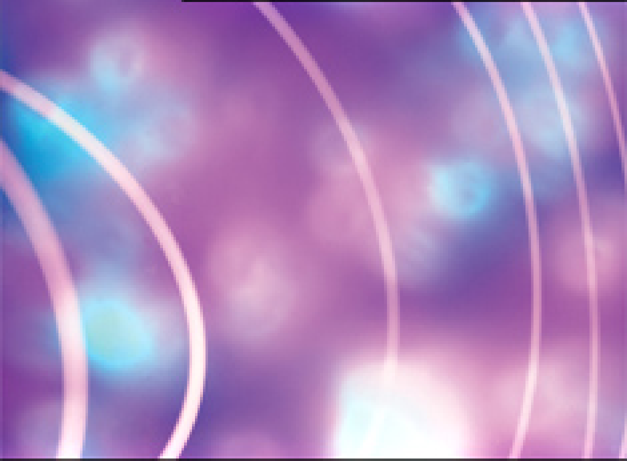
Edited by Dominique Schreurs, Máirtín O'Droma, Anthony A. Goacher, and Michael Gadringer



THE CAMBRIDGE RF AND MICROWAVE ENGINEERING SERIES

RF Power Amplifier Behavioral Modeling

TARGET

CAMBRIDGE

www.cambridge.org/9780521881739

This page intentionally left blank

RF POWER AMPLIFIER BEHAVIORAL MODELING

If you are an engineer or RF designer working with wireless transmitter power amplifier models, this comprehensive and up-to-date exposition of nonlinear power amplifier behavioral modeling theory and techniques is an absolute must-have. Including a detailed treatment of nonlinear impairments, as well as chapters on memory effects, simulation aspects for implementation in commercial system and circuit simulators, and model validation, this one-stop reference makes power amplifier modeling more accessible by connecting the mathematics with the practicalities of RF power amplifier design. Uniquely, the book explains how systematically to evaluate a model's accuracy and validity, compares model types, and offers recommendations as to which model to use in which situation.

DOMINIQUE SCHREURS is Associate Professor in the ESAT-TELEMIC Division, Department of Electrical Engineering, Katholieke Universiteit Leuven, where she also gained her Ph.D. in Electrical Engineering in 1997. She is a Senior Member of the IEEE and was Chair of the IEEE MTT-11 technical committee on microwave measurements. She was a steering committee member of TARGET (Top Amplifier Research Groups in European Team).

MÁIRTÍN O'DROMA is Director of the Telecommunications Research Centre, and Senior Lecturer in the Department of Electronic and Computer Engineering at the University of Limerick. A Fellow of the IET and Senior Member of the IEEE, he was a founding partner and steering committee member of TARGET and a section head of the RF power linearization and amplifier modeling research strand.

ANTHONY A. GOACHER is Research Projects Manager of the Telecommunications Research Centre, University of Limerick. He has an MBA, is a Member of the IET, an Associate Member of the Institute of Physics, and has held a senior management position in the electronics industry for 20 years.

MICHAEL GADRINGER is a Research Assistant in the Institute of Electrical Measurements and Circuit Design, Vienna University of Technology. He is currently involved with power amplifier modeling, linearization and device characterization.

The Cambridge RF and Microwave Engineering Series

Series Editor,
Steve C. Cripps

Peter Aaen, Jaime A. Plá, and John Wood, *Modeling and Characterization of RF and Microwave Power FETs*
Enrico Rubiola, *Phase Noise and Frequency Stability in Oscillators*
Dominique Schreurs, Máirtín O'Droma, Anthony A. Goacher, and Michael Gadringer, *RF Amplifier Behavioral Modeling*
Fan Yang and Yahya Rahmat-Samii, *Electromagnetic Band Gap Structures in Antenna Engineering*

Forthcoming
Sorin Voinigescu and Timothy Dickson, *High-Frequency Integrated Circuits*
Debabani Choudhury, *Millimeter Waves for Commercial Applications*
J. Stephenson Kenney, *RF Power Amplifier Design and Linearization*
David B. Leeson, *Microwave Systems and Engineering*
Stepan Lucyszyn, *Advanced RF MEMS*
Earl McCune, *Practical Digital Wireless Communications Signals*
Allen Podell and Sudipto Chakraborty, *Practical Radio Design Techniques*
Patrick Roblin, *Nonlinear RF Circuits and the Large-Signal Network Analyzer*
Dominique Schreurs, *Microwave Techniques for Microelectronics*
John L. B. Walker, *Handbook of RF and Microwave Solid-State Power Amplifiers*

RF POWER AMPLIFIER BEHAVIORAL MODELING

DOMINIQUE SCHREURS
Katholieke Universiteit Leuven

MÁIRTÍN O'DROMA
University of Limerick

ANTHONY A. GOACHER
University of Limerick

MICHAEL GADRINGER
Vienna University of Technology

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521881739

First published in print format 2008

ISBN-13 978-0-511-43426-6 eBook (Adobe Reader)

ISBN-13 978-0-521-88173-9 hardback

Contents

Notation

$f(\cdot)$	general nonlinear function
$f_{\mathrm{R}}(\cdot)$	nonlinear extension of an IIR digital filter
$f_{\mathrm{D}}(\cdot)$	nonlinear extension of an FIR digital filter
t	continuous-time variable
s	discrete-time variable
T_{s}	sampling time
x	RF input signal or general input signal
y	RF output signal or general output signal
$\tilde{x}$	low-pass complex-envelope input signal
$\tilde{y}$	low-pass complex-envelope output signal
r	envelope amplitude
P	in-phase component of the envelope amplitude
Q	quadrature component of the envelope amplitude
ϕ	envelope phase
ω	angular frequency variable
f	frequency variable
ω_0	carrier angular frequency
f_0	carrier frequency
$G(\cdot)$	RF memoryless nonlinearity
$g(\cdot)$	AM–AM nonlinearity
$\Phi(\cdot)$	AM–PM nonlinearity
Re	real-part operator
Im	imaginary-part operator
$*$	convolution operator

