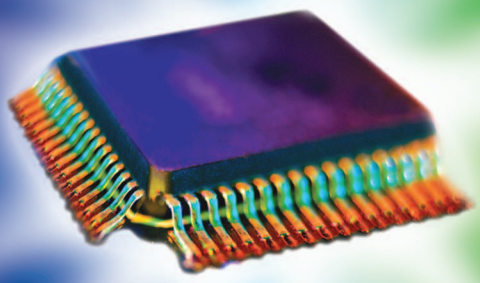
SIMULATE

MEASURE

MODEL



Jan Verspecht
and David E. Root

Polyharmonic Distortion Modeling

For more than a quarter of a century, microwave engineers have had the benefit of a foundation of mutually interacting components of measurement, modeling, and simulation to design and test linear components and systems. This three-legged stool included measurements of S-parameters using a calibrated vector network analyzer (VNA), linear simulation analysis tools (e.g., Touchtone), and models based on S-parameter blocks, which can use measured data or simulated frequency-dependent data.

S-parameters are perhaps the most successful behavioral models ever. They have the powerful property that the S-parameters of individual components are sufficient to determine the S-parameters of any combination of those components. S-parameters of a component are sufficient to predict its response to any signal, provided only that the signal is of sufficiently small amplitude. This follows from the property of superposition, which governs the behavior of linear systems, the systems for which S-parameters apply.

Jan Verspecht (contact@janverspecht.com) is with Jan Verspecht bvba, B-1840 Steenhuffel, Belgium.
David E. Root is with Agilent Technologies, Inc., Santa Rosa, CA 95403 USA.

 1527-3342/06/$20.00©2006 IEEE

The ability to measure, model, and simulate using S-parameters means, in principle, every problem imaginable for linear system design and test is solvable.

Simply put, S-parameters are just complex numbers. At most, they are matrices of complex numbers for devices or circuits with multiple ports. S-parameters are important for several reasons. For one thing, they are easy to measure. Standard vector-corrected network analyzers easily provide the data from real components. The data is reliable and repeatable. The data is also useful because it represents an intrinsic property of the device under test (DUT), independent of the measurement system used to provide the data. In particular, the S-parameters of a two-port system are defined by ratios, which conveniently produce results completely independent of the details of the stimulating signal (such as its phase). That is, the S-parameters of a DUT are invariant with respect to the phase of the incident wave.

Despite the great success of S-parameters, they are severely limited. Conventional S-parameters are defined only for linear systems, or systems behaving linearly with respect to a small signal applied around a static operating point (e.g., fixed bias condition of a transistor). In fact, virtually all real systems are nonlinear. They generate harmonics and intermodulation distortion and cause spectral regrowth. S-parameter theory doesn't apply to such systems. It may be a good approximation over some range of input, but it is incapable of even estimating the nonlinear response of real systems.

Perhaps the most important driven nonlinear system of interest is the power amplifier. Its entire raison d'etre is to amplify a signal. Amplification requires an active, nonlinear device and a time-varying signal. Thus an amplifier, when it is actually amplifying a signal, is a driven nonlinear system, which falls outside the class of systems for which (linear) S-parameters apply. Most previous attempts to treat such systems as linear systems parameterized by drive levels are, in fact, flawed. They ignore new phenomena and terms that appear only when nonlinear systems are driven but for which there is no analog in linear systems. It is not surprising that those ad hoc attempts to generalize S-parameters result in inferences (e.g., test results, simulations, and designs) that are unreliable, nonrepeatable, or flat out don't meet specs.

This article explains and then goes beyond the limitations of simple-minded (and incorrect) generalizations of S-parameters to driven nonlinear systems. We will show how a simple yet rigorous framework, with corresponding fully interoperational nonlinear model, measurement hardware, and nonlinear simulation environment, can circumvent these problems at a very modest additional cost. If we are willing to consider only the addition of a second complex number, (or a second matrix of complex numbers for devices with multiple ports), it is possible to do for driven nonlinear systems what S-parameters do for linear systems. Moreover, it has already been demonstrated that there are now interoperable tools of measurement systems, nonlinear models, and large-signal simulation environments ready to provide the infrastructure for nonlinear design and test.

Imagine one could describe driven nonlinear systems in a way similar to S-parameters for linear systems. That is, imagine there was a way to measure, model, and simulate nonlinear driven systems that would allow correct, reliable, and repeatable inferences of what an arbitrary arrangement of such systems would do under drive. This capability is actually necessary for multistage amplifiers, where the input stage of the second amplifier, for example, is not perfectly matched at the fundamental or generated harmonics and injects signals back into the output of the prior stage. We will demonstrate a new nonlinear model, called the polyharmonic distortion (PHD) model, which is perfectly mated to existing nonlinear simulation capabilities that can be identified with advanced nonlinear measurements as seamlessly as S-parameters for the linear case.

PHD modeling is a black-box frequency domain modeling technique.

On one hand, this might seem surprising, since nonlinear problems are hard. Nonlinear systems respond to signals of different shapes and sizes in an infinite number of ways. Common questions include: "What kind of signals should I use to stimulate the DUT?," "What kinds of useful inferences can I make with this data?," "How can I analyze this data to make predictions of DUT behavior?," and "Can I measure what I need for the job with existing commercially available equipment?"

PHD Modeling

PHD modeling is a black-box, frequency-domain modeling technique. The annotation *black box* refers to the fact that no knowledge is used nor required concerning the internal circuitry of the DUT. All information needed to construct a PHD model is acquired through externally stimulating the signal ports of a DUT and measuring the response signals. The frequency domain formulation means that the approach is well suited for distributed (dispersive) high-frequency applications. This is true for both the measurement techniques and the modeling approach. Note that these considerations are true for conventional linear S-parameters, which can also be considered as a black-box frequency-domain modeling

technique. The advantage of using a black-box approach is that it is truly technology independent. It does not matter whether one is dealing with silicon bipolar technology or compound semiconductor field-effect transistors. Another advantage is that a black-box model, unlike a circuit schematic, can be shared with and used by other people without revealing the details of the internal circuit. In other words, it provides complete and fundamental protection of intellectual property. This characteristic is highly appreciated in a business environment. Of course, with black-box modeling, as with all engineering solutions, there are tradeoffs to consider in practical use conditions. Black-box models are, by definition, only valid for signals that are close to the signals that were used to simulate the DUT to produce the responses used for model identification (extraction). If the model needs to be valid across a wide range of signals, then a wide range of excitation signals is needed and, as a result, the measurement time will be long and the resulting model will be complex.

The PHD model is identified from the responses of a DUT stimulated by a set of harmonically related discrete tones, where the fundamental tone is dominant and the harmonically related tones are relatively small. As such, it is typically applied for modeling microwave amplifiers with narrowband input signals. Note that the narrowband constraint is not on the amplifier itself but on the input signal. It is perfectly possible, for example, to describe the distortion of a narrowband input signal for a wide range of carrier frequencies.

The basic idea is that the PHD modeling approach can be used as a natural extension of S-parameters under large-signal conditions. One connects a DUT to a large-signal network analyzer (LSNA) instrument, and a model is automatically extracted that accurately describes all kinds of nonlinear behavior such as amplitude and phase of harmonics, compression characteristics, AM-PM, spectral regrowth, amplitude dependent input, and output match. The real beauty of the approach is that it provides much more than a bunch of plots of the aforementioned characteristics. One PHD model can be used in a computer-aided design (CAD) environment to consistently describe many different nonlinear characteristics. As with S-parameters, the PHD approach works both ways. It not only provides the necessary information for an accurate automatically extracted CAD model, it also provides a consistent framework for experimentally verifying (testing to specifications) the large-signal behavior of a nonlinear component under drive once it has been produced. It can't be overemphasized that S-parameters are simply inadequate for both the modeling and the characterization of driven nonlinear components; S-parameters are incomplete once nonlinear effects are present in driven systems. A nice characteristic is that a PHD model reduces to classic S-parameters for small input amplitudes. As such, an LSNA instrument equipped with the means to measure a PHD model performs a superset of the measurements possible with a classic VNA. As such, LSNA instruments will gradually replace all VNAs that are in use today to characterize semiconductor devices all the way from R&D to manufacturing—easily a multimillion-dollar business.

