Linearization: Reducing Distortion in Power Amplifiers

Allen Katz

Our society's need to exchange greater and greater amounts of information has created an unprecedented demand for highly linear power amplifiers (PAs). High linearity is required for the spectrally efficient transmission of information.

This article discusses techniques for the cancellation of distortion (linearization). Different methods of linearization are introduced and compared. The linearization of solid-state power amplifiers (SSPAs), traveling-wave-tube amplifiers (TWTAs) and klystron-power amplifiers (KPAs) are considered. Although the focus of this article is on power amplifiers, many of the techniques are applicable to other components as mixers, low-noise amplifiers, and even photonic components, such as lasers and optical modulators.

Amplifier Linearity

Technological developments are rapidly changing the communication business. In the past, the bulk of satellite transmissions was single-carrier video signals. Digital compression now allows many television signals to be transmitted in the frequency space previously occupied by a single signal. Nonvideo, broadband very small aperture terminals (VSATs) and mobile telephone/Internet services are altering traditional satellite loading. New terrestrial microwave services for the transmission of video, data, cellular telephone, and personal communications are appearing daily. Bandwidth-efficient modulation (BEM) schemes are becoming common. Virtually all of these services involve the transmission of multiple signals and/or large quantities of information at high data rates. For such signals,

Allen Katz is with The College of New Jersey, Ewing, New Jersey.

ISSN 1527-3342/01/$10.00©2001 IEEE

whether transmitted by frequency-division multiple access (FDMA), code-division multiple access (CDMA), or time-division multiple access (TDMA), amplifier linearity is a major consideration.

At high power levels (>100 W) TWTAs and KPAs offer the best microwave performance in terms of size, cost, and efficiency but lag behind SSPAs in linearity. The use of linearization can yield TWTA and KPA performance comparable or superior to conventional SSPAs. At lower powers, the advantage switches to SSPAs. As a result of new stringent linearity requirements, even relatively linear SSPAs can benefit from linearization.

Nonlinear Distortion

Nonlinear distortion can be thought of as the creation of undesired signal energy at frequencies not contained in the original signal. Distortion is produced by a loss of linearity. Amplitude linearity can be considered a measure of how closely the input-output transfer response of an amplifier resembles a straight line. When an amplifier's input level increases by a certain percent, its output level should increase by the same percent. A deviation from a straight line can be represented by a power series

$$V_{out} = K_1 V_{in} + K_2 V_{in}^2 + K_3 V_{in}^3 \cdots + K_n V_{in}^n. \tag{1}$$

When a single-carrier input signal, represented by a sine wave, is substituted into this expression, the output waveform will contain the original sine wave and harmonic distortion products. The harmonics can be eliminated by filtering and do not pose a problem except for wideband-communications applications of an octave or greater bandwidth. However, when more than one carrier is present, beat products are produced in the vicinity of the input signals. These new signals are known as intermodulation-distortion (IMD) products. They are located at frequencies above and below the input carriers and at frequency intervals equal to the separations of the input carriers (Figure 1). Filtering cannot easily eliminate IMD products since they are located on the same frequency or near to the desired input signals.

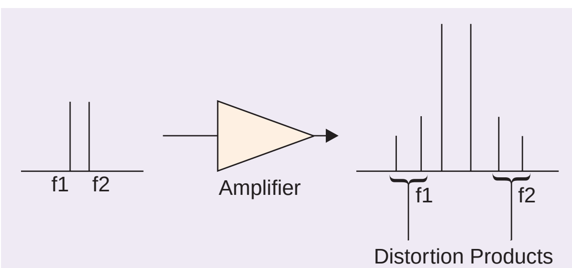

Figure 1. *When ≥ 2 signals are amplified, distortion products appear in the vicinity of the desired signals.*

Distortion is also produced by phase nonlinearity. The shift in phase angle that a signal encounters in passing through an amplifier is a measure of the time delay. Ideally, this phase shift, or time delay, should be constant for all power levels. A change in time delay with frequency, known as phase delay, envelope delay, or group delay, causes linear distortion and can be corrected with a phase equalizer

$$\theta(P_{in}) = \text{constant}, \tag{2}$$

where P_{in} is the instantaneous-input power level. In practical amplifiers, there can be a substantial change in phase with power level

$$\theta = f(P_{in}). \tag{3}$$

This change in phase with amplitude converts variations in signal level to phase modulation (PM). For a sinusoidal signal envelope

$$P_{in}(t) = k\left(A\cos[\omega_m t]\right)^2,$$

the resulting spectrum resembles that of a sinusoidal modulated PM signal

$$A_c \cos\left(\omega_c t + M\cos[\omega_m t]\right) = A_c \sum_{n=-\infty}^{n=\infty} J_n(M)\cos\left([\omega_c + n\omega_m]t\right), \tag{4}$$

where ω_c is the carrier frequency, ω_m is the modulation frequency (frequency of the envelope), and M is the modulation index (proportional to A). The PM sidebands are the IMD. Thus, phase nonlinearity produces IMD products in a similar fashion to amplitude nonlinearity. In some systems, phase nonlinearity is the principal cause of distortion.

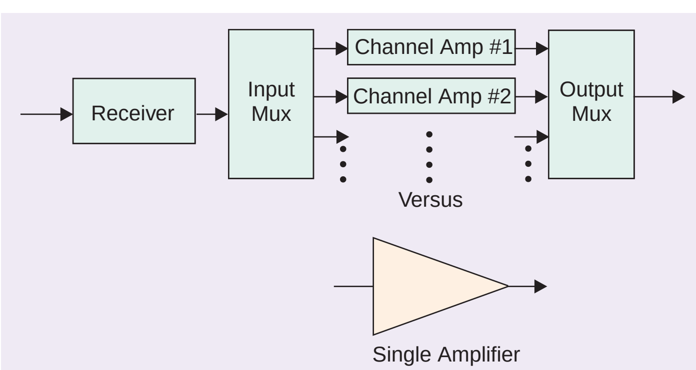

Figure 2. *In cellular telephony, sending several carriers through one amplifier is more cost effective.*

When multiple signals are sent through a communications system, an amplifier must be operated at a reduced power level (backed off) in order to keep distortion at an acceptable level. Distortion is often measured as the ratio of the carrier-to-IMD power level. This ratio is known as C/I. An acceptable level of IMD or C/I usually depends on the carrier-to-noise ratio (CNR) required at the receiver. IMD products can be considered to add to a receiver's noise level on power basis. For a carrier to IMD ratio,

- If C/I = CNR, the resultant CNR degrades by approximately 3 dB.
- If C/I = CNR + 6 dB, the resultant CNR degrades by approximately 1 dB.
- If C/I = CNR + 10 dB, the resultant CNR degrades by approximately 0.05 dB.

Thus, if the IMD products are to have a negligible affect on system performance, they should be at least 10 dB smaller than the carrier level.

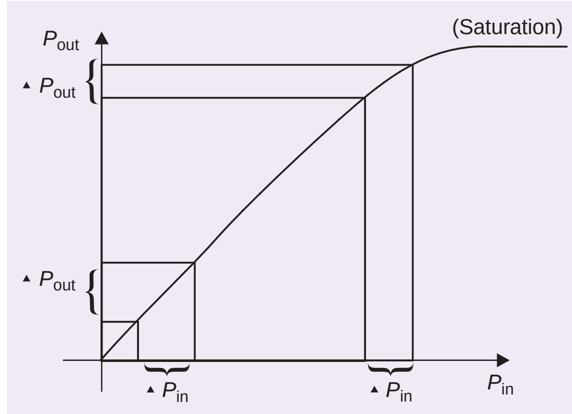

Figure 3. *As an amplifier is driven closer to SAT, its output level will increase by a smaller amount.*

In the case of cellular telephony, it is often more convenient and cost effective to transmit several carriers through a common amplifier rather than to use multiple amplifiers and a lossy multiplexer (Figure 2). To avoid unacceptably high IMD, the common amplifier must be highly linear.

