A DSP CONTROLLED ADAPTIVE FEEDFORWARD AMPLIFIER LINEARIZER

by

Stephen James Grant

B.A.Sc., University of British Columbia, 1993

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in the School

of

Engineering Science

SIMON FRASER UNIVERSITY

July 1996

APPROVAL

NAME: Stephen James Grant

DEGREE: Master of Applied Science

TITLE OF THESIS: A DSP Controlled Adaptive Feedforward Amplifier Linearizer

Examining Committee:

Chair: Dr. John Jones

_____

Dr. James Cavers
Senior Supervisor

_____

Dr. Paul Goud
Senior Supervisor

_____

Dr. Paul Ho
Supervisor

_____

Dr. Shawn Stapleton
Examiner

Date Approved: _____

ABSTRACT

Currently, the only available wideband (multiple MHz), accurate, amplifier linearization method is feedforward. Feedforward, though, requires automatic adaptation of key parameters for reliable distortion cancellation as operational and environmental conditions vary. In this thesis, previous analysis on adaptive feedforward linearization is extended to include an alternative placement of the adaptive signal cancellation coefficient. In contrast to previous analysis, this placement results in a non-quadratic error surface. Consequently, two available criteria for the optimization of the signal cancellation coefficient result in different optimal values. This result can have practical implementation consequences under certain operating conditions. A new analysis is presented that shows that various inaccuracies in the implementation of baseband correlation, such as frequency and phase offsets, filter mismatches, and incomplete image suppression, do not affect the final converged coefficient values. With a novel and appropriate use of DSP, a feedforward linearizer has been implemented with adaptation driven by easily computed gradient signals. This overcomes the difficulties, such as DC offsets at the output of analog mixers and masking of weak signals by stronger ones, that slow and/or cause incorrect convergence of many previously reported implementations. The result is 40 dB reduction of intermodulation spectra over a bandwidth of 7 MHz. Coefficient convergence occurs within 50 msec of start-up, and following a 6 dB change in input power, reconvergence occurs in 3 msec with no loss in distortion suppression.

ACKNOWLEDGEMENTS

Nothing is ever done alone; thus, I would like to thank certain people who helped me along the way with this project. Thanks to my Senior Supervisors, Jim Cavers and Paul Goud, who, I feel, were responsible for advancing my knowledge and confidence in the area of wireless communications to a new and hopefully much higher level than before. Deserving of thanks as well, is my wife, Irma, for her constant support the whole way through the project, from start to finish.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

All wireless radio transmitters contain RF amplifiers which are nonlinear to some degree. The primary consequence of amplifier nonlinearities is the generation of intermodulation distortion (IMD) if the signal to be amplified has a non-constant envelope, such as for linear or multicarrier modulation formats. Not only does IMD corrupt the amplified signal itself, but more seriously it causes adjacent channel interference due to spectral regrowth. In interference limited systems such as cellular radio, strict limits are usually placed upon allowable intermodulation power in adjacent channels; consequently, some form of amplifier linearization is usually required. Several linearization techniques employed to combat IMD are feedback, predistortion, and feedforward. Adaptive feedforward is the scheme studied in this thesis.

1.1 Characterization of Amplifier Nonlinearities

Nonlinear RF amplifiers are characterized by measurement of their AM/AM (amplitude dependent gain) and AM/PM (amplitude dependent phase shift) characteristics. These measurements may be performed using a network analyzer in power sweep mode. Specifically, the gain and phase of the amplifier are measured at a single frequency as the input power level is varied. Not only are RF amplifiers nonlinear, but they also possess memory: the output signal depends on the current value of the input signal as well as previous values spanning the memory of the amplifier. If the reciprocal of the bandwidth of the input signal is much larger than the memory of the amplifier, as is the case for most RF amplifiers driven with narrowband signals, the amplifier can be modeled as memoryless

for that particular input signal. Thus, for complex baseband analytical and simulation purposes, the AM/AM and AM/PM measurements can be summarized in a single frequency-independent memoryless function, namely complex voltage gain

$$G(x) = g(x)e^{j\phi(x)} \tag{1}$$

The magnitude and phase of $G(x)$ are simply the measured gain and phase of the amplifier as functions of x—the instantaneous power in the input signal. For wideband signals, the memory of the amplifier becomes a significant fraction of the reciprocal of the bandwidth of the input signal; thus, it must be considered for accurate analytical and simulation studies. In this case, a single frequency-independent function is not sufficient to model the amplifier nonlinearity, and more powerful modeling techniques must be used such as a Volterra series approach. For the analysis presented in this thesis, though, the power amplifier is assumed to be memoryless.