Theory

Describing Functions: A Unifying Framework for Frequency-Domain Nonlinear Behavioral Models

We will now introduce the theoretical foundations of the PHD model. Similar to S-parameters, the basic quantities we are working with are traveling voltage waves. The waves are defined as in the case of classic S-parameters: they are linear combinations of the signal port voltage, V, and the signal port current, I, whereby the current quantity is defined as positive when flowing into the DUT. The incident waves are called the A-waves and the scattered waves are called the B-waves. They are defined as follows:

$$A = \frac{V + Z_c I}{2}, \tag{1}$$

$$B = \frac{V - Z_c I}{2}. \tag{2}$$

The default value of the characteristic impedance Z_c is 50 Ω. For certain applications, however, the choice of another value may be more practical. One example is power transistor applications where it may be simpler to use a value that comes close to the output impedance of the transistor, e.g., 10 Ω. Note that the waves are defined based on a pure mathematical transformation of the signal port voltage and current and are not associated with a physical wave transmission structure. Therefore, the wave quantities are more accurately called pseudowaves [1]. Also note that other wave definitions are in use and that the convention that we use, as described by (1) and (2), is compatible with commercial harmonic balance simulators.

In general, we will be working with nonlinear functional relationships between the wave quantities. This is very different from S-parameters that can only describe a linear relationship. The PHD approach assumes the presence of discrete tone signals (multisines) for the incident as well as for the scattered waves. In general, these discrete tones may appear at arbitrary frequencies, as explained in [3]. In this article, however, we will limit ourselves to the simpler and, from a research point of view, more mature case where the signals can be represented by

a fundamental with harmonics. In other words, the signals are periodic or they are narrowband modulated versions of a fundamental with harmonics. In that case, each carrier frequency can easily be denoted by using the harmonic index, which equals zero for the dc contribution, one for the fundamental, and two for the second harmonic. In our notation for indicating the wave variables, a first subscript refers to the signal port and a second subscript refers to the harmonic index. The problem we are solving can then be formulated as follows: For a given DUT, determine the set of multivariate complex functions $F_{pm}(.)$ that correlate all of the relevant input spectral components A_{qn} with the output spectral components B_{pm}, whereby q and p range from one to the number of signal ports, and whereby m and n range from zero to the highest harmonic index. This is mathematically expressed as

$$B_{pm} = F_{pm}(A_{11}, A_{12} \ldots, A_{21}, A_{22}, \ldots). \tag{3}$$

Note that we assume for now that the fundamental frequency is a known constant. The functions $F_{pm}(.)$ are called the describing functions [2]. The concept is illustrated in Figure 1.

The spectral mapping (3) is a very general mathematical framework from which practical models can be developed in the frequency domain. The PHD model is a particular approximation of (3), which involves the linearization of (3) around the signal class discussed previously. Less-restrictive approximations are possible and are needed to describe additional nonlinear interactions such as intermodulation distortion of mixers, which is beyond the scope of this article. The point is that starting from (3), a systematic set of approximations, experiment designs, and model identification schemes can be combined to produce powerful and useful behavioral models of driven nonlinear components. The LSNA instruments are already capable of characterizing components under excitations more complicated than those needed to identify the PHD model described here. This work will be rapidly developing in the next several years.

In 1995, a breakthrough occurred when we started to exploit certain mathematical properties of these functions $F_{pm}(.)$ [6].

A first property is related to the fact that $F_{pm}(.)$ describes a time-invariant system. This implies that applying an arbitrary delay to the input signals, in our case the incident A-waves, always results in exactly the same time delay for the output signals, the scattered B-waves. In the frequency domain, applying a time delay is equivalent to the application of a linear phase shift (proportional to frequency), and as such this fact can mathematically be expressed as

$$\forall\theta: \; B_{pm}e^{jm\theta} = F_{pm}(A_{11}e^{j\theta}, A_{12}e^{j2\theta}, \ldots, \times A_{21}e^{j\theta}, A_{22}e^{j2\theta}, \ldots). \tag{4}$$

A second property, which is of a totally different nature, is related to the nonanalyticity of the functions $F_{pm}(.)$.

Phase Normalization and Linearization

In the following, both of the aforementioned properties are exploited to derive the PHD model equations. Since (4) is valid for all values of θ, we can make θ equal to the inverted phase of A_{11}, the incident fundamental. Note that other choices of θ are also possible [3]. Our choice is most natural for power transistor and power amplifier applications, where A_{11} is the dominant large-signal input component.

For notational elegance, we introduce the phasor P, defined as

$$P = e^{+j\varphi(A_{11})}. \tag{5}$$

Substituting $e^{j\theta}$ by P^{-1} in (4) results in

$$B_{pm} = F_{pm}(|A_{11}|, A_{12}P^{-2}, A_{13}P^{-3}, \ldots, \times A_{21}P^{-1}, A_{22}P^{-2}, \ldots)P^{+m}. \tag{6}$$

The advantage of (6), when compared to (3), is that the first input argument will always be a positive real number, namely the amplitude of the fundamental component at the input port 1, rather than a complex number. This greatly simplifies further processing.

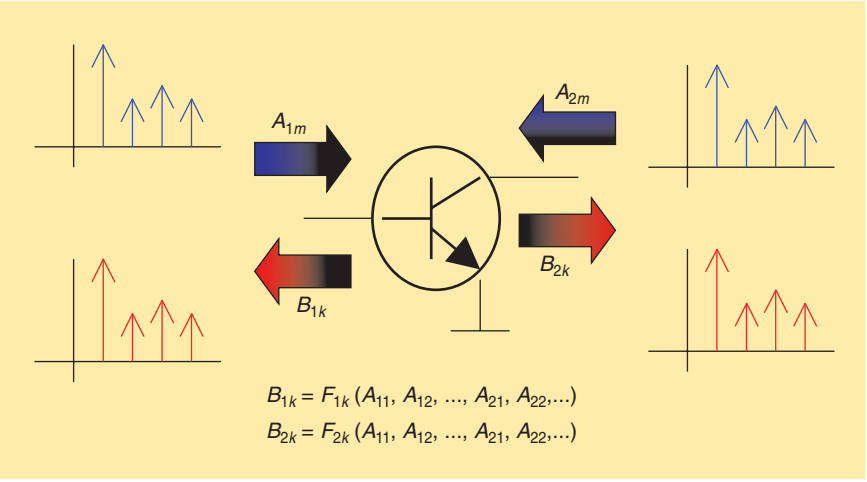


Figure 1. *The concept of describing functions.*

In general, we are working under large-signal, nonlinear operating conditions, and the superposition principle is not valid. In many practical cases, however, such as in power amplifiers stimulated with a narrowband input signal, there is only one dominant large-signal input component present (A_{11}) whereas all other input components (the harmonic frequency components) are relatively small. In that case, we will be able to use the superposition principle for the relatively small input components. This is called the harmonic superposition principle [8].

The harmonic superposition principle is graphically illustrated in Figure 2. To keep the graph simple, we only consider the presence of the A_{1m} and B_{2n} components and we neglect the presence of the A_{2m} and B_{1n} components. First, let us consider the case where only A_{11} is different from zero. The input spectral components A_{1m} and the output spectral components B_{2n} corresponding to this case are indicated by black arrows. Note the presence of significant harmonic components for the B_{2n} components. Now leave the A_{11} excitation the same and add a relatively small A_{12} component (second harmonic at the input). This will result in a deviation of the output spectrum B_2, indicated by the red arrows. The same holds of course for a third (green) and a fourth (blue) harmonic. The harmonic superposition principle holds when the overall deviation of the output spectrum B_2 is the superposition of all individual deviations. This conjecture was experimentally verified, as described in [8], and appeared to be true for all practical power amplifier design cases, whatever the class of the amplifier. The harmonic superposition principle is the key to the PHD model. Linearization of (6) versus all components besides the large signal A_{11} leads to

$$\begin{aligned} B_{pm} =& K_{pm}(|A_{11}|)P^{+m} \\ &+ \sum_{qn} G_{pq,mn}(|A_{11}|)P^{+m}\mathrm{Re}(A_{qn}P^{-n}) \\ &+ \sum_{qn} H_{pq,mn}(|A_{11}|)P^{+m}\,\mathrm{Im}(A_{qn}P^{-n}), \end{aligned} \tag{7}$$

whereby

$$K_{pm}(|A_{11}|) = F_{pm}(|A_{11}|, 0, \ldots 0)\ , \tag{8}$$

$$G_{pq,mn}(|A_{11}|) = \left.\frac{\partial F_{pm}}{\partial \mathrm{Re}(A_{qn}P^{-n})}\right|_{|A_{11}|,0,\ldots 0}, \tag{9}$$

$$H_{pq,mn}(|A_{11}|) = \left.\frac{\partial F_{pm}}{\partial \mathrm{Im}(A_{qn}P^{-n})}\right|_{|A_{11}|,0,\ldots 0}. \tag{10}$$

Note that the real and imaginary parts of the input arguments are treated as separate and independent entities. In mathematical terms, it is said that the spectral mapping function $F_{pm}(.)$ is nonanalytic. The appearance of a nonanalytic function may seem strange since it is so often the case in engineering and physics that we deal with analytic functions (e.g., exponential functions of complex arguments and causal response functions of complex frequencies). In fact, classic S-parameters, when considered as a behavioral model for a linear system, result in a spectral mapping function that is analytic. This fact is clearly demonstrated in Figure 3. In this figure, we depict the measured amplitude of $H_{22,11}(.)$ and $G_{22,11}(.)$ as a function of $|A_{11}|$ for an actual RFIC power amplifier. When $|A_{11}|$ is low, the S-parameter model is valid and the amplitudes of $H_{22,11}(.)$ and $G_{22,11}(.)$ are identical, indicating that the spectral mapping is analytic. At higher input amplitudes, $H_{22,11}(.)$ and $G_{22,11}(.)$ start to move apart, proving that the spectral mapping becomes nonanalytic. A mathematical proof of the existence of such nonanalytic behavior, which is based on a simple exercise, is given in the "On the Origin of the Conjugate Terms" sidebar.