For the transmission of a single carrier, IMD is usually not a limitation. However, with digitally modulated signals, spectral regrowth (SR) can be a serious problem. SR manifests itself in a form equivalent to IMD. It is not unique to digital signals but an aspect of angle modulation (FM and PM). Angle-modulated signals have a theoretically infinite bandwidth; for example, the spectrum of a sinusoidal modulated-PM signal of (3) contains an infinite number of sidebands. In practice, the bandwidth is limited to a finite frequency band beyond which sideband amplitude drops off rapidly. Analog PM has an approximate bandwidth given by Carson's rule

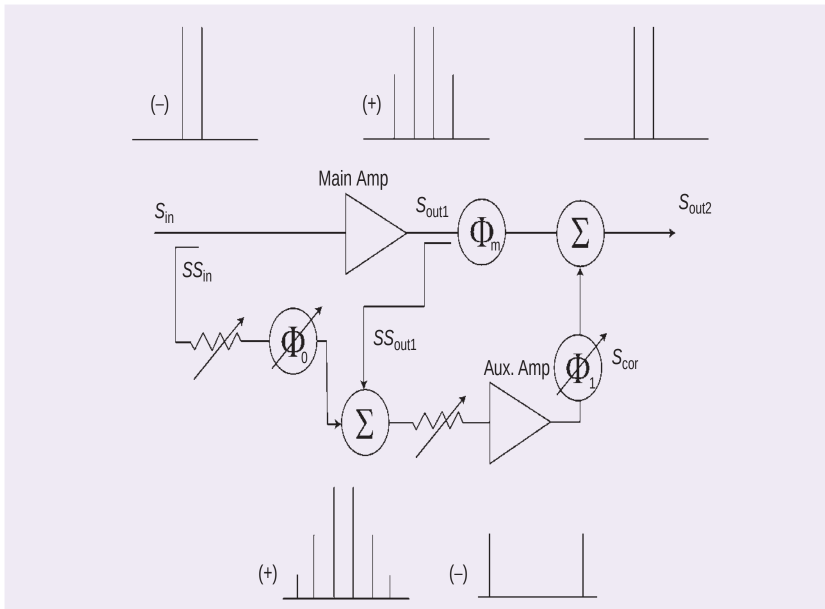

Figure 4. *Feedforward linearization employs two loops for the cancellation of IMD.*

$$BW = 2(\Delta f + f_m), \tag{5}$$

where Δf is the peak frequency deviation and f_m is the modulation frequency. The effective bandwidth of angle-modulated digital signals can be much greater than predicted by (5) due to the high-frequency components of the modulating waveform. To reduce their bandwidth to a more acceptable value, digital waveforms are normally low-pass filtered before modulation. Because of the mechanics of most digital modulators, which are not true angle modulators, the amplitude of the carrier is also modulated by this process. In addition, any "band-limiting" filtering of an angle-modulated signal will introduce amplitude modulation. It is primarily this incidental amplitude modulation that causes the SR when a digital signal is passed through a nonlinear amplifier. The distortion of the induced-amplitude waveform produces IMD products that increase the signal's spectrum.

The change in phase with amplitude (3) converts the variations in signal level to angle-modulation sidebands. These new sidebands further broaden the signal bandwidth. Amplitude and phase-induced spectral products add as vectors and are classified, in general, as IMD.

The summation of the IMD terms in an adjacent channel is referred to as the adjacent-channel power level (ACPL),

$$\text{ACPL} = \Sigma\text{IMDs}\,\big|\ \text{in an adjacent channel.}$$

The ratio of the adjacent-channel power to the carrier power is known as the adjacent-channel power ratio (ACPR).

ACPL is a major concern in personal-communications systems (PCS) since transmission often occurs on a channel adjacent to one in which reception of a much weaker distant signal may be taking place. To ensure freedom from interference, transmitter IMD products must be below the C/I by anywhere from 35 to >65 dB, depending on the application. These levels of linearity are considerably higher than had been required of communications amplifiers in the past, except for some special applications.

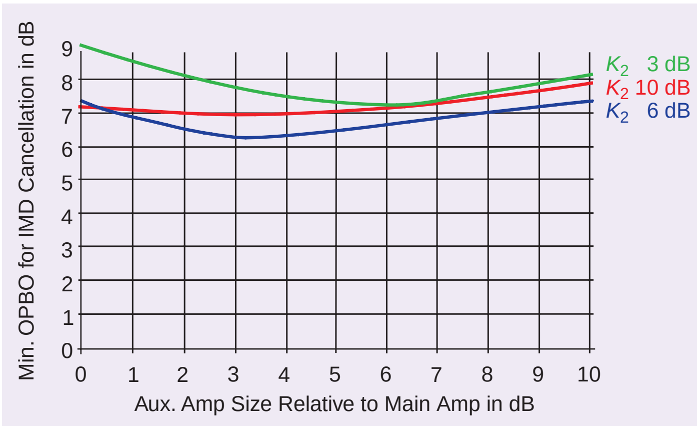

Figure 5. *The minimum OPBO for cancellation of IMD by a FF amplifier depends on the aux-amplifier size and output coupler coefficient.*

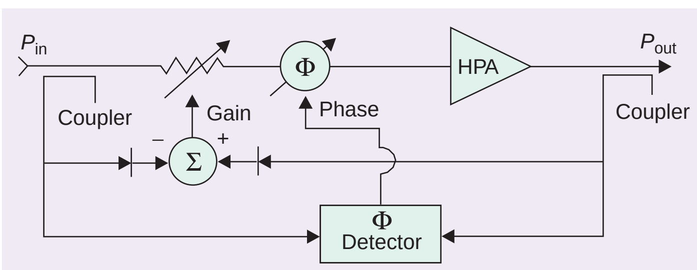

Figure 6. *IFB compares an amplifier's output and input and uses the detected difference to minimize distortion.*

Saturated Power

All amplifiers have some maximum output-power capacity, referred to as *saturated power* or simply *saturation* (SAT) (Figure 3). Driving an amplifier with a greater input signal will not produce an output above this level. As an amplifier is driven closer to SAT, its deviation from a straight-line response will increase. Its output level will increase by a smaller amount for a fixed increase in input signal, as shown in Figure 3. Thus, the closer an amplifier is driven to SAT, the greater the amount of distortion it normally produces.

The SAT point of TWTAs and KPAs is clearly defined as the output power normally decreases beyond SAT. Many SSPAs are sensitive to overdrive and can be easily damaged by operation at or beyond SAT. In addition, SSPAs tend to approach SAT exponentially. These factors make engineers reluctant to measure and use SAT as a reference for comparison of SSPA performance. They prefer to use the power at which an amplifier's gain compresses by 1 dB as the reference (REF) for amplifier comparison.

$$\text{REF} = \text{1-db CP} = \text{SAT} - D. \tag{6}$$

For SSPAs with reasonable linearity, the difference (D) in output level between SAT and the 1 dB compression point (CP) is

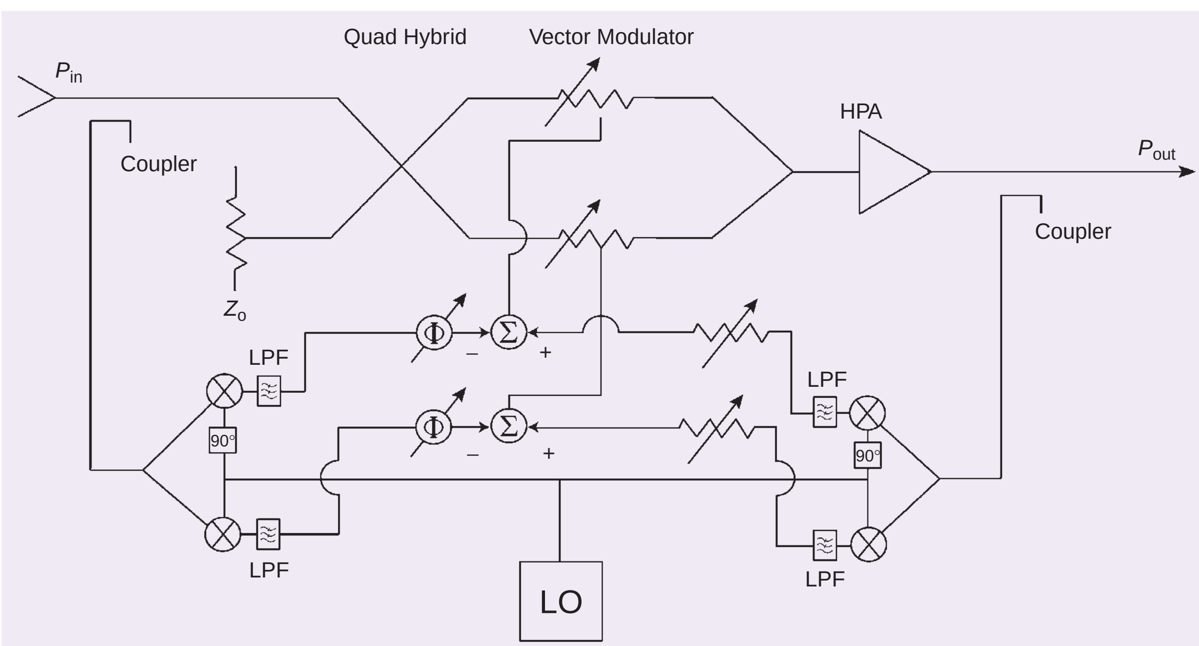

Figure 7. *Cartesian feedback eliminates the need for phase correction components by using the difference between in-phase and quadrature signals to control attenuators in a vector modulator.*

about 1 dB. Unfortunately, *D* varies from amplifier to amplifier. Generally, amplifiers with high linearity will have a smaller difference (*D* < .25 dB), while amplifiers with poor linearity can have a difference of several dB (*D* > 2 dB).

