Linearizers - Distortion Reduction in High Power Amplifiers

Allen Katz

The College of New Jersey/Linearizer Technology, Inc.

Abstract -- Linearization has become of enormous interest because of the need to send greater and greater amounts of information over a limited spectrum, and the need to transmit signals more efficiently. As a result there have been numerous innovative papers that have greatly advanced the technology. At the same time, there seems to be loss of appreciation of the fundamentals upon which these advances are based. This paper will review the basics of linearization. Both analog and digital techniques will be included. It will set limits on the performance that can be achieved. It will consider the links between linearity and efficiency. It will also discuss memory effects and their impact on linearizer performance, and conclude with the state of current linearizer development particularly with respect to limits on upper frequency and bandwidth.

***Index Terms* — linearizer, linearization, predistortion, high power amplifier, TWTA, millimeter-wave.**

I. Introduction

Technological developments have changed the communication business. Virtually all communications now use bandwidth efficient digitally modulated signals, and often involve the transmission of large quantities of information at high data rates. For such signals amplifier linearity is a major concern.

At high power levels, the best microwave performance in terms of size, cost and efficiency is still offered by TWTAs, but GaN based SSPAs are gaining ground. Both of these types of high power amplifiers (HPAs) can benefit greatly from linearization – typically > 3 dB of additional output power for comparable linearity and for conventional designs > a doubling in efficiency. The ultimate decision to linearize is basically economic. For most HPAs, output power can be traded for linearity. The poorer the linearity of a HPA, the greater is the advantage of linearization. The size of the HPA is also an important factor. The larger the power of a HPA, the easier it is to justify linearization, since the cost of linearization is independent of the HPA size. For most HPAs, the closer it is operated to maximum (saturated) power, the greater is its efficiency. A consequence of this better efficiency and reduced power overhead made possible by linearization is smaller size, weight and thermal load. Also the greater the level of linearity specified, the greater is the advantage gained by linearization.

II. Need for Linearity

Non-linear distortion can be thought of as the creation of undesired signal energy at frequencies not contained in the original signal. This distortion is produced by a loss of linearity. The elimination of the interference caused by these spectral products to adjacent channel signals is the principal reason linearity is required. There is a secondary reason for eliminating distortion, which is related to actual *distortion* of the signal. If a signal is sufficiently distorted, it will not be decoded correctly and errors in copy will result. This type of distortion is created by both linear and nonlinear processes, and is not subject to the same restrictions. In most HPA applications, it is the adjacent channel distortion that is the dominant concern. This type of distortion is the focus of this paper.

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.

$$V_{out}=K_1V_{in}+K_2V_{in}^2+K_3V_{in}^3....+K_nV_{in}^n \qquad (1)$$

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. This condition is illustrated in Figure 1. Filtering cannot easily eliminate IMD products, since they are located on the same frequency or near to the desired input signals.

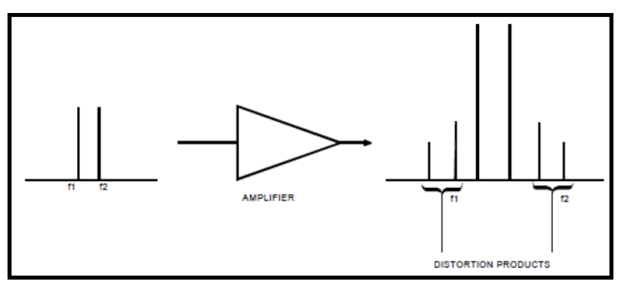

Fig. 1 - When ≥2 signals are amplified, distortion products appear in the vicinity of the desired signals

Distortion is also produced by phase non-linearity. 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 [1].

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

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

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

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

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

the resulting spectrum resembles that of a sinusoidal modulated PM signal

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

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 non-linearity produces IMD products in a similar fashion to amplitude non-linearity. In some systems phase non-linearity is the principal cause of distortion.

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:

C/I = CNR, the CNR degrades by approx. 3 dB
C/I = CNR+6 dB, the CNR degrades by approx. 1 dB
C/I = CNR+10 dB, the CNR degrades by approx. 0.05 dB

Thus if the IMD products are to have a negligible effect on system performance, they should be on the order of 10 dB smaller than the noise level. 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 equation (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.

$$BW = 2(\Delta f+f_m) \qquad (5)$$

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 equation (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, which causes the SR when a digital signal is passed through a non-linear amplifier. The distortion of the induced amplitude waveform produces IMD products, which increase a 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 ratio (ACPR).

$$ACPR = \Sigma IMDs \big|_{\text{in an adjacent channel}}$$

The ratio of the adjacent channel power to the carrier power is known as the adjacent channel power ratio or ACPR.

ACPR 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 carrier (C/I ratio) by anywhere from 35 to greater than 65 dB, depending on the application.

III. Saturated Power and Efficiency

The concept of maximum or saturated (SAT) power is of great importance. All amplifiers have some maximum output power capacity - see Figure 2. Driving an HPA 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 and its output level will increase by a smaller amount, for a fixed increase in input signal. The closer an HPA is driven to SAT, the greater the amount of distortion it produces.