The PHD model equation is derived by substituting the real and imaginary parts of the input arguments in (7) by a linear combination of the input arguments and their corresponding conjugates. Since

$$\mathrm{Re}(A_{qn}P^{-n}) = \frac{A_{qn}P^{-n} + \mathrm{conj}(A_{qn}P^{-n})}{2}, \tag{11}$$

$$\mathrm{Im}(A_{qn}P^{-n}) = \frac{A_{qn}P^{-n} - \mathrm{conj}(A_{qn}P^{-n})}{2j}, \tag{12}$$

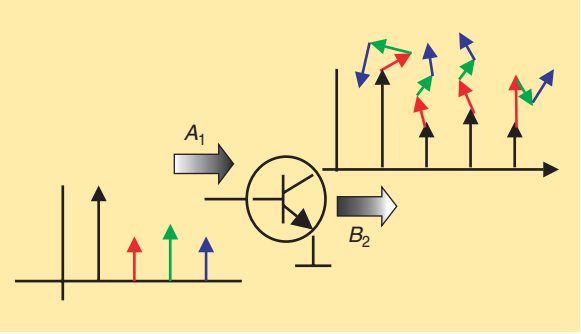


Figure 2. *The harmonic superposition principle.*

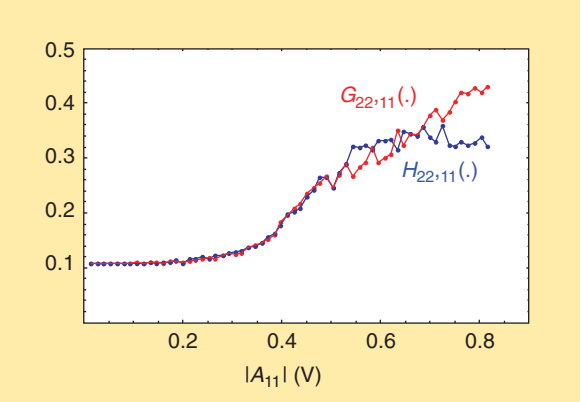


Figure 3. *Amplitudes of $G_{22,11}(.)$ and $H_{22,11}(.)$.*

one can write

$$B_{pm} = K_{pm}(|A_{11}|)P^{+m} + \sum_{qn} G_{pq,mn}(|A_{11}|)P^{+m} \times \left(\frac{A_{qn}P^{-n} + \text{conj}(A_{qn}P^{-n})}{2}\right) + \sum_{qn} H_{pq,mn}(|A_{11}|) \times P^{+m}\left(\frac{A_{qn}P^{-n} + \text{conj}(A_{qn}P^{-n})}{2j}\right). \quad (13)$$

Rearranging the terms finally leads to the relatively simple PHD model equation

$$B_{pm} = \sum_{qn} S_{pq,mn}(|A_{11}|)P^{+m-n}A_{qn} + \sum_{qn} T_{pq,mn}(|A_{11}|)P^{+m+n}\,\text{conj}(A_{qn}). \quad (14)$$

Note that two new functions, $S_{pq,mn}(.)$ and $T_{pq,mn}(.)$, are introduced. They are defined as

$$S_{p1,m1}(|A_{11}|) = \frac{K_{pm}(|A_{11}|)}{|A_{11}|}, \quad (15)$$

$$T_{p1,m1}(|A_{11}|) = 0 \quad (16)$$

$$\forall\{q,n\} \neq \{1,1\}: \quad S_{pq,mn}(|A_{11}|) = \frac{G_{pq,mn}(A_{11}|) - jH_{pq,mn}(|A_{11}|)}{2}, \quad (17)$$

$$\forall\{q,n\} \neq \{1,1\}: \quad T_{pq,mn}(|A_{11}|) = \frac{G_{pq,mn}(A_{11}|) + jH_{pq,mn}(|A_{11}|)}{2}. \quad (18)$$

All of the functions $T_{p1,m1}(.)$ are defined in (16) as being equal to zero. This can be explained by the fact that the terms in (14) with n and q equal to one are degenerate since

$$P^{+m-1}A_{11} = P^{+m+1}\text{conj}(A_{11}) = |A_{11}|. \quad (19)$$

As a result, it is only the sum $S_{p1,m1}(|A_{11}|) + T_{p1,m1}(|A_{11}|)$ that matters in (14) and not the individual functions. To define a unique value for these functions, the value of $T_{p1,m1}(|A_{11}|)$ is defined as zero by convention.

Intuitive Interpretations

The basic PHD model (14) simply describes that the B-waves result from a linear mapping of the A-waves, similar to classic S-parameters. Some significant differences with S-parameters are explained in the following.

First of all, the right-hand side of (14) contains a contribution associated with the A-waves as well as the conjugate of the A-waves. The conjugate part is not present at all with S-parameters. That is the case since, with S-parameters, the contribution of an A-wave to a particular B-wave is not a function of the phase of that A-wave. Any phase shift in A will just result in the same phase shift of the contribution to the particular B-wave. This is no longer the case, however, when a large A_{11} wave is present at the input of the DUT. In that case, the large signal A_{11} wave creates a phase reference point for all of the other incident A-waves, and the contribution to the B-waves of a particular A-wave depends on the phase relationship between this particular A-wave and the large signal A_{11} wave. This relative phase dependency is expressed in (14) through the presence of the conjugate A-wave terms. This is clarified with the following example. Consider (14) restricted to the simple case of a B_{21} (fundamental at the output) depending on A_{21} (reflected fundamental at the output) and A_{11} (fundamental incident at the input). In that case, (14) is reduced to

The PHD modeling approach can be used as a natural extension of S-parameters under large-signal conditions.

$$B_{21} = S_{21,11}(|A_{11}|)A_{11} + S_{22,11}(|A_{11}|)A_{21} + T_{22,11}(|A_{11}|)P^2\text{conj}(A_{21}). \quad (20)$$

The contribution of A_{21} to B_{21} will be noted as $\Delta_{21}B_{21}$ and is given by the two rightmost terms

$$\Delta_{21}B_{21} = S_{21,11}(|A_{11}|)A_{21} + T_{22,11}(|A_{11}|)P^2\text{conj}(A_{21}). \quad (21)$$

Dividing the left- and right-hand sides of (21) by A_{21} results in the large signal equivalent of the classic S-parameter S_{22}

$$\frac{\Delta_{21}B_{21}}{A_{21}} = S_{22,11}(|A_{11}|) + T_{22,11}(|A_{11}|)P^2\frac{\text{conj}(A_{21})}{A_{21}}. \quad (22)$$

Using (5), this can be written as

$$\frac{\Delta_{21}B_{21}}{A_{21}} = S_{22,11}(|A_{11}|) + T_{22,11}(|A_{11}|)e^{-j2(\varphi(A_{21})-\varphi(A_{11}))}. \quad (23)$$

The large-signal S_{22}, as calculated in (23), has two terms. The first term is a function of the amplitude of A_{11} only and behaves exactly like a classic S_{22} (except for the fact, of course, that it depends on the input signal amplitude). The second term is more peculiar. It depends not

On the Origin of the Conjugate Terms

There are several ways to understand the nonanalyticity of the spectral mappings $F_{pm}(.)$. Perhaps the simplest is just to take the example of a static algebraic nonlinearity (e.g., polynomial) in the time domain and compute the mapping in the spectral domain.

