Survey of Characterization Techniques for Nonlinear Communication Components and Systems

Christopher P. Silva, Christopher J. Clark, Andrew A. Moulthrop, and Michael S. Muha
Electronic Systems Division
The Aerospace Corporation
2350 East El Segundo Blvd.
El Segundo, CA 90245-4691, USA
+1-310-336-7597
chris.p.silva@aero.org

Abstract—This paper focuses on the measurement and modeling techniques needed for the construction and validation of system-level simulation tools, as well as the diagnosis and specification validation of hardware components and subsystems. A comparison of frequency- and time-domain measurement approaches will first be made, especially as they apply to nonlinear contexts. Based on a key *vector network analyzer* (VNA)-based technique developed for characterizing frequency-translating devices, a new baseband time-domain measurement system will be described that provides the state-of-the-art accuracy needed for the characterization, verification, and troubleshooting of emerging wideband communication systems. The paper's second part addresses modeling topics such as fidelity requirements determination, n-box models for power amplifiers, and the polyspectral method, a powerful technique for system-level modeling based on time-domain measurements that can consist of operational digitally modulated signals. Concrete applications of the method to power amplifier modeling will be made, demonstrating the highest predictive fidelity levels of any approach known to the authors.

TABLE OF CONTENTS

0-7803-8870-4/05/$20.00 ©2005 IEEE
IEEEAC paper # 1201, Version 4, Updated December 20, 2004

1. INTRODUCTION

In the last few decades, the demand for greater information throughput and device power efficiencies has escalated rapidly in both terrestial and satellite RF communications contexts. The former broadband need is a consequence of desired information services and capabilities that often now include imagery and video. In the wireless communications arena, this requirement is especially complicated by the fact that there is an increasing number of users for these services and only a limited amount of bandwidth to accommodate them. This naturally leads to the problems of channel reuse and bandwidth efficiency. The reuse problem is being primarily addressed with various multiple-access schemes, such as those based on orthogonal codes [termed *Code Division Multiple Access* (CDMA)], separated time slots [termed *Time-Division Multiple Access* (TDMA)], separated frequency slots (termed *Frequency Division Multiple Access* (FDMA)], and separated geographical cells. The efficiency issue requires the use of higher-order digital modulation schemes which have signals with multi-amplitude levels [e.g., *quadrature-amplitude modulation* (QAM)]. In a similar way, the power efficiency need has always been important for satellite communications, because of the natural preciousness of the power available in satellite vehicles. This need has again, however, been further intensified by the demands for service portability that involve handheld devices with likewise limited power resources. This issue has been addressed along several fronts, ranging from improved battery technologies (solar and portable) through device and system integration (e.g., system on chip). One of the primary users of power in a base station, handheld transmitter, or satellite transponder is the *high-power amplifier* (HPA) that comes in both *solid-state amplifier* (SSA) and *traveling-wave tube amplifier* (TWTA) forms. It is well known that such HPAs are most power efficient (that is, with respect to transmit output power versus input DC power) when they are operated nonlinearly. Unfortunately, the added efficiency gained here must be traded against the nonlinear distortion (either deterministic or stochastic) that produces detrimental performance effects on the communications signal passing through the power amplifier. The latter effects are normally offset by the use of some sort of nonlinear

compensation placed in the channel to make the overall link more linear. Because this compensation must essentially "invert" the power amplifier response, its design requires an accurate power amplifier characterization. There is currently an intense amount of activity investigating both the distortion versus efficiency tradeoff in power amplifiers [1], as well as the development of nonlinear distortion mitigation technology [2]. This activity is currently directed towards personal communication services that must meet especially stringent performance requirements.

In addition to the critical HPA component, there are several other important link components that must be accurately characterized in order to meet the evermore stringent demands on modern communication systems design. The classical *frequency translating device* (FTD), most commonly manifested as a mixer, is pervasively used and must now be better characterized in order to ascertain their distortion effects on complex digitally modulated signals. These FTDs are normally operated in their small-signal regime and well matched to their communication chain environment, so that their main distortion is manifested in a non-perfect linear transmission response, the phase component of which has been the primary challenge to ascertain. This component is important because it increasingly impacts the performance of most modern digital modulations that have a phase aspect to them. Other commonly used devices are limiters and frequency multipliers, which likewise will have both linear and nonlinear imperfections that must be captured in order to ascertain their system performance impact. The majority of the other components in the communications chain will operate linearly, and have the same input/output frequency range, so that traditional *vector network analyzer* (VNA)-based techniques can be used to characterize them. Another important element of the communications link that must be addressed is any atmospheric transmission medium, which in contrast produces only stochastic distortion on the propagated RF signal. This is a whole discipline to itself, and will not be covered in this paper (but see [3], for example, which addresses the RF satellite-to-ground propagation portion of this subject). Finally, there is the stochastic distortion that occurs everywhere along the hardware communications chain that manifests itself in various noise types produced by the components. This broad and detailed subject will likewise not be treated in this paper.

As is the case for any modern communications engineering problem, *computer-aided analysis* (CAA) is performed up front, addressing the above tradeoffs and scenarios, in order to arrive at viable and effective designs that will perform properly when implemented and fielded. There are several important reasons for this now firmly established practice. The primary driver has been the rapid development of powerful and economical computational capabilities that enable the use of these tools on very complex systems. The intense competition found in the telecommunications industry, with its requirements for shorter development and implementation cycles, as well as the increasing demands by users in difficult channel contexts, has made these tools imperative. With their use, engineers can readily perform design trades over a wide range of scenarios, thereby allowing for performance optimization that would otherwise be impossible with the classical and costly hardware prototype-and-test practices of the past. In addition, such simulations, which range from the hardware device to the entire system level, can ensure that the actual implemented system will function properly from the start. Behind these many powerful advantages, however, lies a very crucial assumption: *the simulation models that make up these tools accurately reflect hardware/transmission media reality.* If they do not, the outcomes of these simulations, which are too often taken for granted as automatically valid, can be misleading and result in costly engineering mistakes. A remedy of this situation for the hardware portion of the communications simulation—which must be an integral part of the application of these tools—is the use of measurements for the development and evaluation of component simulation models.

There are two basic classes of hardware component/subsystem models that must be addressed: linear and nonlinear [4]. The former class of models are mature and accurate, relying on the fact that they are formally and globally identified either by a simple constitutive relationship (in the case of individual circuit devices) or a CW frequency sweep (in the case of systems). This means that once these characterizations are established, they will be accurate for all signals of interest, provided that the component or system remains in a linear regime of operation. The frequency-domain measurements here can be either performed with a spectrum analyzer, which only provides transmission amplitudes with limited accuracies, or the preferred VNA instrument, which obtains both the amplitude and phase components of the transmission response, as well as characterizing the imperfect match provided by the device.

In contrast, nonlinear models are much less mature and are essentially the primary motivation for the development of time-domain measurement techniques. Without this class of models, the above marriage between simulation fidelity and hardware measurement reality would be simple and straightforward to affect. The need for high-fidelity nonlinear system models is further exacerbated in the context of wideband modulated signals. Here the use of standard linear-based modeling approaches that rely on simple tonal measurements is adequate only for relatively narrowband signals. This is because tones cannot adequately capture the inherently complicated intermodulation and memory effects nonlinear systems exhibit for input signals with large bandwidth and possible amplitude variation. Furthermore, a nonlinear system can be thought of mathematically as an input/output operator that has a continuity with respect to its input signals. Hence if the input signal of interest is close to the ones used to construct the model in the first

place, continuity of the operator will dictate that the model will provide good output predictions. Depending on the "steepness" of the operator in the vicinity of the construction signals, this predictive fidelity will break down gracefully or quickly as the input signal deviates from the construction signal (with respect to its bandwidth, for example). The bottom line is that nonlinear systems, unlike linear ones, can only be locally *identified* in a manner that generalizes the classical linear impulse response, and such representations *cannot be* obtained from simple tonal measurements [5].

This paper provides a description and discussion of the measurement and modeling techniques needed to characterize modern communication system hardware components/subsystems, with an emphasis on covering the wideband time-domain measurement and systematic nonlinear system modeling techniques developed by the authors since 1996. These tools will help overcome the weaknesses and challenges described above for computer simulation modeling, design validation, and system troubleshooting. The outline of the paper is as follows. Section 2 will provide a short introduction to frequency- and time-domain measurements, with a comparison made based on their suitability for the characterization of linear and nonlinear components/subsystems. Section 3 will cover the FTD characterization problem, highlighting the various approaches that have been developed, with an emphasis on the VNA-based FTD measurement techniques developed by the authors. The latter techniques became the crucial enabler of the *baseband time-domain measurement technique* and its implementation as the *Aerospace Time Domain Measurement System* (ATDMS) [6, 7] (covered in Section 4), whose accuracy rivals that normally attributed to frequency-domain measurements. Subsection one of Section 5 will begin by describing the basics of system modeling, addressing the important questions of how to quantify model predictive fidelity, as well as how to determine the fidelity level needed for a given communications system. The subsection will end with indications of how to use the ATDMS in model construction and evaluation. Subsection two of Section 5 will provide an introduction to power amplifier modeling, including a survey of traditional, but not necessarily well-known, nonlinear n-box power amplifier models. Section 6 of the paper will focus on the *polyspectral method*, which provides a powerful new set of systematic analysis, evaluation, and design tools for modeling and compensation, with the modeling application primarily emphasized here. Subsection one of Section 6 will provide a brief description of the polyspectral approach, including its origins, theoretical basis, and basic model architectures and analytical features. The model architectures to be described will first consist of an enhanced 3-box power amplifier model based on the important *optimal filter* concept. This will be followed by the general and true form of the polyspectral model that can be thought of as an nonlinear operator series. Concrete application of the approach to power amplifier modeling will be taken up in subsection two of Section 6. In particular, a baseband variant of a standard polyspectral model architecture will be proposed that is especially accurate for power amplifiers. Modeling results will then be reported for a 20-GHz TWTA under a variety of wideband 16-*Amplitude Phase Keying* (APK) digital modulation excitations. Section 7 will provide a summary and some conclusions concerning the characterization of communication system components and subsystems, making the strong case that time-domain techniques should be considered a necessary and valuable option, especially for wideband nonlinear contexts. Assessments and plans for future polyspectral-based developments for wideband nonlinear communications will also be provided. Given its survey genre, the exposition in this paper will necessarily be narrative and representative, with in-depth technical details left for the reader to investigate in the cited published references.

2. FREQUENCY- AND TIME-DOMAIN MEASUREMENTS

Not only are accurate measurements needed to keep simulation models grounded in reality as argued above, they are also required for diagnostic and specification testing purposes that are part and parcel of checking out the implemented system. In either case, there are two basic types of measurements that can be made: frequency-domain or time-domain [8]. Each type has its advantages and disadvantages, with the former being the most widely used and accepted, while the latter has lately begun to emerge as a viable and powerful alternative, as this paper will advocate. With both of these measurements developed to a sufficient degree of fidelity and maturity, an effective arsenal of tools will be available to measure and model both linear and nonlinear systems, thereby resulting in CAA packages that can be confidently applied and relied upon for important design decisions. It will be instructive now to briefly describe the characteristic features of these two basic approaches to hardware measurement.

Frequency-domain measurements are primarily accomplished with a VNA that provides very accurate calibrated gain and phase CW measurements using a 12-term error correction scheme. The calibration is done with well-established standards and techniques that are traceable to a standards laboratory. This instrument is excellent for the measurement of linear components and systems with equal input/output frequency ranges. However, very commonly used FTDs, such as mixers, or frequency multipliers cannot be individually measured with a VNA since the input/output frequency ranges do not match. Likewise, intermodulation products and harmonics produced by nonlinear devices cannot be captured by standard VNAs. A nonlinear VNA has been developed that can capture such information [9]. This is afforded by its ability to make calibrated amplitude and phase measurement at harmonics of the input frequency. A commercial version of this new instrument, also called a *large-signal network analyzer* (LSNA), is currently being offered by Maury Microwave [10]. A much more common alternative for captur-

ing nonlinear effects is the spectrum analyzer, which exhibits less accuracy in power measurements than a standard VNA, and does not provide any information about phase. Very accurate power meter measurements can also be made to arrive at nonlinear transfer characteristics, although phase information is again still not captured. Only the less accurate spectrum analyzer and nonlinear VNA can capture signals more general than single or multiple tones.

To date, time-domain measurements are less accurate than frequency-domain ones, but have no component/system-type limitations. These measurements have traditionally been made directly at RF, using either the more common *digital storage oscilloscope* (DSO) or an Agilent *microwave transition analyzer* (MTA), usually incorporating a subsampling technique that requires a periodic input stimulus. In contrast to frequency-domain measurements, these instruments can capture all nonlinearity effects over wide bandwidths, whether deterministic (such as intermodulation products, harmonics, etc.) or stochastic (that is, noise). Such measurements are especially useful for nonlinearities with memory that exhibit frequency-dependent effects. The important advantage here is that one can employ operational modulated signals as the stimuli, instead of being restricted to simple tonal signals that can differ significantly from these realistic signals. As a result, such measurements will capture the real behavior one would encounter in the actual system, and this is important in view of the general mathematical operator locality discussion made above regarding nonlinear components and systems. In addition, there are several formal modeling approaches that can locally identify a nonlinear operator based on such time-domain measurements, assuming they are of sufficient accuracy. The classical challenge here has been to overcome the accuracy limitation of these measurements, which degrade with frequency of operation in a manner much more severe than with a calibrated VNA. In fact, calibration standards and practices for time-domain measurements are very immature at best. The basic purpose of this paper is to inform the aerospace technical community of an effective means of overcoming this classical measurement challenge.