For this reason, in this article the relative amplifier performance will be referenced to (single-carrier) SAT. Output-power backoff (OPBO) will be relative to an amplifier's single-carrier SAT. (For most SSPAs, SAT can be safely determined using a network analyzer in a rapid power-sweep mode. For amplifiers that are especially thermally sensitive, pulsed power-sweep techniques may be used.) When comparing the data presented here with that of SSPAs based on a 1-db CP REF, an appropriate correction factor should be assumed.

Generally the greatest efficiency of a high power amplifier (HPA) will occur at or near SAT. Similarly, the closer to SAT a linear amplifier (class-A and, to a large extent, class-AB) is driven, the greater the amount of distortion it produces. For a satellite system, if a CNR of 16 dB (10 dB FM threshold + 6 dB for rain fading) is required and the IMD products are to have a negligible effect, then a C/I ≥ 26 dB is needed. To satisfy this requirement, a TWTA would typically have to be backed off 5-7 dB and sometimes more. This is about a 4-to-1 reduction in usable power. For TDMA applications, the backoff is less, usually 2-4 dB, to keep distortion in the form of SR from interfering with adjacent-channel communications. To satisfy cellular/PCS adjacent-channel IMD requirements, a (class-A) SSPA would have to be backed off about 6.5 dB for a C/I = 35 dB and by more than 15 dB for a C/I = 65 dB. These are huge reductions in usable output power. Therefore, it is desirable to look at various linearization techniques.

Linearization Techniques

Linearization is a systematic procedure for reducing an amplifier's distortion. There are many different ways of linearizing an amplifier. Usually, extra components are added to the design of a conventional amplifier. These extra components can often be configured into a subassembly or box that is referred to as a linearizer. Linearization allows an amplifier to produce more output power and operate at a higher level of efficiency for a given level of distortion. Feedforward, feedback, and predistortion are the most common forms of linearization. Besides these, there are a variety of other approaches that are being investigated. Most of these approaches use special techniques to obtain a linear output signal from highly nonlinear amplifiers. None of these alternate methods have been widely applied in wireless or microwave applications.

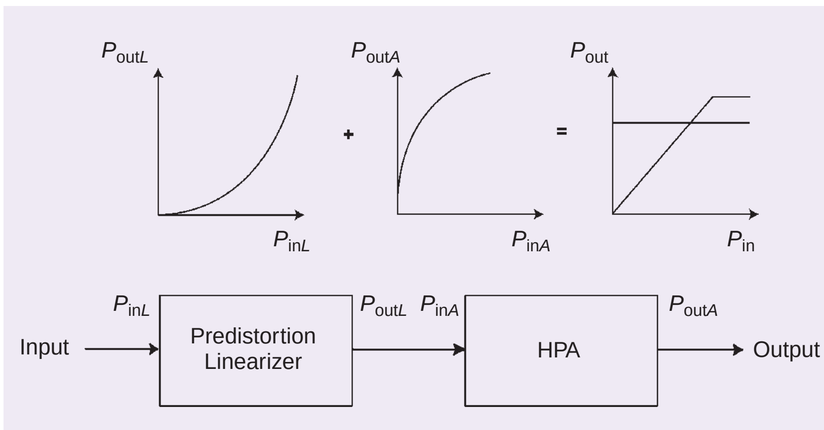

Figure 8. *PD linearizers generate a response opposite to an HPA's response in magnitude and phase.*

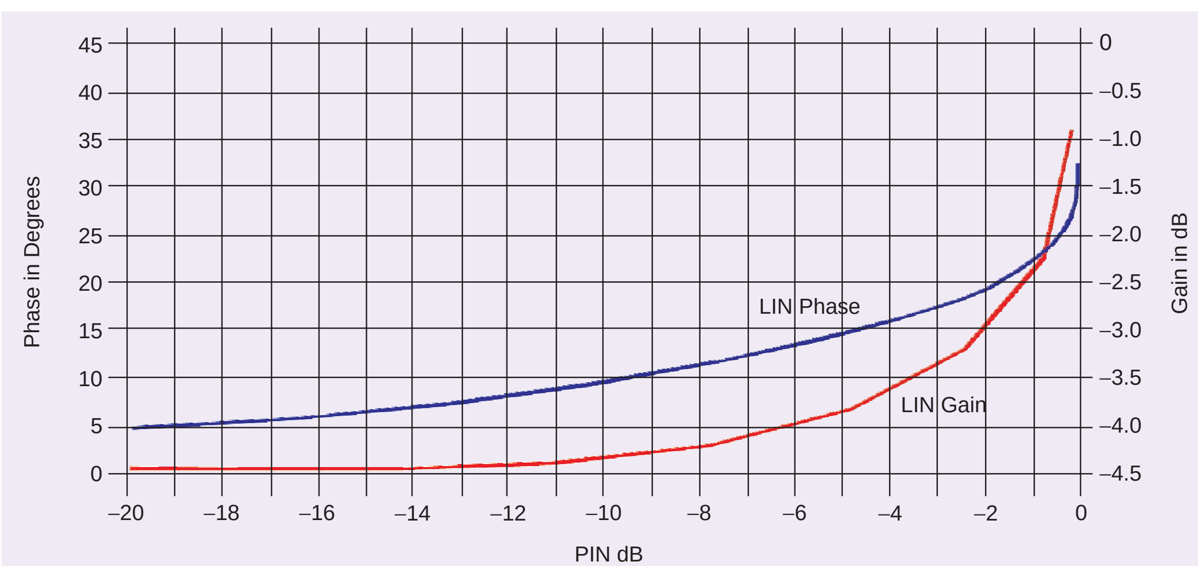

Figure 9. *An ideal PD-linearizer response requires the gain-and-phase slope to become infinite as SAT is approached.*

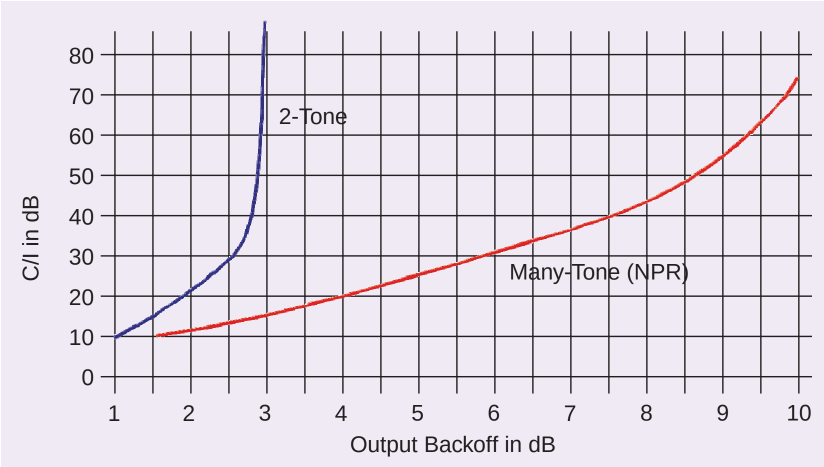

Figure 10. *C/I of an ideal linearizer for two and an infinite NPR.*

Feedforward Linearization

Feedforward (FF) has been extensively used with SSPAs and functions well with TWTAs and KPAs but is rather complex to implement and not easily added to an existing amplifier. A block diagram of a basic FF system is shown in Figure 4. This system consists of two loops. The first loop subtracts samples of the input signal (S_{in}) from the output signal (S_{out1}) to produce a sample of the main amplifier's distortion. S_{out1} consists of the amplified input signal plus any distortion introduced by the amplifier

$$S_{out1} = GS_{in}\angle\Phi_{amp} + \mathrm{IMD}, \tag{7}$$

where G is the gain and $\angle\Phi_{amp}$ is the phase shift introduced by the main amplifier. The samples of S_{in} (SS_{in}) and S_{out1} (SS_{out1}) are

$$SS_{in} = K_0 S_{in} \text{ and}$$
$$SS_{out1} = K_1 S_{out1},$$

where K_0 and K_1 are the coupling coefficients of the directional couplers used to sample S_{in} and S_{out1}, respectively. If SS_{in} is attenuated and delayed in phase such that

$$A_0 SS_{in}\angle\Phi_0 = -SS_{out1} \text{ or}$$
$$A_0 K_0 S_{in}\angle\Phi_0 = GK_1 S_{in}\angle\left(\Phi_{amp} + 180°\right), \tag{8}$$

then S_{in} is canceled and the output of loop 1 is K_1IMD. A_0 and Φ_0 are, respectively, the attenuation and phase shift introduced in loop 1 for adjustment of the carrier cancellation.