The SAT point of TWTA is clearly defined as the output power normally decreases beyond SAT. SSPAs tend to approach saturation exponentially, which make SAT more difficult to define. Often the point at which an HPA compresses by X dB is used as the reference (REF) for SAT. For example X = 3 dB. Sometimes the power at which an HPA's gain compresses by 1 dB is used for the REF.

$$\text{REF} = 1 \text{ dB CP} = \text{SAT} - \text{D} \qquad (6)$$

For HPAs with reasonable linearity, the difference (D) in output level between SAT and the 1 dB compression point (CP) is about 1 dB. Unfortunately D varies from amplifier to amplifier. HPAs with high linearity will have a smaller difference (D < .25 dB), while amplifiers with poor linearity can have a difference of several dB (D > 3 dB). The new GaN HPAs can have a D > 5 dB. (For comparison purposes in this paper output power backoff (OPBO) will be relative to an HPA's single carrier SAT power).

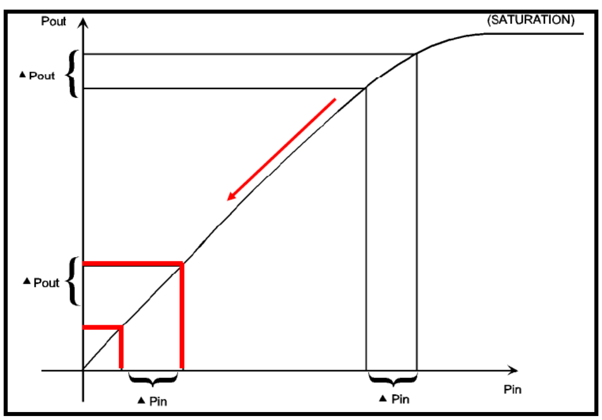

Fig. 2 - As an HPA is driven closer to SAT, its output level will increase by a smaller amount.

Linearization is a systematic procedure for reducing an amplifier's distortion. There are many different ways of linearizing an HPA, but in all cases the end goal is an HPA with a constant gain and phase at all power levels up to to SAT as shown in Figure 3. Obviously once SAT is reached, the gain cannot remain constant. If the HPA's output power remains constant, then the gain must decrease at 1 dB/dB. In a so called ideal amplifier, the 1 dB CP occurs at 1 dB in input power beyond SAT.

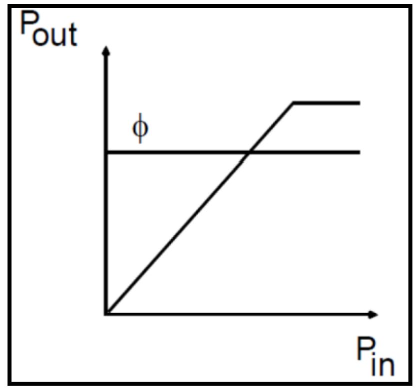

Fig. 3 - *Ideal* HPA characteristics – want constant gain and phase to SAT

The *ideal* HPA sets a performance bound on linearity that cannot be exceed by a *real* HPA. The performance of an *ideal* HPA for two common measures of linearity, 2-tone C/I and noise power ratio (NPR) – distortion produced by an infinite number of carriers (of random amplitude and phase - noise) is also show in Figure 3. For example, the graph shows that you can get virtually any 2-tone C/I required for an OPBO greater than 3 dB, but for an OPBO = 2 dB, the best you can achieve is about 2.1 dB. Figure 10 shows the 2-tone C/I achievable by an ideal transfer characteristic. The C/I goes to infinity for OPBO greater than 3 dB. This result occurs because the peak-envelope-power (PEP) of a 2-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 subject to only a linear response. However to achieve 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:

$$PEP = NP_{av} \qquad (7)$$

where N is the number of carriers and P_{av} is the average power of the overall signal. For 4 carriers the OPBO for no IMD increases to 6 dB. The same limitation applies single carrier signals with a high peak-to-average modulation such as WCDMA.

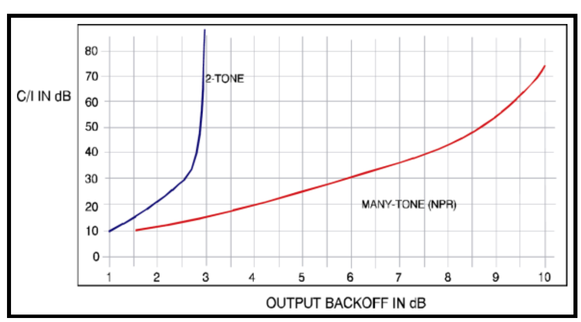

Fig. 3 - 2-tone C/I and NPR achievable with an ideal HPA

IV. LINEARIZATION

It should be clear that linearization does not directly improve efficiency. Linearization does allow a HPA to operate closer to saturation for a given level of distortion, and since generally the closer a HPA is operated to saturation to higher is its efficiency, linearization indirectly improves efficiency. It often enables a HPA to double or even triple its efficiency. Numerous amplifier designs have been developed that allow HPA to achieve better efficiency higher OPBO. Doherty and Envelope Tracking (where the drain voltage tracks the signal envelope) amplifiers are two examples of HPAs that can maintain efficiency at higher backoff. Both these HPAs require linearization for linear operation, and have benefited from the availability of improved linearization techniques.