Figure 1 shows the characteristics of a typical class AB RF power amplifier measured at a single frequency within the passband of the amplifier [1]. Note that the gain and input power are normalized such that saturation occurs at unity output power for unity input power. Consequently, normalized input power also represents input backoff from saturation. For example, 6 dB backoff corresponds to an input power of 0.25.

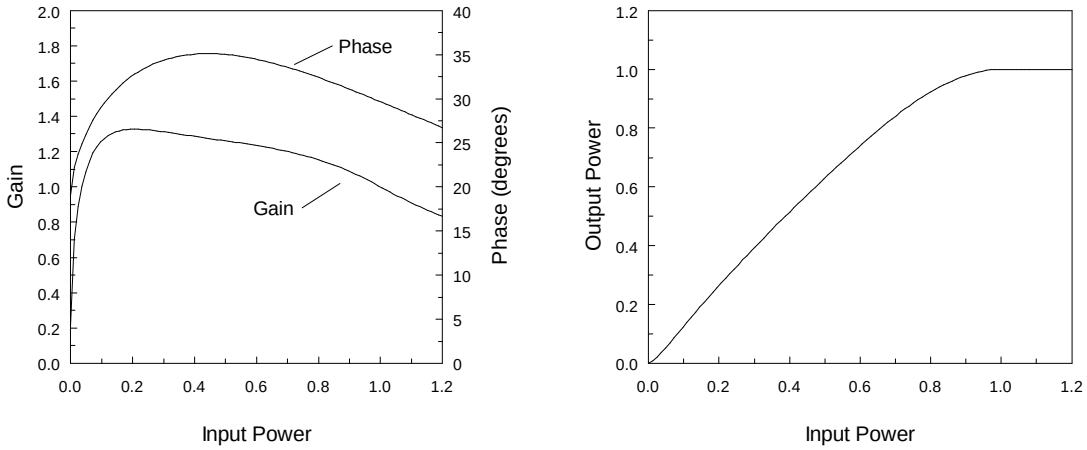


Figure 1. Measured characteristics of a typical class AB power amplifier

Evident in Figure 1 is a strong amplitude dependence of the amplifier gain and phase. Because the transistors in the amplifier are biased in class AB rather than class A, they cut off at low signal voltages. This causes the gain to fall off at low input power. The gain also rolls off at high input power due to saturation of the transistors. The variation in phase with input power is due to voltage-dependent device capacitances.

Figure 2 shows simulated power spectra of a narrowband $\pi/4$-DQPSK signal before and after amplification with the amplifier whose nonlinear characteristics are shown in Figure 1. For simulation purposes, $G(x)$ is represented by polynomials fitted to the measured gain and phase curves over the range of input powers extending from 0 to 1 (saturation); $G(x)$ is then extrapolated further into saturation. If $v_m(t)$ is the complex envelope of the amplifier input signal, then the complex envelope of the distorted amplifier output signal is given by $v_a(t) = v_m(t)G[|v_m(t)|^2]$. 35% rolloff root raised cosine filtering is used for the simulation which results in a peak-to-average power ratio of approximately 2.5 dB; input backoff is 3 dB. Note that frequency has been normalized by the symbol rate. As can be seen, the amplifier nonlinearity causes significant spectral regrowth.