The second loop subtracts the amplified sampled distortion of loop 1 from a delayed S_{out1} to ideally produce a distortion-free output signal (S_{out2}). The loop 1 output signal is amplified by an auxiliary (aux) amplifier of gain GA and phase shift Φ_{aux} to provide a correction signal (S_{cor}) of sufficient level to cancel the distortion introduced by the main amplifier. S_{cor} is combined with the main amplifier signal at a final directional coupler of coefficient K_2. If

$$\begin{aligned} S_{cor} &= A_1 GAK_1K_2\mathrm{IMD}\angle(\Phi_{aux} + \Phi_1) \\ &= \mathrm{IMD}\angle(\Phi_{in} + 180°), \end{aligned} \tag{9}$$

then the HPA output will be distortion free. A_1 and Φ_1 are, respectively, the attenuation and phase shift introduced in loop 2 for adjustment of the distortion cancellation. Φ_m is a delay added after the main amplifier to equalize the delay introduced by the aux amplifier

$$S_{out2} = S_{out1}\angle\Phi_m + S_{cor}. \tag{10}$$

From this discussion, it may appear that undistorted output can be obtained from a FF amplifier right up to SAT. Saturated output power can never be obtained from a FF amplifier because of the losses in the phase shifter and couplers which must be located after the main amplifier. The main signal S_{out1} is reduced in amplitude by a factor (R_1) due to passing through the K_1 coupler. In dB,

$$R_1 = 10\log\left(1 - 10^{-(k1/10)}\right) + L_1, \tag{11}$$

where L_1 is the dissipation loss of the coupler in dB. K_1 can be made very small, provided the main amplifier has sufficient gain (a K_1 of –30 dB is not unusual). The K_2 of the final directional coupler must also be relatively small to minimize the loss of output power (R_2). Since the two signals, carriers, and distortion being

Table 1. Trade summary for DSP-based PD linearization.

Advantages	Disadvantages
Accurate correction over wide dynamic range and for irregular nonmonotonic characteristics.	Correction bandwidth limited by sampling rate: SR = 2 CBW = (4 N + 2)*BW*.
Easy to modify and update.	Cost can be higher than analog.
Simple to implement as adaptive system.	Power consumption can be high.

combined are not at the same frequency, power will be split between the load and the coupler's dump port. The R_2 power loss in dB as function of K_2 is described by (10) with 2 substituted for 1 in the variable names. The overall loss in SAT is

$$\Delta SAT = 10\log\left(1-10^{-(k1/10)}\right)+10\log\left(1-10^{-(k2/10)}\right) + L_1 + L_2 + L_m, \tag{12}$$

where L_m is the loss of the delay line (Φ_m). In practice, it is very difficult to achieve a ΔSAT of less than 1 dB. ΔSAT can be considered the minimum OPBO of a FF amplifier. In actuality, ΔSAT must be added to the difference between the SAT of an amplifier with single and multicarrier signals. This factor can vary from about .5 to >1.5 dB for HPAs. Furthermore, the amplifier's true SAT should not be considered, only that of the main amplifier. A FF amplifier combines both the power of the main and the aux amplifier. The sum of the SAT of both these amplifiers should be considered when comparing the relative OPBO performance of different methods of linearization.

Practical considerations limit the size of the aux amplifier. This limits S_{cor} and, in turn, the undistorted FF-output level. The smaller the K_2 is set, the larger in power the aux amplifier must be sized. The aux amplifier must also be operated relatively linear so as not to distort the distortion signal, thus introducing distortion of its own. Figure 5 shows the relationship between minimum OPBO (referenced from single-carrier SAT of the main amplifier and the aux amplifier) and aux-amplifier size (relative to the main power amplifier) for cancellation of IMD. Minimum OPBO is given for different values of output-coupler coefficient K_2. These results depend on the linearity of the main and aux amplifiers and on the resistive loss of the couplers and delay line. Linear characteristics typical of a class-A gallium arsenide (GaAs) field-effect transistor (FET) SSPA were assumed for both amplifiers, and resistive losses of 1 dB were assumed for the passive output components. With an aux amplifier of half the size of the main amplifier (3 dB), cancellation of IMD can be achieved only up to about –6.3 dB from SAT with a K_2 of 6 dB (Figure 5). If only the SAT of the main amplifier is considered, the minimum corrected OPBO is –4.2 dB but occurs for an aux amplifier equal in size to the main amplifier and a K_2 of about 3 dB. (A minimum IMD cancellation of 20 dB was assumed. If only 10 dB is acceptable, an additional 1-2 dB increase in output level can be achieved.) In practice, other factors limit IMD reduction, and perfect cancellation can never be achieved. Figure 5 reveals why FF is not a good choice for linearization of amplifiers near SAT. Other linearization methods can

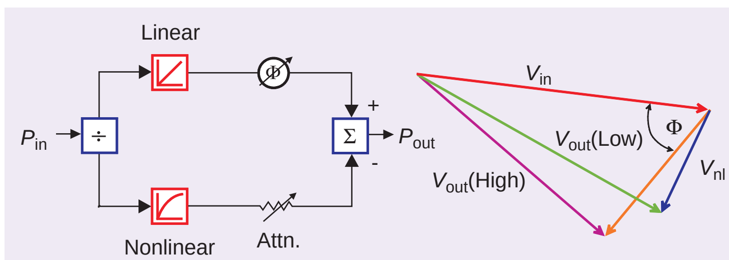

Figure 11. *Gain expansion can be produced by subtracting a linear path from a nonlinear path.*

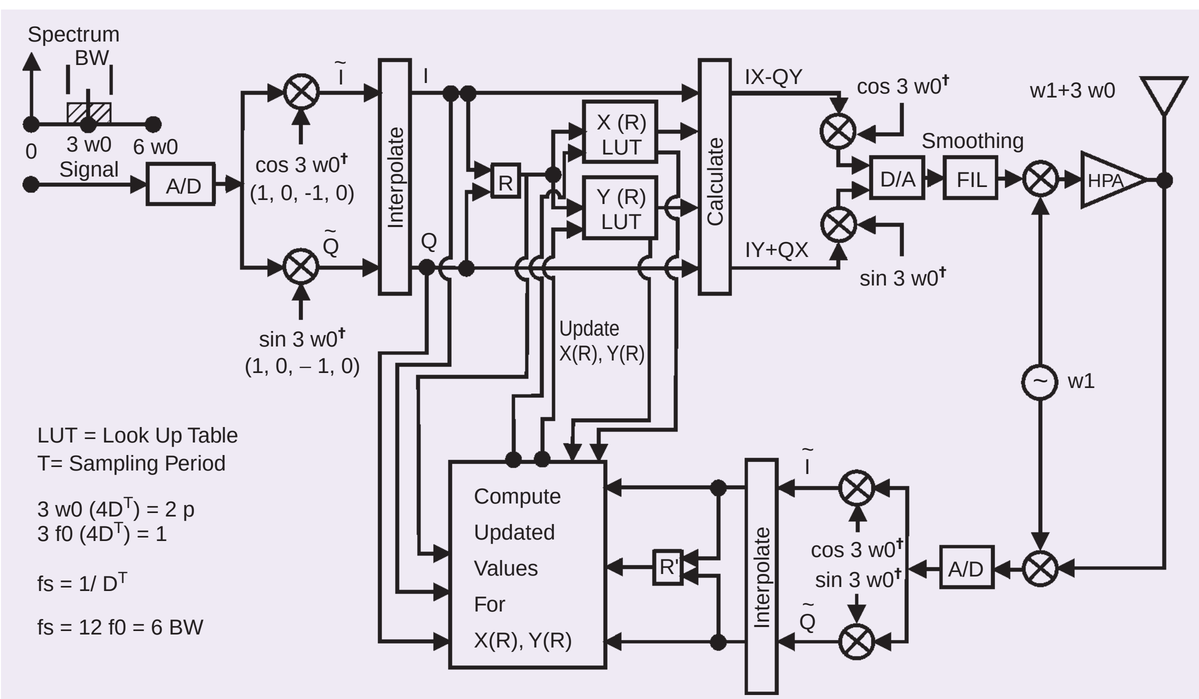

Figure 12. *DSP-linearization system using Cartesian predistortion and adaptive correction.*