Given that current time-domain recording instruments have no means of internal calibration, are not internally accessible for user intervention to accomplish this, and always exhibit degraded accuracy with increasing frequency, the only possible means to improve measurement accuracy is to lower the frequency range through downconversion. This has several important advantages that need to be noted here. First, it can be readily shown with a power meter, for example, that such a lowering of the frequency range can greatly enhance the recording instrument accuracy. An example of such a phenomenon is depicted in Figure 1 for an MTA. Note that a 4-GHz bandwidth signal would require only a 2-GHz bandwidth to record at baseband, over which the MTA has a gain variation of only ± 0.1 dB. In a likewise manner, the phase deviation from linearity can be as high

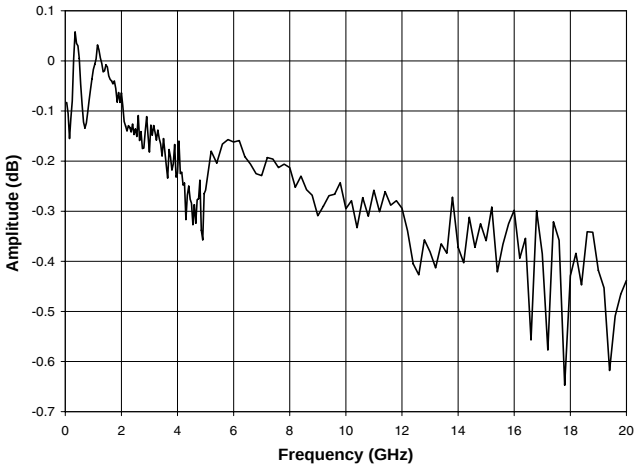


Figure 1 – Microwave transition analyzer amplitude response calibration using an HP8485D Power Sensor as a "gold standard."

as 40° for a typical DSO at 20 GHz, but it is negligible below 5 GHz [11]. Second, with downconversion, signals beyond the measurement capability of current instruments (for example, the 40-GHz limit of the MTA) would be recordable in the time domain. If this downconversion is done coherently with the upconversion, the phase noise of the carrier can also be eliminated, and the accuracy of the measurement will be independent of the carrier frequency. In addition, there is a very large reduction in the sampling requirements needed to capture the signal, because it can approach the finite and much smaller information bandwidth of the communications signal. This allows for either longer time records or higher time resolution for the measured data. If the downconversion goes all the way down to baseband, the above advantages will be maximized, and the recorded waveforms will be complex in nature, becoming the so-called *lowpass equivalent* (LPE) envelope of the original RF waveform. In addition, this latter form of the data dovetails precisely with the standard industry practice of employing LPE models in communication system simulation tools (commercial and in-house) [4]. As a result, such measured data can be employed in the direct construction and evaluation of these models, thus providing an LPE measurement counterpart to the modeling. The fundamental linchpin of this approach, however, is that the imperfection of the real downconversion process must be precisely characterized and calibrated out of the measurement. This characterization must be both in amplitude and phase, and represents a classical problem in measurement practice that has existed since the development of FTDs. This challenge has been recently overcome, as will be discussed next, serving as the basic foundation for the time-domain measurement technique to be covered in this paper.

3. FREQUENCY-TRANSLATING DEVICE CHARACTERIZATION

Frequency-translating devices (FTDs), which can range from simple microwave mixers to entire channel segments, are fundamental elements of virtually every communication system. There are two basic classes of FTDs. One class, called *single-sideband* (SSB) FTDs, are characterized by the fact that only the upper or lower sideband, of the two sidebands normally generated in the frequency-translation process, appears at their output. These FTDs are primarily employed as frequency converters between RF and IF ranges in typical communication channels. The second class, naturally termed *double-sideband* (DSB) FTDs, retain both sidebands in their output and involve a baseband signal. These FTDs commonly appear in modulators and demodulators.

The need for the full transmission response of FTDs, that is, both their amplitude and phase frequency response, has become increasingly important in modern communication systems. With the prevalent use of complex phase modulations in these systems, it becomes necessary to obtain the phase response of FTDs since this may have a detrimental effect on performance. Likewise, as noted above, the accurate transmission characterization of FTDs provides the enabling vehicle for the establishment of time-domain measurements as a useful and viable approach for wideband RF/microwave components and systems. Even without these current motivations, the full characterization of FTDs has been a classical measurement problem for decades, with several methods developed to date. Note that although the amplitude response (also known as conversion loss) of an FTD is straightforward to understand, the definition of the phase response for an FTD must be carefully formulated since the input and output frequency ranges typically do not overlap. In essence, one has to think of the FTD as consisting of a perfect frequency translator, preceded or followed by a linear filter that represents its imperfections. For standardization purposes, this filter is taken to be at baseband and in the form an LPE filter.

The VNA-based method to be highlighted here, which infers the full response for an individual FTD using pairwise measurements amongst three FTDs, is more accurate than any of the previous measurement methods. For example, in the method that employs a scalar network analyzer, only the conversion loss is found, and hence the characterization is incomplete [12]. In the case of previous VNA-based methods that involve the measurement of single nonfrequency-translating FTD pairs, there is a non-testable assumption that the FTDs are essentially perfectly matched or that one of the FTDs is a "gold standard" that itself has been accurately characterized in some way [13]. A more recent VNA-based approach involving a single FTD and a reflective termination harbors inaccuracies because the undesired mixing products generated by the FTD are not removed before the reflected signal re-enters the FTD [14]. An improved version of this single-mixer approach that provides for both the transmission response and input/output matching of the FTD was reported in [15]. An MTA-based modulation technique has also been proposed to obtain a full individual FTD response, but the instrument capabilities limit the speed, accuracy, and frequency coverage of the measurement [16]. Finally, it is also conceivable that the nonlinear VNA could be used to characterize FTDs since it can be calibrated over a frequency grid that covers the differing input/output frequency ranges.

The FTD measurement procedure covered in [17, 18, 19] offers a significant improvement in speed and accuracy for SSB FTDs, and is the only method known by our group for the full characterization of DSB FTDs. The technique makes use of a VNA and three FTDs: two test mixers besides the FTD under test (FTDUT), where one of the test mixers must have a *reciprocal* response, that is, have the same upconversion/downconversion frequency response. It has been found that commonly used double- and triple-balanced mixers exhibit this property closely when operated linearly and their ports are properly terminated [20]. In addition, the FTDUT must have an accessible *local oscillator* (LO) port because of the phase coherency required by the method. In essence, several measurements are made of nonfrequency-translating FTD pairs, after which some simple algebra provides the desired individual response of the FTDUT. For SSB FTDs, at least three such measurements are required to arrive at the desired response. For DSB FTDs, there are two ways of proceeding, the less accurate one again requiring three measurements, while the more accurate method needs a minimum of six measurements to accomplish its characterization. It is the latter approach that serves as the basis for the baseband time-domain measurement technique and system to be described in Section 4.

Figure 2 provides the general test configuration for both SSB and DSB FTDs. Besides the VNA and two test mixers

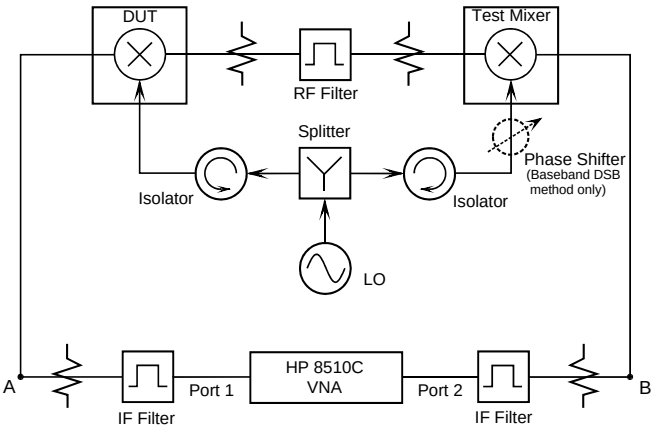


Figure 2 – General FTD test configuration.

already mentioned above, an LO source with a splitter, some isolators, attenuators, filters, and a phase shifter are needed, the latter used only for the baseband DSB approach.

The added components are needed to reduce leakages and spurious signals to prevent corruption of the measurements. It must be emphasized that only a transmission response is obtained here, so that these results will apply only to similar matched environments. In addition, this response is a baseband LPE, which essentially means that the bandpass filter of the actual FTD model has been shifted down to baseband after its negative frequency counterpart has been removed [21]. The data collection and analysis using this test setup has been fully automated in LabVIEW® [22] to reduce operator error.

For the SSB FTD measurement procedure, the VNA port IF filters in Figure 2 are bandpass. Figure 3 provides a representative flow diagram of the minimal procedure that involves three pairwise measurements of three FTDs. Here TM1 and TM2 represent test mixers of a frequency

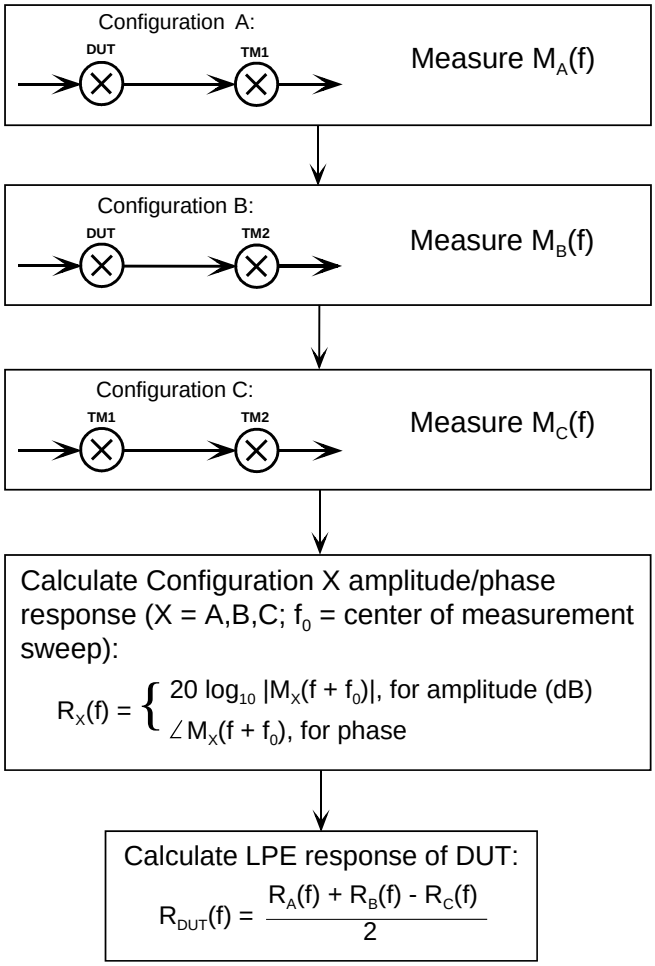


Figure 3 – SSB and bandpass DSB FTD measurement flow diagram for the case of one device under test (DUT) and two test mixers (TM1 and TM2), where one of the test mixers (TM1) possesses a reciprocal transmission response. Other reciprocity cases would be measured similarly. The resulting total frequency response for the DUT is in LPE form.

plan compatible with the FTDUT, and such that test mixer TM1 is reciprocal since it is seen to be used both as an upconverter and downconverter. Once the minimal three pairwise frequency responses $M_X(f),\ X = A, B, C$ are obtained using the VNA, they are shifted down to baseband by the common center frequency f_0 of the sweeps. Observe that these LPE responses will generally not be symmetric about DC. The three amplitude (in dB) and phase components of these responses are then used to de-embed the total, individual LPE response of the FTDUT. Note that both an upconverter-downconverter or downconverter-upconverter configuration is allowed here, and additional reciprocal test mixers and corresponding measurements can be added to increase the FTDUT characterization accuracy. The latter feature comes from the fact that an overdetermined set of measurements will result, so that the FTDUT response instead becomes an average over multiple valid responses. Figure 4 makes a validating comparison of the new technique with the aforementioned MTA-based modulation approach for a 20-to-8 GHz SSB downconverter. The measurement

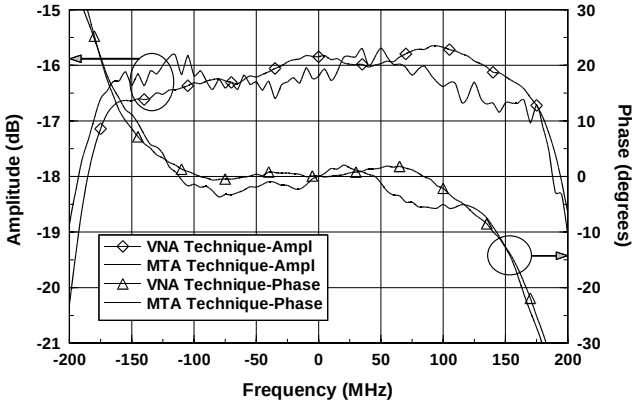


Figure 4 – Comparison of MTA-based modulation and SSB FTD measurement techniques for a 20-to-8 GHz SSB downconverter. The latter curves are an average of eight inferred responses.

bandwidth was 500 MHz, with the results displayed over a 400-MHz range. In this case it was found that all three FTDs were reciprocal, leading to eight valid responses for the downconverter that were subsequently averaged and compared with the MTA result. The agreement between the amplitude and phase response curves was found to be within a very respectable value of 1.15 dB and 6.14°, respectively. Note that the new method provides a smoother result, since the ripple structure found with the MTA method, mainly caused by VSWR interaction at the measurement ports, cannot be removed.

For the DSB FTD measurement procedure, there are two choices dictated by the frequency range (baseband or bandpass) for the VNA sweep of the FTD pairs. For the bandpass choice, the RF filter between the FTDs in Figure 2 becomes a lowpass filter, the VNA port IF filters become RF bandpass, and the same flow diagram of Figure 3 applies. For the baseband choice, the two VNA port IF filters become lowpass, and two sets of measurements

have to be made, differentiated by a 90° change of the phase shifter in the downconverting LO path. In this case, the minimal procedure requires six pairwise measurements of three mixers, as indicated in the representative flow diagram of Figure 5. In particular, the minimal three sets

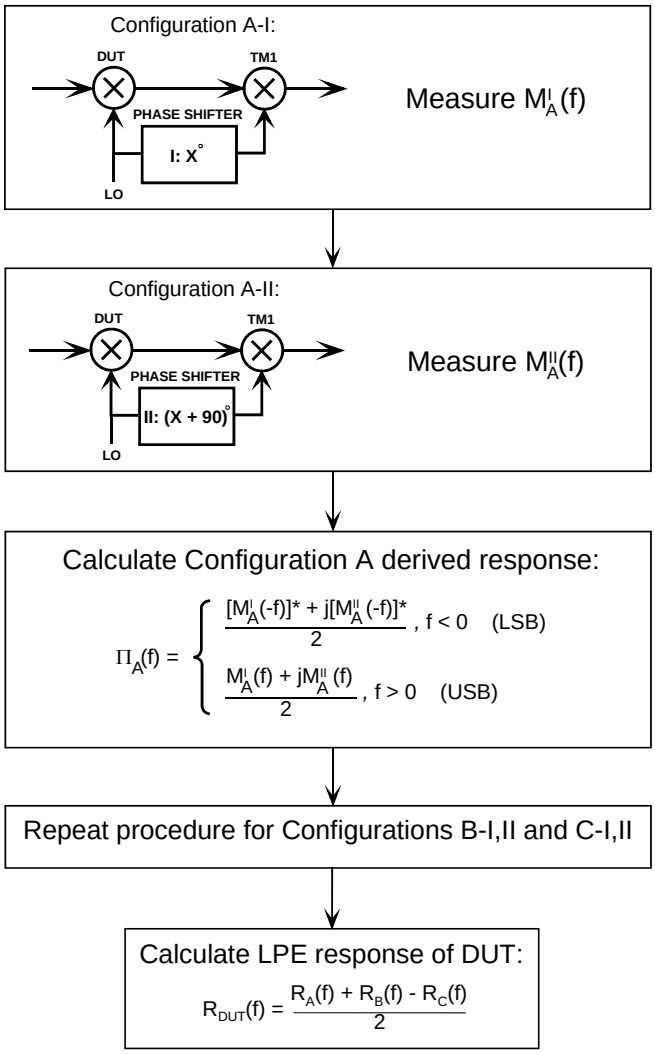


Figure 5 – Baseband DSB FTD measurement flow diagram for the same conditions as in Figure 3.

of quadrature VNA measurements results in three derived baseband responses $\Pi_X(f),\ X = A, B, C$, consisting of the de-embedded upper and lower sideband responses (denoted as USB and LSB, respectively). The two phase shifter settings are needed to de-embed these sideband responses, since they combine in the downconversion to baseband and hence cannot be separated with one measurement. Similar to the SSB case, the amplitude (in dB) and phase components of the derived responses are represented by the notation $R_X(f),\ X = A, B, C$ and are used to arrive at the total, individual LPE of the FTDUT. The results for this baseband version of the DSB procedure will usually be more accurate than the ones obtained with the bandpass DSB technique.