Abbreviations

ACEPR	adjacent-channel error power ratio
ACI	adjacent-channel interference
ACLR	adjacent-channel leakage ratio
ACPR	adjacent-channel power ratio
ADC	analogue-to-digital converter
ADS	advanced design system
AM–AM/AM–PM	AM–AM and AM–PM model or characterisation
ANN	artificial neural network
APK	amplitude phase-shift keying
ARMA	autoregressive moving average
AWG	arbitrary waveform generator
AWR-MO	Applied Microwave Research's Microwave office
AWR-VSS	Applied Microwave Research's Visual system simulator
BBACS	broadband amplifier characterisation setup
BER	bit error rate
BF	Bessel–Fourier
BPSK	binary phase-shift keying
C/I	carrier-to-intermodulation ratio
C3IM	carrier-to-third-order intermodulation product ratio
CAD	computer-aided design
CCDF	complementary cumulative density function
CDMA	code-division multiple access

CW	continuous-wave
DAC	digital-to-analogue converter
DIDO	dual-input–dual-output
DSO	digital storage oscilloscope
DSP	digital signal processing
DTD	direct-time domain
DUT	device under test
EIRP	equivalent isotropic radiated power
ESDA	electronic system design automatisation
EVM	error-vector magnitude
FCC	Federal Communications Commission
FDMA	frequency-division multiple access
FET	field-effect transistor
FFT	fast Fourier transform
FIR	finite impulse response
FOBF	Fourier-series-optimised Bessel–Fourier
FOM	figure of merit
FPGA	field-programmable gate array
GSM	global system for mobile communications
HB	harmonic balance
HEMT	high-electron-mobility transistor
IBO	input power backoff
IC	integrated circuit
IF	intermediate-frequency
IFFT	inverse fast Fourier transform
IIR	infinite impulse response
IMD	intermodulation
IMP	intermodulation product
IP	intercept point
IRF	impulse response function

I–Q	in-phase–quadrature
IS-95	Interim Standard 95, also known as cdmaOne
LDMOS	laterally diffused metal oxide semiconductor
LF	low-frequency
LNA	low-noise amplifier
LS	least-squares
LSNA	large-signal network analyser
LTI	linear time-invariant
MDL	minimum description length
MESFET	metal–semiconductor field-effect transistor
MFTD	mixed frequency- and time-domain signal representation
MIMO	multiple-input–multiple-output
MLP	multilayer perceptron
MS	modified Saleh
NARMA	nonlinear autoregressive moving-average
NARMAX	nonlinear autoregressive moving-average with exogenous input
NIM	nonlinear integral model
NL	nonlinear
NMSE	normalised mean-square error
NOCEM	non-constant envelope modulated
NPR	noise power ratio
NVNA	nonlinear vector network analyser
OBO	output power backoff
ODE	ordinary differential equation
OFDM	orthogonal frequency-division multiplexing
OOK	on/off keying
PA	power amplifier
PAE	power-added efficiency
PAPR	peak to average power ratio
PCB	printed circuit board

PDF	probability density function
PL	percentage linearisation
PSB	Poza–Sarkozy–Berger
PSD	power spectral density
QAM	quadrature amplitude modulation
QPSK	quadrature phase-shift keying
RF	radio-frequency
RMS	root-mean-square
RRC	root-raised cosine
SAW	surface acoustic wave
SER	symbol-error rate
SISO	single-input–single-output
SNR	signal-to-noise ratio
SSPA	solid-state power amplifier
SVD	singular-value decomposition
SWANS	scalable wireless ad hoc network simulator
TDNN	time-delay neural network
TWTA	travelling-wave tube amplifier
UML	universal modelling language
VAF	variance accounted for
VCO	voltage-controlled oscillator
VDHL	very high speed IC hardware description language
VIOMAP	Volterra input–output map
VNA	vector network analyser
VSA	vector signal analyser
VSG	vector signal generator
WCDMA	wideband code-division multiple access

Preface

This book provides a comprehensive treatment of radio-frequency (RF) nonlinear power amplifier behavioural modelling, from the fundamental concepts and principles through to the range of classical and, especially, current modelling techniques.

The continuing rapid growth of wireless communications and radio transmission systems, with their ever increasing sophistication, complexity and range of application, has been paralleled by a similar growth in research into all aspects of electronic components, systems and subsystems. This has given rise to a great variety of new and advanced technologies catering for the breadth of frequencies, bandwidths and powers expected in new and existing air interfaces and in the mobile wireless world, for the ever increasing integration of widely differing interfaces into single devices, with the future likelihood that these devices will be active on two or more interfaces simultaneously. For radio communications, or simply radio transmission systems, from the high-frequency (HF) band to the microwave and millimetrewave bands, the transmitter power amplifier (PA) is a pivotal enabling component. This is especially apparent when setting and satisfying air-interface specifications, the correct transmitted signal power levels and tolerable levels of in-band and out-of-band signal impairment. The reason for this high-profile role of the PA is that it is the major source of signal distortion and spurious signal generation, harmonics and intermodulation products. Further, it is by far the greatest energy-consuming component in the radio transmission path. Depending on the class of amplifier and the operating conditions dictated by the complexity of the signals to be amplified, its DC to RF power-conversion efficiency is generally poor, resulting in power wastage. As this wastage occurs mostly through heat dissipation, in many situations active extraction of this heat through cooling systems is necessitated, which in turn leads to further energy costs. Hence, in all applications, a reduction in energy consumption and heat dissipation through improved efficiency of the PA is a desired goal.

Technically, this increase in PA efficiency is usually achieved at the expense of increased nonlinear distortion effects. Predicting, assessing and quantifying the impact of these detrimental effects on the transmitted signals and on the radio environment requires accurate behavioural models of power amplifiers on the one hand and a detailed knowledge of the radio characteristics of the environment on the other. Accurate behavioural models are also required to support research into nonlinear impairment-reduction techniques (such as power amplifier linearisation), efficiency-improvement techniques, full transmitter and communications-link system design, investigation into new wireless communications systems and concepts (with new signal-modulation techniques and multiple-access techniques) and so forth. For such reasons RF nonlinear PA behavioural modelling has grown to become a topic of great interest for all those involved in radio communications engineering.