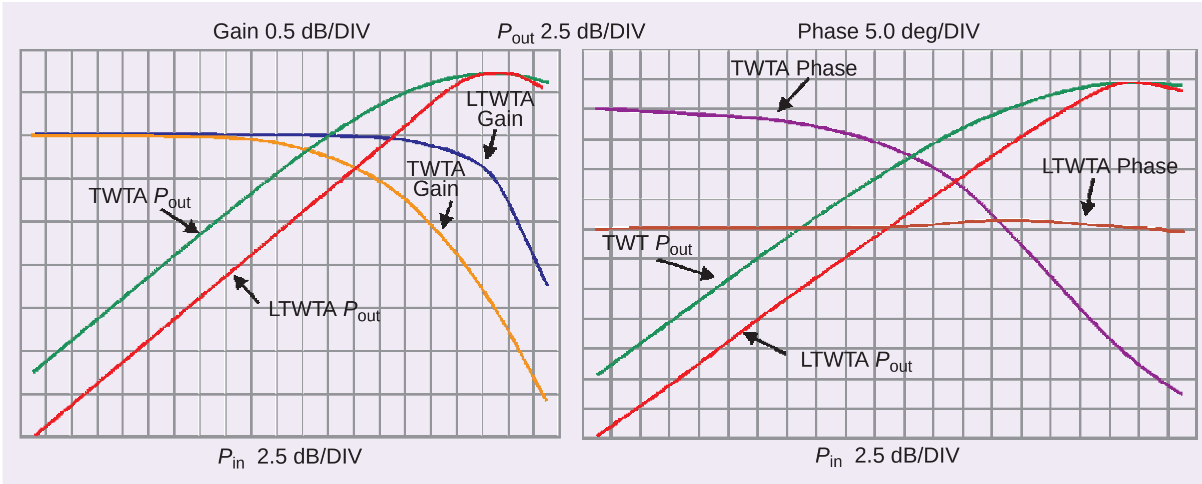

Figure 13. *Transfer characteristics of TWTA and linearized TWTA.*

provide superior IMD cancellation with considerably less complexity; however, for OPBOs greater than ~ 6-7 dB, FF becomes competitive, and for high linearity may be the system of choice.

Feedback Linearization

There has been considerable work on the use of feedback for the linearization of RF and microwave amplifiers. Feedback techniques can be divided into several distinct branches. The use of linear networks for feedback is well documented but has seen little use at microwave frequencies. The reason for this reluctance is probably due to concerns with amplifier stability and the difficulty in making networks with nonideal components function over wide frequency bands.

Indirect feedback (IFB) techniques have been more widely applied. In this approach an amplifier's input and output signals are detected and lowpass filtered, and the resulting baseband signals compared. The error signal (V_e) is used to modify the amplifier's characteristics to minimize distortion.

$$V_e = |DS_{out} - DS_{in}| \tag{13}$$

where DS_{out} and DS_{in} are, respectively, the detected output and input signals. V_e can be used to control the gain of the amplifier by means of a voltage variable attenuator. Dynamic electronic bias systems (DEBS), in which an amplifier bias is changed in response to the output signal, can be considered a variation on this form of linearization. The most widely known form of DEBS uses the input signal as the reference without comparison to produce an indirect form of FF linearization. Superior linearity can be obtained by correcting both amplitude and phase. The magnitude and phase error signals can be determined as illustrated in Figure 6. The resulting voltages are used to control an attenuator and a phase shifter to minimize signal error.

An alternate approach, known as Cartesian feedback, separates the signal into in-phase and quadrature components. This eliminates the need for phase-shift components and still allows the correction of gain and phase by adjusting the amplitudes of

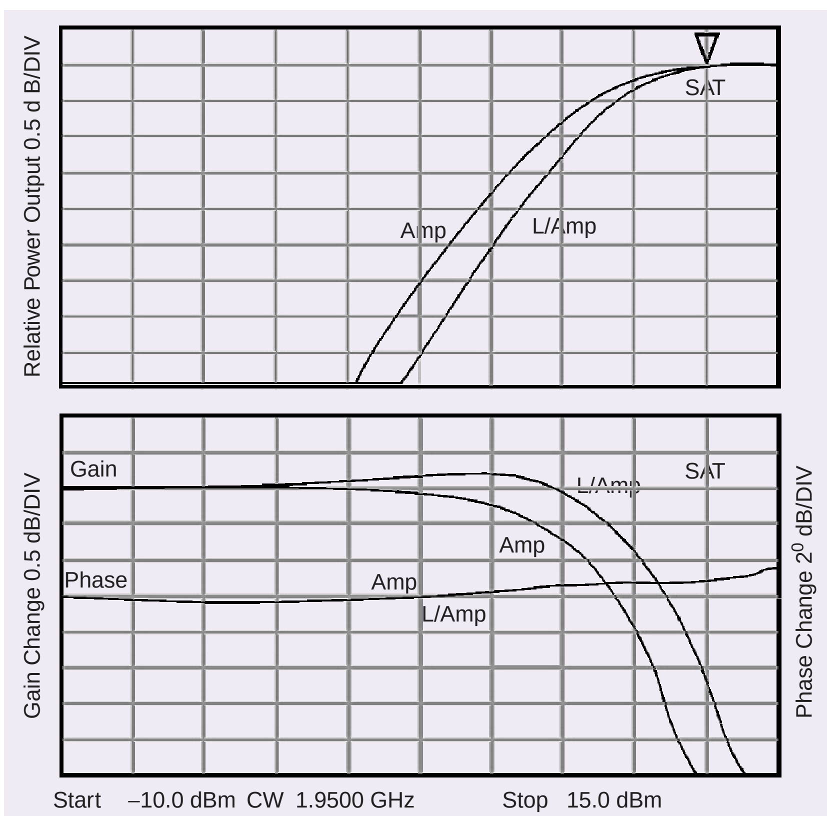

Figure 14. *Transfer characteristics of class-A, S-band SSPA and linearized SSPA.*

two orthogonal components. Figure 7 shows an example of a Cartesian-feedback system. The baseband in-phase and quadrature components are compared and used to control the attenuators in a vector modulator. Detection must be done synchronously (quadrature detection).

Cartesian feedback is most often used with quadrature phase-shift-keyed (QPSK) modulation. In this case, the output-side demodulated in-phase and quadrature components are subtracted directly from the respective in-phase and quadrature modulation signals at the input. This eliminates the need to demodulate on the input side. The correction at baseband is often done in the digital domain using digital signal processing (DSP) techniques.

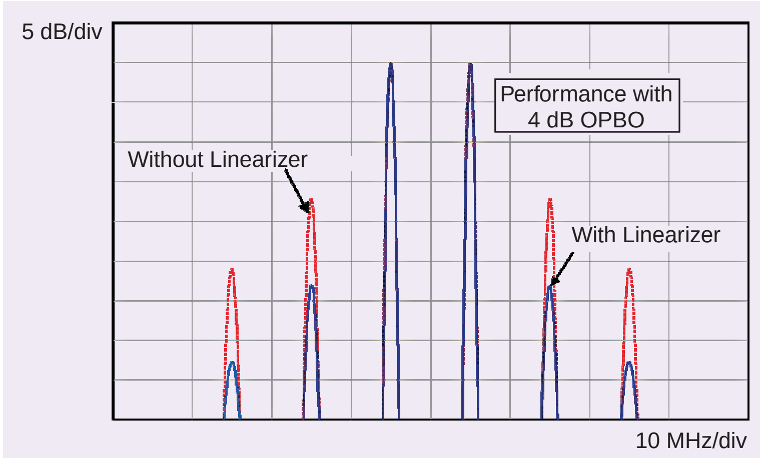

Figure 15. *For a TWTA, a two-tone C/I improvement of > 15 dB at 4-dB OPBO is common.*

Very high linearity can be achieved by using IFB, which is self-correcting for changes due to environmental and aging effects. IFB's principal limitation is an inability to handle wideband signals. In practice, it is difficult to make a feedback system respond to signal-envelope changes much greater than several MHz, because of the delay (Δt_S) of the amplifier and associated signal-processing components. The signal bandwidth must satisfy

$$BW < 1/(4\Delta t_S) \tag{14}$$

for significant correction. Thus, the total delay must be less than 25 ns for a 10 MHz *BW*. Microwave amplifiers can have delays of 10-20 ns. An advantage of Cartesian feedback is that the *BW*s of the in-phase and quadrature components are approximately equal, while, in Polar feedback systems, the *BW* of the phase component is much greater than the *BW* of the amplitude component.