To show the consistency of the two possible DSB approaches, they were applied to a 20 GHz-to-baseband downconverter over a 4-GHz measurement bandwidth. Figure 6 shows the results of the comparison for the phase response, with a similar result found for the corresponding amplitude response. There is good agreement seen here, with the

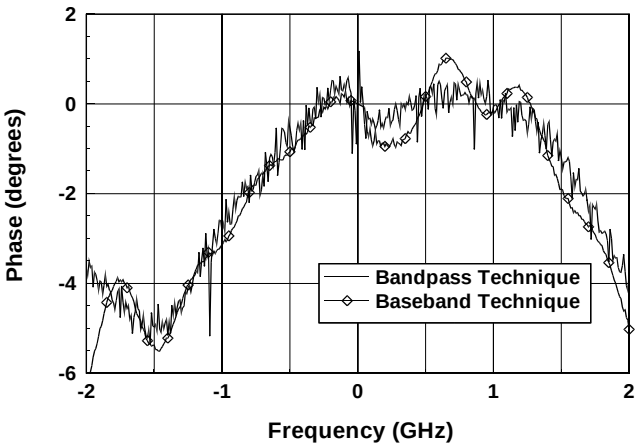


Figure 6 – Phase response comparison of the bandpass and baseband DSB FTD measurement techniques for a 20 GHz-to-baseband downconverter. The close agreement of the curves indicates the consistency of the two approaches.

bandpass approach exhibiting more broadband noise and spikes, while the baseband approach had some VSWR-induced ripple that could be reduced with further attenuation. Because there are no known methods to independently measure the full response of this mixer, a validation test was devised in which the response of a bandpass filter placed between the mixer pair was de-embedded and compared with a direct VNA measurement of the filter. In addition, the center frequency of the filter was 600 MHz offset from the LO frequency of the mixers, which themselves had asymmetric responses. Figure 7 makes a comparison of the amplitude responses between the de-embedded and direct measurement approach. A similar agreement was found for the phase response, thereby validating the baseband DSB technique.

4. BASEBAND TIME-DOMAIN MEASUREMENT TECHNIQUE AND SYSTEM

With the unique baseband DSB FTD technique of the previous section, a calibrated baseband time-domain measurement technique was developed and implemented as a complete *Aerospace time-domain measurement system* (ATDMS) [6, 7, 23]. With this system, accurate measurements of modulated microwave or millimeter-wave signals having several GHz of bandwidth can now be made. As indicated above, a coherent LO is needed for the downconversion, and the signals measured must be repetitive since the two components of the complex baseband waveform, termed the *in-phase* and *quadrature* components, must be recorded consecutively with appropriate phase shifts of the downconverter LO. The stabil-

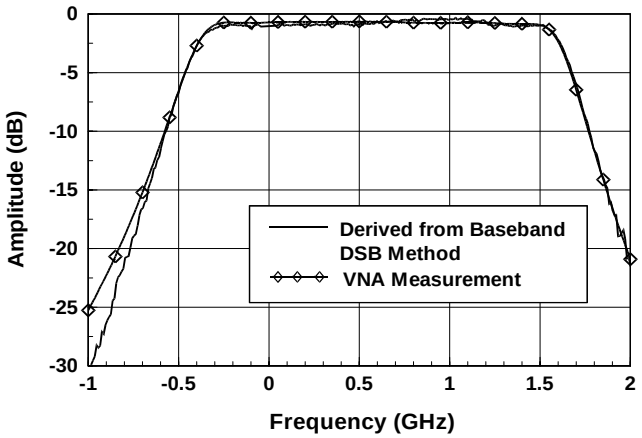


Figure 7 – Amplitude response of a bandpass filter derived from baseband DSB FTD measurements compared to a direct VNA measurement. The close agreement found here and for the phase response served to validate the baseband DSB FTD measurement technique.

ity of the LO is not critical since the coherency of the down-conversion removes any phase noise that may be present. The unique measurements made here can be used to characterize anything from individual modulators through entire linear or nonlinear communication channel segments. A representative set of examples will be taken up in Section 5.

There are several useful features worth noting for the general ATDMS. First, this system provides a transmitter that may be used as a test signal source, as well as a wideband (multi-GHz) test mixer with a reciprocal response characteristic that is needed for calibration purposes. Second, the ATDMS contains a *downconverting receiver* (DCR) that is calibrated using the baseband DSB FTD measurement technique discussed in the previous section. A typical implementation of the DCR component of the ATDMS is presented in Figure 8. It consists of a wideband downconverting mixer

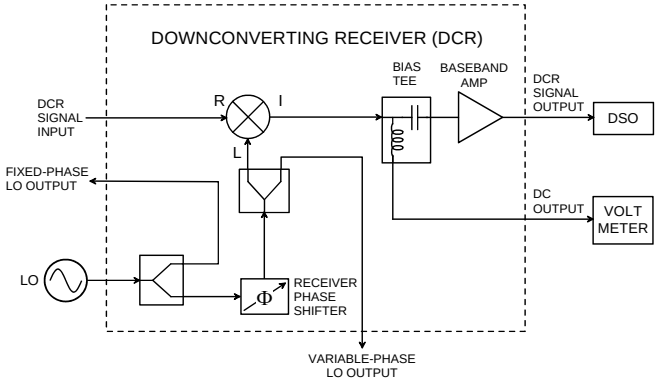


Figure 8 – Block diagram of the core downconverting receiver (DCR) component of the ATDMS using a digital storage oscilloscope (DSO) as a recording instrument.

with an LO network that provides for phase shifting (needed for the DSB FTD method) and LO outputs to lock with the transmitter side of the ATDMS setup. Note that a baseband amplifier follows the downconverting mixer to bring up the signal level for the baseband recording instrument. This instrument can be either a DSO or an MTA, where the latter can be more accurate than the former because its timebase can be locked to that of the signal generator, but suffers from a limited amount of memory. Because the baseband amplifier normally blocks the DC voltage (corresponding to the carrier of the RF waveform), a bias tee and voltmeter are also needed to obtain this frequency component. The receiver response of the ATDMS (including any interconnect cables to the DUT) is analytically removed from the recorded input/output time-domain data. Finally, the system is fully automated in its operation under LabVIEW® control. More recently, there have been several other refinements developed for the ATDMS that increases its repeatability and accuracy even further [24, 25].

Two basic measurement considerations must also be noted when using the ATDMS. First, the DCR introduces both linear and nonlinear distortion to the recorded signal. The nonlinear distortion is minimized by filtering any mixing products produced, and keeping the mixers in a linear mode of operation as much as possible. The linear distortion is minimized by using DCR components with much wider bandwidth than the DUT, and by reducing VSWR interactions. When the DCR components have a comparable bandwidth to the DUT, then their linear distortion must be calibrated out. Second, there are still significant DC offset errors that enter the recorded signals due to intrinsic mixer imbalances. These intrinsic mixer offsets are determined using four different LO phase shifter settings that are 90° apart, and then subsequently subtracting these offsets from the DC component of the measured waveforms.

The details of the calibration, measurement, and post-data processing procedures are given in [7], and the latest enhancements to the system are presented in [24, 25]. Only a summary for the original system will be provided here. In essence, the DCR calibration procedure consists of two parts: (1) using the baseband DSB FTD characterization technique described above to arrive at the full DCR response outside the DC frequency, and (2) a two-step process to obtain the LPE DCR DC response (involving the four phase settings mentioned above, along with a separate CW calibration). The baseband waveform and its DC component are recorded separately at four phase shifter settings, the former is corrected using the previously determined DCR response, its DC component is restored using the LPE DCR DC response, and finally a power normalization factor is applied.

To illustrate the validity of the ATDMS, a comparison with a direct MTA measurement is shown in Figure 9 for a microwave pulse of 0.35-ns duration and 0.5-V amplitude, and using an LO frequency of 19.6 GHz. Both the corrected and uncorrected baseband envelopes are shown together with

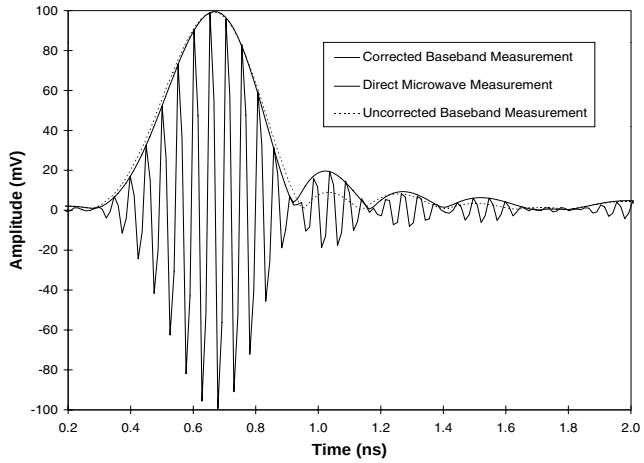


Figure 9 – Comparison of the ATDMS to a direct RF microwave transition analyzer (MTA) measurement for a narrow microwave pulse.

the direct MTA measurement. Note the excellent agreement between the corrected measurement and the direct one, which indicates the validity of the new measurement system. In the frequency domain, over a comparison bandwidth of 4.4 GHz that does not include the DC region (where significant accuracy improvements were made since this initial comparison), the agreement was found to be within ± 0.4 dB for the amplitude, and $\pm 3^\circ$ for the phase.

Representative applications of the ATDMS to the characterization of power amplifiers will be covered in the next section, and as indicated above, the technique can certainly be applied to several contexts other than modeling. For example, the nonlinear noise characterization of TWTAs has been enabled by the ATDMS [26].

5. SURVEY OF SYSTEM MODELING ISSUES, PROCEDURES, AND n-BOX ARCHITECTURES

This section consists of two subsections, the first addressing some important and generic modeling issues such as the metrics needed to quantitatively calculate model predictive fidelity, the determination of model fidelity required for a given communications system, and the construction and validation of system models with the ATDMS as the prime example. The second subsection will focus on the modeling of power amplifiers by first introducing the three classes of models for these components, followed by a survey of the n-box system-level power amplifier models that have been developed, the more sophisticated of which have not been readily available in the literature. It will be demonstrated that these models may still not be adequate for wideband systems, thus requiring the application of the polyspectral method that is the subject of Section 6.

System Model Construction, Evaluation, and Fidelity

In general, the modeling procedure consists of two basic steps: construction and validation, both of which require measurements of sufficient accuracy for the context at hand. These measurements can be in the frequency or time domain, although the latter is more appropriate for nonlinear regions of operation as demonstrated previously. The parameters of the construction measurements (frequency or time domain, tonal or digitally modulated stimulus signals, number of measurements, etc.) are dictated by the particular modeling methodology involved, as will be detailed in subsection two of Section 5, and Section 6, below. The basic purpose of the validation measurements is to determine quantitatively the error in the model's predictive fidelity. It is clear in both cases that the accuracy and repeatability of the measurements must necessarily be better than the fidelity requirements for the model with some comfortable margin. Thus there must be some metric formulated to gauge this accuracy/repeatability and predictive fidelity, of which there are several alternatives dictated by the nature of the measurements.

The first metric that could be proposed is the commonly used and universally accepted digital communications system performance measure termed the *bit-error rate* (BER). It is essentially a direct measure of how well a given end-to-end communications system is operating with respect to information transmission. It is a scalar that dictates the probability of error in the decision of a received bit compared to the originally transmitted bit. This error will of course depend on several parameters in the communications link, and is normally plotted against the primary parameter that is the received *signal-to-noise ratio* (SNR), denoted by E_b/N_0, where E_b represents the bit energy and N_0 represents the spectral density of the equivalent *additive white Gaussian noise* (AWGN) in the received signal. The other parameters upon which BER depends will include the particular digital modulation used in the communications system, as well as any signal corruptions that have occurred along the transmission chain, whether deterministic (such as linear filtering and power amplifier nonlinear distortion), or stochastic (such as thermal noise and atmospheric effects). In any case, BER is almost universally used to assess system performance in simulation parameter trades conducted with an end-to-end link model. Note that BER should be used only to assess the fidelity of an entire end-to-end link model, since it is basically a summary of the entire link imperfections (see discussion below). In this case, the measured BER would have to be compared with the link model prediction to determine this fidelity for a given E_b/N_0. In order to accomplish such a validation, a faithful hardware emulation of the link (without the atmosphere) would have to be constructed and measured.

Despite the importance of the BER metric for performance assessment, there are several difficulties for its application as a fidelity measure. First, the fundamental way to develop an accurate end-to-end link model is to make sure that

each of the component models are individually accurate. To do the latter really requires a more direct metric in the frequency or time domain, since to apply the BER metric would require the building of a complete modem to accompany each component. This exercise would either be a very expensive undertaking, or one that is not practical (as in the case of a frequency converter, for example). Even if such a modem could be built, there would be the problem of having to model the modem itself, so that any errors here would be indistinguishable from the ones for the component, that is, the BER prediction error would include inseparable contributions from the modem and component modeling. This inseparability would extend to the entire end-to-end link model evaluation if BER is used as the metric. Second, the BER metric is based on a finite set of discrete samples of the received time-domain waveform, and these samples are in turn quantized to a practical level in order to make a bit or symbol decision. For both simulation models and hardware measurements, the sample times and amplitude levels are on a much finer scale than what is used to determine BER. Thus the BER more coarsely reflects the nature of the received waveform compared to what an error metric based on simulated versus measured waveforms could achieve. Finally, because the receiver is typically much more bandlimited than a measurement instrument, the BER metric will be addressing a smoothed version of the received time-domain waveform relative to what a measurement instrument would capture to calculate other error metrics.

The next commonly used metric is associated with standard frequency-domain measurements, consisting of a simple comparison between a measured value and one from a standard or a model prediction. These measurements usually involve the use of a single tone input that is swept in frequency, with the output also measured at the same single frequency, whether it be its power (as in the case of a power meter), its amplitude (as in the case of a spectrum analyzer), or both its amplitude and phase (as in the case of a VNA). Of course, for the spectrum analyzer, arbitrary signals can be measured as well, although internally tunable filters are used to sweep across the spectrum using a very narrow gate of frequencies. A common frequency-domain test for the case of a nonlinear HPA is to use a two-tone input (at f_1 and f_2) and measure the third-order intermodulation output tones found at $2f_1 - f_2$ or $2f_2 - f_1$ that will lie close to or in the frequency band $[f_1,\ f_2]$.