The three common forms of linearization are Feed-forward, Feedback and Predistortion. All three of these forms of linearization can be implemented by both analog and digital techniques. Besides these there are a variety of approaches that are often associated with linearization, but are really unique amplifier architectures designed to improve HPA efficiency. Most of these approaches require linearization to achieve satisfactory linearity.

Feed-Forward Linearization

Feed-forward (FF) has been around for a long time and has been extensively used with HPAs, but is rather complex to implement. 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.

$$S_{out1} = GS_{in}\angle\Phi_{amp}+IMD \qquad (8)$$

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 respectively

$$SS_{in} = K_0S_{in} \text{ and } SS_{out1} = K_1S_{out1}$$

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

$$A_0SS_{in}\angle\Phi_0 = -SS_{out1} \quad \text{or}$$

$$A_0K_0S_{in}\angle\Phi_0 = GK_1S_{in}\angle(\Phi_{amp}+180°) \qquad (9)$$

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 produce *ideally* 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

$$S_{cor} = A_1GAK_1K_2IMD\angle(\Phi_{aux}+\Phi_1) = IMD\angle(\Phi_m+180°) \quad (10)$$

then the HPA output will be distortion free. A1 and Φ1 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.

$$S_{out2} = S_{out1}\angle\Phi_m+S_{cor} \qquad (11)$$

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, Sout1, is reduced in amplitude by a factor (R_1) due to passing through the K_1 coupler. In dB

$$R_1 = 10Log(1\text{-}10^{-(k1/10)})+L_1 \qquad (12)$$

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 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 equation (11) with 2 substituted for 1 in the variable names. The overall loss in saturated power ΔSAT is

$$10Log(1\text{-}10^{-(k1/10)})+10Log(1\text{-}10^{-(k2/10)})+L_1+L_2+L_m \qquad (13)$$

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 saturated power of an amplifier with single and multi-carrier signals. This factor can vary from about 0.5 to >1.5 dB for HPAs. Furthermore, the amplifier's *true* SAT power should not be considered, only that of the main amplifier. A FF amplifier combines both the power of the main and the auxiliary (aux) amplifier. The sum of the saturated power of both these amplifiers should be considered when comparing the relative OPBO performance of different methods of linearization.

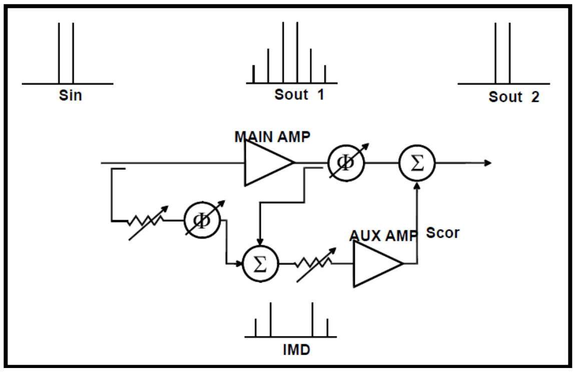

Fig. 4 - Feed-forward linearization employs 2 loops for the cancellation of IMD

Practical considerations limit the size of the aux amplifier. This limits S_{cor} and in turn the undistorted FF output level. The smaller 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 and thus introduce 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 > 20 dB 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 GaAs FET SSPA were assumed for both amplifiers, and resistive losses of 1 dB were assumed for the passive output components. Figure 5 shows that 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. If only the saturated power 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. In practice other factors limit IMD reduction and perfect cancellation can never be achieved. Figures 5 reveals why FF is not a good choice for linearization of amplifiers near SAT. Other linearization methods can provide comparable IMD cancellation with better efficiency and considerably less complexity. Even for OPBOs greater than ~ 6-7 dB, where FF becomes competitive, and for high linearity, it has lost in popularity to less complex and more efficient methods of linearization.

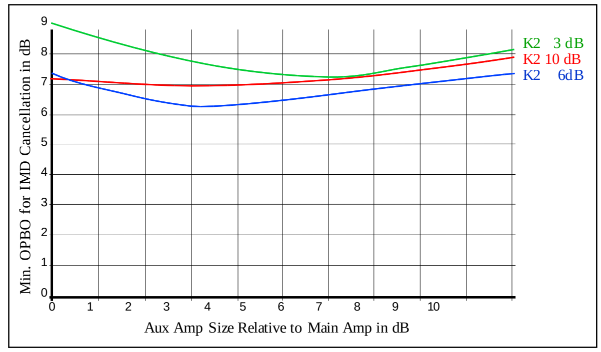

Fig. 5 - The minimum OPBO for cancellation of IMD by a FF HPA depends on aux amplifier size and output coupler coefficient

Feedback Linearization:

HPA Feedback (FB) linearization is not considered enough. FB 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 due to concerns with HPA stability and the difficulty in making networks with non-ideal components function over wide frequency bands. Indirect feedback (IFB) techniques have been more widely applied. In this approach an HPA's input and output signals are detected and lowpass filtered, and the resulting *baseband* signals compared. The error signal (Ve) is used to modify the amplifier's characteristics so as to minimize distortion.

$$Ve = |DSout-DSin| \quad (14)$$

where DSout and DSin are respectively the detected output and input signals. Ve can be used to control the gain of the amplifier by means of a voltage variable attenuator. 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 respectively control an attenuator and a phase shifter so as to minimize signal error. Figure 6. IFB compares an amplifier's output and input, and uses the *detected* difference to minimize distortion.