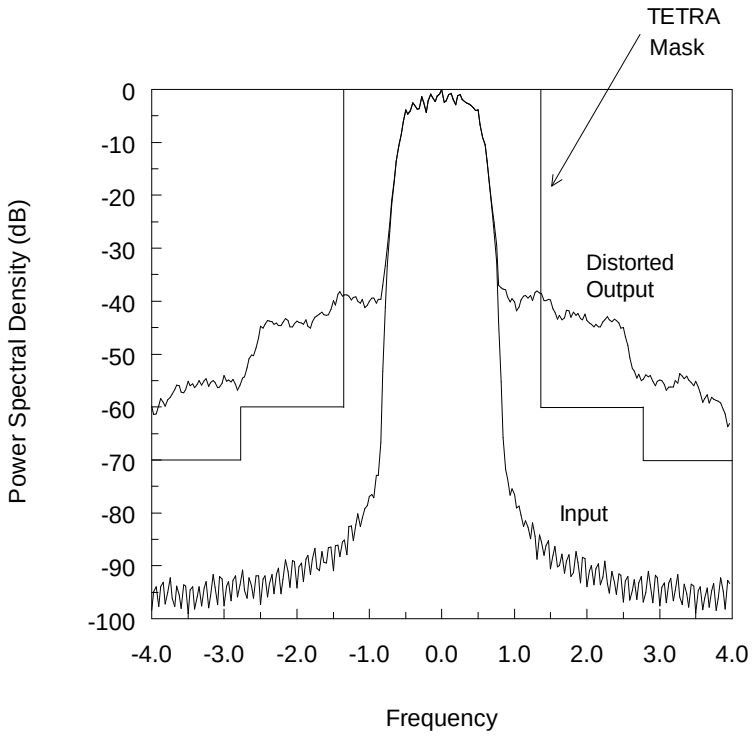


Figure 2. Simulated input and output power spectra for class AB amplifier with a π/4-DQPSK input signal

Regulatory bodies usually specify power spectral density (PSD) masks which define maximum allowable adjacent channel interference (ACI) levels. TETRA [2], for example, uses a π/4-DQPSK modulation format with 35% root raised cosine filtering at a symbol rate of 18 kHz; the channel spacing is 25 kHz. Adjacent channel protection is specified as 60 dB at 25 kHz, and 70 dB at 50 and 75 kHz. The corresponding spectrum mask is shown in Figure 2. Clearly, TETRA limits are exceeded; consequently, some form of linearization for this amplifier is required in order to conform to this standard.

1.2 Overview of Linearization Strategies

Perhaps the simplest method of achieving linear RF amplification without the use of additional hardware is output backoff of an existing class A amplifier such that it

operates completely in its linear region. Typically though, the backoff required to achieve linear operation is high (25 to 30 dB), and the resulting power efficiency is very low (1 to 2%). Also, for a fixed output power requirement, the cost of building an amplifier with output power rating 25 to 30 dB higher than necessary can be high. Moreover, the heat dissipation from the higher power amplifier may become a problem.

Several other techniques to achieve linear RF amplification have been developed. The most popular are feedback, predistortion, and feedforward, all of which make use of additional hardware. Making a choice of which linearization strategy to employ for a particular application involves tradeoffs of complexity, degree of IMD suppression, and bandwidth. The most prominent feedback scheme, namely Cartesian coordinate modulation feedback [3], has relatively low complexity, offers reasonable IMD suppression, but stability considerations typically limit the bandwidth to a few hundred kHz. Digital implementations of predistortion [4,5] have higher complexity than feedback, offer better IMD suppression, but again, possible bandwidths are low (up to a few tens of kHz) due to limited DSP computation rates. Reported implementations of analog predistortion [6], although potentially wideband, have modest complexity, but suffer from limited IMD suppression. In contrast to the above linearization techniques, feedforward [7,8,9] is currently the only linearization strategy that simultaneously offers wide bandwidth and good IMD suppression; the cost is relatively high complexity. Automatic adaptation of key parameters, though, as discussed in the next section, is essential for reliable distortion cancellation as operating conditions vary.