We start by considering a system described by a simple instantaneous nonlinearity containing both a linear and cubic term. We look at the following three cases, for which the analysis can be computed exactly. The first case is the linear response of this nonlinear system around a static operating point. This is the familiar condition for which linear S-parameters apply. The second case is the linearization of the system around a time-varying large-signal operating state, with the time variation and perturbation having the same fundamental period. The third case is a simple generalization of the second where the linear perturbation is at a distinct frequency compared to the fundamental frequency of the periodically driven nonlinear system. The objective is to look at the linearized response of the system in the frequency domain and demonstrate that the relationship between the perturbation phasor and its linear response phasor is not an analytic function in cases 2 and 3, namely when the system is driven. That is, these examples illustrate the simultaneous presence of both a and a* terms in the response of driven nonlinear systems to additional injected signals.

The nonlinearity is described by

$$f(x) = \alpha x + \gamma x^3. \tag{1}$$

The signal is written as the sum of a main signal and an additional perturbation term, assumed to be small.

$$x(t) = x_0(t) + \Delta x(t). \tag{2}$$

The objective is to calculate the linear response of system (1) to signals (2).

Case 1

Consider the signal $x(t)$, given by the sum of a (real) dc component and a small tone at frequency $f = \omega/2\pi$, i.e.,

$$x_0(t) = A$$
$$\Delta x(t) = \frac{\delta e^{j\omega\tau} + \delta^* e^{-j\omega\tau}}{2}.$$

Here, A is real and δ is a small complex number, which allows for the phase of the perturbation tone to take any desired value. The signal is manifestly real.

The linear response in $\Delta x(t)$ can be computed by

$$\Delta(y(t)) = f(x_0(t) + \Delta x(t)) - f(x_0(t)) \approx f'(x_0(t))\Delta x(t) \tag{3}$$

with the approximations becoming exact as $\Delta x(t) \to 0$. For case 1, we evaluate the conductance nonlinearity $f'(x_0)$, from (1) at the fixed value $x_0 = A$ to get

$$f'(A) = \alpha + 3\gamma A^2. \tag{4}$$

Substituting (4) into (3), we obtain

$$\Delta(y(t)) = [\alpha + 3\gamma A^2]\left(\frac{\delta e^{j\omega t} + \delta^* e^{-j\omega t}}{2}\right). \tag{5}$$

If we look at the complex coefficient of the term proportional to $e^{j\omega t}$, we obtain

$$\frac{[\alpha + 3\gamma A^2]}{2}\delta. \tag{6}$$

Since (5) is a linear input-output relationship with constant coefficient, the complex Fourier component at the output frequency is linearly related to the complex Fourier component at the (same) input frequency. That is, $\hat{Y} = G(A)\hat{X}$, where $\hat{X}$ and $\hat{Y}$ are the complex Fourier coefficients of the input and output small-signal phasors, respectively, and $G(A)$ is the gain expression from (4), which depends nonlinearly on the static operating point but is constant in time.

Case 2

$$x_0(t) = A\cos(\omega t)$$
$$\Delta x(t) = \frac{\delta e^{j\omega t} + \delta^* e^{-j\omega t}}{2}.$$

This time we take $x_0(t)$ to be a (periodically) time-varying signal, $x_0(t) = A\cos(\omega t)$.

There is no loss of generality by taking the phase of the large signal to be zero, since the small tone's phase, considered as the relative phase compared to that of the large tone, accounts for all possible differences for a time-invariant system in the absence of a signal. This is a restatement of the time-translation invariance of the system in the absence of drive.

Evaluating the conductance nonlinearity $f'(x_0(t))$ at $x_0(t) = A\cos(\omega t)$, we obtain for this case

$$f'(A\cos(\omega t)) = \alpha + 3\gamma A^2\cos^2(\omega t) = \left(\alpha + \frac{3\gamma A^2}{2}\right) + \frac{3\gamma A^2}{2}\cos(2\omega t). \tag{7}$$

The second form follows from a simple trigonometric identity

$$\cos^2(\omega t) = \frac{1}{2} + \frac{\cos(2\omega t)}{2}.$$

Using (7) to evaluate (3) for this case we obtain

$$\Delta(y(t)) = \left[\left(\alpha + \frac{3\gamma A^2}{2}\right) + \frac{3\gamma A^2}{2}\left(\frac{e^{2j\omega t} + e^{-2j\omega t}}{2}\right)\right] \times \left(\frac{\delta e^{j\omega t} + \delta^* e^{-j\omega t}}{2}\right). \tag{8}$$

This time we get terms proportional to $e^{j\omega t}$ and $e^{j3\omega t}$ and their complex conjugates; four terms in all. If we restrict our attention, as in case 1, to the complex term proportional to $e^{j\omega t}$, we obtain

$$\left(\frac{\alpha}{2}+\frac{3\gamma A^2}{4}\right)\delta+\frac{3\gamma A^2}{4}\delta^*. \tag{9}$$

We observe that the output phasor at frequency ω is not just proportional to the input phasor δ at frequency ω, but has distinct contributions proportional to both δ and δ*.

That is, the linearization of the nonlinear system, around the simple dynamic operating point determined by the large tone, is not analytic in the sense of complex variable theory. If it were analytic, (9) would depend only on the complex variable δ and not both δ and δ*.

If we take a ratio of the complex output Fourier component to the complex input Fourier component, we obtain

$$\frac{\Delta\hat{Y}(\omega)}{\Delta\hat{X}(\omega)}=\left(\frac{\alpha}{2}+\frac{3\gamma A^2}{4}\right)+\frac{3\gamma A^2}{4}e^{-2j\text{Phase}(\delta)}.$$

Therefore, unlike linear S-parameters, the result is not independent of the phase of the small perturbation tone. That is, the large tone creates a phase reference such that the linear response of the system around the large-signal, time-varying state depends explicitly on the relative phase of the perturbation tone and the large tone.

Case 3

$$x_0(t)=A\cos(\omega t) \tag{10}$$

$$\Delta x(t)=\frac{\delta e^{j\omega_1 t}+\delta^* e^{-j\omega_1 t}}{2}. \tag{11}$$

Here we allow the frequency of the large tone ω and the frequency of the perturbation tone ω_1 to be distinct. The time-varying nonlinear conductance is the same as before, with the only difference being the frequency of the small perturbation term in parentheses in the rightmost factor of (12)

$$\Delta(y(t))=\left[\left(\alpha+\frac{3\gamma A^2}{2}\right)+\frac{3\gamma A^2}{2}\left(\frac{e^{2j\omega t}+e^{-2j\omega t}}{2}\right)\right]\times\left(\frac{\delta e^{j\omega_1 t}+\delta^* e^{-j\omega_1 t}}{2}\right). \tag{12}$$

Since ω_1 and ω are distinct, there are more frequency components than in the previous case. We write $\omega_1=\omega+\Delta$, and look at the terms proportional to $e^{j(\omega+\Delta)t}$ and $e^{j(\omega-\Delta)t}$. We obtain

$$\left(\frac{\alpha}{2}+\frac{3\gamma A^2}{4}\right)\delta \tag{13}$$

and

$$\frac{3\gamma A^2}{4}\delta^*, \tag{14}$$

respectively. These terms represent the single-sided spectrum of the lower and upper sidebands of the intermodulation spectrum of the system (1) for excitation (2), defined by (10) and (11) around the fundamental frequency of the drive.

We note that as the tone spacing Δ goes to zero, both these contributions overlap (add) at the center frequency ω of the time-varying drive, and we have the result of case 2.

The isolation of terms proportional to δ from those proportional to δ* that results from this method remains true for the general dynamic nonlinearity, not just the example used in (1). In the general case, the upper and lower sideband phasors depend on the frequency offset, Δ (unlike the simple example here). Case 2, which represents the PHD model, can be recovered using case 3 for each sideband for finite Δ and then taking the limit $\Delta\to 0$. This indicates that it is possible to extract each upper and lower sideband term (per harmonic frequency component) from measurements of the system response to a small tone of a single, arbitrary phase [4] rather than introduce two (or more) distinct phases to extract the two terms of (9) when they appear together.

Examination of case 3 reveals that the complex conjugate term, in both cases 2 and 3, results from an intermodulation or mixing, a result of nonlinearity, and disappears as the size of the drive signal decreases to zero. This is evident by evaluating (13) and (14) [or (9) for case 2] as $A\to 0$ in (1). The term proportional to δ* vanishes in (14), and the terms proportional to δ in (13) reduce to the result we would get for a linear system with gain α. In the limit $\omega\to 0$, case 3 reduces to case 1, corresponding to the system linearized around a static operating point A. This is most easily seen by taking the limit $\omega\to 0$ in (12). Thus, although the PHD model is representative of case 2 (perturbation signals at exact integer multiples of the fundamental drive signal), the origins of the different terms are more obvious by examining the slightly more general case 3.