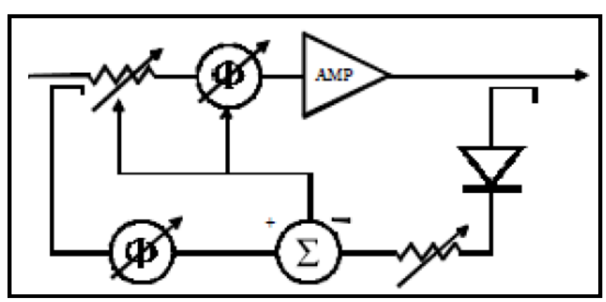

Fig. 6 - IFB compares an amplifier's output and input, and uses the *detected* difference to minimize distortion.

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 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.

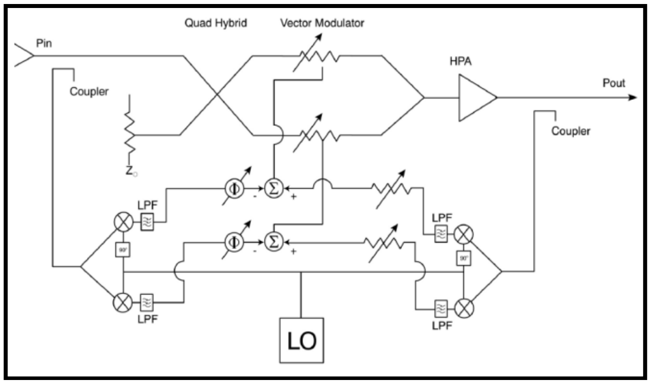

Fig. 7 - Cartesian feedback eliminates the need for phase correction components by using the difference between in-phase and quadrature signals to control attenuators in a vector modulator

Cartesian feedback is most often used with digital 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*. Often the correction at baseband is done in the digital domain using digital signal processing (DSP) techniques.

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. It is "practically" difficult to make a feedback system responds 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

$$BW < 1/(4\Delta t_S) \qquad (15)$$

for significant correction. Thus the total delay must be less than 25 ns for a 10 MHz BW. Microwave HPAs can have delays of 10 to 20 ns. An advantage of Cartesian feedback is that the BWs 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 linearizers (PDLs) have been used extensively in microwave because of their relative simplicity. Today the majority of PDLs are implemented digitally for telecom applications. If an HPA is to be integrated with a digital modulator, and the bandwidth of the signal (BWS) is narrow enough to allow the PDL to be implement in the modulator, then linearization can be added for no or minimal additional cost (the added firmware). The deciding factors for employing linearization are the availability of the modulator and the BWS. For larger bandwidths, both the cost of digital processing and its power overhead (power consumed by digital components) increase sharply. For a C/I reduction > ~ 20 dB, the modulator must be capable of processing over a bandwidth (BWP) several times BWS, typically:

$$BWP \geq 5\sim7\ BWS \qquad (16)$$

For BWS > 25 ~ 50 MHz, analog PDL becomes more cost effective. Another complication with digital PDL is that HPA nonlinearity can change with frequency. For effective linearization over wide bandwidths, a PDL must change its nonlinear characteristics with frequency to match those of the HPA. Digital PDL cannot easily make these changes without going to very complex processing that further limits its capacity to handle higher frequencies. For narrow band signal, it can partially compensate for this degradation by adaptively changing its correction characteristics. The ability of analog PDL to modify its characteristics over even a multi-GHz frequency band has been demonstrated numerous times. However, since analog PDL has an associated cost (hardware must be added), the question of its application depends more on the benefits it can provide, and as discussed on the nonlinearity of the HPA.

PDLs generate a non-linear transfer characteristic that can be thought of as the reverse of the amplifier's transfer characteristics in both magnitude and phase as seen in Figure 8.

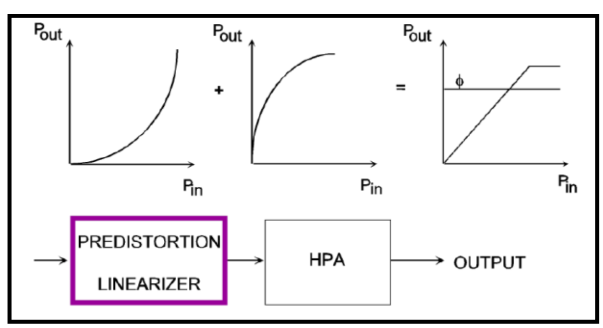

Fig. 8 - PDLs generate a response opposite to an HPA's response in magnitude and phase

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 PDL are made equal in amplitude and 180 degrees out phase with the IMDs generated by the HPA (at its output), the IMDs will cancel. As already discussed, 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.