It is hoped that this book will provide RF research engineers in industry, research institutes and centres and also students and academics with a comprehensive resource covering this major area of wireless communications research and development engineering. Although there is an abundant literature covering different PA behavioural modelling approaches, mainly comprising specialist journals and books of international conference proceedings, there are few works dedicated to their comprehensive treatment, analysis and comparison. This work seeks to fill this gap, bringing together much of the classical treatment and modern conceptual, theoretical and algorithmic developments.

The theoretical foundations for PA behavioural modelling are presented in Chapter 1. This is a systematic overview and comparative assessment of the various approaches to RF power amplifier modelling that have received widespread attention by the scientific community. The chapter is organised into three sections, on power amplifier modelling basics, system-level power amplifier models and circuit-level power amplifier models. In the first section, a theoretical foundation to support the subsequent PA model classification and analysis is set out. The approach is to address the physical and behavioural modelling strategies and then to classify behavioural models as either static or dynamic with varying levels of complexity. Then a distinction is made between heuristic and systematic approaches, hence creating a theoretical framework for comparing different behavioural model formats with respect to their formulation, extraction and, in most cases, predictive capabilities.

Approaches to PA representations for use in system-level simulators are treated in the second section, on system-level power amplifier models. These are analytic signal- or complex-envelope-based techniques, leading to single-input–single-output (SISO) low-pass equivalent models, whose input and output are the complex functions needed to represent the bidimensional nature of amplitude and phase modulation.

The final section of this chapter, on circuit-level power amplifier models, provides an overview of behavioural models intended for use in conventional PA circuit simulators. Representing the voltage, current or power-wave signals as real entities, these models handle the complete, and computationally demanding – because of the different RF and envelope signal time scales involved – input and output signal dynamics. This includes taking into account the signals' harmonic content and, possibly, the input and output mismatches and other physical circuit features.

Having introduced and classified models according to their mathematical structures, in Chapter 2 we address other important properties and classifications of PA behavioural models that, in one way or another, are not directly related to the models themselves. These arise out of experimentally observed PA characteristics and may be grouped into those properties derived from the model structure, those introduced by the PA modelling application and those reflecting the behaviour of the observed amplifier under a specific excitation. Some of these properties may describe the same model characteristic but from different perspectives. This is borne in mind in the approach to their treatment here, where the aim is to provide an integrated and complete overview of behavioural models based on their properties.

Models extracted from amplifier measurements are optimised to mimic the behaviour seen in these measurements, i.e. to produce identical or near identical behavioural results. Such models, therefore, reflect influences of the particular amplifier characterisation technique. Hence an overview of typical amplifier measurement setups together with a compilation of models extracted by these means completes the treatment of PA properties in Chapter 2. The next three chapters deal with memoryless models, models with linear memory and models with nonlinear memory.

In Chapter 3 memoryless nonlinear PA behavioural models are considered; the most popular models presented and investigated are the complex power series expansion, the Saleh model in both polar and quadrature forms and the Bessel–Fourier model. Other models considered are the Fourier model, the Hetrakul and Taylor model, the Berman and Mahle model and the Wiener-based polynomial models. Static envelope characteristics, i.e. the static AM–AM and/or AM–PM characteristics, are taken as the basis for defining a behavioural model as memoryless. Some of these models are well established, though new developments and new insights keep occurring. An example of the latter, included in this chapter, is the new modified Saleh model, developed to overcome some particular weaknesses of the original Saleh model. Generally, a comparative approach is taken in parallel with the exposition of the models. All are applied to a particular memoryless-equivalent AM–AM and AM–PM (AM–AM/AM–PM) characterisation of an LDMOS amplifier amplifying a WCDMA signal, the predicted results being set against actual PA measurements. Other model aspects addressed comparatively include implementation and complexity, intermodulation product decomposition and harmonic handling capacity.

These conventional nonlinear memoryless models, based on static AM–AM and AM–PM representations, are frequency independent and can represent with reasonable accuracy the characteristics of various amplifiers driven by narrowband input signals. However, if an attempt is made to amplify 'wideband' signals, where the bandwidth of the signal is comparable with the inherent bandwidth of the amplifier, a frequency-dependent behaviour will be encountered. This phenomenon is described as a memory effect. The range of memory effects found in modern PA systems, especially higher-power solid state PAs, may be classified as linear or nonlinear or as short or long term. Knowing when and where these arise and how they contribute to system impairment is important to designers and researchers. Approaches to behavioural modelling that take account of both nonlinearities and memory effect phenomena is the theme of Chapters 4 and 5.

In Chapter 4 we focus on investigating those nonlinear models that handle memory effects (i.e. frequency-dependent behaviour) using linear filters. The models described in this chapter are structurally categorised into two-box, three-box or parallel-cascade structure. The two-box models presented are the Wiener and Hammerstein topologies, while the three-box models include the Poza–Sarkozy–Berger (PSB) model and the frequency-dependent Saleh model. These models represent some first attempts at extending the nonlinear static AM–AM models and AM–PM models to cover frequency-dependent effects. The parallel-cascade models presented in this chapter are the Abuelma'atti and polyspectral models. In these a parallel branch structure is used to describe linear memory.

Significantly more challenging is the behavioural modelling of nonlinear PAs that exhibit nonlinear memory effects. Chapter 5 contains a comprehensive overview of this topic and addresses memory polynomial models, the time-delay neural network (TDNN) model, the nonlinear autoregressive moving-average (NARMA) model, the parallel-cascade Wiener model, Volterra-series-based models and the state-space-based model. The simplest modelling approach is the memory polynomial. The introduction of non-uniform time-delay tabs yields better results. The TDNN and NARMA approaches are strongly related to the memory polynomial. In the TDNN model the memoryless nonlinear network is described by an artificial neural network. In the case of the NARMA model, the output depends not only on past values of the input but also on past values of the output. As stability may be an issue, criteria are derived to check for this.