Predistortion Linearization

Predistortion (PD) linearizers have been used extensively in microwave and satellite applications because of their relative simplicity and their ability to be added to existing amplifiers as separate stand-alone units. Unlike FF linearizers, they can provide a viable improvement in linearity near SAT but can be difficult to apply in applications requiring very high linearity (C/I > 50 dB). PD linearizers generate a nonlinear-transfer characteristic that can be thought of as the reverse of the amplifier's transfer characteristics in both magnitude and phase (Figure 8). An alternate way of thinking of a PD linearizer is to view the linearizer as a generator of IMD products. If the IMDs produced by the linearizer are made equal in amplitude and 180° out of phase with the IMDs generated by the amplifier, the IMDs will cancel. This condition occurs when the gain and phase of the linearized amplifier remain constant with change in power level.

In dB, the gain of the linearizer (*GL*) must increase by the same amount the amplifier's gain (*GA*) decreases

$$GL(P_{\text{out}L}) - GL_{SS} = -[GA(P_{\text{in}A}) - GA_{SS}] \mid P_{\text{out}L} = P_{\text{in}A}, \tag{15}$$

where GL_{SS} and GA_{SS} are, respectively, the small signal gains of the linearizer and the amplifier, and $GL(P_{\text{out}L})$ and $GA(P_{\text{in}A})$ are, respectively, these gains as a function of linearizer output and amplifier input levels. Likewise, the phase shift introduced by the linearizer must increase by the same amount the amplifier's phase decreases (or vice-versa, depending on the direction of phase change by the amplifier)

$$\Phi L(P_{\text{out}L}) - \Phi L_{SS} = -[\Phi A(P_{\text{in}A}) - \Phi A_{SS}] \mid P_{\text{out}L} = P_{\text{in}A}. \tag{16}$$

When these conditions are met, the result is the composite linear-transfer characteristic (Figure 8). This is the response of an ideal limiter. Once an amplifier has saturated, it is impossible to obtain more output power

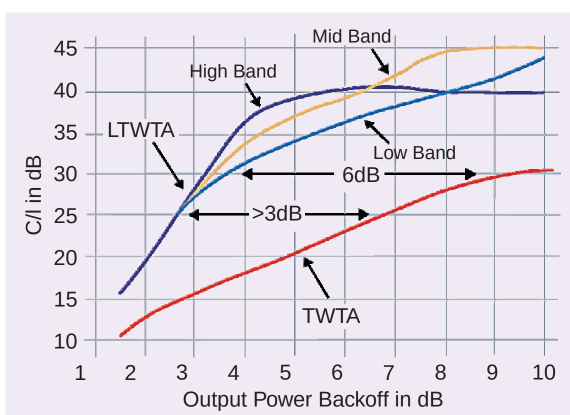

Figure 16. *A 4× increase in power for a two-tone C/I of 30 dB can be obtained by linearizing a TWTA.*

by driving the amplifier harder. Thus, the best a PD linearizer can do is to produce an ideal-limiter characteristic. Despite this limitation, it is possible for a linearizer to provide large benefits in signal quality when output power is reduced from SAT. Some improvement is possible, even at SAT and beyond, as the linearizer can correct for post-SAT phase distortion and power slump, but this improvement is usually very small. Since the power out of the amplifier (in dB) is

$$P_{outA} = P_{inA} + GA = P_{outL} + GA = P_{inL} + GL + GA.$$

Referenced to the power into the linearizer (P_{inL}), (15) and (16) can be rewritten and the desired transfer characteristics of the linearizer expressed as follows:

$$GL(P_{inL}) = GL_{SS} + GA_{SS} - GA\left(P_{inL} + GL(P_{inL})\right) \quad (17)$$

$$\Phi L(P_{inL}) = \Phi L_{SS} + \Phi A_{SS} - \Phi GA\left(P_{inL} + GL(P_{inL})\right). \quad (18)$$

Equations (17) and (18) can be solved iteratively for the ideal linearizer response needed to correct a given amplifier's transfer response. Figure 9 shows the response needed to ideally correct a typical TWTA. As SAT is approached, the rate of gain and phase change become infinite

$$dGL/dP_{in} = \infty \text{ and}$$
$$d\Phi L/dP_{in} = \infty \text{ as } P_{out} \to \text{SAT}.$$

Such a characteristic cannot be achieved in practice. Often, a small amount of gain expansion near SAT due to the finite dGL/dP_{in} available is traded for superior C/I near SAT at the expenses of degraded C/I at higher OPBOs.

Another limitation of PD (and FF) is the dependence of some amplifiers' transfer characteristics on the frequency content of the signal. This phenomenon is sometimes referred to as *memory effects*. Great care must be taken in the design of an amplifier to minimize these effects if the maximum benefit of PD linearization is to be achieved.

The two-tone C/I achievable by an ideal transfer characteristic is shown in Figure 10. The C/I goes to infinity, for OPBO > 3 dB. This result occurs because the peak envelope-power (PEP) of a two-tone signal is 3 dB greater than the average power. A signal backed off by more than 3 dB never experiences clipping at SAT and is only subject to a linear response. However, achieving this same level of performance with a larger number of carriers requires a greater level of OPBO. This is a consequence of the increase in PEP with carrier number

$$\text{PEP} = NP_{av}, \quad (19)$$

where N is the number of carriers and P_{av} is the average power of the overall signal. For four carriers, the OPBO for no IMD increases to 6 dB.

The C/I for an ideal limiter driven by an infinite number of carriers (of random phase) is also shown in Figure 10. The infinite-carrier case is known as noise power ratio (NPR). Although the OPBO required for a given C/I increases with N, the improvement provided by PD linearization also increases with N.

A PD linearizer can be produced by dividing an input signal into two parallel signal paths. One path is linear and can simply be a length of transmission line. The other path is nonlinear with a compression characteristic. This characteristic can be obtained from an amplifier driven into SAT. Subtracting the output signals from the two parallel paths results in a gain expansion (Figure 11)

$$V_{out} = V_{lin} - V_{nl}. \quad (20)$$

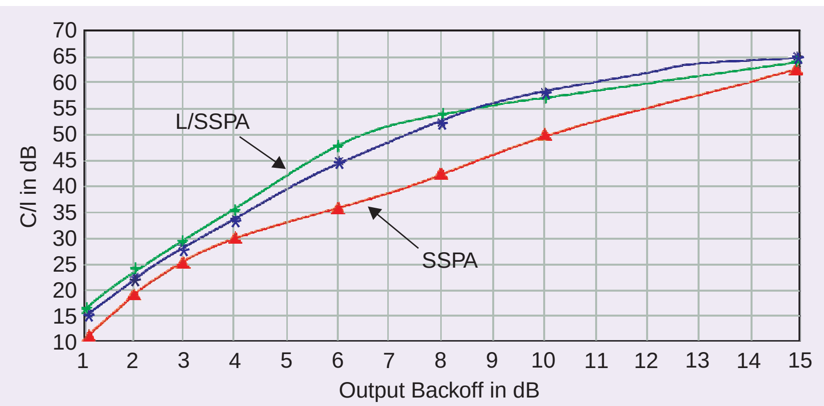

Figure 17. *Linearizing a class-A SSPA gives only 0.5 dB more power at 26-dB C/I, but 2.5 dB at a 50-dB C/I.*

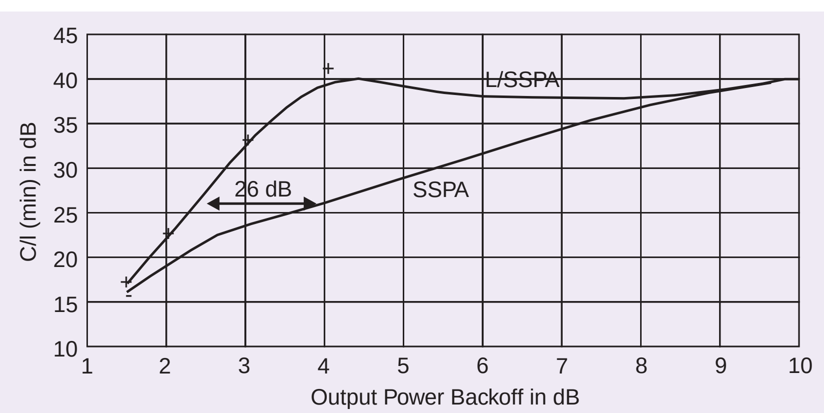

Figure 18. *Linearizing a less linear class-AB SSPA gives >1.5 dB more power at a C/I of 26 dB.*

The gain of the linear path (V_{lin}) remains constant with increasing drive level, while the gain of the nonlinear path (V_{ln}) decreases as SAT is approached. Thus, the overall gain (V_{out}) increases. Adjustment of the angle (θ) between the two paths allows the change of phase with level to be controlled.