In all of the frequency-domain measurements just described, a simple comparison between scalar quantities can be made at each frequency to arrive at an absolute or relative error either between successive measurements of the same phenomena, between what is measured against a standard, or between what is measured and what a given model predicts. In the case of an entire frequency response or power spectrum, however, a single scalar error metric is often desired for the entire "waveform." This is likewise always the case for the less common, but more generally applicable time-domain measurements. In both instances, one arrives at a finite set of data values equally spaced by the frequency resolution or sampling time that was recorded by the measurement instrument, and a metric is needed to quantitatively compare the two waveforms. Such a metric has been previously developed [27], termed the *normalized mean square error* (NMSE), which is essentially the power in the error vector between the waveforms normalized to the total measured or other appropriate power. For the case of two complex baseband waveforms, as would be produced through the use of the ATDMS described above or an LPE simulation model, this metric would be given explicitly by

$$NMSE := 10\log_{10}\left\{\frac{\sum_{k=1}^{N_d}\left[\left(y^1_{I,k}-y^2_{I,k}\right)^2+\left(y^1_{Q,k}-y^2_{Q,k}\right)^2\right]}{\sum_{k=1}^{N_d}\left[\left(y^1_{I,k}\right)^2+\left(y^1_{Q,k}\right)^2\right]}\right\}, \quad \text{(1a)}$$

where

$$y^i_{I(Q)} := \text{In-phase (quadrature) waveform } i \text{ } (N_d \text{ points}) \quad \text{(1b)}$$

and for the case of normalization to the power in the first waveform. In order to gauge the values produced by this metric, observe that an NMSE value of -10 dB would correspond to a relative waveform error of 10%.

Because of the importance of the BER metric and the practicality of the NMSE metric, it would be desirable to relate the two for a given communications system or even in general. Unfortunately, the precise relationship between these metrics is still an open problem, although some qualitative remarks can be made here, and the model NMSE determination method to be described below will also shed some light on this complex connection. First, NMSE is a scalar characterizing the normalized distance between two waveforms. Given a waveform and a value of NMSE, there is an infinite number of waveforms that can produce this value of NMSE. However, each of these waveforms may give rise to a different BER value, hence resulting in a range of BERs. In contrast, BER requires the measurement of a waveform at a discrete set of times (the latter of which can have errors), and then bins the measurement amplitude-wise before a decision can be made. It can thus be concluded that the mapping sought here is between a particular NMSE value and a resulting range of BER, with the mapping depending on the digital modulation used. By continuity, it follows that the lower the value of NMSE, the narrower the range of BER that can result. Hence to have an accurate prediction of BER (that is, with a small range of uncertainty), the NMSE for the system model must be sufficiently low, and this in turn requires any measurement instrument used to construct/validate the model to have an accuracy/repeatability that is in turn lower than the model NMSE by a sufficient margin (a good rule of thumb is to use 10 dB for this margin).

In view of the above conclusions with respect to BER versus NMSE, it essentially becomes imperative to develop

some estimation procedure for the NMSE needed to provide accurate BER predictions, the ultimate product of a system simulation. The issue of the required model fidelity needed for a given simulation has received relatively little attention in the literature. In addition, as it will be shown in the next subsection, the validation of models really needs to take place using operational-like signals as input. This need becomes more prominent in nonlinear contexts, and with operational signals having increasing bandwidth, data rates, and modulation complexity. With the use of operational signals both to construct and validate models, it becomes imperative to have a sufficiently accurate/repeatable time-domain measurement system such as the ATDMS. The required NMSE fidelity needed for both measurements and models can be systematically determined in a recursive manner from BER performance curves and system requirements. This procedure is described below in terms of the example illustrated in Figure 10, and assumes that both the noise that determines the BER and the modeling error are essentially additive white Gaussian in nature.

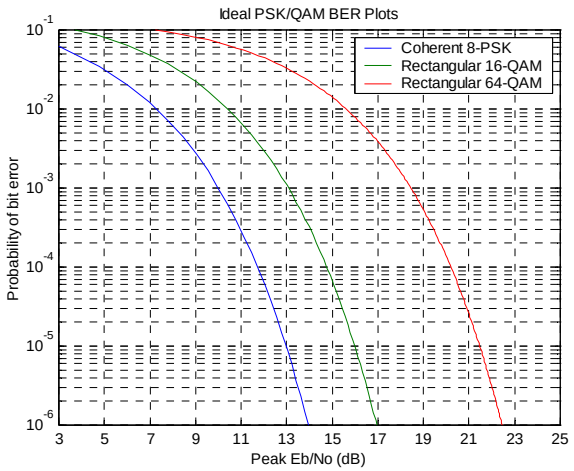


Figure 10 – Ideal BER performance curves for gray-coded PSK and QAM modulations. Such curves can serve as a basis for determining required model NMSE fidelity.

(i) The first estimate would come from the use of an ideal BER performance curve, with or without ideal error correction coding, that reflects the system at hand. Three such curves are shown in Figure 10, where there is no error correction coding applied, and the modulations (8-PSK, 16-QAM, and 64-QAM) are assumed to be gray coded in their symbol assignment. A BER performance value or range must then be chosen for which the model is desired to be accurate. Using the curves, the corresponding value or range of E_b/N_0 value or range is then found. In the case of the example in Figure 10, a value of 10^{-3} for the uncoded BER value can be chosen, resulting in E_b/N_0 values of 10, 13, and 18.4 dB for the 8-PSK, 16-QAM, and 64-QAM modulations, respectively. When using a range of BER values, the E_b/N_0 value corresponding to the lowest desired BER value would be used (this will be the largest E_b/N_0 value and the most demanding BER for the model to predict).

(ii) By negating the determined E_b/N_0 value in the dB scale, one arrives at the N_0/E_b value, which in the case of our example gives -10, -13, and -18.4 dB. Because BER assumes both noise (thermal, phase) and other distortions (deterministic and stochastic) are causing the bit errors, it follows that the modeling error must be sufficiently less than these sources so that the model itself is not contributing to the BER. It thus follows that the noise floor of the model, which is what is measured by the NMSE metric, must be sufficiently less than the practical value for N_0/E_b, which encompasses all the bit-error sources. In particular, a good estimate for the implementation loss between the ideal and practical systems must thus be subtracted from the ideal N_0/E_b value, and another margin must in turn be subtracted to ensure that the modeling error is sufficiently below the practical N_0/E_b value. In the case of our example, a 4 dB implementation loss was chosen, and an additional margin of 10 dB, corresponding to $\approx 10\%$ relative error for true noise estimation, was added. This results in a final model NMSE requirement of -24, -27, and -32.4 dB for 8-PSK, 16-QAM, and 64-QAM modulations, respectively. Another comfortable margin of 10 dB has to be added to these values in order to arrive at the required measurement accuracies/repeatabilities needed to construct/validate these models.

(iii) Using models constructed/validated that meet the required NMSE values from the previous step, an end-to-end link simulation is then conducted that includes all the expected imperfections in the channel, but preserving the use (or not) of error correction coding chosen for the original idealized case. This will result in another more realistic BER performance curve, for which the above steps, without the implementation loss factor, can be repeated. This will result in a more refined estimate of the model fidelity required, which is then compared with the original estimate. If the new estimate requires a higher model fidelity, then the models have to be likewise refined to meet the new requirements, and the exercise is repeated until convergence to a final fidelity requirement is achieved. If the new estimate requires a lower model fidelity, then the models are sufficiently accurate for the system and no further iteration is required.

As mentioned above, the use of accurate time-domain measurements is essentially imperative to properly validate models for operational signals. In order to obtain such measurements, the ATDMS can be applied as representatively shown

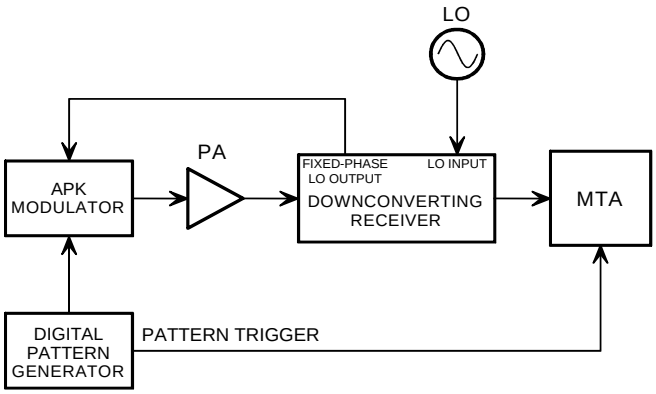


Figure 11 – Typical ATDMS configuration for making input/output time-domain measurements used in power amplifier model construction and validation. This setup used the more accurate MTA as the recording instrument, and an *amplitude phase keyed* (APK) modulator as a source.

in Figure 11 for a power amplifier. Here an *amplitude phase keyed* (APK) modulator is used as the source, with a digital pattern generator providing the artificial and repetitive modulation signal. In this case the more accurate MTA is used as the recording device, with a pattern trigger used to synchronize the signal generating and recording process. Note also that the APK modulator is locked to the DCR LO as needed for coherent downconversion. Because of the high accuracy of the ATDMS, the validation of block models is now possible and preferable in the time domain. Using this setup, a comparison of the simulation model output waveform prediction can be made with what is measured in the laboratory for the same measured input waveform.

Traditional VNA-Based Nonlinear Power Amplifier Block Models

The primary challenging focus in the modeling endeavor has been on power amplifier characterization, since it is the component that is primarily operated in a nonlinear fashion, and since such characterization must necessarily precede any distortion compensation design. The characterization of power amplifiers falls into three basic classes. *Physical models* provide the most fundamental descriptions of power amplifiers, making use of solid-state physics for SSAs or electromagnetics for TWTAs. Such models provide the highest fidelity and broadest signal class applicability. However, this must be traded against their often complex construction, and the fact that they take the longest to execute as a simulation model. Indeed, they are typically out of the question for system performance trades, where many thousands of simulation runs are required to arrive at a statistically representative value for BER. The next *circuit-equivalent model* class attempts to approximate the solid-state physics/electromagnetics with circuits whose parameters are extracted from carefully designed experiments. These models are primarily used for SSAs, but some work can be found for TWTAs that results in ladder-type circuits. This class is still typified by a large number of measurements and lengthy optimizations in order to construct, and are still too long to execute for system simulation purposes. The final *blackbox model* class, also known as *block models*, require only simple input/output measurements to construct. The goal here is to arrive at an accurate description of the operator that is the power amplifier. This is the approach of choice for system modeling and especially TWTA characterization, and will be the class used in the three examples to be presented in this paper. These models feature simple construction and quick simulation execution, but their fidelity capability has not really been fully explored, especially in the context of complex wideband signal environments.

As previously discussed, simulation block models are almost exclusively constructed at baseband using the LPE representation of signals, so that sampling the high-frequency carrier to compute waveforms is avoided. This powerful method is capable of analyzing systems driven by signals ranging from single-tone sinusoids to complex, digitally modulated carriers. A review of LPE theory can be found in the FTD characterization paper [18], as well as the classical text [21]. It suffices to note that for a sinusoid represented by $A\cos(2\pi ft+\theta)$, the complex envelope is given by $A\exp(\jmath\theta)$. In addition, any general bandpass signal can be written in the form of the above sinusoid, where A and θ become functions of time, resulting in a time-varying complex envelope.

Traditional block models are based on simple, single-tone and swept-tone RF measurements that can be readily and quickly made with a standard VNA. The simplest form of this model is the so-called *1-box model* that consists of a single, memoryless nonlinearity that describes how the amplitude and phase of a centerband sinusoid is changed when passing through the power amplifier (see Figure 12). In particular, a VNA measurement is used

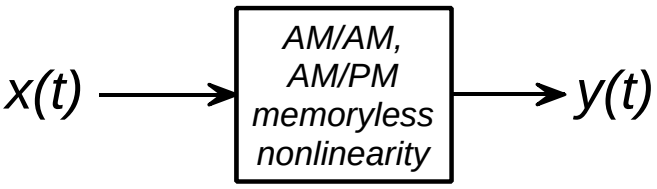


Figure 12 – One-box nonlinear block model.

to determine the output amplitude versus input amplitude transfer characteristic (called the *AM/AM* curve), as well as the output phase shift versus input amplitude (called the *AM/PM* curve). Quantitatively, the definition is given by:

$$\left.\begin{array}{ll} \textit{Input signal:} & x(t)=A\cos(2\pi f_c t+\theta_c) \\ \textit{Output signal:} & y(t)=G(A)\cos[2\pi f_c t+\theta_c+\Theta(A)] \end{array}\right\} \quad (2)$$

where f_c is the carrier frequency, θ_c is an arbitrary phase, $G(A)$ represents the AM/AM conversion, while $\Theta(A)$ is the AM/PM conversion. A generic example of these nonlinear conversions that are typical for a TWTA and similar for an

SSA is given in Figure 13. These conversions are seen to

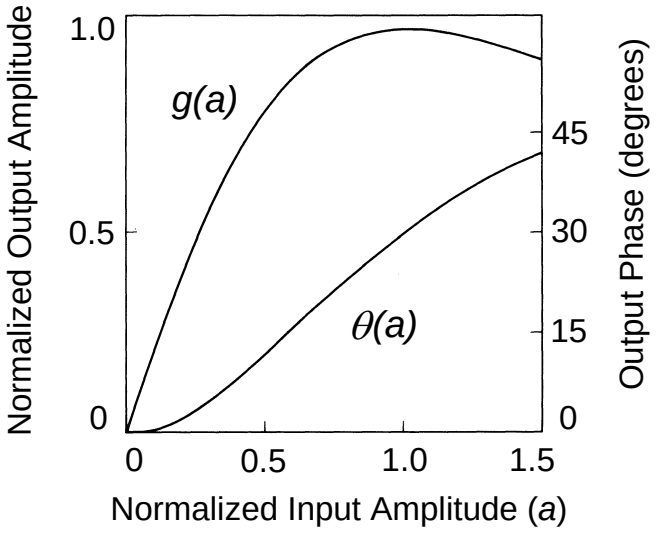


Figure 13 – Generic power amplifier AM/AM and AM/PM conversion curves. The quantities a, g, and θ are normalized versions of A, G, and Θ in Eq. (2).

be of an envelope type since the frequency f_c serves only as a reference. This means that the LPE form of the model will simply act on the envelope of the complex baseband input signal. To apply the model to a general bandpass input signal, the signal is first rewritten as discussed above, and then thought of as instantaneously sinusoidal so that the nonlinear conversions can be directly applied. It is this model that is implemented in commercial computer-aided design programs [28, 29].