Another way to model nonlinear PAs with nonlinear memory effects is by an extension of the Wiener modelling approach. By introducing parallel branches consisting of a linear time-invariant system followed by a memoryless nonlinear system, nonlinear memory effects can be modelled adequately. The Volterra-series-based models form a large class of models with nonlinear memory. The difficulty in computation and optimisation of the fitting parameters, e.g. the Volterra kernels, of the analytical functions for dynamic (envelope-frequency-dependent) input–output measured data is addressed. It is notable how the complexity of the model increases with increasing memory and nonlinearity order, requiring the extraction of an ever larger number of coefficients to achieve an adequate approximation. A number of extended approaches have been developed to overcome this intrinsic disadvantage of Volterra series models. A parallel FIR-based model has a reduced computational complexity. The Laguerre–Volterra modelling approach yields a reduction in the number of model parameters. The modified or dynamic Volterra model aims to handle higher levels of nonlinearity. Finally, a relationship between Volterra models and TDNN models is presented.

Memory polynomial models and Volterra-series-based models of a lower degree are only really efficient for systems with memory but which are weakly nonlinear. However, state-space-based behavioural models are not so restricted. The dynamics of the PA are determined directly from time-series data, resulting in a compact, accurate and transportable model. Ways in which multisine excitations can render model development more efficient are also presented.

In Chapter 6 PA model validation and comparison are addressed. As PAs are in general complex dynamic systems that combine both short- and long-term memory effects with nonlinear phenomena, there is quite a variety of ways to approach questions of validation and comparison. In contrast with linear systems with memory, where superposition holds, nonlinear dynamical systems must be 'locally' modelled and validated. Therefore, test signals and model comparison criteria must be carefully chosen to suit a particular set of typical operating conditions. In this chapter we set down suitable figures or characteristics of merit (metrics) for model performance comparison in different telecommunication-application contexts. Our overall goal is to present concepts in ways that will help a reader to formulate suitable figure(s) of merit for his or her application.

A two-part approach is taken. General figures of merit (FOMs) are presented first and the main concepts regarding their applicability are explained. Although most of the proposed metrics can be generalised for sampled and or stochastic signals, only deterministic continuous-time signals are considered here. Starting from a general time-domain metric, several variants are proposed each of which is specially suitable for a certain measurement setup. Then more realistic applications are considered. Here most of the proposed figures of merit are formulated for sampled (i.e. discrete-time) signals and in terms of statistical measures such as the covariance and the power spectral density. The stochastic-process approach is seen as potentially useful for modern measurement instruments and system simulators, where complex telecommunication standards test signals are usually characterised statistically. This part of Chapter 6 includes an application example in which different figures of merit are compared.

Simulation tools are widely used for designing and analysing complex communications systems. In Chapter 7, the final chapter, an overview of aspects of system simulation is provided with a view to the integration of RF power amplifier behavioural models into such simulations. Generally communications simulation tools seek to describe the operating characteristics and performances of a complete communications link, whether simple or complex, and to mimic through mathematical models all the analogue and digital signal-processing activities that occur in the real system, whether at baseband, intermediate or radio frequencies. Here distinctions between the different forms of simulation encountered in the telecommunication field are made and examples of the associated software products, mainly commercial ones, are presented. In this way the kind of full-system simulations that are relevant to the behavioural modelling of RF power amplifiers is highlighted.

Following this, in Section 7.3, a general overview of analogue signal behavioural simulators for wireless communication systems, together with figure of merit considerations in behavioural simulations, is presented. First, an explanation of some relevant simulation terminology is given. In this overview distinctions are made between circuit-level and system-level simulations, both of which are closely allied in RF PA behavioural modelling. For the former, harmonic-balance simulation, circuit-envelope simulation and mixed-signal high-frequency IC circuit-level simulation are briefly described and the respective contexts of their application set out.

As this book's focus is on system-level simulation, the latter part of Chapter 7 is concerned mainly with aspects relevant to the theme of the book. These include analogue signal representation, sampling and processing considerations, sampling rate issues – including multirate sampling – and signal decomposability. Continuous-in-time and finite-time-window time-domain simulation modes are also considered. An example of a general schema for the computation flow and execution of a communications-link simulation at system level is also given. This could be considered to be a heterogeneous simulation, or in this case 'co-simulation', as it integrates digital-logic and analogue signal system-level models of computation.

This book is the product of a significant integrated collaborative effort by many researchers from a wide range of research centres and universities across Europe. This was possible because of the proactive infrastructural support provided under TARGET (Top Amplifier Research Groups in a European Team), one of the European Networks of Excellence 2004–2007 (www.target-org.net), headed by Professor Gottfried Magerl of the Vienna University of Technology. Naturally, the book editors and all the contributors acknowledge this invaluable support. Full details of all authors are listed at the end of the book. The editors would like to express their thanks to the book reviewers, and to all the authors for their patient detailed revision of texts and other contributions.

1 Overview of power amplifier modelling

1.1 Introduction

This chapter presents an overview and comparative assessment of the various approaches to RF power amplifier (PA) modelling that have received widespread attention by the scientific community. The chapter is organised into three sections: power amplifier modelling basics, system-level power amplifier models and circuit-level power amplifier models.

Section 1.2 on power amplifier modelling basics provides the basic knowledge to support the subsequent PA model classification and analysis. First, physical and behavioural modelling strategies are addressed and then behavioural models are classified as either static or dynamic with varying levels of complexity. Then, a distinction is made between the heuristic and systematic approaches, hence creating a theoretical framework for comparing different behavioural model formats with respect to their formulation, extraction and, in most cases, predictive capabilities.

In Section 1.3, dedicated to system-level power amplifier models, PA representations intended to be used in system-level simulators are considered. These are analytic signal- or complex-envelope-based techniques; they do not represent the RF carrier directly and RF effects are not specifically included. They are single-input–single-output (SISO) low-pass equivalent models, whose input and output constitute the complex functions needed to represent the bidimensional nature of amplitude and phase modulation.

The final section, on circuit-level power amplifier models, provides an overview of behavioural models intended for use in conventional PA circuit simulators. These models handle the complete input and output RF modulated signals, which are real entities, at two different time scales, one, very fast, for the RF carrier and another, much slower, for the modulating envelope. So, in contrast with system-level models, they also take into account the signals' harmonic content and, possibly, the input and output mismatches. For that, they need to represent the voltage and current, or incident and reflected power waves, of the PA input and output ports, thus becoming two-input–two-output model structures.