$$GL(P_{outL})-GL_{ss} = -[GA(P_{inA})-GA_{ss}]\,\big|\,P_{outL}=P_{inA} \quad (17)$$

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

$$\Phi L(P_{outL})-\Phi L_{ss} = -[\Phi A(P_{inA})-\Phi A_{ss}]\,\big|\,P_{outL} = P_{inA} \quad (18)$$

Some improvement is possible even at SAT and beyond as the linearizer can correct for post-saturation phase distortion and power slump – but this improvement is small. Since the power out of the HPA (in dB) is

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

Equations (17) and (18) can be rewritten referenced to the power into the linearizer (P_{inL}), and the desired transfer characteristics of the linearizer expressed as follows:

$$GL(P_{inL}) = GL_{ss}+GA_{ss}-GA(P_{inL}+GL(P_{inL})) \quad (19)$$

$$\Phi L(P_{inL})=\Phi L_{ss}+\Phi A_{ss}-\Phi A(P_{inL}+GL(P_{inL})) \quad (20)$$

Equations (19) and (20) can be solved iteratively for the ideal linearizer response needed to correct a given amplifiers transfer response. Figure 9 shows the response needed to ideally correct a typical HPA. As SAT is approached the rate of gain and phase change become infinite.

$$dGL/dP_{in} = \infty \;\text{ and }\; d\Phi L/dP_{in} = \infty \;\text{ as } P_{out} \rightarrow \text{Sat}$$

For analog PDL, such a characteristic cannot be achieved in practice. Often a small amount of gain *expansion* near saturation, 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 PDL is the dependence of some amplifier's transfer characteristic's on the frequency content of the signal. This phenomenon is sometimes referred to as *memory effects*. The change in nonlinearity of an amplifier is considered one form of memory effects. Great care must be taken in the design of an HPA to minimize these effects, if the maximum benefit of PD linearization is to be achieved.

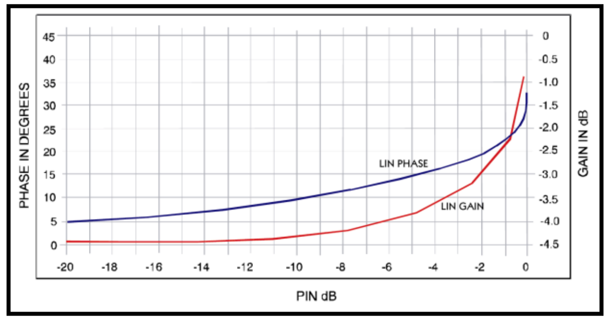

Fig. 9 - An ideal PDL response requires the gain and phase slope to become infinite as SAT is approached

Memory Effects

Memory effects are the dependence of the non-linear behavior of an HPA not only on the *present* amplitude of the signal, but also on its past values.

$$Vo = f(Vin, time) \quad (21)$$

Memory effects can be caused by a variety of problems. Some of the most important are a) frequency memory effects due to change in an HPA's nonlinear characteristics with frequency; b) bias memory effects due to changes in a device's bias (at both the drain and gate of FETs) caused by changes in the envelope of a signal; and c) temperature memory effects caused by the change in an HPA's nonlinear characteristics with temperature. With the digital PDL, often the solution to memory effects is the application of complex signal processing and multi dimensional lookup tables that weight both the present and recent past events. For example, frequency memory effects, illustrated in Figure 10, can be compensated using pre- and de-emphasis linear filtering before and after predistortion. With analog predistortion, the linearizer's nonlinear characteristics can usually be tailored to match the HPA's nonlinearity over frequency. TWTAs, for example, generally need a greater phase change at the high end of their operating band than the low end.

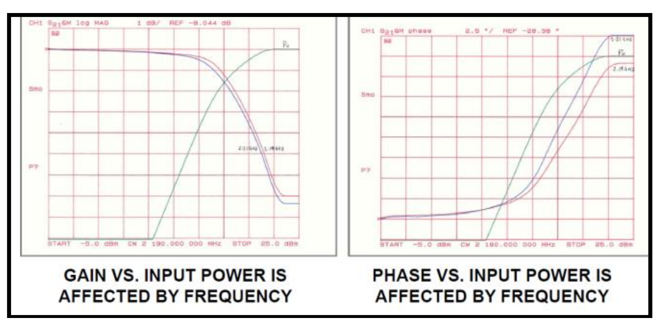

Fig. 10 – Real HPA nonlinearity changes with frequency

A better solution that is often overlooked is to minimize the memory effects in the design of the HPA. For example, it is possible to reduce frequency memory effects when designing the matching networks of an HPA.

The biggest contributor to the nonlinearity of memory effects is most often the modulation in the drain and gate bias voltages of the HPA due to a signal's changing envelope. A simplified circuit diagram of a solid state HPA is shown in Figure 11. The measured variation in drain voltage (ripple) due to the envelope of a 2-tone test signal is shown in Figure 12. As the envelop increases, more drain current is drawn, and the resulting voltage drop across the impedance of the inductor (decoupling circuitry use to isolate the RF path from the dc supply) causes the drain voltage to decrease. This ripple effectively *amplitude modulates* (AM/AM) the signal to produce sidebands at the same frequencies as the IMD.

$$SB = (f1+f2)/2 \pm 2\Delta f \qquad (22)$$

where SB is the frequency of the modulation sidebands, f1 and f2 are the carrier frequencies and Δf is the carrier spacing. If the voltage drop was purely resistive, then the ripple would be 180° out of phase with the envelop, and the SBs would subtract from the IMDs; in effect *linearizing* the HPA. In the example the SBs are at 120° and only partially subtract, but the situation is more complex. The changing voltage also causes phase modulation (AM/PM) as result of change to device junction capacitance.