Wide bandwidth capability makes feedforward an attractive scheme for several applications. At cellular base stations, rather than using one amplifier per channel

followed by lossy high power combiners, it is more efficient to combine channels at low power and use a single wideband linearized amplifier for the resultant signal. Another potential use of feedforward is for emerging PCS applications, including wideband CDMA, in which the bandwidth requirements place feedback and predistortion out of the league of viable linearization schemes.

1.3 Development of Feedforward Linearization

In 1927, H.S. Black of Bell Telephone Laboratories invented the concept of negative feedback as a method of linearizing amplifiers for the Bell Telephone system [10]. Lesser known is that four years earlier, in the search for a linearization method, he invented the concept of feedforward. His idea for feedforward was simple: reduce the amplifier output to the same level as the input and subtract one from the other to leave only the distortion generated by the amplifier. Amplify the distortion with a separate amplifier and then subtract it from the original amplifier output to leave only a linearly amplified version of the input signal. Black's idea for negative feedback was spawned from his simple feedforward concept: feed an attenuated version of the amplifier output signal back to the input in anti-phase and combine it with the input signal. Use the same amplifier (rather than a separate amplifier as in feedforward) to amplify the difference signal thus producing a linearly amplified version of the input signal. The advantage of the feedback solution, due to the fact that it operated closed-loop, was that it was automatic and required no manual adjustment as operating conditions changed. Its disadvantage, of course, was its potential for instability.

The basic operating principles of a feedforward amplifier as shown in Figure 3 are now described. The feedforward configuration consists of two circuits, the signal cancellation circuit and the error cancellation circuit. The purpose of the signal cancellation circuit is to suppress the reference signal from the main amplifier (or PA) output signal leaving only amplifier distortion, both linear and nonlinear, in the error signal. Linear distortion, in contrast to nonlinear distortion described already, is due simply to deviations of the amplifier's frequency response from flat gain and linear phase [11]. Note that the feedforward technique can also compensate for memory effects, since distortion due to memory in the main amplifier is also included in the error signal and thus ultimately canceled in the linearizer output. The fact that the PA output can be decomposed into two components—a linear term and a distortion term—is discussed in Section 2.