Although there is an abundant literature on the various different PA behavioural modelling approaches, there are only a few works dedicated to their analysis and comparison. A widely known reference in this field is the book of Jeruchim *et al.* [1]. More recently, a book edited by Wood and Root [2] and the papers of Isaksson *et al.* [3] and of Pedro and Maas [4] have appeared. This introduction draws from all these four references but follows the last most closely.

1.2 Power amplifier modelling basics

Power amplifiers have a major effect on the fidelity of wireless communications systems, which justifies the large number of studies undertaken to understand their limitations and then to optimise their performance. Although some earlier studies simply consisted of empirical observations of PA input–output behaviour, later works have applied scientific theories to account for the observed behaviour and, hence, to justify the resulting PA models [1–8]. Seen from the more general context of system identification, PA models can be divided into two major groups according to the type of data needed for their extraction: physical models and empirical models [9].

Physical models require knowledge of the electronic elements that constitute the PA, their relationships and the theoretical rules describing their interactions. They use nonlinear models of the PA active device and of the other, passive, components (these models may themselves be of a physical or empirical nature) to form a set of nonlinear equations relating the terminal voltages and currents. Using an equivalent-circuit description (typically having an empirical nature) of the PA, these models are appropriate to circuit-level simulation and provide a result accuracy that is, nowadays, limited almost only by the quality of the active device model. Unfortunately, such precision has a high price in simulation time and the need for a detailed description of the PA internal structure.

When such a PA equivalent circuit is not available, or whenever a complete system-level simulation is desired, PA behavioural models are preferred. Since they are solely based on input–output (behavioural) observations, their accuracy is highly sensitive to the adopted model structure and the parameter extraction procedure. So, it is no surprise that distinct model topologies and different observation data sets may lead to a large disparity in model applicability and simulation results. In fact, though such a behavioural-modelling approach may guarantee the accurate reproduction of the data set used for its extraction, or, possibly, of some other set pertaining to the same excitation class, it is not obvious that it will also produce useful results for a different data set, a different PA of the same family or a PA based on a completely different technology. That is, in contrast with the physical-modelling alternative, the generalisation of the predictive capability of a behavioural model should always be viewed with circumspection.

1.2.1 *Nonlinear system identification background*

In order to establish a theoretical framework with which to analyse the various approaches to PA behavioural modelling, it is convenient to recall some basic results of system identification theory.

In that framework, our power amplifier is described either by a nonlinear function or a system operator; it is assumed to be either static or dynamic respectively. In the static case its output $y(t)$ can be uniquely defined as a function of the instantaneous input $x(t)$, and the model reduces to

$$y(t) = f(x(t)) \tag{1.1}$$

or

$$y = f(x), \tag{1.2}$$

since the dependence with time is, in this case, immaterial.

When the PA presents memory effects to either the modulated RF signal or the modulating envelope, it is said to be dynamic. The output can no longer be uniquely determined from the instantaneous input. It now depends also on the input past and/or the system state. The relation between $y(t)$ and $x(t)$ cannot be modelled simply by a function but becomes an operator that maps a function of time $x(t)$ onto another function of time $y(t)$. Thus the input–output mapping of our PA is represented by a forced nonlinear differential equation,

$$f\left(y(t), \frac{d\,y(t)}{d\,t}, \ldots, \frac{d^p\,y(t)}{d\,t^p}, x(t), \frac{d\,x(t)}{d\,t}, \ldots, \frac{d^r\,x(t)}{d\,t^r}\right) = 0. \tag{1.3}$$

This states that the output and its time derivatives (in general, the system state) may be nonlinearly related to the input and its time derivatives. Since our PA behavioural models have to be evaluated in a digital computer, i.e. a finite-state machine, it is convenient to adopt a discrete-time environment, in which the time variable becomes a succession of uniform time samples of convenient sampling period T_s; thus the time and the continuous time signals may be translated as $t \to sT_\mathrm{s}$, $x(t) \to x(s)$ and $y(t) \to y(s)$, $s \in Z$. In this way, the solution of the nonlinear differential equation in Equation (1.3) can be expressed in the following recursive form [10]:

$$y(s) = f_\mathrm{R}(y(s-1), \ldots, y(s-Q_1), x(s), x(s-1), \ldots, x(s-Q_2)). \tag{1.4}$$

Here $y(s)$, the present output at time instant sT_s, depends in a nonlinear way, dictated by f_R, the nonlinear function, on the system state (herein expressed by $y(s-q), q = 1, \ldots, Q_1$), the present input $x(s)$ and its past values, $x(s-q)$. This nonlinear extension of infinite impulse response digital filters [10] (nonlinear IIR) is assumed to be the general form for recursive PA behavioural models.

System identification results have shown that, under a broad range of conditions [10–12] (basically operator causality, stability, continuity and fading memory), such a system can also be represented with any desirable small error by a non-recursive, or direct, form, where the relevant input past is restricted to $q \in \{0, 1, 2, \ldots, Q\}$, the so-called system memory span [10]:

$$y(s) = f_\mathrm{D}(x(s), x(s-1), \ldots, x(s-Q)) \tag{1.5}$$

in which $f_\mathrm{D}(\cdot)$ is again a multidimensional nonlinear function of its arguments. This nonlinear extension of finite impulse response digital filters [10] (nonlinear FIR), is again the general form that a direct, or feedforward, behavioural model should obey.

Various forms have been adopted for the multidimensional functions $f_{\mathrm{R}}(\cdot)$ and $f_{\mathrm{D}}(\cdot)$, although two of these have received particular attention in nonlinear system identification. This is due to their formal mathematical support and because they lead directly to a canonical realisation and so to a certain model topology. These two forms are polynomial filters [10–14] and artificial neural networks (ANNs) [15–17].