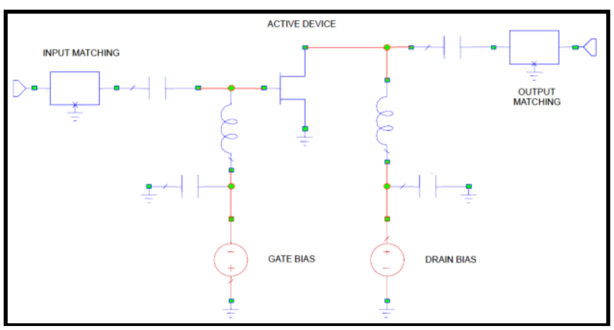

Fig. 11 – Basic solid-state power amplifier

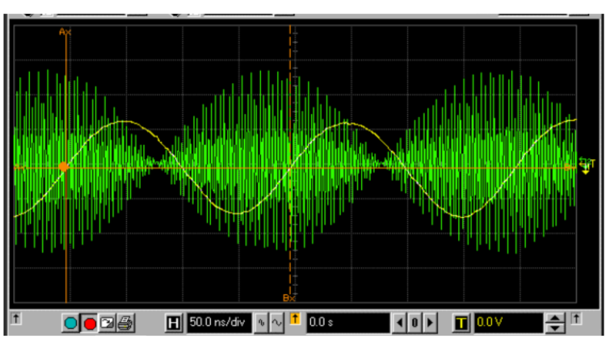

Fig. 12 – RF envelope (green) is ~ 140° out of phase with the dc voltage at the drain (ripple in yellow)

The resulting modulation can be written as

$$V_o(t) = \overbrace{\left[\frac{1}{2} + \frac{A}{2}\cos\,\omega_m t\right]}^{\text{AM}} \cdot \cos\left[\omega_c t + \overbrace{\varphi\cos(\omega_m t + \theta)}^{\text{PM}}\right] \qquad (23)$$

Where the AM and PM SBs are

$$V_{AM}(t) = \frac{1}{2}\cos(\omega_c t) + \frac{A}{4}\left[\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t\right]$$

$$V_{PM}(t) = J_0(\varphi)\cos(\omega_c t) - J_1(\varphi)\sin((\omega_c - \omega_m)t - \theta) - J_1(\varphi)\sin((\omega_c + \omega_m)t + \theta)$$

Depending on $\boldsymbol{\Theta}$, the AM and PM upper and lower SBs can add non-symmetrically. Values of $\boldsymbol{\Theta}$ other than 0 and π will generate asymmetric IMDs, which cannot be completely canceled by standard PDL.

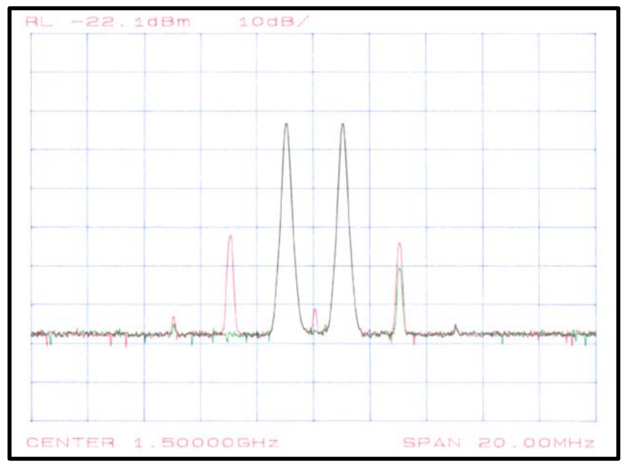

Fig. 13 – A memoryless predistortion linearizer cannot completely cancel asymmetric IMD.

Bias modulation can be minimized by proper design of the bias circuitry. Biasing networks should be used with minimum reactance at the envelope frequency, while still maintaining a high impedance at the RF band. This goal is not always easily achieved. A worst case scenario will occur for high power, low drain voltage HPAs, operating at low frequencies with very wide signal bandwidth (high envelope frequency). New GaN HPAs are easier to design for lower drain bias modulation because of the higher voltage and consequently lower current required for the same power level as a GaAs HPA. An improved bias network design is shown in Figure 14.

Variations in the envelope of a signal can also produce rapid changes in temperature in an HPA's power devices. If the envelope frequency is high (>100 kHz, but sometimes much higher), thermal inertia will cause the device's temperature to be constant. If the envelope frequency is low, the temperature of the active device will vary as a function of the envelope, and the change in the temperature affect its nonlinear characteristic. The best design solution is to be aware of the power devices' thermal characteristics and select devices with a large thermal time constant.