For the more general nonlinear system, the degenerate case 2, where upper and low sidebands overlap, the two different contributions that land on the same frequency—necessarily a harmonic of the driven system—come from different modulation indices. The separation of the two terms by frequency offset allows these distinct mechanisms related to the Fourier coefficients of the conductance nonlinearity to be independently identified from an experiment using a single small tone at arbitrary phase, relative to the large signal drive tone.

only on the input signal amplitude through the function $T_{22,11}(.)$ but also on the phase difference between A_{21} and A_{11} through the complex exponential. Note that it does not depend on the amplitude of A_{21}. As such, one can state that the large signal S_{22} is described by a set of two complex functions (with the input amplitude as argument): a first function $S_{22,11}(.)$, which represents the part independent from the phase relationship between A_{21} and A_{11} and a second function $T_{22,11}(.)$, which represents the part that depends on the phase relationship between A_{21} and A_{11}.

The basic PHD model simply describes that the B-waves result from a linear mapping of the A-waves.

The significance of the $T_{22,11}(.)$ term is nicely demonstrated by the measured results of Figure 4. The figure represents a polar plot of the real and imaginary part of the B_{21} phasors, whereby a set of small A_{21}s depicting a smiley is injected into port 2, and whereby this experiment is done for seven different amplitudes of A_{11}. As such, each of the smileys corresponds to one A_{11} amplitude and can be considered as a representation of the $\Delta_{21}B_{21}$ in (21). The smiley looks undistorted at low A_{11} amplitudes, but gets squeezed at high A_{11} amplitudes. The squeezing is a direct consequence of the presence of the $T_{22,11}(.)$ term since the $S_{22,11}(.)$ term only describes a rotation and a scaling of the smiley (the graphical equivalent of multiplying a set of phasors by a fixed complex number).

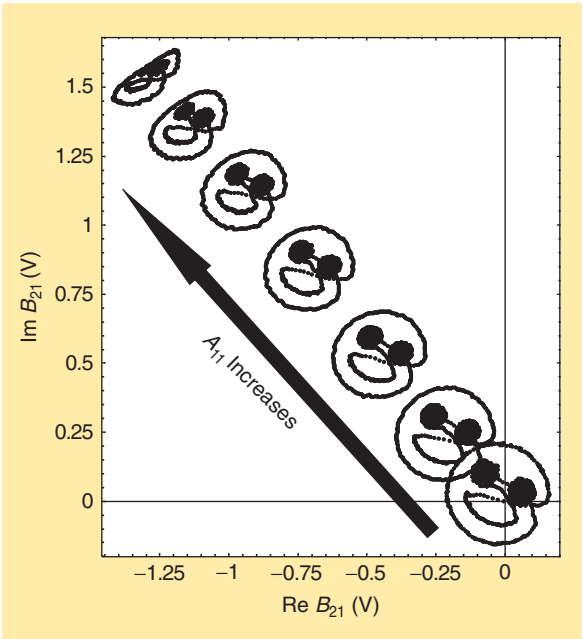


Figure 4. *Conjugate term distorts the smiley face.*

Besides the relative phase dependency, the PHD model has another unique feature when compared to S-parameters, namely that it relates input and output spectral components that have different frequencies. It describes, for example, how A_{13}, the third harmonic of the incident wave, will contribute to a change in B_{22}, the second harmonic at port 2. This corresponds to the concept of the conversion matrix well known to mixer designers [7]. Finally, a word on the significance of the Ps in (14). The Ps ensure that the whole of (14) represents a time-invariant DUT. Consider, for example, (14) and apply a delay τ to all of the A-waves. Define a new phasor Q, whereby f stands for the fundamental frequency

$$Q = e^{-j2\pi f\tau}. \tag{24}$$

Next, denote all delayed wave quantities by a superscript D. One can then write

$$A^D_{qn} = A_{qn}Q^n, \tag{25}$$

$$P^D = PQ. \tag{26}$$

Now calculate the B^D-wave corresponding with the delayed A-waves by substituting (26) and (25) into (14). This results in

$$B^D_{pm} = \sum_{qn} S_{pq,mn}(|A_{11}|)(PQ)^{+m-n}(A_{qn}Q^n) + \sum_{qn} T_{pq,mn}(|A_{11}|)(PQ)^{+m-n}\text{conj}(A_{qn}Q^n). \tag{27}$$

This can be simplified to

$$B^D_{pm} = \left(\sum_{qn} S_{pq,mn}(|A_{11}|)P^{+m-n}A_{qn} + \sum_{qn} T_{pq,mn}(|A_{11}|)P^{+m-n}\text{conj}(A_{qn})\right)Q^m \tag{28}$$

or simply

$$B^D_{pm} = B_{pm}Q^m = B_{pm}e^{-j2\pi mf\tau}. \tag{29}$$

In other words, the B-waves have been delayed by the same amount τ, as one expects from a time-invariant DUT. Note that this is no longer the case if one omits the Ps in (14). The most important consequence of the Ps is that the functions $S_{pq,mn}(.)$ and $T_{pq,mn}(.)$ are time-invariant properties of the DUT. Neither the amplitude nor the phase of the functions $S_{pq,mn}(.)$ and $T_{pq,mn}(.)$ changes as a function of time. Although this might seem trivial, many people get confused when they are dealing with relationships between tones that have different frequencies, especially when they are looking at phase characteristics. The PHD model, as represented by (14), provides an elegant mathematical

and experimental framework to deal with the aforementioned phase problem.

Nonlinear DUT Characteristics from the PHD Model

The following illustrates how the PHD model encapsulates and describes different nonlinear DUT characteristics. This is done by considering highly simplified versions of (14) and demonstrating the relationship of these simplified versions with existing nonlinear concepts.

Consider, for example, a highly simplified model containing exclusively the $S_{21,11}(.)$ term

$$B_{21} = S_{21,11}(|A_{11}|)A_{11}. \tag{30}$$

Division of both sides of (30) by A_{11} reveals that the amplitude of the function $S_{21,11}(.)$ corresponds to the compression characteristic of the DUT, while the AM-PM conversion characteristic is given by the phase of $S_{21,11}(.)$

$$S_{21,11}(|A_{11}|) = \frac{B_{21}}{A_{11}}. \tag{31}$$

Figure 5 shows the measured amplitude and phase of $S_{21,11}(.)$ of an Agilent Technologies' HMMC-5200 wideband microwave IC amplifier with a fundamental frequency of 9.9 GHz. Note that, unless specified otherwise, all measurement examples in this paragraph correspond to the same device and fundamental frequency. Defining the resulting compression and AM-PM conversion characteristic by means of a simplified PHD model implicitly assumes that it is independent from harmonic components and from the fundamental component incident to port 2. This is different from classic compression and AM-PM characteristics that are being measured on systems having imperfect matching characteristics. As a result, classic measurements of these characteristics differ from measurement system to measurement system. The $S_{21,11}(.)$ numbers returned by a PHD model measurement setup are compensated for the nonideal instrument port matches. For advanced measurement setups [4] even the effects of reflected harmonics can be included. This is actually similar to S-parameter measurements on a classic VNA; although the port match of two VNAs may significantly differ, the S-parameters returned by the instrument are not affected. As such, the measured $S_{pq.mn}(.)$ and $T_{pq,mn}(.)$ functions are true device characteristics, not disturbed by instrument imperfections.