In the first case, $f_{\mathrm{D}}(\cdot)$ is replaced by a multidimensional polynomial approximation, so that Equation (1.5) takes the form

$$\begin{aligned} y(s) &= P_{\mathrm{D}}\left(x(s), x(s-1), \ldots, x(s-Q)\right) \\ &= \sum_{q=0}^{Q} a_1(q)x(s-q) + \sum_{q_1=0}^{Q}\sum_{q_2=0}^{Q} a_2(q_1, q_2)x(s-q_1)x(s-q_2) + \cdots \\ &\quad + \sum_{q_1=0}^{Q} \cdots \sum_{q_N=0}^{Q} a_N(q_1, \ldots, q_N)x(s-q_1)\cdots x(s-q_N). \end{aligned} \tag{1.6}$$

This form shows that the nonlinear system is approximated by a series of multilinear terms. Although simple in concept, this ‘polynomial FIR’ model architecture is known for its large number of parameters.

The function $f_{\mathrm{R}}(\cdot)$ can also be replaced by a multidimensional polynomial leading to recursive polynomial IIR structures. These provide similar approximation capabilities for many fewer parameters than the direct topology. However, the polynomial IIR is significantly more difficult to extract than the direct topology; this has impeded its application in the PA modelling field.

Indeed, the comparative ease of extraction of the polynomial FIR, in comparison with other PA models, provides its particular and attractive advantage. Since the output is linear in respect of the model parameters, i.e. the kernels $a_n(q_1, \ldots, q_n)$, and dependent only on multilinear functions of the delayed versions of the input, it can be extracted in a systematic way using conventional linear identification procedures.

If $f_{\mathrm{D}}(\cdot)$ or $P_{\mathrm{D}}(\cdot)$ is approximated by a Taylor series then this FIR filter is known as a Volterra series or Volterra filter [10–14]. This Volterra series approximation is particularly interesting as it produces an optimal approximation (in a uniform-error sense) near the point where it is expanded. Therefore it shows good modelling properties in the small-signal, or mildly nonlinear, regimes. However, it shows catastrophic degradation under strong nonlinear operation.

In fact, $f_{\mathrm{D}}(\cdot)$ can be replaced by any other multidimensional polynomial. For example, the Wiener series is orthogonal for white Gaussian noise as an excitation signal [13, 14]; other orthogonal polynomials have been proposed for other excitations [10, 13, 18, 19]. In these cases, the respective series produce results that are optimal (in a mean-square-error sense) in the vicinity of the power level used and for the particular type of input used in the model extraction. These representations are, therefore, amenable to the modelling of strong nonlinear systems when the excitation bandwidth and statistics can be considered close to those used in extraction experiments. A presentation of Wiener series expansions and their orthogonality under white Gaussian noise excitation is given in Section 3.11.

Such polynomial FIR filters can be realised in the form indicated in Figure 1.1. The multiplicity of nth-order cross products between all delayed inputs may be noted; it is to these that the nonlinear filter owes its notoriously complex, although general, form. In a similar way, polynomial IIR filters can be realised. A bilinear, recursive, nonlinear IIR filter implementation is shown in Figure 1.2 [4].

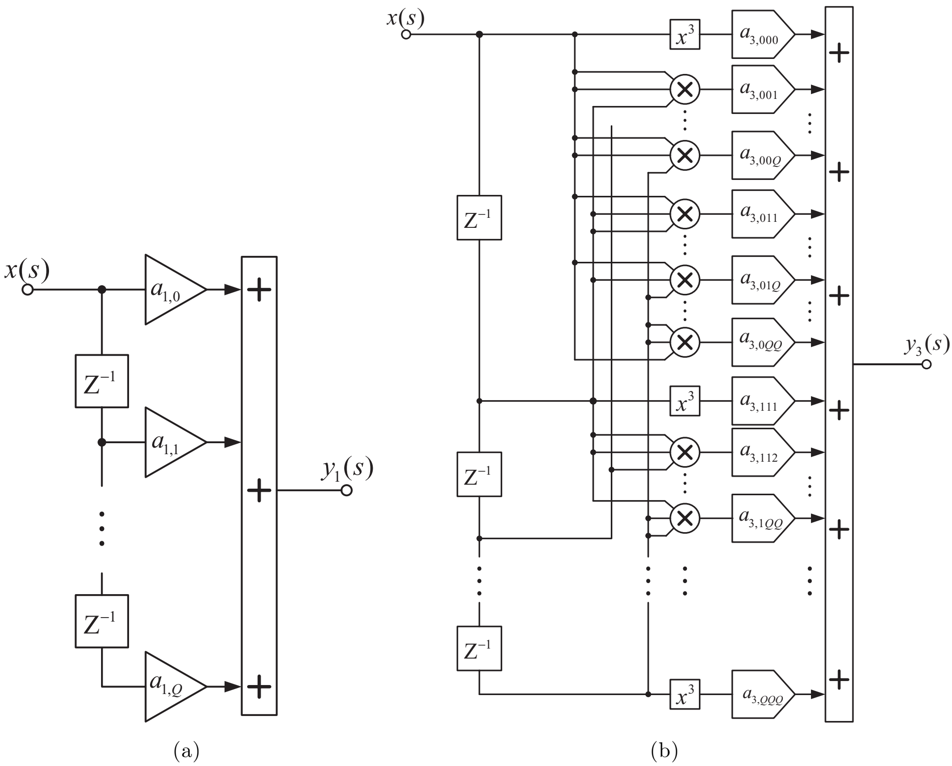


Figure 1.1 Examples of canonical forms of nonlinear FIR filters. (a) Canonical FIR filter of first order, (b) canonical FIR filter of third order. The operator Z^{-1} indicates a unit delay tap (see subsection 5.2.1).