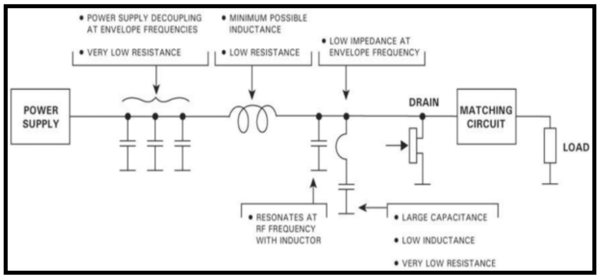

Fig. 14 – Improved drain biasing network in which the inductance has been reduced to minimize memory effects, and the envelope is terminated in a short circuit

V. EXTENDING THE LIMITS ON LINEARIZATION

Recent development work has focused on making linearizers operate over very wide bandwidths and on increasing the highest frequency where linearization has been applied. For wideband applications a PDL's gain and phase transfer characteristics must change over frequency as well as with input level to match the changes of the HPA's nonlinearity at different frequencies. In addition, small signal (linear) gain and phase (time delay) must be maintained over the frequency range of interest for the linearizer to properly function. If an HPA's small signal gain changes with frequency, the gain correction required from the PDL will be shifted and the improvement in linearity degraded. Consequently HPA gain ripple can be a major problem. It can limit the level of correction achievable even with an ideal linearizer. The relationship between gain ripple and degradation in IMD was investigated by simulating a typical linearizer-HPA system. The simulation showed that if the gain ripple can be kept below ±2 dB, more than a 5 dB improvement in 2-tone carrier-to-intermodulation ratio (C/I) can be achieved over backoff and frequency. Figure 15 shows the degradation in 2-tone C/I caused by HPA ripple. With ±2 dB ripple at a 30 dB C/I, the reduction can be almost 3 dB and increases to about 7.5 dB with a ±6 dB ripple.

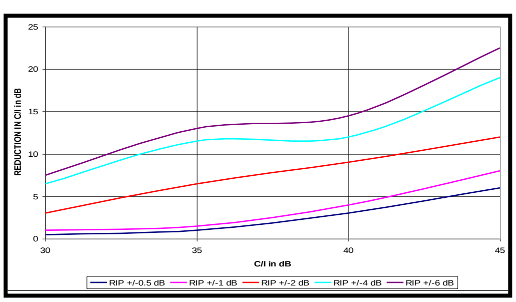

Fig. 15 Worst case degradation caused by gain ripple on the C/I of a linearized HPA

Over narrow bandwidths, phase delay is normally not a problem, but with wideband linearization systems care must be taken to insure that distortion correction products are not significantly shifted in phase relative to the carriers that generated them. Ripple equalization is an ideal application for digital filtering. The problem is that for the bandwidths of interest digital processing becomes prohibitively expensive and power consumptive.

Analog predistortors have been demonstrated that can produce controllable tailored nonlinear responses over multi-GHz bandwidths. Figure 16 shows the response of a wideband predistortor that displays with input power level the increasing gain and decreasing phase typically needed to correct a GaN HPA over a 2 to 20 GHz) frequency range.

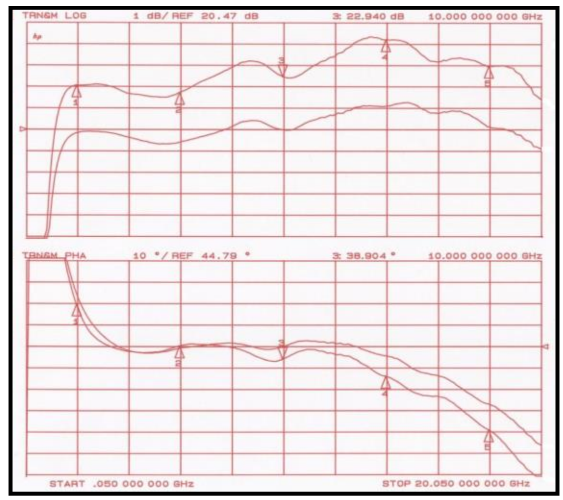

Fig. 16 - Wideband predistortor with useful 2 to 20 GHz frequency response

The lower curve in the top graph and the upper curve in the bottom graph are respectively the small signal gain and phase frequency responses. As the PD is driven with greater input power, these responses move apart to produce the increasing gain change and the deceasing phase change needed to correct the HPA's distortion. Similar responses have been produce up into millimeter-wave frequencies. For most linear HPA applications only odd order IMD products need be considered. For HPAs of an octave or greater bandwidth, both even and odd order distortion products and both IMD and harmonic distortion are present. Thus the linearizer must correct (or suppress) both to achieve linear performance. To date linearized HPAs have been produced with more than 10 GHz of bandwidth and that operate over 2.5 octaves of bandwidth. Efforts are presently underway to extend these limits.