There are several important observations to be made here. First, this model assumes that inputting a sinusoid into a power amplifier simply results in another sinusoid at the output with the same frequency. This is not correct unless a so-called *zonal filter* is also placed at the output of the power amplifier, since it is well known that harmonics will be generated in the nonlinear region of operation. Note that the measurement automatically accomplishes this operation since the VNA detects only signals at the same frequency as the input. This model also assumes, for the case of more general signals with finite frequency content, that the above conversions measured at centerband do not vary across the band of interest. This is often not the case and is thus one of the causes for the model's fidelity breakdown for wideband signals. This is a manifestation of the finite memory that power amplifiers have, especially TWTAs, which is approximated to be zero in this model. Another cause for breakdown is the fact that for such general signals, unique nonlinear intermodulation and memory effects arise because of the simultaneous presence of multiple frequencies at the input. As already discussed above, sinusoids cannot identify nonlinear systems, so that it is only by the continuity of the nonlinear power amplifier operator that the model works adequately for relatively narrowband signals that are close to sinusoids.

In a step up from the 1-box model, perhaps the simplest nonlinear model with memory is the so-called *2-box model*, in which a filter is placed in front of the 1-box model (see Figure 14). This model is especially used for TWTAs

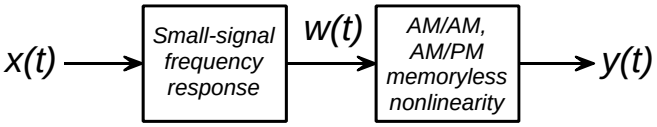


Figure 14 – Two-box nonlinear block model.

because of their significant memory and the fact that it reflects the natural physical operation of the tube, where its linear behavior occurs in the input section, while its nonlinear behavior takes place in the output section. The most common form of the filter derives from a simple VNA sweep of the power amplifier in its small-signal range of operation. In order not to count the small-signal response twice in the overall model, a normalization step has to be applied to either the filter or the nonlinear conversion curves. For the filter, this normalization consists of scaling/shifting the amplitude/phase frequency response until it has unity gain and zero phase at centerband. Equivalently, the normalization of the conversion curves would consist of scaling the AM/AM curve to have unity slope in its linear, small-signal region, and to shift the AM/PM curve vertically until it has a zero phase shift in the same small-signal region (where the phase shift would be constant). Note that the normalization of the filter is more precise than that of the conversion curves.

Mathematically, it can be shown that the pre-filter amplitude response actually translates (with respect to the instantaneous input frequency) both the centerband AM/AM and AM/PM conversion curves horizontally along their identical input axes, while the pre-filter phase response translates the centerband AM/PM conversion curves vertically along its output phase shift axis. Although this frequency-dependent conversion curve translation is an improvement over the 1-box model, it still fails to capture changes in the shape of these conversion curves with respect to frequency. It is for this reason, together with those cited above for the 1-box model, that the 2-box model will again break down for wideband signals.

Continuing in the same theme, the *3-box* model places an output filter on the 2-box model to provide more "fitting capability," as illustrated in Figure 15. This model is

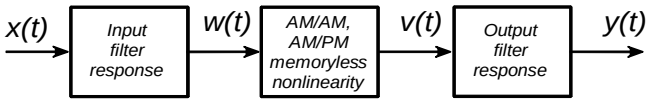


Figure 15 – Three-box nonlinear block model.

relatively unknown and various ways have been proposed to populate the two filter boxes. The original form of the model is precisely specified as follows. First, in addition to the small-signal VNA sweep used for the 2-box model, a complementary large-signal VNA sweep of the power amplifier is made using a tone whose constant power level saturates the device at centerband. Denoting the small-signal response by $H_{ss}(f)$ and the large-signal response by $H_{ls}(f)$, it follows that the input filter response $H_{in}(f)$ and the output filter response $H_{out}(f)$ are given by

$$H_{in}(f) = \frac{H_{ss}(f)}{|H_{ls}(f)|}, \quad H_{out}(f) = |H_{ls}(f)|, \qquad (3)$$

so that the large-signal phase response is not used. Note that Eq. (3) affects a partial normalization of the overall model since $H_{in}(f)H_{out}(f) = H_{ss}(f)$ in the small-signal region. The remaining normalization can again take place in either of two manners. Like the 2-box model, the AM/AM and AM/PM curves can be made to have unity slope and zero phase shift, respectively, in their small-signal regions. Alternatively, both $H_{in}(f)$ and $H_{out}(f)$ can be made to have unity gain and zero phase at centerband, which once more is the less ambiguous approach.

Mathematically what occurs in this case is that $H_{in}(f)$ translates the centerband conversion curves with respect to the instantaneous input frequency just like the pre-filter did for the 2-box model, while $H_{out}(f)$ translates the AM/AM curve vertically along its output axis and the AM/PM curve vertically along its output phase shift axis, with the former translation being new compared to the 2-box model. However, the shape of the conversion curves still remains the same, as in the case of the 2-box model. As expected, this model will perform better than the 2-box model, but again will break down for wideband signals since it is still based on simple tonal measurements.

There are several refinements and variations of the above n-box models that have been proposed, most of which change the memoryless nonlinearity from its simple centerband AM/AM and AM/PM conversion curves. One set of these changes involves the use of dynamic measurements to arrive at generalized forms of the tonal nonlinear conversion curves (such as based on two-tones and modulated signals instead of sinusoids; see [30, 31], for example). The other set of modifications involves the measurement and fitting of the tonal AM/AM and AM/PM conversion curves with respect to frequency, accounting for their change in shape [32, 33]. With respect to the filters, one variation that is sometimes applied involves a more sophisticated VNA sweep done in the presence of a saturating centerband tone. The idea here is to try to arrive at a linear approximation to the nonlinearity in the hopes of improving the model's predictive fidelity. Another variation of the 3-box model that will be discussed in subsection one of Section 6 will make use of the first step of the polyspectral method to arrive at an improved estimation of $H_{out}(f)$. More advanced techniques, such as using a nonlinear form of the autoregressive moving average method, as well as general neural network approaches have also been explored [34, 35, 36]. Although these models may provide higher predictive fidelity compared to the baseline n-box models described above, they can be much more tedious/difficult to construct and still remain just attempts at fitting the nonlinear operator that is the power amplifier, and thus are not of a formal system identification nature. The latter nature, however, is precisely the basis for the polyspectral method to be described in the next section.

To give concrete insight into how the baseline n-box models compare, especially in the context of wideband modulated signals, 1- and 2-box models were constructed and evaluated for a commercial GaAs FET 20-GHz SSA. The small-signal response of the SSA is shown in Figure 16 across an 8-GHz bandwidth, while Figure 17 provides the nonlinear conversions that have been fitted with Bessel series curves for model execution. Two types of digitally modulated signals

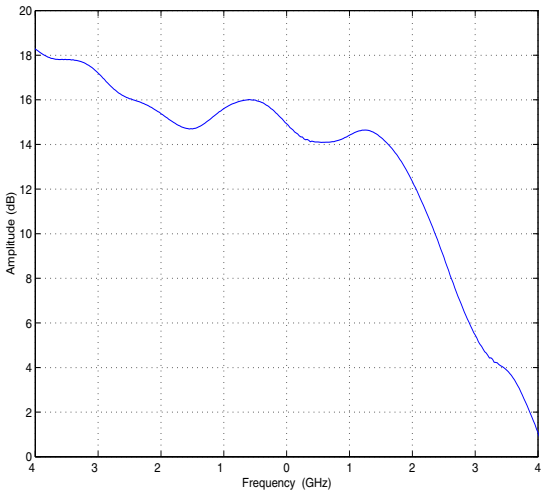
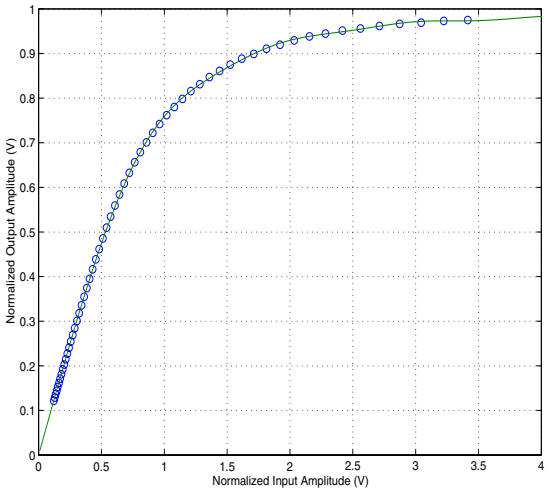


Figure 16 – Small-signal response of a commercial 20-GHz SSA.

were used to exercise the models, which were recorded with the ATDMS (input and output), and the model fidelity was evaluated in terms of the NMSE metric defined above. The operating point for the signals was chosen to be at the 3-dB compression point on the output versus input power curve. The simpler constant envelope *bi-phase phase shift keying* (BPSK) signals used ranged in data rate from 150 Mbps to 2.4 Gbps. A more complex multi-level 16-*quadrature amplitude modulation* (QAM) signal was used at a fixed rate of 9.6 Gbps, but whose spectral nature was modified on the input of the SSA by applying pre-filters of 3-, 4-, and 5-GHz bandwidths, as well as at full bandwidth with no pre-filter. Tables 1 and 2 provide the results found in the BPSK and 16-QAM cases, respectively. Three important basic conclusions can be drawn from these results that could have been anticipated from the above discussions. First, the model fidelity decreases as the input signal bandwidth

(a)



(b)

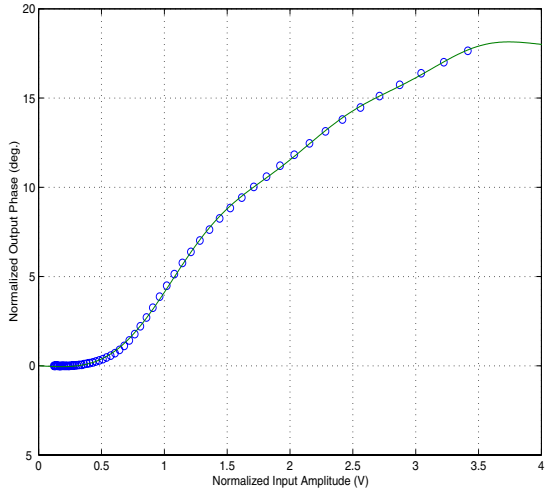


Figure 17 – Nonlinear conversions with Bessel series curve fits for the SSA in Figure 16. (a) AM/AM characteristic. (b) AM/PM characteristic.

increases, whether caused by increasing the data rate in the BPSK case or widening the pre-filter in the 16-QAM case. Second, the 2-box model is superior to the 1-box model for all signals because of its capture of natural SSA filtering effects. Third, the model fidelity for both models when applied to the constant-envelope BPSK signals was superior to that for the multi-level 16-QAM signals. Note that the 16-QAM signals vary further than the BPSK signals from the sinusoids used to obtain the model parameters. It could readily be extrapolated from these results that if a 3-box model had been constructed, it would have been superior to the 2-box model in all cases, and would still preserve the fidelity trends for bandwidth and signal complexity. The next section will demonstrate a modeling approach that will be a significant improvement over all of these tonal-based

Table 1 – Amplifier model NMSE for BPSK signals.

Data Rate (Mbps)	*One-Box Model NMSE (dB)*	*Two-Box Model NMSE (dB)*
150	−27.92	−33.69
600	−22.64	−28.34
1200	−19.84	−26.24
2400	−18.72	−24.31

Table 2 – Amplifier model NMSE for 9.6 Gbps 16-QAM signals.

Bandlimiting Filter (GHz)	*One-Box Model NMSE (dB)*	*Two-Box Model NMSE (dB)*
3.0	−20.04	−24.30
4.0	−19.86	−23.61
5.0	−19.31	−23.19
None	−15.62	−19.44

methods.

6. THE POLYSPECTRAL METHOD

Theoretical Basis, Basics, and Features

The polyspectral model approach is a well-established technique in the spectral analysis community and has been used successfully for low-frequency linear and nonlinear mechanical systems for some time [37, 38]. The application to a communications context has remained theoretical at best until the development of the ATDMS, which has enabled the accurate recording of input/output time-domain waveforms that lies at the heart of its construction. It is possible, especially given the excellent results already found for this approach and its many powerful features (to be highlighted here), that the polyspectral method will become a major tool in the nonlinear modeling and distortion mitigation design of modern communication systems. More details about the method can be found in the announcement publications made by our group [39, 40, 41].

The theoretical basis for the polyspectral approach lies in the formal input/output identification of nonlinear systems [42, 43], so as to arrive at a characterization that will be accurately predictive for a large class of input signals. One of the primary means of achieving this identification is to treat the given nonlinear system as a mathematical input/output operator (with memory) that is then subsequently represented locally as a series. This formal approach is in stark contrast to the traditional block models discussed above. Among the several general forms of the operator series that have been proposed [5, 44], the original and most well known is the *Volterra series* that provides a generalization of the familiar linear impulse response concept. Recall that the two basic

classes of time-invariant systems with memory are: linear systems that obey the superposition principle, and nonlinear systems that do not. Assuming only causal systems that correspond to physical systems, recall that a linear system is globally identified by its impulse response $h(\tau)$, in that for any general input signal $x(t)$, the output $y(t)$ is given by

$$y(t) = \mathcal{L}[x(t)] = \int_0^t h(\tau)x(t-\tau)d\tau, \tag{4a}$$

where $\mathcal{L}[\cdot]$ represents the system operator, and the integral is assumed to be well defined. In the case of a nonlinear system, the classical form of the Volterra series representation for the system operator $\mathcal{N}[\cdot]$ is given by

$$y(t) = \mathcal{N}[x(t)] = h_0 + \int_0^T h_1(\tau_1)x(t-\tau_1)d\tau_1 + \sum_{n=2}^{\infty} \int_0^T \cdots \int_0^T h_n(\tau_1, \ldots, \tau_n) \prod_{j=1}^{n} x(t-\tau_j)d\tau_j, \tag{4b}$$

where $h_i(\tau_1, \ldots, \tau_i)$, $i = 0, \ldots,$ represents the i^{th}-order Volterra kernel that is the generalized version of $h(\tau)$. The basic distinction to make between the linear and nonlinear representations is that the latter is only locally applicable, and it is a series, which has several important consequences. First, this means that such a representation will necessarily be only an approximation at best, since not all of the terms of the series can be determined or retained. Second, care must also be exercised with respect to the region of convergence for the series, which itself is difficult to determine. Third, the block models presented earlier can be thought of as ad-hoc attempts at fitting the subject operator, given that they have equivalent series representations that may happen to have sufficient degrees of freedom to do the job. Finally, note from Eq. (4b) that the Volterra model can be represented as a series of parallel branches emanating from the input and summed on the output, where the i^{th} branch is represented by the i^{th}-order Volterra kernel.

Although many approaches (both time- and frequency-domain) have been proposed for the determination of the Volterra kernels [5, 38, 44], all of them become prohibitively complex and unacceptably inaccurate as the number of terms increases. As indicated in [38], for example, the frequency-domain construction approaches require multi-dimensional Fourier transforms and a very large number of time-domain measurements in order to arrive at statistically accurate estimates of the kernels. Nevertheless, the Volterra representation provides many important insights into nonlinear system behavior that warrant its construction when at all possible. Such insights include the natural quantification of distortion into its components [offsets, *linear intersymbol interference* (ISI), nonlinear ISI], a similar beneficial breakdown of nonlinear memory effects and their consequences, and its natural application to nonlinear equalizer design, as well as the development of analytical link models useful for bounding BER performance [2].