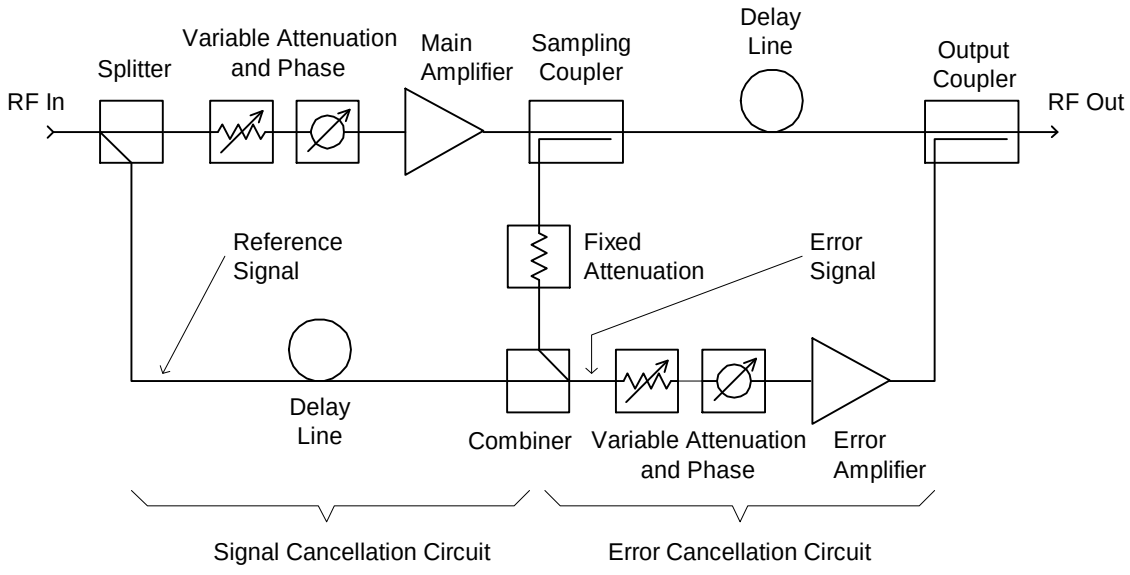


Figure 3. RF circuit configuration of a typical feedforward amplifier

In order to suppress the reference signal, the values of the sampling coupler and fixed attenuation are chosen to match the gain of the main amplifier so that the PA output signal is reduced to approximately the same level as the reference signal. The variable phase shifter ahead of the PA is then adjusted to place the PA output in anti-phase with the reference. The variable attenuation serves the fine tuning function of precisely matching the level of the PA output and the reference. The delay line in the reference branch, necessary for wide bandwidth operation, compensates for the group delay of the main amplifier by time aligning the PA output and reference signals before combining.

The purpose of the error cancellation circuit is to suppress the distortion component of the PA output signal leaving only the linearly amplified component in the linearizer output signal. In order to suppress the error signal, the gain of the error amplifier is chosen to match the sum of the values of the sampling coupler, fixed attenuator, and output coupler so that the error signal is increased to approximately the same level as the distortion component of the PA output signal. The variable phase shifter ahead of the error amplifier is then adjusted to place the error in anti-phase with the PA output. The variable attenuation, again, serves the fine tuning function of precisely matching the level of the error signal and the distortion component of the PA output. The delay line serves the same purpose as in the signal cancellation circuit. The error amplifier must be chosen such that it linearly amplifies the error signal while still providing the required output power, otherwise uncorrectable IMD shows up in the linearizer output. This usually dictates the use of a linear class A amplifier with sufficient backoff. Note that any bandwidth limit, manifested as incomplete distortion suppression, is imposed either by

imperfect delay matching or by linear distortion in the error amplifier, the variable attenuators/phase shifters, or the couplers and combiners.

The crux of the proper operation of the feedforward circuit is the proper adjustment of the attenuation and phase in the signal and error cancellation circuits such that good IMD suppression is maintained over time. Variations of component characteristics with temperature and time as well as changes in operating conditions such as input power level and supply voltage all necessitate readjustment. For these reasons Black, himself, essentially abandoned feedforward in favour of feedback. He found, using vacuum tube amplifiers, that "every hour on the hour—24 hours a day—somebody had to adjust the filament current to its correct value. In doing this, they were permitting plus or minus 1/2- to 1-dB variation in amplifier gain, whereas, for my purpose, the gain had to be absolutely perfect. In addition, every six hours it became necessary to adjust the B battery voltage, because the amplifier gain would be out of hand. There were other complications too, but these were enough!" [10]. Even with modern solid state amplifiers, changes with temperature and time are still significant enough, with respect to the accuracy requirements of the feedforward circuit, to necessitate adaptation.

After its invention in 1923, feedforward was essentially ignored until Seidel, also at Bell Laboratories, investigated the use of feedforward for microwave frequency TWT amplifiers in the late sixties and early seventies [7]. Seidel constructed a feedforward amplifier which achieved distortion suppression of 38 dB over a 20 MHz band. The setup employed an automatic control scheme for the variable attenuation and phase in the error cancellation circuit. The control scheme was based on driving a mechanical attenuator/phase shifter with an error signal derived by comparing the amplitude and phase

at two different points in the error cancellation circuit of a pilot tone inserted after the main amplifier. No control scheme was utilized for the variable attenuation and phase in the signal cancellation circuit. In this way he was able to maintain time independent distortion suppression over a period of several months.

Several patents concerned with adaptive feedforward systems then started to appear in the mid-eighties, and many more appeared in the early nineties. These patents deal with two general methods of adaptation both with or without the use of pilot tones, namely adaptation based on power minimization [12,13] and adaptation based on gradient signals [14,15]. The control scheme for the former attempts to adjust the attenuation and phase in the signal cancellation circuit in such a way to minimize the measured power of the error signal in the frequency band occupied by the reference signal. Minimum power in the error signal is equivalent to suppression of the reference signal. The attenuation and phase in the error cancellation circuit are adjusted in such a way to minimize the measured power of the linearizer output signal in a frequency band occupied only by distortion. Minimum power in the output signal is equivalent to suppression of distortion. Once the optimal parameters are found, deliberate misadjustment is required over time to assess whether or not the respective powers are indeed still minimized. This deliberate perturbation periodically reduces IMD suppression—an undesirable side effect.