Design advances have greatly simplified PD linearizers. Past linearizers were limited in bandwidth and dynamic range and difficult to adjust. New designs offer greater-than-octave frequency performance, and the complex nonmonotonic transfer responses needed by some HPAs. They are much smaller in size and provide enhanced performance with easy alignment and excellent stability.

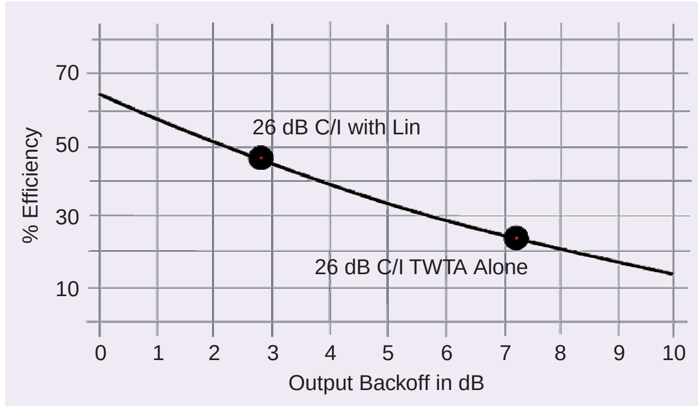

Figure 19. *A very significant increase in TWTA efficiency can be achieved for a C/I of 26 dB.*

Advances in DSP have caused great interest in synthesizing PD-transfer characteristics digitally. Such systems offer the potential of creating complex curves not readily produced by analog means. DSP PD's principal limitation is the need for the processing to be done at baseband—requiring up-and-down conversion for use with a microwave amplifier. The correction bandwidth (CBW) is also limited by the speed of the digital processor. The time between signal sampling is related not just to the signal BW, but also the number (N) of signal BWs on either side of the signal where distortion reduction is required. Table 1 shows a summary of some of the advantages and disadvantages of DSP-based PD. A DSP-linearization system employing Cartesian predistortion and adaptive correction is illustrated in Figure 12. Today, speeds adequate for many personal-communications applications are achievable. In the near future, CBWs of several-hundred MHz will be practical.

Adaptive Linearization

For high-linearity applications (C/I > ~ 50 dB), the adjustment and maintenance of the optimal linearizer settings become very critical. A change in phase of less than a degree can move a linearized amplifier out of specification. As a result of this parameter sensitivity, much effort has been devoted to the development of linearizers that can automatically adapt to environmental and stimulus change. DSP-based linearization is particularly suitable for an adaptive approach.

Adaptive linearizers can be considered a form of IFB linearization in which the feedback control is applied to PD and FF linearizers. A measure of the linearizer's performance is generated. This performance measure (V_{PM}) can take many forms but is always based on measurements over a time period greater than $2/BW$. V_{PM} can be derived from the difference between input and output waveforms (Figure 6) or the integrated IMD present in an unoccupied portion of spectrum near the desired signal. A microcomputer is normally used to analyze V_{PM} and determine the optimum linearizer settings.

In a FF linearizer, the microcomputer could control A_0 and Φ_0 in the signal loop and A_1 and Φ_1 in the cancellation loop [respectively, (8) and (9) and Figure 4]. Using a search algorithm, the computer would vary these parameters so as to keep V_{PM} at a minimum value. Adaptive correction is particularly important in FF systems as the balance is only correct for a specific power level.

In the PD linearizer of Figure 10, the microcomputer could control the attenuator and phase shifter to maintain V_{PM} at a minimum value as in the FF example. Alternately, the desired nonlinearity could be produced using a power series:

$$V_{out} = k1V_{in} + k2V_{in}^2 + k3V_{in}^3. \tag{21}$$

V_{in}^2 and V_{in}^3 can be generated using double-balanced mixers. V_{in} is applied to both ports of the mixer to obtain an output of V_{in}^2. A second mixer is used to obtain V_{in}^3. If needed, additional mixers can be used to obtain even higher powers. The values of coefficients k1, k2, and k3 could be controlled by the microcomputer. Two nonlinear PD elements can be combined in an arrangement similar to a Cartesian-feedback system to keep both gain and phase optimal.

These adaptive linearizers do not have the frequency-response limitations of feedback linearizers since they do not attempt to correct for changes in the signal's envelope. These linearizers respond slowly to gradual changes in the system's characteristics. Their principal disadvantage is complexity.

Linearizer Advantage

The transfer characteristics of a typical TWTA and the corrected response provided by a contemporary PD linearizer are shown in Figure 13. Note how the shape of the linearized P_{out}/P_{in} curve approaches the desired ideal-limiter characteristic of Figure 8. The separation of the 1-db CP from SAT is a good indicator of linearizer performance. Ideally, the 1-db CP is located 1 dB in input power beyond SAT. It is not unusual for TWTAs to have the 1-db CP occur 10-12 dB before SAT. In Figure 11, the

1-db CP is moved from 6.25 dB before SAT for the TWTA, to just about SAT for the linearized TWTA. The linearizer also reduces the change in phase with power level from more than 40° for the TWTA alone to a near flat line.

Figure 14 shows the transfer characteristics of a SSPA and the corresponding corrected response resulting from linearization. The characteristics are for a class-A-power metal-semiconductor field-effect transistor (MESFET) amplifier. Although the change in 1-db CP is not as great as for a TWTA, the benefit can still be substantial.

The need to exchange greater amounts of information has created a demand for highly linear power amplifiers

An example of the two-tone output spectrum of a typical TWTA, with and without linearization at 4-dB OPBO, is shown in Figure 15. A reduction in IMD of greater than 15 dB is common at this OPBO level. The improvement in two-tone C/I as a function of OPBO achieved by using a PD linearizer with a TWTA, a class-A MESFET SSPA, and a class-AB MESFET SSPA are depicted in Figures 16-18, respectively. For a TWTA at a C/I of 26 dB, the linearizer can provide a greater than 3-dB increase in output power. If a C/I ratio of 30 dB is required, the TWTA would have to be backed off at least 10 dB, but with the linearizer, it need only be backed off 4 dB. This is a 6-dB increase in output power.

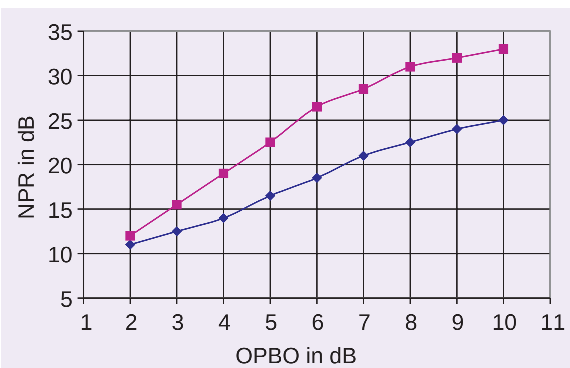

Figure 20. *NPR predicts amplifier performance with many carriers.*

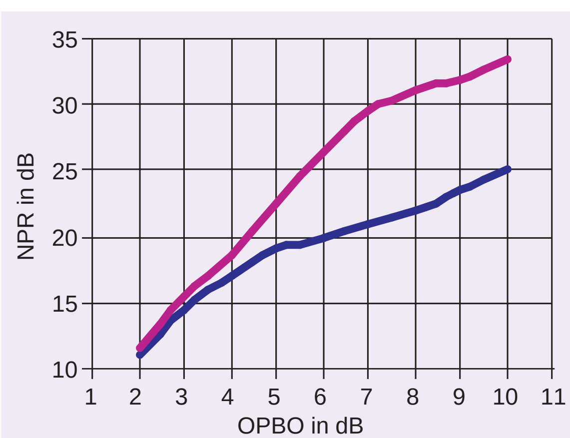

Figure 21. *NPR of class-AB SSPA.*

The advantage of linearizing SSPAs varies greatly with bias class and device type. The class-A (MESFET) amplifier of Figure 16 shows only about a 0.5-dB increase in output power for a C/I of 26 dB, but a 2.5-dB power increase for a 50-dB C/I. The class-AB SSPA (Figure 17) shows about a 1.5-dB increase in output power for a C/I of 26 dB. Ordinarily, the more linear an SSPA, the less the advantage of linearization. When designing an HPA to be linearized, emphasis should be placed on optimizing factors other than linearity.