To overcome the aforementioned model construction difficulties, the polyspectral method essentially makes use of pre-chosen memoryless nonlinearities cascaded with unknown filter parameters to arrive at a single-frequency, simplified version of the Volterra model. Although there is some sacrifice in the range of representable operators compared to the Volterra approach, this simplification greatly reduces the model construction measurement and calculation burden, yet provides sufficient degrees of freedom to accurately model power amplifiers (see the concrete application presented in subsection two of Section 6, below). The unknown filter parameters are determined from a statistically representative set of measured input/output time-domain records. In particular, these time-domain measurements employ random-like stimuli with well-defined distribution properties, such as can be provided by pseudo-randomly modulated signals, for example. The additional measurements needed over traditional frequency-domain models is more than offset by the efficient ATDMS automation in LabVIEW®discussed previously, not to mention the significant enhancement of model predictive fidelity, as well as several other powerful and desirable features that the method provides (as discussed next). As a consequence of its construction procedure, the polyspectral model can be considered to be a time-domain model, an additional contrast with traditional VNA-based models.

Before going on to a description of the basic model construction steps and architectures of the polyspectral method, there are several key advantageous features of this approach to point out for both modeling and distortion mitigation analysis/design applications. First of all, the model filter parameters are determined by closed-form spectral expressions that globally minimize the model error for the chosen architecture. This is an extremely important feature since the difficulties with false solutions in nonlinear optimization procedures are well known. Second, the method provides for a complete and quantitative error analysis, both with respect to the model fidelity and the bias and random errors in the model filter parameters. The latter error estimates provide for a systematic design of experiment for the data gathering needed to construct the model (such as the number N_r of input/output records to measure). Another very important feature is that parallel branches can be added to the model, like the terms of a series, and through a decorrelation procedure this will lead to a guaranteed increase in model fidelity. Fourth, the method is also very relevant to the development of nonlinear compensator designs, since these are simply realizations of an inverse model that can be readily constructed in the same way, and result in structures that are natural for hardware implementation. In particular, the first linear step in this method, to be described below, can itself be useful for deriving an optimal linear equalizer and provides a means for quantifying distortion mitigation effectiveness (see [39] for more details about polyspectral compensation applications). Finally, once such a model is constructed for the critical power amplifier that is found in all communications links, it can be combined with other linear or nonlinear channel component models to arrive at an entire analytical link model,

which can subsequently be used in conjunction with the Volterra series method to bound BER and develop nonlinear equalizer designs (beyond the scope of this paper).

The first polyspectral model architecture to be described here, which is essentially an enhancement of the 3-box model discussed above, rests on a fundamental step and concept of the polyspectral method itself, that of the so-called *optimal filter*. In a nutshell, the optimal filter, denoted by $H_o(f)$, provides the best linear approximation to a given nonlinear system in the sense illustrated by the model shown in Figure 18 and explained below. In particular, the error

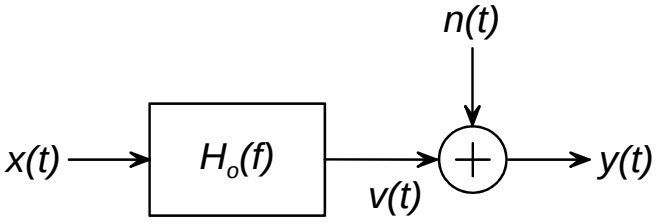


Figure 18 – Optimal linear model for a nonlinear system.

power between the optimal filter's output $v(t)$ and the actual system output $y(t)$, captured in the signal $n(t)$, is minimized with respect to the given class of input signals $x(t)$. As a result, this filter will in general vary with operating point and signal type. The explicit expression for the optimal filter response is given by:

$$H_o(f) = \frac{S_{xy}(f)}{S_{xx}(f)}\,, \tag{5}$$

where S_{xy} is the *cross-spectral density function* between the input and output, while S_{xx} is the *autospectral density* of the input. The calculation of these spectral densities is performed using measured input/output time-domain data records of finite duration, where the input signal is of a stationary random nature. Because our application of this approach used baseband measurements, it follows that all the signals (and hence records) here are complex in nature and of the form $\tilde{z}(t) = z_I(t) + \jmath z_Q(t)$, where the real part is the usual in-phase component and the imaginary part is the quadrature component. Assuming the length of the data record is T, and the number of data records is N_r, these spectral densities are smoothly estimated from:

$$S_{\tilde{x}\tilde{y}}(f) \;=\; \frac{1}{N_r T}\sum_{k=1}^{N_r} X_k^*(f,T)Y_k(f,T) \tag{6a}$$

$$S_{\tilde{x}\tilde{x}}(f) \;=\; \frac{1}{N_r T}\sum_{k=1}^{N_r} |X_k(f,T)|^2 \tag{6b}$$

where $X_k(f,T)$ and $Y_k(f,T)$ are the finite Fourier transforms of the kth input ($\tilde{z}$) and output ($\tilde{z}$) baseband data records, respectively, and are easily calculated using the standard FFT algorithm. Formally, the response $H_o(f)$ minimizes the autospectral density of $n(t)$ with respect to all possible filter responses. In addition, it can be shown that with this response, both the input $x(t)$ and output $v(t)$ will *not* correlate with the modeling noise $n(t)$. This means that all the linearity between the original measured input and output has been extracted into $H_o(f)$ and only a strictly nonlinear relationship (if any) remains between $x(t)$ and $n(t)$. It is the next steps of the polyspectral modeling approach that seek to capture this residual relationship.

The sources and quantification of the random and bias errors of these spectral density estimates have been worked out in detail [37]. It is found that with a careful design-of-experiment and a sufficient number of data records, these errors can be held to adequately low values for satisfactory modeling fidelity and compensation design. The frequency resolution of the resulting filter is given by $\Delta f = 1/T$. The error of the resulting optimal linear model is quantitatively captured in the frequency-dependent *linear coherence function* (LCF), which is denoted by $\gamma_{xy}^2(f)$ and defined as:

$$\gamma_{xy}^2(f) := \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} = 1 - \frac{S_{nn}(f)}{S_{yy}(f)}\,. \tag{7}$$

Specifically, the relative modeling error is the unity complement of $\gamma_{xy}^2(f)$ as seen in Eq. (7). The LCF lies between zero and one and provides a measure of the linearity of the given system. In particular, a completely linear system will have an LCF identically equal to one for all frequencies (and hence an identically zero modeling error). This property also makes the LCF a useful waveform metric for the assessment of a given nonlinear compensation approach, since it can be used to calculate the increase in linearity of the system after its insertion.

There are several ways in which the optimal filter could be used to improve current nonlinear block models. For example, one could augment the standard 2-box model with an optimal output filter to arrive at a new 3-box model, or one could replace the output filter of the 3-box model with the optimal filter. Because the baseline 3-box model is more accurate than the baseline 2-box model, it follows that the optimally enhanced 3-box model will be more accurate than the optimally augmented 2-box model. In both cases, the optimal filter is constructed between the output of the memoryless nonlinearity of the original model and the measured output of the subject power amplifier.

A concrete application of this approach was made using the augmented 2-box model case for a 20-GHz TWTA operating near saturation with 9.6 Gbps, 16-APK input signals. The construction data set used consisted of 100 data records each of length 40 ns (corresponding to a frequency resolution of 25 MHz) with a sampling interval of 26 ps. Each input data record was produced from a 20-GHz 16-APK source modulated with a 96-symbol pseudo-random sequence. Constructing the model described resulted in the output filter amplitude response and LCF shown in Figure 19. The amplitude response exhibits an approximately

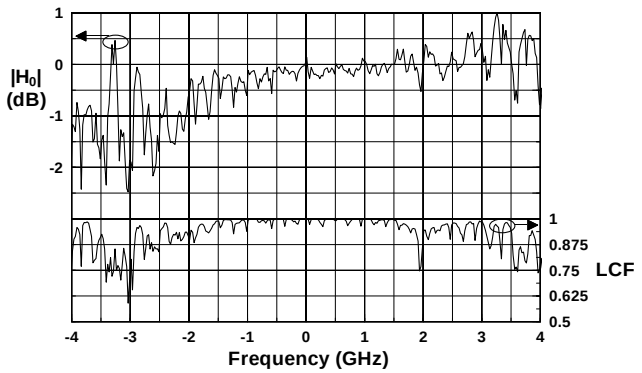


Figure 19 – Enhanced 3-box model output filter magnitude and associated LCF for a 20-GHz TWTA.

2 dB positive gain slope over the 8 GHz bandwidth. The LCF indicates that the two box-model captures a significant amount of the nonlinearity of the TWTA since it is very close to unity over most of the band. The accuracy of the 2-box and 3-box models was also compared using the NMSE metric defined in Eq. (1). A histogram of the NMSE calculated for each data record is shown in Figure 20 for the two models. The mean NMSE found for the 2-box model

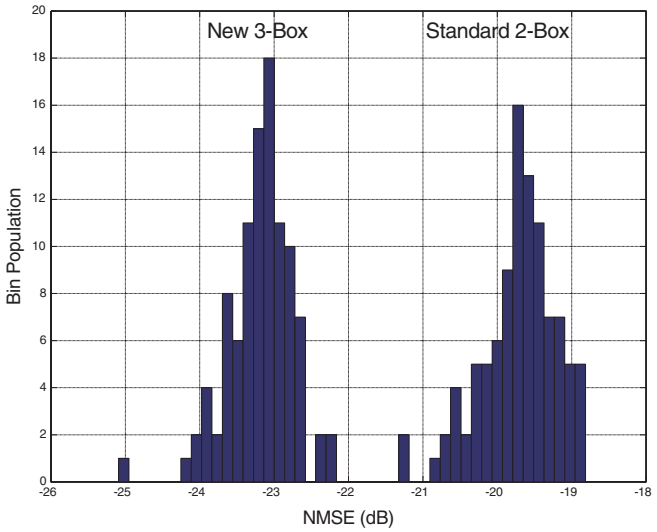


Figure 20 – Histogram NMSE comparison of the standard 2-box and new 3-box models for a 20-GHz TWTA.

was -19.72 dB compared to -23.19 dB for the 3-box model, an improvement of approximately 3.5 dB.

General Polyspectral Model Architectures and Power Amplifier Applications

The general and formal architecture of the polyspectral model consists of a parallel-path structure made up of the user-chosen memoryless nonlinearities and unknown parameter filters discussed above. In all cases, there is one totally linear path, and at least one nonlinear path that differentiates the two basic classes of the model (see Figure 21). These classes are dictated by the placement of the parameter filter before

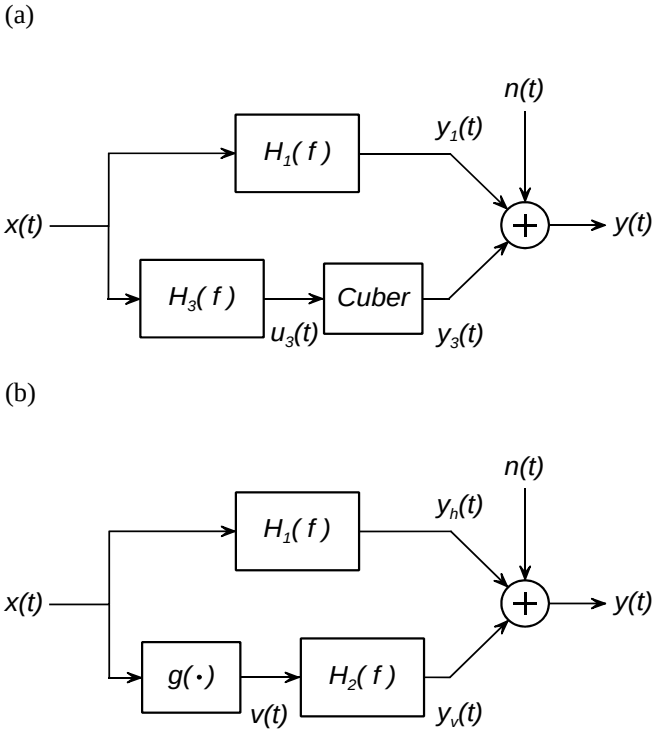


Figure 21 – Two-branch examples of the two basic classes of polyspectral models. The signal $n(t)$ represents the error between the model prediction and the measured output $y(t)$. (a) Filter-nonlinearity model structure. (b) Nonlinearity-filter model structure.

the nonlinearity, termed the *filter-nonlinearity* (FN) class, or after the nonlinearity, termed the *nonlinearity-filter* (NF) class. Hybrid versions of the model can also be proposed that involve both types of nonlinear branches. One can interpret the branches as terms in the operator series being developed, with the filters acting as frequency-dependent coefficients for the series that are determined explicitly by the method. In their original form, these models applied to real signals, so that some modification was necessary to allow them to be baseband models as well. The chosen nonlinearities are limited to simple monomials for the FN type, while the NF type allows for general memoryless nonlinearities (such as the traditional AM/AM and AM/PM envelope conversion curves standardly used for power amplifiers). In the models illustrated in Figure 21, the error signal $n(t)$ represents the error between the model output (given by $y_1 + y_3$ for the FN type, and $y_h + y_v$ for the NF type) and measured output that is represented by the signal $y(t)$. The polyspectral approach provides for a unique quantification of this error, and the simulation version of these models will of course leave this signal out when used in their predictive capacity. Also note that the FN type representation is quite similar to the power series look of the Volterra series, whereas the NF type allows for the consolidation of many such terms into one branch. It is this property, together with a simpler construction procedure, that makes the NF type of model more desirable for application; and it is indeed the basis for our own proposed power amplifier polyspectral model

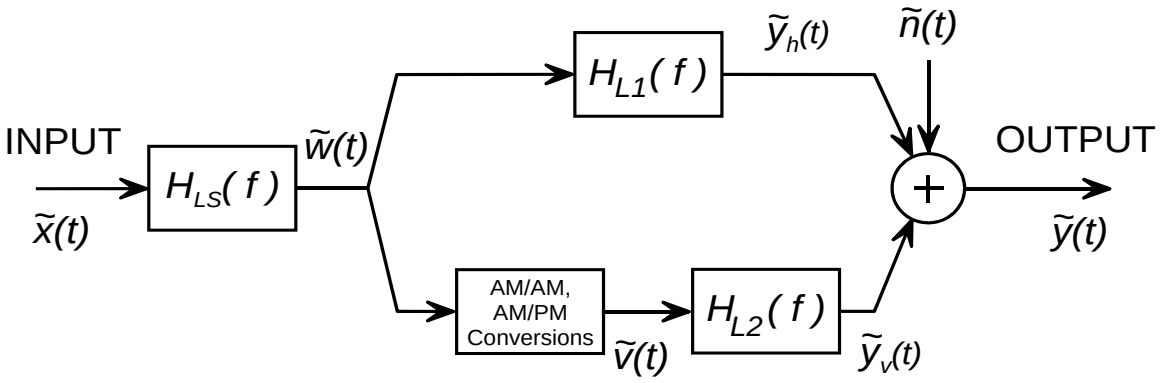


Figure 22 – Specific modified baseband two-branch NF polyspectral model found to be especially effective for power amplifier characterization.

discussed next.