In a similar way, the $S_{11,11}(.)$ function can be interpreted as the large-signal input reflection coefficient:

$$S_{11,11}(|A_{11}|) = \frac{B_{11}}{A_{11}}. \tag{32}$$

Figure 6 shows the amplitude and phase of $S_{11,11}(.)$. Note that the amplitude curve is expansive rather than compressive. This can be explained by the fact that the input matching circuitry has been designed for small signals and is based on classic small-signal S-parameters. When a large signal is being applied, the input impedances of the transistors inside the

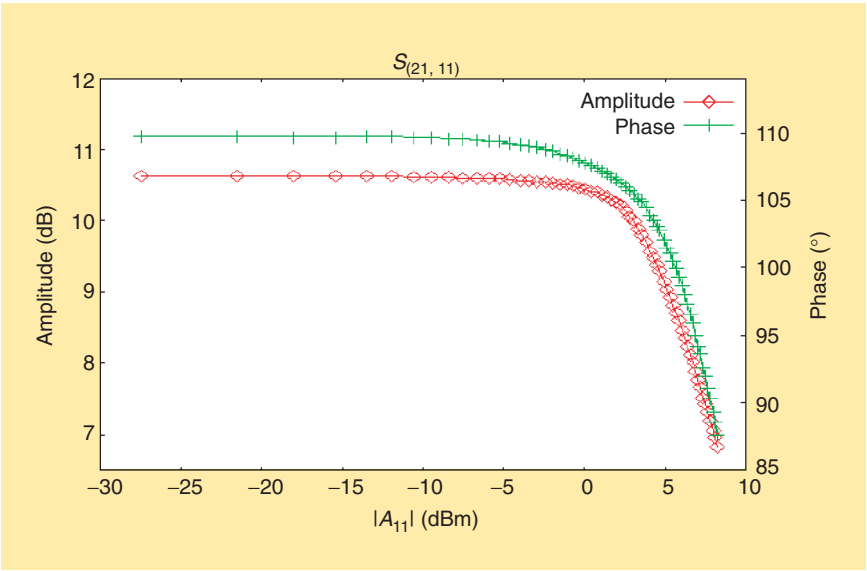


Figure 5. *Compression and AM-PM: $S_{21,11}(.)$.*

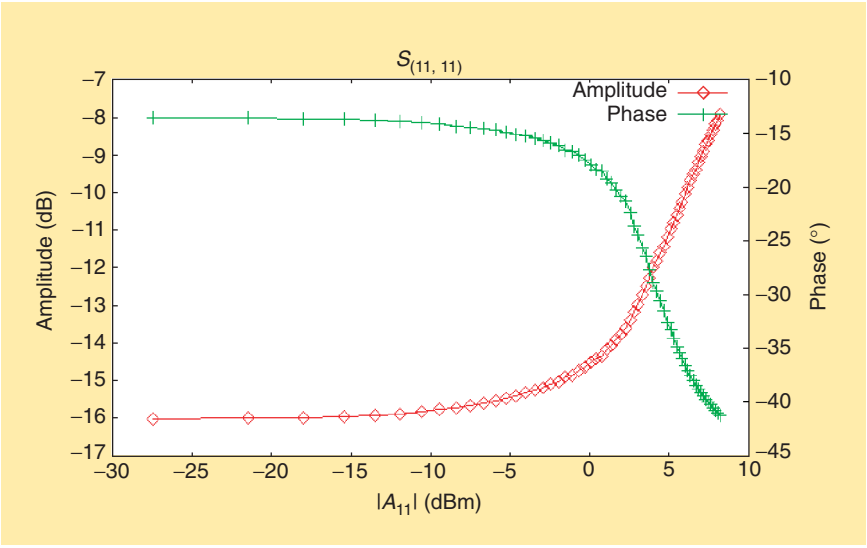


Figure 6. *Large-signal reflection: $S_{11,11}(.)$.*

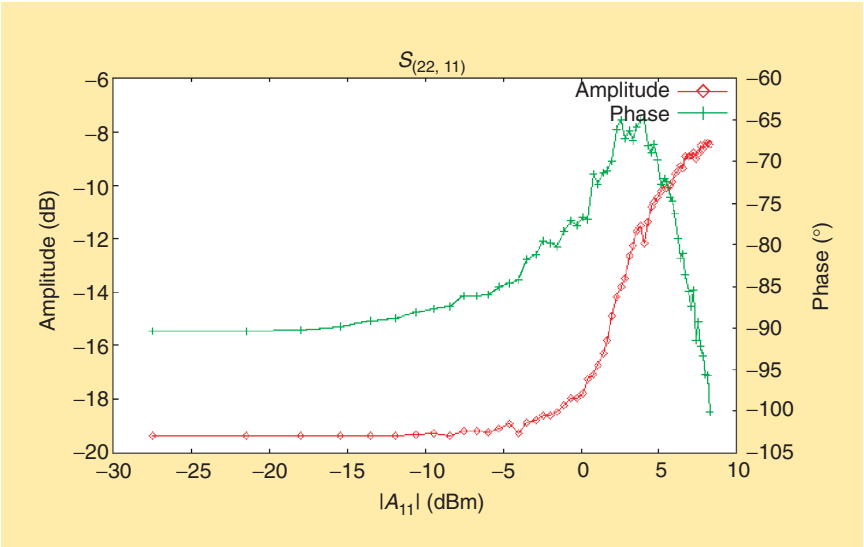


Figure 7. $S_{21,11}(.)$.

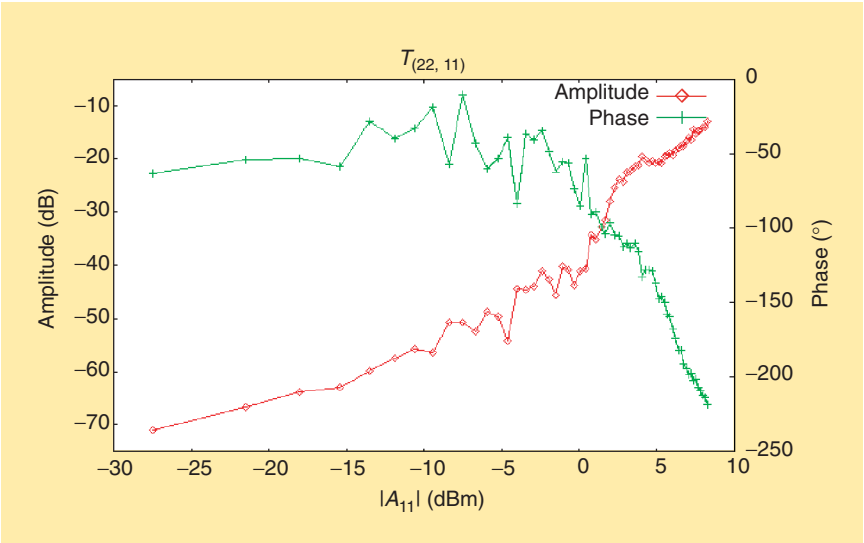


Figure 8. $T_{22,11}(.)$.

circuit change because of nonlinear effects while the matching circuits are typically linear and remain constant. As a result, the matching circuits are suboptimal under large signal conditions, and the amount of power reflected increases.

A similar result can be obtained for the output match. $S_{22,11}(.)$ and $T_{22,11}(.)$ provide an original and scientifically sound description of large signal output match, sometimes referred to as Hot S_{22}. The simplified PHD model equation for this case is a fundamental-only description of the B_{21} wave, as described by (20). Hot S_{22} behavior is tackled in a scientifically sound way by using the combination of $S_{22,11}(.)$ and $T_{22,11}(.)$. To our knowledge, this is an original result. Classic Hot S_{22} approaches completely ignore the existence of $T_{22,11}(.)$ [9]. In Figures 7 and 8, we show measured values of the amplitude and phase of $S_{22,11}(.)$ and $T_{22,11}(.)$ as a function of the amplitude of A_{11}, respectively.

As can be seen in the figures, $S_{22,11}(.)$ behaves similar to $S_{11,11}(.)$, the large signal input match. For small A_{11} amplitudes, the output match is pretty good, and at large A_{11} amplitudes, the characteristic expands and the output match begins deteriorating. For small A_{11} amplitudes, $S_{22,11}(.)$ and $S_{21,11}(.)$ approach the classic S-parameters s_{21} and s_{22}. $T_{22,11}(.)$ behaves very differently. Its amplitude becomes arbitrarily small when the amplitude of A_{11} approaches zero. This illustrates the fact that the component $T_{22,11}(.)$ is only visible under large-signal (nonlinear) operating conditions. The amplitude of $T_{22,11}(.)$ becomes significant when compression kicks in. As such, problems can be expected with classic Hot S_{22} approaches, as explained in [9], since those approaches completely neglect the existence of this component. Although it is not the case in our example, the amplitude of $T_{22,11}(.)$ can become even larger than the amplitude of $S_{22,11}(.)$, as described in [3].