Work is also taking place to extend the upper frequency limit on linearization. In this regard, it should be noted that PDL can be performed at a lower frequency including digitally at the modulator and mixed up to virtually any frequency. There are, however, practical limitations on such systems. Generally the principal reason for moving to a higher frequency is bandwidth. The lower in frequency the predistortion is generated, the larger is the percentage bandwidth of the predistortor and the more difficult it will be to achieve the desired performance over wideband. In addition, it is desirable to locate the predistortor as close to the HPA as possible to minimize variation in gain with frequency (ripple) between the PDL and the HPA. Consequently, analog PDLs are under development for the upper millimeter-wave bands. A demonstration PDL has been developed for 96 GHz downlink satellite use – see Figure 17.

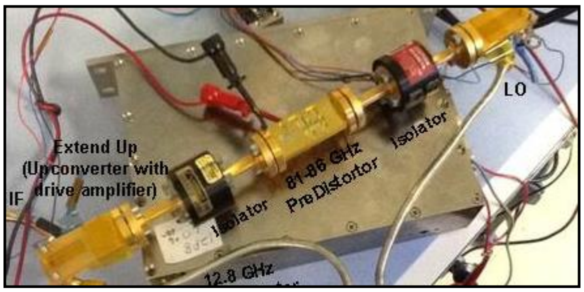

Fig. 17 –W-band (92 – 96 GHz) linearizer in test

It was tested with a TWTA operating in the 92 to 96 GHz range. The C/I of the linearized TWTA is shown in Figure 18. For a C/I of 25 dB, linearization increased the available TWTA output power by > 5 dB. PDLs have been produced for Q, V, E and W-band and the possibility G band is being considered.

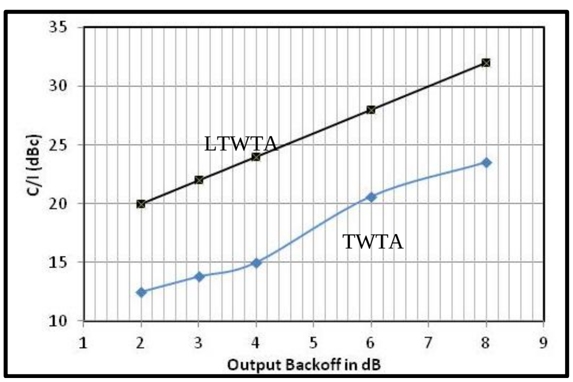

Fig. 19 – C/I of LTWTA at 94 GHz

VI. Summary

This paper reviews the basics behind the many HPA linearizers in use today and considered the trades involved in the decision to linearize. Digital PDL is widely applied when HPAs are used with a modulator into which the linearization can be incorporated. It is necessary that the signal bandwidth be narrow enough to economically allow linearization using digital signal processing. For wider bandwidth signals, analog PDL is often preferred depending on the nonlinearity of the HPA, the output power required and the allowed level of distortion. GaN HPAs because of their higher nonlinearity and wide bandwidth capability are a good candidate for analog PDL. The application of FF and FB linearization was also considered and shown to be limited relative to PDL. The impact of memory effects on linearization was considered and the importance of minimizing their effect in an HPA's design stressed. The progress being made to produce linearizers of very wide bandwidth and to extend linearization to the upper millimeter-wave bands was also covered. Multi-octave linearizers with more than 10 GHz of bandwidth and linearizers that operate to 100 GHz have been produced.

References

[1] A. Katz, "Linearization: Reducing Distortion in Power Amplifiers," IEEE Microwave Mag., pp. 37-49, Dec. 2001.

[2] S. Cripps, "Advanced Techniques in RF Power Amplifier Design", Artech House, 2002.

[3] Vuolevi and Rahkonen, "Distortion in RF Power Amplifiers", Artech House, 2003.

[4] A. Katz, and et al, "Wide/Multi-band Linearization of TWTAs Using Predistortion," IEEE Trans. on Electron Devices, Vol. 56, pp. 959-964, May, 2009.

[5] D. Goebel, et al, "Development of Linear Traveling Wave Tubes for Telecommunications Applications," *IEEE Transactions on Electron Devices*, Vol. 48, No. 1, pp. 74-81, Jan 2001.

[6] M. Franco, and et al, "Intermodulation Distortion Products in Radio Frequency Power Amplifiers with Digital Predistortion Linearization," MTT-S International Microwave Sym. Digest, San Francisco, CA, June 11-16, 2006.

[7] W. Bosch, and G. Gatti, "Measurement and Simulation of Memory Effects in Predistortion Linearizers", Transactions on Microwave Theory and Techniques, Vol. 37, Dec., 1989.

[8] A. Katz and R. Dorval, "Evaluation and Correction of Time Dependent Amplifier Non-Linearity," IEEE MTT-S International Microwave Sym. Digest, pp. 839-842, June 1996.

[9] M. Franco, and et al, "Minimization of Bias-Induced Memory Effects in UHF Radio Frequency High Power Amplifiers with Broadband Signals," 2007 IEEE Radio and Wireless Sym., Long Beach, CA, pp. 369-372, Jan. 9-11, 2007.

[10] H. Ku, and J. Kenney, "Behavioral Modeling of RF Power Amplifiers Considering IMD and Spectral Regrowth Asymmetries", IEEE MTT-S, 2003.

[11] S. Boumaiza, F. Ghannouchi, "Thermal Memory Effects Modeling and Compensation in RF Power Amplifiers and Predistortion Linearizers", IEEE MTT, December 2003