A concrete example of a developed polyspectral model is shown in Figure 22, which is a baseband modification of the standard NF class model that has been successfully applied to wideband SSAs and TWTAs. The modification consisted of the addition of a pre-filter $H_{LS}(f)$ that again matches more closely the physical operation of the tube discussed above. This pre-filter was taken to be that for the 3-box model discussed in subsection two of Section 5, above, with response given explicitly by Eq. (3). Note that the memoryless nonlinearity was still chosen to be the one used in the standard 1-box model. This has been done for want of a better choice, which is actually another motivation for using this series approach. Also note that because of the pre-filter and this nonlinearity, the model actually contains the front 2-box portion of the 3-box model in its nonlinear branch. In this case, there are two unknown parameter LPE filter responses to determine, denoted as $H_{Li}(f),\ i = 1, 2$ in the figure. Explicit expressions for these filters and the frequency-dependent modeling error provided by the polyspectral method can be found in [41].

The details of a specific wideband 20-GHz helix TWTA application were as follows. The stimuli used to make the construction measurements came from an in-house fabricated, 12 Gbps, 16-APK source that modulated 240-symbol pseudo-random sequences. The TWTA was operated with the outer 16-APK constellation points at a saturated power level. There were $N_r = 200$ input/output time-domain records taken, with a duration of $T = 80$ ns, and containing $N_d = 4096$ samples per record. The former duration gave rise to a 12.5-MHz resolution in the parameter filters, while the latter number of samples corresponded to about 17 samples per modulation symbol. The sequences were chosen to have a period of T so that no windowing would be needed in the discrete Fourier transforms that have to be taken. In addition to the polyspectral model, a 3-box model as described in subsection two of Section 5 above was also constructed and served as the basis for the former model (with respect to its first two boxes). Figure 23 illustrates the 3-box model components, consisting of the amplitude responses for the two filter boxes over an 8-GHz bandwidth, and the normalized AM/AM and AM/PM curves for the TWTA. Note how the pre-filter amplitude response

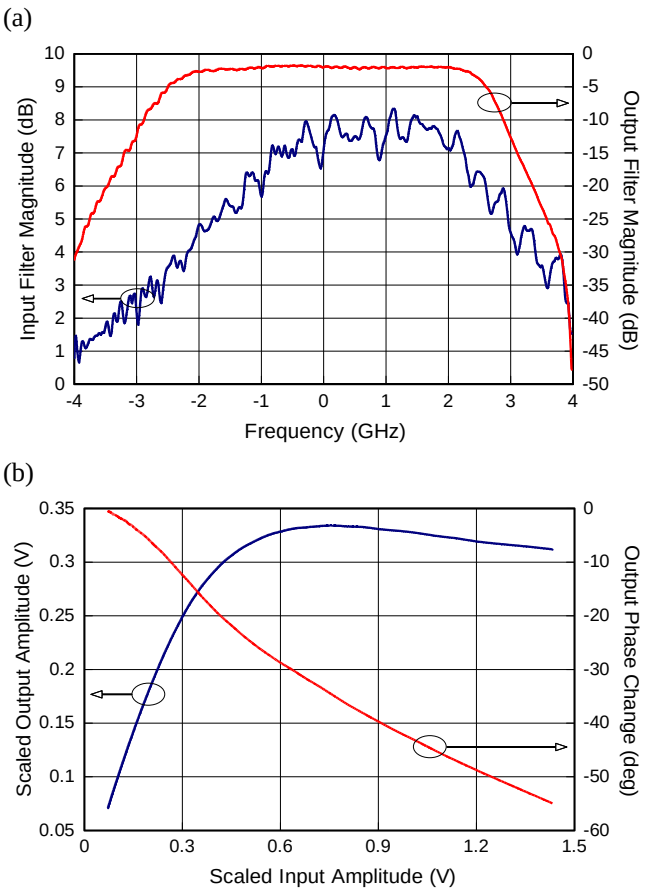


Figure 23 – Three-box model components for a 20-GHz helix TWTA. (a) Pre- and post-filter magnitude responses. (b) Normalized AM/AM and AM/PM conversion curves.

of the TWTA [blue curve in Figure 23(a)] exhibits a strong parabolic amplitude that is characteristic of a tube, while the AM/AM and AM/PM curves are similar in appearance

to the ones for the SSA shown above in Figure 16. Both of these components were used in the construction of the polyspectral model in Figure 22. Figure 24 shows the obtained amplitude responses for the two parameter filters, whose small magnitudes come from the fact that there was an attenuating output coupler added to the tube, so that the output signals were again at a level that could be measured with an MTA or DSO. The sharp spikes seen in

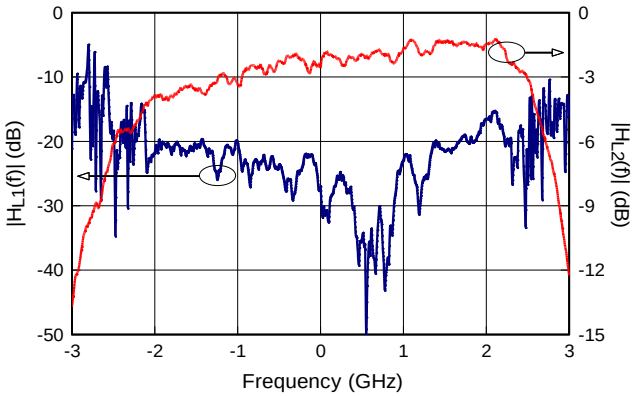


Figure 24 – Polyspectral model branch filter magnitude responses for a 20-GHz helix TWTA, where H_{L1} (H_{L2}) corresponds to the linear (nonlinear) branch.

the linear branch filter response [$H_{L1}(f)$, blue curve], and to a lesser degree in the nonlinear branch filter response [$H_{L2}(f)$, red curve], are a result of the fact that the 16-APK source had only one data stream feeding it instead of the four independent ones normally needed for this type of modulation. The spikes correspond to unwanted correlations caused by the delays used on the single data stream.

To validate the model, a large set of other input/output time-domain measurements were made for various operating conditions, including three separate operating points (saturation like the construction records, 2 dB and 6 dB input backoff), short and long record lengths (40 and 80 ns), two data rates (9.6 and 12 Gbps), and four 16-APK constellation geometries (two corresponding to each data rate). For each set of measurements, $N_r = 100$, $N_d = 2048$ or 4096 for the short and long records, respectively, and the sampling interval was $\Delta t = 19.5$ ps. For the short (long) records, the pseudo-random sequences were of length 96 and 120 symbols (192 and 240 symbols) for the low and high data rates, respectively. As usual, the NMSE metric of Eq. (1a) was used to determine the model fidelity. Despite the unwanted correlations in the signals mentioned above, Figure 25 shows the excellent NMSE results found compared to the 3-box model (the best of the baseline n-box models), with a 5.6 dB average improvement found for 100 records measured at the lower data rate (thus different 16-APK signal constellation), the saturating operating point, and over a 6-GHz bandwidth. Measurements and evaluations at other operating points, data rates, constellation geometries, and output bandwidths resulted in similar gains for the polyspectral model and also

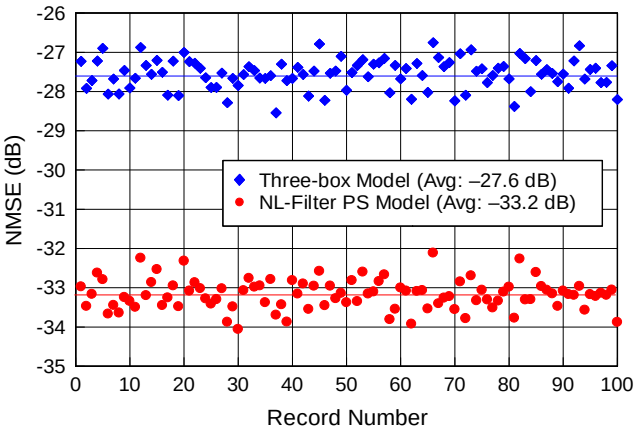


Figure 25 – NMSE comparison of polyspectral and three-box models for a 20-GHz helix TWTA under saturating, 16-APK, and 9.6 Gbps excitation. The horizontal lines represent the average NMSE.

strongly indicated its identification of the TWTA operator in the vicinity of the chosen construction point. Specifically, average NMSE values were in the -32 and -34 dB range (corresponding to a relative error range of 0.04 to 0.06%) over a 6-GHz bandwidth, and in the -33 to -35 dB range (0.03 to 0.05% relative error) over a 3-GHz bandwidth. These values are much more than sufficient for this modulation complexity, and are even quite suitable for 64-QAM simulations.

7. SUMMARY AND CONCLUSIONS

This paper has provided a comprehensive survey of both measurement and modeling techniques as they impact the challenges of modern communication systems analysis and design, with an emphasis on the time-domain approaches developed by the authors. The abundant motivation for more accurate computer-aided engineering models, especially in the context of wideband nonlinear components and channels was provided, with a fundamental need indicated for construction and validation measurements, as well as system diagnostic tools. A comparison was provided between the more established frequency-domain measurements and the less developed time-domain measurements, pointing out the advantages of the latter for the new challenging communication scenarios that are emerging. A strong argument for the performance of time-domain measurements at lower frequencies (including the baseband extreme) was made, leading to the fundamental problem that must be solved to accomplish this practice: the accurate and full characterization of the downconversion process. Among the advantages indicated for this strategy were the elimination of RF carrier phase noise afforded by coherent downconversion, longer time records or finer time sample resolution for a given recording instrument's memory allocation, usage of the more accurate baseband portion of the instrument's frequency range, and the production of LPE data that is directly insertable into current commercial and in-house computer-aided engineering simulation packages.

A solution to this fundamental measurement challenge was provided by describing several characterization methods for frequency-translating devices that are common to all communication systems. Features, assumptions, and procedures were provided for the two basic sideband classes of frequency-translating devices. Validating examples were provided for both classes, with the baseband double sideband method providing the enabling basis for the new Aerospace Time-Domain Measurement System (ATDMS) discussed next. The implementation of the baseband time-domain measurement technique was described, including the necessary assumptions and measurement considerations associated with its successful application. Calibration, measurement, and post-data processing procedures were briefly covered for the original system, which has since been enhanced to further increase its repeatability and absolute accuracy [24, 25]. The system was then validated by comparison with standard direct RF measurements.

The second part of the paper addressed the model construction and validation applications of the state-of-the-art time-domain measurement system. The treatment began with a careful and detailed discussion of the modeling process and its several associated important issues, such as which signals to use to validate a model for a given context, which metric to use to compare a model's predictions to hardware measurements, and what level of predictive fidelity is needed to accurately simulate a given system. With respect to the first issue, it was argued that operational signals need to be used to best accomplish model validation, and that this requires accurate time-domain measurements such as provided by the ATDMS. For the second issue, both the well known BER performance metric and a new NMSE waveform metric were described and compared, with the latter shown to be superior for component model assessment. The complex relationship between these metrics was discussed, and an iterative procedure involving standard BER performance curves was presented to determine the NMSE model fidelity required for a given communications system. An ATDMS-based setup was then described for use in both the construction and validation of system-level models.

The next set of discussions centered on the modeling of power amplifiers, beginning first with the three basic classes of models (physics-based, equivalent circuits, and blackbox or block). It was shown that blackbox models are really the only viable alternative for the simulation trades needed to analyze and design modern communication systems. This is because each trade typically requires thousands of simulation runs to arrive at statistically meaningful BERs, for example, and only block models can execute fast enough to make such runs acceptable in duration. Traditional, and not so traditional, n-box system models were presented next, including details about their architectures, construction procedures, and qualitative features. It was concluded that the 3-box model provides the highest predictive fidelity, although in all cases no system identification is provided by the approaches, only an attempt at fitting the nonlinear input/output operator (with memory) using tonal input signals. This sentiment remained even for the several variations of the models discussed. To illustrate these points, and to further compare the models, an example was provided that concerned the time-domain validation of the baseline nonlinear 1-box and 2-box models that used simple sinusoidal measurements. With their application to a 20-GHz wideband SSA for both BPSK and 16-QAM modulations, the expected limits of these models were uncovered, suggesting that improved models are needed for emerging communication contexts. Specifically, it was found that the predictive fidelity was superior for the 2-box model, the constant-envelope BPSK modulation signals, and for narrower bandwidths. The results also indicated that these models can become inadequate in their NMSE fidelity for wideband, non-constant envelope signals.

In order to overcome the limitations of the n-box models and their variants, a formal nonlinear system identification method was introduced from the mechanical systems community for applications in a communications context. This so-called polyspectral method arrives at a true local identification of the nonlinear operator that represents the power amplifier (or any channel segment). The Volterra series basis for the method was described, followed by an enumeration of the various unique beneficial features of the method, including its systematic series-like approach to approximating the nonlinear operator (with added terms assured of improving fidelity), its closed-form optimization of a given model architecture, its associated quantitative error analysis useful for design of experiments, its natural applicability to the important task of designing effective distortion mitigation, and its use in the development of an analytical link model that has several important advantages of its own. The method requires a large sample of accurately measured time-domain input/output records, which is made possible by the ATDMS. The important optimal filter concept was introduced that serves as the first step in the polyspectral modeling approach, with implications for both the improvement of the baseline 2- and 3-box models, as well as the design of linear equalizers for nonlinear channels. To illustrate the former benefit, an optimally augmented 2-box model was developed for a 20-GHz TWTA and shown to provide significant NMSE improvement over its constituent 2-box model for wideband modulated signals.

The basic two families of the polyspectral model were next described, providing their architectures and basic features. A modified two-branch, nonlinearity-filter variant of the model was then presented that provides state-of-the-art NMSE predictive fidelities for power amplifiers (SSA or TWTA). In fact, it was shown that such a model gave a full 5.6 dB NMSE improvement over the standard 2-box model for a 20-GHz helix TWTA, and 16-APK signals of 9.6-GHz bandwidth, and that this improvement was repeated for many evaluation data sets, representing a wide range of operating

conditions that differed from the construction data set.