All of the examples above refer to a fundamental only PHD model. In general, the approach can also describe the generation of harmonics. The simplest illustration is the capability to predict harmonic distortion analysis (HDA) characteristics. This is illustrated in Figure 9, which shows the HDA up to the fourth harmonic. The equations are simply

$$B_{21} = S_{21,21}(|A_{11}|)PA_{11}, \tag{33}$$

$$B_{23} = S_{21,31}(|A_{11}|)P^2A_{11}, \tag{34}$$

$$B_{24} = S_{21,41}(|A_{11}|)P^3A_{11}. \tag{35}$$

An important but more sophisticated application is the prediction of fundamental and harmonic loadpull behavior. In this case, we want to predict the B_{2h} waves (particularly B_{21}) as a function of the matching conditions at the output, both for the fundamental and the harmonics. To predict the component harmonic loadpull behavior, one needs to solve the following set of equations:

$$B_{2k} = S_{21,k1}(|A_{11}|)|A_{11}| + \sum_h S_{22,kh}(|A_{11}|)A_{2h} + \sum_h T_{22,kh}(|A_{11}|)conj(A_{2h}) \quad (36)$$

$$A_{2h} = \Gamma_h B_{2h}. \quad (37)$$

The first set of equations represents the PHD model; the second set is the mathematical representation of the matching conditions. Note that the set of equations is linear in the real and imaginary parts of A_{2h} (considered as separate variables) and is as such easy to solve. In the above load-pull example, it is assumed that A_{11} has zero phase, such that P equals one.

Measurement Setup and Experiment Design

The experiment design to extract the actual values of the PHD functions $S_{pq,mn}(.)$ and $T_{pq,mn}(.)$ is conceptually straightforward. Assume that we want to determine $S_{21,11}(.)$, $S_{22,11}(.)$, and $T_{22,11}(.)$ as they appear in (20), and this for a particular amplitude of A_{11}. The function extraction process is illustrated in Figure 10. We apply the particular A_{11} amplitude, and we keep it constant during the rest of the experiment. First, we do not apply any other incident wave besides A_{11} (this experiment is represented by the red square). This results in the knowledge of $S_{21,11}(|A_{11}|)$. Next, we perform two independent experiments, one applying an A_{21} with a zero phase and one applying an A_{21} with a 90° phase (corresponding to the blue and green square, respectively). Having those two additional measurements, we have sufficient information to calculate $S_{22,11}(|A_{11}|)$ and $T_{22,11}(|A_{11}|)$. A typical measurement setup is shown in Figure 11. An LSNA (Figure 12), measures all relevant A_{mk} and B_{mk} components. One synthesizer (source 1) is used for the generation of the A_{11} component. Since we are typically working in a large signal regime, the signal of this synthesizer is often amplified before being injected towards the input signal port of the DUT. A second synthesizer (source 2), combined with a switch, is used for the generation of the harmonic small signal components A_{mk}. These signals are called tickler signals. Although three measurements are theoretically sufficient to extract the PHD model functions, one usually performs many more measurements in combination with a linear regression technique. The presence of redundancy in the measurement set offers many possibilities in the framework of system identification, e.g., gathering information on noise errors and residual model errors.

An alternative approach, requiring fewer measurements, is the offset-tone algorithm described in [4] (see also the "On the Origin of the Conjugate Terms" sidebar).

Link with CAD Tools

The PHD model can be linked to harmonic balance and envelope simulators that are capable of implementing black-box frequency-domain models. In fact, the mathematical structure of the equations fits these simulators like a glove. This results in reduced memory requirements and fast simulations. Model accuracy is ensured by the fact that the PHD model is directly derived from measurements. The accuracy statement holds as far as the DUT is stimulated under conditions for which the assumed harmonic superposition principle holds.

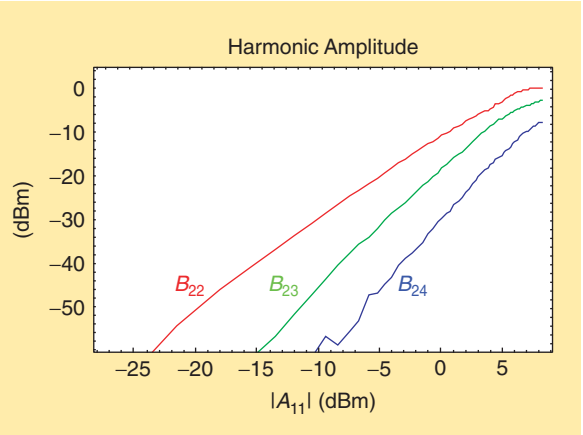


Figure 9. *Harmonic distortion analysis.*

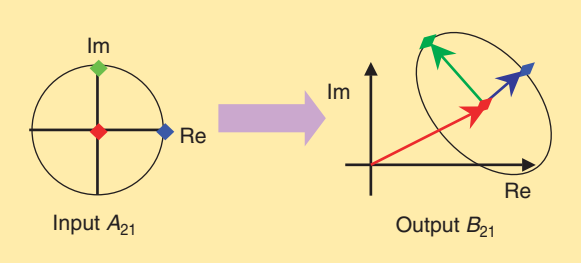


Figure 10. *Parameter extraction procedure.*

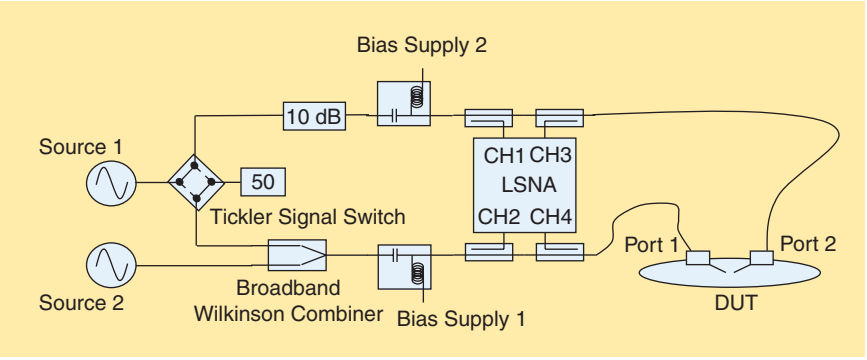


Figure 11. *Measurement setup.*

Figure 13 represents a comparison between the measured and modeled (by means of the PHD model) time domain current and voltage waveforms at the terminals of the HMMC-5200 under load-pull conditions. Note that the load-pull condition was arbitrarily chosen and was not part of the experimental data used to extract the scattering functions. As one can see, the correspondence is striking and should clearly be sufficient for practical power amplifier design. The modeled waveforms were calculated by evaluating the PHD model in Agilent ADS, a commercial harmonic balance simulator.

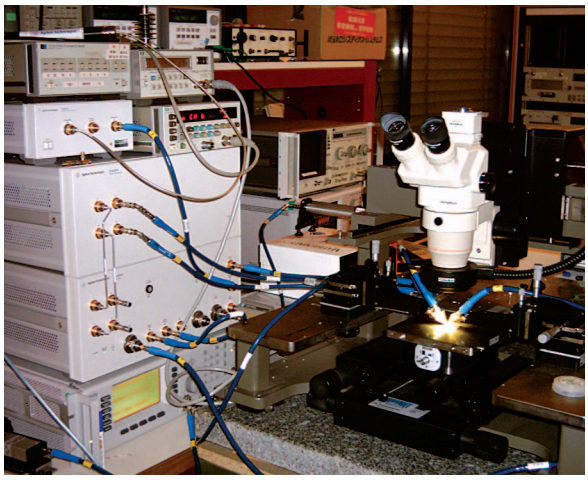


Figure 12. *Large-signal network analyzer (XLIM, France).*

Complex Modulation

The PHD model, as it was presented in the above, describes how discrete tone signals are interacting with devices. In practice, the input signal is often not a set of discrete tones but rather a modulated carrier. Depending on the application, the modulation can have many different formats. In the following, we will show how the PHD model can be applied with signals that are represented as a modulated carrier.