When $f_{\mathrm{R}}(\cdot)$ and $f_{\mathrm{D}}(\cdot)$ are approximated by ANNs, Equations (1.4) and (1.5) take the following pairs of forms [15]:

$$\begin{aligned} u_k(s) &= \sum_{q=1}^{Q_1} wy_k(q)y(s-q) + \sum_{q=0}^{Q_2} wx_k(q)x(s-q) \; + b_k, \\ y(s) &= b_{\mathrm{o}} + \sum_{k=1}^{K} wy_{\mathrm{o}}(k) f(u_k(s)) \end{aligned} \tag{1.7}$$

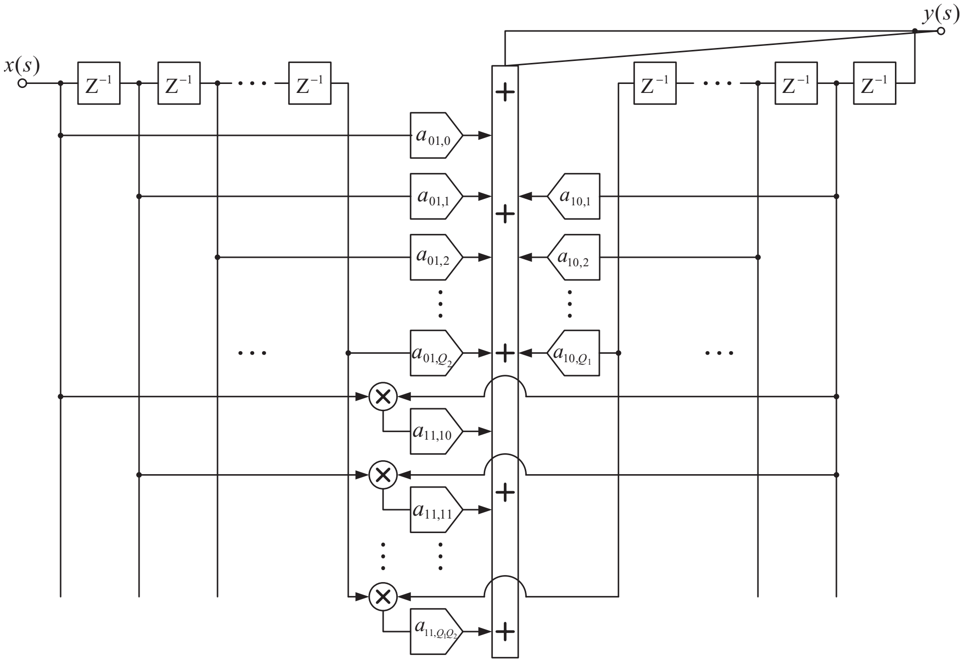


Figure 1.2 General structure of a bilinear recursive nonlinear filter.

and

$$
\begin{aligned}
u_k(s) &= \sum_{q=0}^{Q} w_k(q)x(s-q) + b_k, \\
y(s) &= b_o + \sum_{k=1}^{K} w_o(k)f(u_k(s)),
\end{aligned}
\tag{1.8}
$$

where $wy_k(q)$, $wx_k(q)$, $wy_o(k)$, $w_k(q)$ and $w_o(k)$ are weighting coefficients, b_k and b_o are bias parameters and $f(\cdot)$ is a predefined nonlinear function (the ANN activating function) of its argument [15]. As in the case of polynomial filters, these ANNs have universal approximation capabilities meaning that they are capable of an arbitrarily accurate approximation to arbitrary mappings [16, 17]. This aspect is dealt with in more detail in subsection 5.3.2.

These recursive and feedforward dynamic ANNs can be realised in the forms of Figures 1.3 and 1.4 respectively.

A close look at the feedforward ANN model of Equation (1.8) and Figure 1.4 shows that the model output is built from the addition of the activation functions $f(u_k(s))$ and the weighted outputs plus a bias and that the $u_k(s)$ are biased sums of the various delayed versions of the input, weighted by the coefficients $w_k(q)$. Each $u_k(s)$ can thus be seen as the biased output of a linear FIR filter whose input is the signal $x(s)$ and whose impulse response is $w_k(q)$. So the non-recursive ANN model is actually equivalent to a parallel connection of K branches of linear filters

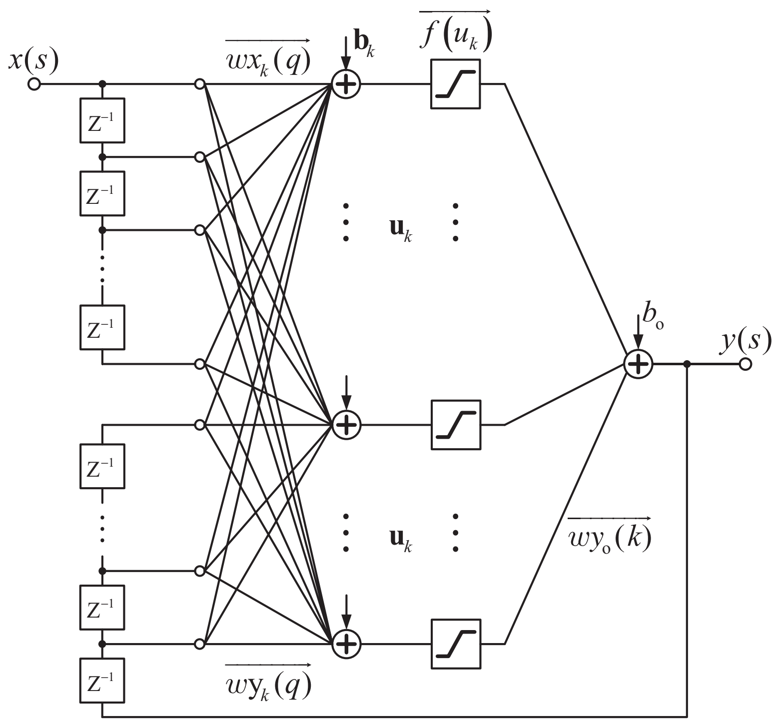


Figure 1.3 General structure of a recursive single-hidden-layer dynamic artificial neural network.