Adaptation using gradient signals is based on continually computing estimates of the gradient of a 3-dimensional power surface which depends on two parameters—the variable attenuation and phase in either the signal or error cancellation circuits. The gradient signal is then used to adapt the parameters in each circuit always in a direction towards the global minimum of the surface. The surface for the signal cancellation circuit

is the power in the error signal; the power is minimized when the reference signal is completely suppressed, leaving only the amplifier distortion in the error signal. The surface for the error cancellation circuit is the power in the linearizer output signal; the power is minimized when the distortion is completely suppressed from the PA output signal. The advantage of adaptation based on gradient signals over that based on power minimization, is that since the gradient signals are continually computed, the control scheme constantly searches for the optimum operating point. No algorithm for deliberate misadjustment is required.

Adaptation using either of the above methods plus pilot tones is based on inserting a pilot both at the input to the feedforward linearizer as artificial signal and at the output of the main amplifier as artificial distortion. The control scheme for the signal cancellation circuit either attempts to null the first pilot tone in the error signal if using the power minimization approach, or uses it to derive a gradient signal if using the gradient approach. The same is true for the second pilot tone, except the observation point is the linearizer output. When both pilots are canceled, so is the amplifier distortion. As is the case for other components in a communication system, it is desirable to avoid pilot tones, if possible, and use traffic signals only.

Other than in the patent literature, very little has been published on implementations or analysis of adaptive feedforward linearizers since Seidel's work. Two publications of note on adaptive feedforward linearization, though, are by Cavers [16] and Narahashi and Nojima [17]. Cavers' work is the first published analysis of adaptation behaviour of a feedforward linearizer and is intended as a benchmark and analytical framework for others developing such linearizers.

Narahashi and Nojima report an implementation of an adaptive feedforward linearizer for multicarrier signals with adaptation based on a power minimization technique using pilot tones. A pilot tone is inserted at the output of the main amplifier and its level is measured in the linearizer output signal by means of a narrowband energy detector. A microprocessor is used to adjust the attenuation and phase in the error cancellation circuit using a perturbation technique. Adaptation in the signal cancellation circuit is performed using the same method, except that one of the carriers is used as the pilot signal, rather than an explicit pilot tone. With this setup, it is reported that 30 dB distortion improvement is obtained in a stable manner for a 100 W, 1.5 GHz, GaAs-FET power amplifier.

1.4 Project Goals

Based on an increasing need for a wide bandwidth, adaptive linearization technique and a lack of published work on adaptive feedforward, it was decided to implement, in contrast to [17], a gradient driven adaptive feedforward linearizer without the use of pilot tones. Gradient adaptation was selected because of the advantages it offers over the power minimization technique as discussed above. Based on available equipment, a 5 Watt, 815 MHz, class AB amplifier was chosen as the main amplifier.

A number of patents, e.g. [14], propose a gradient adaptation technique that relies on analog bandpass correlation requiring the mixing of two modulated RF signals. This method, elaborated on later, suffers from accuracy problems such as mixer DC offsets and the generation of additional IMD—both highly undesirable effects. To overcome these problems, the current work demonstrates a novel and appropriate use of DSP to perform

accurate baseband correlation. The first stage of the project demonstrates gradient adaptation for a narrowband (27 kHz) $\pi/4$-DQPSK input signal. Although the feedforward circuit is inherently wideband, adaptation is demonstrated initially for only a narrowband signal. The next stage, a logical extension of the current work but not included in this thesis, will demonstrate the DSP computation of the necessary gradient signals for wideband (multiple MHz) input signals. The concepts involved in this extension are discussed in more detail in a later section.