The key idea is to use a complex envelope domain representation of the A-wave and B-wave signals and to write the relationship between the A-waves and the B-waves as if it is a quasistatic relationship. The idea of the envelope domain is shown in (37), which describes the relationship between a time domain signal $x(t)$ and its complex envelope representation by a series of time-varying complex functions $X_h(t)$

$$x(t) = \mathrm{Re}\left(\sum_h X_h(t) e^{j2\pi h f_c t}\right). \tag{38}$$

Note that f_c represents the carrier frequency and that there is an envelope representation for the fundamental as well as for the harmonics. When this envelope representation is applied to the A-waves and the B-waves, one can rewrite the PHD model (14), whereby all wave quantities are replaced by the corresponding time-dependent envelope representations

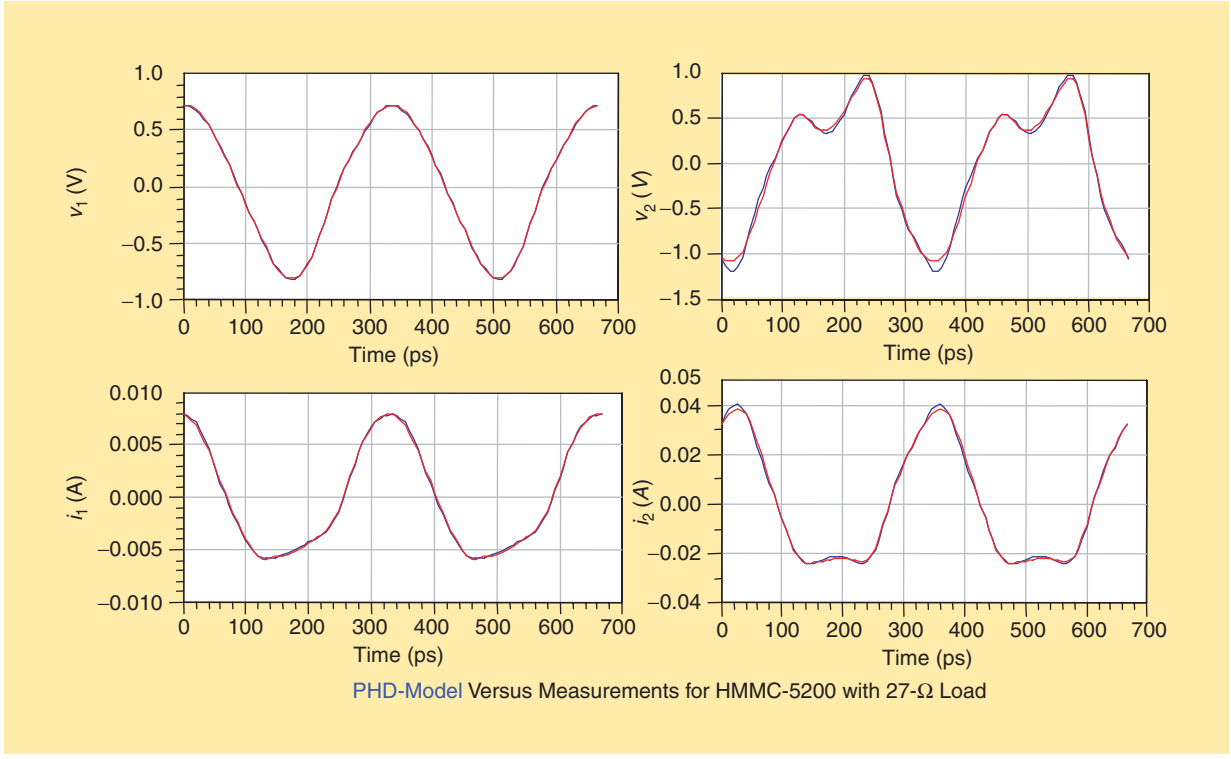


Figure 13. *Time domain waveforms.*

$$B_{pm}(t) = \sum_{qn} S_{pq,mn}(|A_{11}(t)|)P(t)^{+m-n}A_{qn}(t) + \sum_{qn} T_{pq,mn}(|A_{11}(t)|)P(t)^{+m-n}\mathrm{conj}(A_{qn}(t)). \quad (39)$$

Equation (39) can then be used to calculate the amplitude and phase of the *B*-wave complex envelopes as a function of the *A*-wave complex envelopes. The resulting time-dependent *B*-wave complex envelopes can be transformed into the frequency domain by a Fourier transform, whereby the resulting spectra are used to calculate typical nonlinear parameters such as adjacent-channel-power-ratio (IP3, IP5,…). Figure 14 shows an overlay of the output spectrum of an amplifier excited by a North American digital cellular signal, as predicted by a simulation and as predicted by a PHD model. Contrary to the previous examples, the PHD model was not derived from measurements but from harmonic balance circuit simulations, as explained in [4]. Note the excellent agreement between both characteristics.

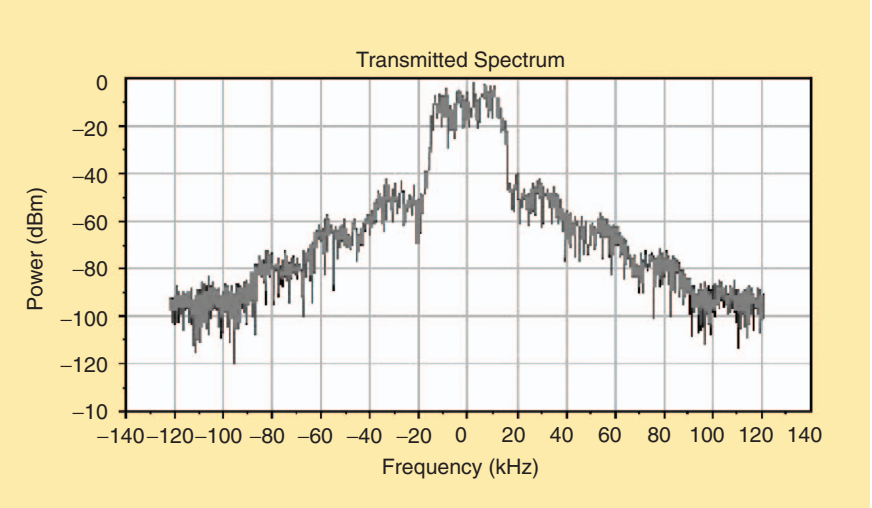


Figure 14. *Prediction of spectral regrowth.*

The question is, of course, when and to what degree the quasistatisticity principle, as used to derive (39), holds. Obviously, the principle will always hold if the modulation occurs slowly enough. But how slow is slow enough? The answer lies in the physics of the DUT. As long as any significant change in the modulation takes a longer time than the physical time constants governing the behavior of the system, the approach will work. These physical time constants are typically related to thermal issues, internal bias circuitry dynamics, and semiconductor material trapping effects. For a particular wideband RFIC, measured on wafer, the quasistatisticity principle was tested and proven to be valid up to a modulation bandwidth of about 1 GHz, implying that there were no significant time constants in the system larger than about 1 ns. This result can, of course, no longer be guaranteed once the RFIC is packaged and all kinds of parasitics are introduced.

Conclusions

We have presented the PHD modeling approach. It is a black-box frequency-domain model that provides a foundation for measurement, modeling, and simulation of driven nonlinear systems. The PHD model is very accurate for a wide variety of nonlinear characteristics, including compression, AM-PM, harmonics, load-pull, and time-domain waveforms. The PHD model faithfully represents driven nonlinear systems with mismatches at both the fundamental and harmonics. This enables the accurate simulation of distortion through cascaded chains of nonlinear components, thus providing key new design verification capabilities for RF and microwave modules and subsystems.

References

[1] R. Marks and D. Williams, "A general waveguide circuit theory," *J. Res. Nat. Inst. Standards Technol.*, vol. 97, no. 5, pp. 533–562, Sep.–Oct. 1992.

[2] J.C. Peyton Jones and S.A. Billings, "Describing functions, Volterra series, and the analysis of non-linear systems in the frequency domain," *Int. J. Contr.*, vol. 53, no. 4, pp. 871–887, 1991.

[3] J. Verspecht, D.F. Williams, D. Schreurs, K.A. Remley, and M.D. McKinley, "Linearization of large-signal scattering functions," *IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 4, pp. 1369–1376, Apr. 2005.

[4] D.E. Root, J. Verspecht, D. Sharrit, J. Wood, and A. Cognata, "Broad-band poly-harmonic distortion (PHD) behavioral models from fast automated simulations and large-signal vectorial network measurements,"*IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 11, pp. 3656–3664, Nov. 2005.

[5] J. Wood and D.E. Root, *Fundamentals of Nonlinear Behavioral Modeling for RF and Microwave Design*. Norwood, MA: Artech House, 2005, pp. 119–133.

[6] J. Verspecht, "Describing functions can better model hard nonlinearities in the frequency domain than the volterra theory," Annex Ph.D. thesis, Vrije Universiteit Brussel, Belgium, Sept. 1995 [Online]. Available: http://www.janverspecht.com

[7] S. Maas, *Microwave Mixers*. Norwood. MA: Artech House, 1992.

[8] J. Verspecht and P. Van Esch, "Accurately characterizing hard nonlinear behavior of microwave components with the nonlinear network measurement system: Introducing 'nonlinear scattering functions,'" in *Proc. 5th Int. Workshop Integrated Nonlinear Microwave Millimeterwave Circuits*, Germany, Oct. 1998, pp. 17–26.

[9] J. Verspecht, "Hot S-parameter techniques: 6 = 4 + 2," in *66th ARFTG Microwave Measurements Conf. Dig*, Dec. 2005., pp. 7–15 [Online]. Available: http://www.janverspecht.com

www.agilent.com
Product specifications and descriptions in this document subject to change without notice.

Printed in USA, August 12, 2008
5989-9574EN