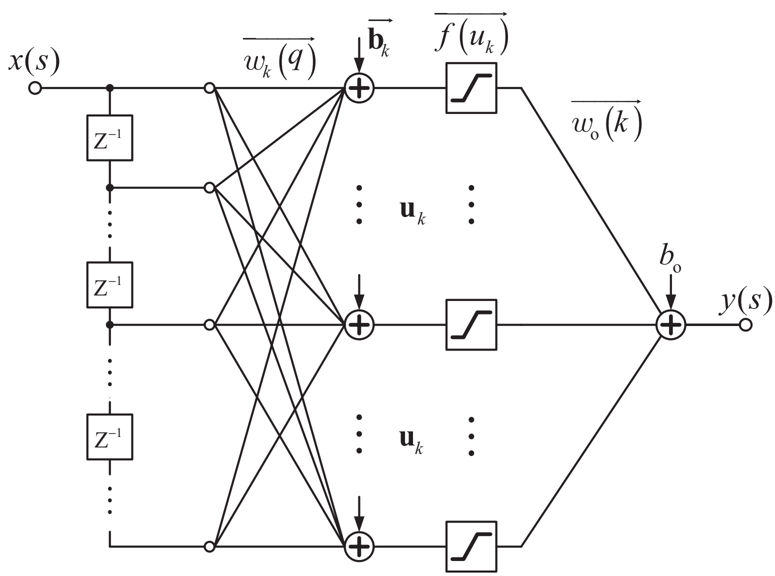


Figure 1.4 General structure of a feedforward single-hidden-layer dynamic artificial neural network.

followed by a memoryless nonlinearity, as shown in Figure 1.5.

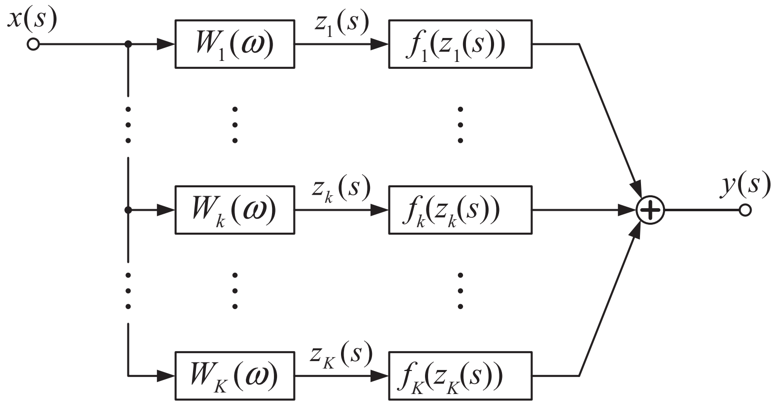


Figure 1.5 Equivalent structure of a feedforward single-hidden-layer perceptron ANN. Note that here the combinations of the branch biases b_k, the activation functions $f(u_k(s))$, the branch gains $w_o(k)$, and the final bias b_o are here represented by different branch memoryless nonlinearities $f_k(z_k(s))$.

If the branch memoryless nonlinearities were now approximated by polynomial functions we would end up again with a polynomial filter. This shows that there is essentially no distinction between a feedforward time-delay ANN and a non-recursive polynomial filter. They simply constitute two alternative ways of approximating the multidimensional function $f_D(\cdot)$, of Equation (1.5). There are, however, some slight differences in these two approaches that will be addressed below. These are worth mentioning because of their impact on PA behavioural modelling activities.

The series form of polynomial filters enables certain output properties to be related to each polynomial degree, and this can be used to guide the parameter extraction procedure. This is especially true if the polynomial series is orthogonal for the input used in the model identification process. For example, the relationship between the intermodulation content of the system's response to a multisine (a signal consisting of several sinusoidal tones) and the coefficients of an appropriate multidimensional orthogonal polynomial have recently been found [18, 19] (the structure and design of multisine signals will be discussed in subsection 2.5.6). However, since in an ANN all memoryless nonlinearities share a common form, there is no way to identify such relationships. Consequently, while polynomial filters can be extracted in a direct way, ANN parameters can be obtained only from some nonlinear optimisation scheme.

Moreover, despite the universal approximation properties of ANNs, there is no way of knowing *a priori* how many hidden neurons are needed to represent a specific system, nor is there any way of predicting the modelling improvement gained when this number is increased. It cannot even be ensured that the extracted ANN is unique or that it is optimal for a certain number of neurons. This can obviously pose some potential problems for the ANN's predictability, especially for inputs outside the signal class used for the identification, i.e. the ANN training process.

However, in contrast with the intrinsically local approximating properties of polynomials, ANNs behave as global approximates, an important advantage when one is modelling strongly nonlinear systems. Also, since the sigmoidal functions used in ANNs are bounded in output amplitude, ANNs are, in principle, better than polynomials at extrapolating beyond the zone where the system was operated during parameter extraction.

1.2.2 Nonlinear dynamic properties of microwave PAs

We now turn our attention to some typical nonlinear effects presented by practical microwave and wireless PAs. Considering the variety of available PA technologies, it is not easy to give a completely comprehensive view. Nevertheless, the technical literature in this subject indicates that a few effects at least are commonly observed in a fairly wide range of devices.

Both solid-state PAs (SSPAs) and travelling-wave tube PAs (TWTAs) have been frequently represented by cascade combinations of linear filters and a memoryless nonlinearity [20–22], the so-called two-box and three-box models. These structures introduce linear memory effects at the input and output that can be physically related to the PA's input and output tuned networks.

Beyond these linear memory effects, there are also some dynamic effects that show up only in the presence of nonlinear regimes. This is the case for the so-called long-term memory effects commonly attributed to the active device's low-frequency dispersion and electrothermal interactions and the interactions of the active device with the bias circuitry [23–29] (compare also subsection 2.4.1). Described by the dynamic interaction of two or more nonlinearities through a dynamic network, these long-term memory effects manifest nonlinear dynamics that cannot be modelled by any non-interacting linear filter and memoryless nonlinearity box models. Indeed, Pedro *et al.* [26] showed that such effects can be represented by a memoryless nonlinearity and a filter in a feedback path, as depicted in Figure 1.6, while Vuolevi *et al.* [25] and Vuolevi and Rahkonen [27] used a cascade connection of two nonlinearities with a linear filter in between.

As a common basis for the following behavioural-model discussion, we will assume that a general PA has the form shown in Figure 1.6. Through $H(\omega)$ and $O(\omega)$, this feedback model can account for linear memory effects not only in the carrier but also in the information envelope; these occur whenever the PA characteristic is not flat within the operating signal's bandwidth. In addition, the model is also capable of describing nonlinear memory effects in the carrier (AM–PM) and/or the envelope whenever the feedback filter $F(\omega)$ exhibits dynamic behaviour at the carrier frequency, the carrier harmonics frequencies or the demodulated envelope frequency [4, 26, 27].