Both the time-domain measurement and modeling techniques reviewed in this paper have become relatively mature, and are beginning to experience more widespread and validating application. For example, the baseband time-domain measurement technique, and its various accuracy improvements, has been completely implemented as a hardware measurement instrument (namely, the ATDMS), with several units built, tested, and applied under complete automation employing LabVIEW® control. Just recently, a higher-frequency version of the ATDMS has been successfully implemented and applied. The technique has been thoroughly validated and some very excellent repeatability results have been demonstrated (approaching −55 dB NMSE). However, there is still an important issue that remains to be addressed: the reciprocity assumption made for the FTD characterization that lies at the heart of the ATDMS. More investigation is needed here to quantify and reduce this effect that ultimately limits the accuracy of the ATDMS. Likewise with respect to the modeling, the nonlinear system characterization challenge for power amplifiers, especially TWTAs, has essentially been overcome with the application of the polyspectral method. However, more validating application to SSAs is still needed before total victory can be declared. In addition, there are still several evolutionary embellishments to be studied, such as the addition of a third branch to the current two-branch model, as well as the evaluation of other candidates for its pre-filter. Recent work has now also begun on the use of the polyspectral method for the system-level modeling of limiters and frequency multipliers, which are another common component of modern communication systems. In fact, several new model architectures have been formulated for these devices, which are currently undergoing evaluation using hardware measurements. Finally, the design and evaluation of wideband nonlinear distortion compensation will be the next major challenge, for which the ATDMS and the accompanying polyspectral analysis tool presented here will again prove to be critical in overcoming.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of this work by The Aerospace Corporation through one of its mainstay program customers, as well as its Mission-Oriented Investigation and Experimentation program. The former program provided the context for the measurement and modeling challenges we addressed and eventually overcame. The stable and sustaining long-term support we received made possible all of our accomplishments, and we would like to personally acknowledge our Aerospace program managers, Arthur H. Shapiro, followed by Dr. Wayne R. Fenner, for their commitment and guidance through the years this work was undertaken. In addition, we would also like to extend our gratitude to our local technical staff management in the Electronic Systems Division for their continued support, encouragement, and interest in the development of this work, especially Drs. Jerry D. Michaelson, Allyson D. Yarbrough, Keith M. Soo Hoo, Joe Straus, Wanda Austin, and Sumner Matsunaga. A special thanks goes to Herbert J. Wintroub, Distinguished Engineer, who was always an inspiration and advocate of this work, as well as a very close friend and father figure to all of us. We dedicate this paper to his memory. It is only with this visionary and continued support that such challenging undertakings led to fruitful and complete outcomes.

REFERENCES

[1] P. B. Kenington, *High Linearity RF Amplifier Design*. Norwood, Mass.: Artech House, 2000.

[2] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. New York: Kluwer Academic/Plenum, 1999.

[3] J. E. Allnut, *Satellite-to-Ground Radiowave Propagation*. London: Peter Peregrinus, Ltd., 1989.

[4] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology, and Techniques*, 2nd ed. New York: Kluwer Academic/Plenum, 2000.

[5] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Analysis*. New York: Wiley-Interscience, 2000.

[6] A. A. Moulthrop, M. S. Muha, C. P. Silva, and C. J. Clark, "A new time-domain measurement technique for microwave devices," in *IEEE MTT-S Intl. Microwave Symposium Digest*, vol. 2, pp. 945–948, (Baltimore, Maryland), Jun. 1998.

[7] M. S. Muha, C. J. Clark, A. A. Moulthrop, and C. P. Silva, "Accurate measurement of wideband modulated signals," *Microwave Journal*, vol. 43, no. 6, pp. 84–98, Jun. 2000.

[8] G. H. Bryant, *Principles of Microwave Measurements*. London: Peter Peregrinus, Ltd., 1993.

[9] T. Van den Broeck and J. Verspecht, "Calibrated vectorial "nonlinear network" analyzers," in *IEEE MTT-S Intl. Microwave Symposium Digest*, pp. 1069–1072, (San Diego, California), May 1994.

[10] "Large-Signal Network Analyzer—Bringing Reality to Waveform Engineering," *Technical Data 4T-090 rev A*, Maury Microwave Inc., Ontario, California, Jun. 2003.

[11] J. Verspecht and K. Rush, "Individual characterization of broadband sampling oscilloscopes with a nose-to-nose calibration procedure," *IEEE Trans. Instrument. Measurement*, vol. IM-43, no. 2, pp. 347–354, Apr. 1994.

[12] Thomas S. Laverghetta, *Modern Microwave Measurements and Techniques*. Norwood, Mass.: Artech House, 1988.

[13] "Amplitude and phase measurements of frequency translating devices using the HP 8510B Network Analyzer," *Product Note 8510-7*, Hewlett-Packard Co., Sep. 1987.

[14] M. E. Knox, "A novel technique for characterizing the absolute group delay and delay linearity of frequency translating devices," in *53rd Automatic RF Techniques Group Conference Digest*, pp. 50–56, (Anaheim, California), Jun. 1999.

[15] J. Dunsmore, "Novel method for vector mixer characterization and mixer test system vector error correction," in *2002 IEEE MTT-S Intl. Microwave Symposium Digest*, vol. 3, pp. 1833–1836, (Seattle, Washington), Jun. 2002.

[16] "HP 71500A Microwave Transition Analyzer Group Delay Personality," *Product Note 70820-10*, Hewlett-Packard Co., Apr. 1994.

[17] C. J. Clark, A. A. Moulthrop, M. S. Muha, and C. P. Silva, "Transmission response measurements of frequency translating devices," in *IEEE MTT-S Intl. Microwave Symposium Digest*, pp. 1285–1288, (San Francisco, California), Jun. 1996.

[18] C. J. Clark, A. A. Moulthrop, M. S. Muha, and C. P. Silva, "Network analyzer measurements of frequency translating devices," *Microwave Journal*, vol. 39, no. 11, pp. 114–126, Nov. 1996.

[19] C. J. Clark, A. A. Moulthrop, M. S. Muha, and C. P. Silva, "Transmission response measurements of frequency translating devices using a vector network analyzer," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-44, no. 12, pp. 2724–2737, Dec. 1996.

[20] A. A. Moulthrop, M. S. Muha, and C. P. Silva, "On the frequency response symmetry of diode mixers," *IEEE Trans. Instrument. Measurement* (submitted).

[21] M. Schwartz, W. R. Bennet, and S. Stein, *Communication Systems and Techniques*, pp. 35–43. New York: IEEE Press, 1996.

[22] *LabVIEW® Virtual Instrumentation Software*, National Instruments Corp., Austin, Texas.

[23] C. J. Clark, G. Chrisikos, M. S. Muha, A. A. Moulthrop, and C. P. Silva, "Time-domain envelope measurement technique with application to wideband power amplifier modeling," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-46, no. 12, pp. 2531–2540, Dec. 1998.

[24] A. A. Moulthrop and M. S. Muha, "Enhancements to the measurement accuracy of wideband modulated signals," *Intl. J. RF & Microwave Computer-Aided Engineering*, vol. 13, no. 1, pp. 32–39, Jan. 2003.

[25] A. A. Moulthrop, M. S. Muha, and C. P. Silva, "Minimizing distortions in the time-domain measurement of microwave communications signals," in *61st Automatic RF Techniques Group Conference Digest*, pp. 133–141, (Philadelphia, Pennsylvania), Jun. 2003.

[26] M. S. Muha, A. A. Moulthrop, and C. P. Silva, "Nonlinear noise measurement on a high-power amplifier," in *55th Automatic RF Techniques Group Conference Digest*, pp. 102–105, (Boston, Massachusetts), Jun. 2000.

[27] M. S. Muha, C. J. Clark, A. A. Moulthrop, and C. P. Silva, "Validation of power amplifier nonlinear block models," in *IEEE MTT-S Intl. Microwave Symposium Digest*, vol. 2, pp. 759–762, (Anaheim, California), Jun. 1999.

[28] *Advanced Design Systems*, Agilent EESof EDA, Agilent Technologies, Westlake Village, Calif.

[29] *SPW™, Signal Processing Worksystem*, Cadence Design Systems, Inc., San Jose, Calif.

[30] A. A. Moulthrop, C. J. Clark, C. P. Silva, and M. S. Muha, "A dynamic AM/AM and AM/PM measurement technique," in *IEEE MTT-S Intl. Microwave Symposium Digest*, vol. 3, pp. 1455–1458, (Denver, Colorado), Jun. 1997.

[31] C. J. Clark, C. P. Silva, A. A. Moulthrop, and M. S. Muha, "Power-amplifier characterization using a two-tone measurement technique," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 6, pp. 1590–1602, Jun. 2002.

[32] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. on Commun.*, vol. COM-29, no. 11, pp. 1715–1720, Nov. 1981.

[33] M. T. Abuelma'atti, "Frequency-dependent nonlinear quadrature model for TWT amplifiers," *IEEE Trans. on Commun.*, vol. COM-32, no. 8, pp. 982–986, Aug. 1984.

[34] G. Chrisikos, C. J. Clark, A. A. Moulthrop, M. S. Muha, and C. P. Silva, "A nonlinear ARMA model for simulating power amplifiers," in *IEEE MTT-S Intl. Microwave Symposium Digest*, vol. 2, pp. 733–736, (Baltimore, Maryland), Jun. 1998.

[35] M. Ibnkahla, N. J. Bershad, J. Sombrin, and F. Castanié, "Neural network modeling and identification of nonlinear channels with memory: Algorithms, applications, and analytic models," *IEEE Trans. on Signal Processing*, vol. 46, no. 5, pp. 1208–1220, May 1998.

[36] V. Rizzoli, A. Neri, D. Masotti, and A. Lipparini, "A new family of neural network-based bidirectional and dispersive behavioral models for nonlinear RF/microwave subsystems," *Intl. J. RF and Microwave CAE*, vol. 12, pp. 51–70, 2002.

[37] J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*. New York: John Wiley & Sons, 1993.

[38] J. S. Bendat, *Nonlinear Systems Techniques and Applications*. New York: John Wiley & Sons, 1998.

[39] C. P. Silva, C. J. Clark, A. A. Moulthrop, and M. S. Muha, "Optimal-filter approach for nonlinear power amplifier modeling and equalization," in *IEEE MTT-S Intl. Microwave Symposium Digest*, vol. 1, pp. 437–440, (Boston, Massachusetts), Jun. 2000.

[40] C. P. Silva, A. A. Moulthrop, M. S. Muha, and C. J. Clark, "Application of polyspectral techniques to nonlinear modeling and compensation," in *IEEE MTT-S Intl. Microwave Symposium Digest*, vol. 1, pp. 13–16, (Phoenix, Arizona), May 2001.

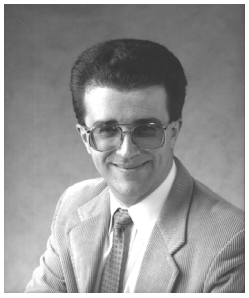
[41] C. P. Silva, A. A. Moulthrop, and M. S. Muha, "Introduction to polyspectral modeling and compensation techniques for wideband communications systems," in *58th Automatic RF Techniques Group Conference Digest*, pp. 1–15, (San Diego, California), Nov. 2001.

[42] S. A. Billings, "Identification of nonlinear systems—A survey," *IEE Proceedings*, vol. 127, Pt. D, pp. 272–285, Nov. 1980.

[43] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems Part I: Deterministic non-linear systems," *Intl. J. Control*, vol. 41, no. 2, pp. 303–328, 1985.

[44] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Malabar, Florida: Robert E. Krieger Publishing Company, 1989.

BIOGRAPHIES



Christopher P. Silva *(S'81–M'89–SM'98–F'98) was born on March 17, 1960, in Fortuna, California. He received the B.S., M.S., and Ph.D. degrees, all in electrical engineering, in 1982, 1985, and 1993, respectively, from the University of California at Berkeley. His graduate work was supported mainly by a National Science Foundation Fellowship and a Lockheed Leadership Fellowship.*

He joined the Electronics Research Laboratory of The Aerospace Corporation, Los Angeles, California, in 1989, and is currently a Senior Engineering Specialist in the Communication Electronics Department, Electronic Systems Division. He has been the principal investigator on several internally funded research projects addressing nonlinear microwave CAD, secure communications by means of chaos, and compensation of nonlinear satellite communications channels. In addition, he is currently heading a study on the nonlinear modeling and compensation of high-data-rate microwave communications systems in support of national advanced satellite programs. He is also involved with the application of wavelets analysis to CAD and communications signaling. Previously, he worked as a Post-Graduate Researcher at the Electronics Research Laboratory, University of California, Berkeley, where he investigated the chaotic dynamics of nonlinear circuits and systems.

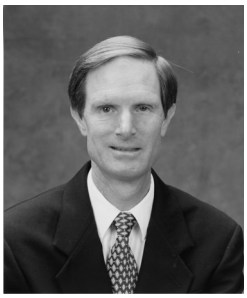
Dr. Silva is also a member of Eta Kappa Nu, Tau Beta Pi, Phi Beta Kappa, the American Association for the Advancement of Science, the American Institute of Aeronautics and Astronautics (Senior Member), the Society for Industrial and Applied Mathematics, the American Mathematical Society, and the Mathematical Association of America.

Christopher J. Clark *was born in Washington, DC, in 1960. He received the B.S.E.E. and M.S.E.E. degrees from the University of Maryland, College Park, in 1983 and 1986, respectively.*

From 1984 to 1986 he worked for the Watkins-Johnson Company, where he was responsible for the design and development of RF receiving systems. From 1986 to 1992 he worked for TRW, Inc., where he developed microwave components for satellite applications. His primary focus was in the design of MESFET, HEMT, and HBT GaAs MMICs for phased-array systems. From 1992 to 1999 he worked at The Aerospace Corporation, where he was involved with the design and development of space communication systems. From 1999 to 2003 he was a Principal Engineer at Multilink Technology Corp., Santa Monica, where he was involved with the design and development of high-speed electronics for use in fiber optic communication systems.

He is currently a Senior Engineering Specialist at The Aerospace Corporation in the Communication Electronics Department, Electronic Systems Division. His work again involves design, analysis, and hardware prototype support for several space communication systems.

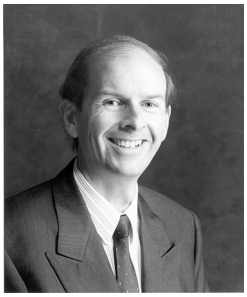
Mr. Clark is a member of Eta Kappa Nu and Tau Beta Pi.



Andrew A. Moulthrop *was born in San Francisco, California, in 1955. He received the B.A. degree in physics in 1977 from the University of California, San Diego, and the Ph.D. Degree in physics in 1984 from the University of California, Berkeley.*

He is currently a Senior Engineering Specialist in the Communication Electronics Department, Electronic Systems Division, at The Aerospace Corporation, Los Angeles. His work involves the design, development, and measurement of microwave components and systems. Previously, he worked as a Research Assistant at the University of California, Berkeley, where he co-developed a Josephson junction parametric amplifier, as well as measured various properties of superfluid helium.

Dr. Moulthrop is a member of Phi Beta Kappa.



Michael S. Muha *was born in Inglewood, California, on February 21, 1958. He received the B.S. Degree in physics in 1979 from the University of Southern California, Los Angeles, and the M.S. Degree in applied physics in 1981 from the California Institute of Technology, Pasadena, where he worked on the design and development of millimeter-wave and far-infrared imaging antenna arrays.*

He is currently a Senior Engineering Specialist in the Communication Electronics Department, Electronic Systems Division, at The Aerospace Corporation, Los Angeles. His work involves the design, development, and characterization of high-data-rate microwave communications systems.

Mr. Muha is a member of Phi Beta Kappa and Phi Kappa Phi.