2. PRINCIPLES OF ADAPTATION

In order to analyze the adaptation behaviour of a feedforward linearizer, it is convenient to work with the complex baseband model shown in Figure 4 in which all RF signals are replaced by their complex envelopes. Note that in this thesis the following convention is used: $v(t)$ is the complex envelope of the bandpass signal $\tilde{v}(t) = \mathrm{Re}[v(t)e^{j2\pi f_c t}]$ centred about the carrier frequency f_c; $x(t) = |v(t)|^2$ is the instantaneous power of $v(t)$; and $P = E[x(t)]$ is the average power of $v(t)$. Subscripts on $v(t)$, $x(t)$, and P are used to distinguish different measurement points in the complex baseband model. Thus, observing this convention, $v_m(t)$ is the modulated linearizer input signal, $v_a(t)$ is the power amplifier output signal, $v_e(t)$ is the error signal, and $v_o(t)$ is the resultant linearizer output signal.

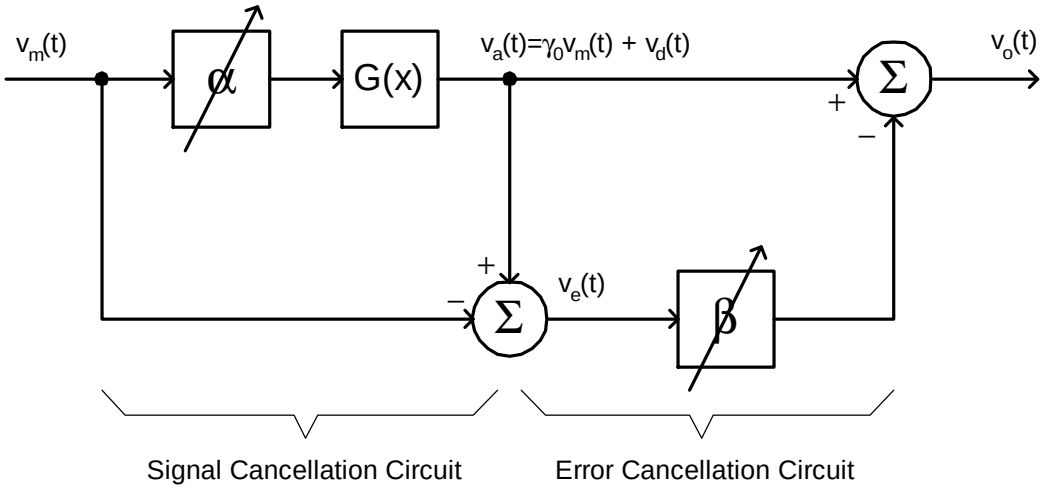


Figure 4. Complex baseband model of adaptive feedforward linearizer.

Since attenuation and phase shift of a bandpass signal corresponds to multiplication of its complex envelope by a single complex coefficient, the variable attenuation and phase in the signal and error cancellation circuits is represented simply by

the variable coefficients α and β respectively. The nonlinear power amplifier (PA) is represented by its complex voltage gain $G(x)$ which is a function of the instantaneous power of the PA input signal.

The assumptions made in deriving the complex baseband model are that the power amplifier is memoryless, and that all the components in the RF feedforward circuit other than the PA, namely the variable attenuators/phase shifters, the splitters, combiners, and couplers, and the error amplifier, are assumed to have flat gain and linear phase across the frequency band of interest. Although the feedforward linearization technique inherently overcomes memory effects, the memoryless assumption is made in order to simplify subsequent analysis. Additional assumptions are accurate delay matching in the signal and error cancellation circuits and a perfectly linear, unity gain error amplifier. Although accurate delay matching is assumed here, the effect of delay mismatches is shown in Section 2.4. Modeling the error amplifier as linear is realistic since in practice a class A amplifier is used with sufficient backoff to ensure linear operation. Linear operation of the error amplifier is necessary in practice, since any IMD introduced in the error cancellation circuit is not correctable in the linearizer output.

