model in the context of digital predistortion.

REFERENCES

- [1] A. Brihuega et al., "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4000–4017, 2020.
- [2] S. Afsardoost et al., "Digital predistortion using a vector-switched model," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 4, pp. 1166–1174, April 2012.
- [3] A. Zhu et al., "Digital predistortion for envelope-tracking power amplifiers using decomposed piecewise Volterra series," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 10, pp. 2237–2247, 2008.
- [4] A. Zhu, "Decomposed Vector Rotation-Based Behavioral Modeling for Digital Predistortion of RF Power Amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 737–744, 2015.
- [5] L. Xu et al., "An alternative model for mixtures of experts," *Advances in Neural Information Processing Systems* 7, pp. 633–640, 1995.
- [6] S. E. Yuksel et al., "Twenty Years of Mixture of Experts," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1177–1193, 2012.
- [7] A. P. Dempster et al., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [8] J. Pang et al., "Broadband RF-input continuous-mode load-modulated balanced power amplifier with input phase adjustment," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 10, pp. 4466–4478, 2020.
- [9] D. R. Morgan et al., "A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.