An even greater HPA-output-power increase should be achieved for signals of more than two carriers, although a higher level of OPBO will be required for the same C/I level as the number of carriers is increased. Generally, the greater the required linearity, the greater the benefit of using a linearizer. Conversely, the closer an HPA is operated to SAT, the smaller the benefit of using a linearizer.

There can be other reasons, besides increased output power, for linearizing an HPA. For example, thermal considerations can place major constraints on the design of an HPA. Linearization increases an HPA's efficiency by allowing it to operate closer to SAT. Increased efficiency reduces thermal loading. Figure 19 shows how efficiency is related to OPBO for a modern high-efficiency TWTA. For a C/I of 26 dB, the use of a linearizer can provide greater than a 70% efficiency increase. In the case of an SSPA, linearization may allow operation at a more efficient bias than would have been otherwise possible.

The performance of a linearized HPA with many carriers (>10) is normally tested using a NPR measurement. In this test, white noise is used to simulate the presence of many carriers of random amplitude and phase. The NPR of a typical TWTA and a linearized TWTA are shown in Figure 20. In Figure 21, similar NPR measurements are shown for a class-AB SSPA.

With single-carrier modulated signals, a linearizer can often be of great value, especially with BEM. For example, HPAs transmitting single-carrier QPSK and offset QPSK (OQPSK) signals are usually operated at a reduced-output level. They are backed off to prevent SR, which can interfere with adjacent-channel signals. Linearization can reduce this spreading to an acceptable level (>25 dB) for OPBOs of .25-.5 dB from SAT. Figure 22 shows an illustration of the improvement provided by a linearizer for a QPSK (or OQPSK) satellite signal. At a 4-dB OPBO, about a 10-dB decrease in interference level is achieved. Figure 23 shows the reduction in spectral regrowth achieved by linearization of a TWTA.

It has been found empirically that a 30-dB SR corresponds to a two-tone C/I ratio of about 25 dB. The SR of

QPSK and OQSPK modulation are similar adjacent to the modulation bandwidth; however, OQPSK gives improved SR performance at greater frequency separation. Generally, the SR of binary phase-shift-keyed (BPSK) and 8-PSK are close to that of QPSK/OQPSK with BPSK having slightly poorer (–1 dB) and 8-PSK providing slightly better performance. In most cases linearization can also improve the bit-error rate (BER) of digital modulated signals.

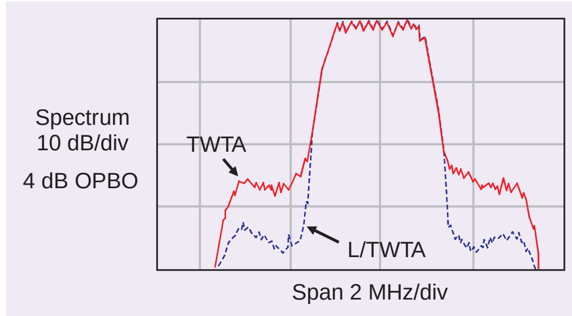

Figure 22. *Bandwidth/noise reduction of QPSK signal achieved by linearization.*

Summary

Linearizers are needed to increase HPAs' power capacity and efficiency when handling multicarrier and BEM traffic. New linearizer designs have greatly enhanced performance and bandwidths, made alignment easier, and provide excellent stability and reliability. These linearizers can deliver up to a four-fold increase in TWTA power capacity and more than double TWTA efficiency. They can increase SSPA-power capacity and efficiency when high linearity is required. The greatest benefit is accrued for class-B and -AB amplifiers in applications requiring a high linearity. In these cases, linearizers can deliver a >3 dB increase in power capacity and more than double SSPA efficiency. Generally, FF and adaptive linearization are most valuable for applications requiring very high linearity. IFB methods work well but are limited in bandwidth. PD has the advantage of relative simplicity. It works over wide bandwidths and is viable for applications requiring both low- and moderate-to- high linearity.

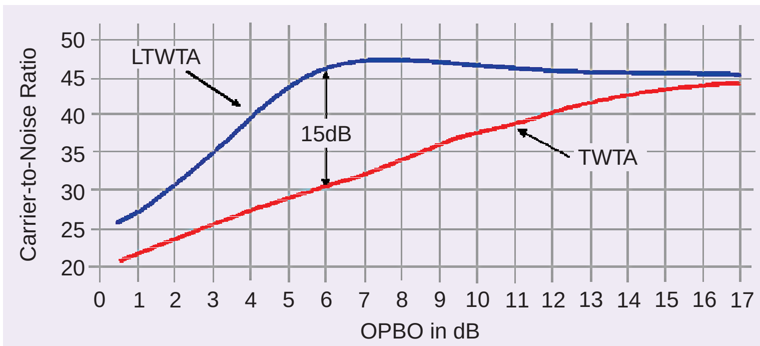

Figure 23. *Reduction in spectral regrowth provided by linearization of a TWTA.*

References

[1] J. Balicki, E. Cook, R. Heidt, and V. Rutter, "The AR6A single-sideband microwave radio system: The traveling tube amplifier," *Bell Syst. Tech. J.*, vol. 62, no. 10, pp. 3429-3445, Dec. 1983.

[2] E. Ballesteros and F. Perez, "Microwave power amplifiers linearized by active feedback networks," presented at the IEEE MTT-S Symp. Workshop on Linearizers, Jun. 1988.

[3] W. Bosch and G. Gatti, "Measurement and simulation of memory effects in predistortion linearizers," *IEEE Trans. Microwave Theory Tech.*, vol. 37, pp. 1885-1890, Dec. 1989.

[4] C. Bremenson, et al., "Linearizing TWT amplifiers in satellite transponders, systems aspects and practical implementation." in *Proc. AIAA 8th Communication Satellite Systems Conf.*, 1980, pp. 80-89.

[5] D. Cahana et al., "Linearized transponder technology for satellite communications, Part 1: Linearizer circuit development and experimental characterization," *COMSAT Tech. Rev.*, vol. 15, no. 2A, pp. 277-308, Fall 1985.

[6] R. Inada et al., "A compact 4-GHz linearizer for space use," *IEEE Trans. Microwave Theory Tech.*, vol. 34, pp. 1327-1332, Dec. 1986.

[7] A. Katz and R. Dorval, "Evaluation and correction of time dependent amplifier non-linearity," *IEEE MTT-S IMS Dig.*, pp. 839-842, Jun. 1996.

[8] A. Katz et al., "A reflective diode linearizer for spacecraft applications," *IEEE MTT-S IMS Dig.*, pp. 661-664, Jun. 1985.

[9] P. Kenington, "Methods linearize RF transmitters and power amps, Part 1," *Microw. RF*, pp. 103-116, Dec. 1998.

[10] P. Kenington, "Methods linearize RF transmitters and power amps, Part 2," *Microw. RF*, pp. 79-89, Jan. 1999.

[11] A. M. Khilla et al., "Multi-band predistortion linearizers for satellite transponders and ground stations," *AIAA J.*, pp. 927-937, 1992.

[12] Y.S. Lee, I. Brelian, and A. Atia, "Linearized transponder technology for satellite communications; Part 2: System simulation and performance assessment," *COMSAT Tech. Rev.*, vol. 15, no. 2A, pp. 309-341, Fall 1985.

[13] M. Kumar et al., "Predistortion linearizer using GaAs dual-gate MESFET for TWTA and SSPA used in satellite transponders," *IEEE Trans. Microwave Theory Tech.*, vol. 33, pp. 1479-1484, Dec. 1985.

[14] J Minkoffm, "Intermodulation noise in solid-state power amplifiers for wideband signal transmission," in *Proc. AIAA 9th Communications Satellite Systems Conference*, 1982.

[15] S.S. Moochalla et al., "An integrated Ku-band linearizer driver amplifier for TWTAs with gain and wide bandwidth," presented at the AIAA 14th Int. Communications Satellite Systems Conf., 1992.

[16] A. Saleh, "Intermodulation analysis of FDMA satellite systems employing compensated and uncompensated TWTs," *IEEE Trans. Commun.*, vol. 30, no. 5, pp. 1233-1242, May 1982.

[17] G. Satoh and T. Mizuno, "Impact of a new TWTA linearizer upon QPSK/TDMA transmission performance." *IEEE J. Select. Areas Commun.*, vol. 1, pp. 39-45, Jan. 1983.

[18] J. Steck and D. Pham. "A new TWT linearizer for satellite-communications transmitters," *Microw. Syst. News Commun. Tech.*, vol.17, no.9, pp. 28-42, Jun. 1987.