The simplifications made are the following: the splitters, combiners, couplers, and delay lines are lossless; the coupling factor of the sampling and output couplers is unity; and the fixed attenuator provides zero attenuation. No loss of generality is incurred here, since modeling the various losses and coupling factors would only change the optimal values of α and β. If, however, it is desired to analyze the power efficiency of the feedforward circuit, it would be necessary to include the various losses and coupling factors.

Note that an alternative placement of α is in the reference branch of the signal cancellation circuit rather than in the main branch. This placement is undesirable for practical reasons: if any distortion is generated by the attenuator/phase shifter, it is injected into the error cancellation circuit and appears at the linearizer output as uncorrectable distortion. In contrast, for placement in the main branch, any additional distortion generated by the attenuator/phase shifter is lumped with that generated by the PA and is ultimately canceled in the linearizer output.

Analysis of the feedforward circuit begins by considering the amplifier output signal $v_a(t)$. The nonlinear complex gain function may be expanded in a power series in x with complex coefficients c_i as follows

$$G(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots \tag{2}$$

Observing the complex baseband model, the PA output signal is

$$v_a(t) = \alpha v_m(t) G\left[|\alpha|^2 x_m(t)\right] \tag{3}$$

Expanding $G[\cdot]$ as in (2) gives

$$\begin{aligned} v_a(t) &= c_0 \alpha v_m(t) + \alpha v_m(t)\left[c_1 |\alpha|^2 x_m(t) + c_2 |\alpha|^4 x_m^2(t) + c_3 |\alpha|^6 x_m^3(t) + \ldots\right] \\ &= c_0 \alpha v_m(t) + \alpha\left[c_1 |\alpha|^2 v_m^2(t) v_m^*(t) + c_2 |\alpha|^4 v_m^3(t) v_m^{*2}(t) + c_3 |\alpha|^6 v_m^4(t) v_m^{*3}(t) + \ldots\right] \end{aligned} \tag{4}$$

which shows that the PA output signal is composed of a linear term plus a collection of odd-ordered nonlinear terms which generate the intermodulation products IM3, IM5, IM7, etc. For example, if $v_m(t)$ is the following two-tone signal

$$v_m(t) = e^{j2\pi f_1 t} + e^{j2\pi f_2 t} \tag{5}$$

then the expansion of the c_1 term in the power series representation of $v_a(t)$ is

$$c_1\alpha|\alpha|^2\left[3e^{j2\pi f_1 t}+3e^{j2\pi f_2 t}+e^{j2\pi(2f_1-f_2)t}+e^{j2\pi(2f_2-f_1)t}\right] \tag{6}$$

which is composed of tones at the original frequencies f_1 and f_2, plus IM3 products at the frequencies $2f_1 - f_2$ and $2f_2 - f_1$. The fact that each nonlinear term generates tones at the original frequencies f_1 and f_2 illustrates that both the linear and nonlinear terms may be correlated with the original input signal. More generally, calculation of the covariance of any nonlinear term and the original input signal results in even-ordered moments of the instantaneous input power which may be nonzero.

Using principles of linear estimation, the PA output signal may be expressed as the sum of a linearly amplified component plus intermodulation distortion:

$$v_a(t)=\gamma_o v_m(t)+v_d(t) \tag{7}$$

The value of the linear gain γ_o is that which results in the greatest correlation between $v_a(t)$ and $v_m(t)$ or zero correlation between $v_d(t)$ and $v_m(t)$. Using the latter, the linear gain can be determined by setting the covariance of $v_d(t)$ and $v_m(t)$ to zero and solving for γ_o which, employing (3) and (7), gives

$$\gamma_o=\frac{1}{P_m}\alpha E\left[x_m G\left(|\alpha|^2 x_m\right)\right] \tag{8}$$

where $E[\cdot]$ denotes expectation with respect to the probability density function (pdf) of the instantaneous power of the input signal $v_m(t)$. Clearly, γ_o is dependent on the modulation format and the average input power, both of which affect the pdf of $x_m(t)$. Moreover, in contrast to the case where α is placed in the reference branch of the signal cancellation circuit, both γ_o and $v_d(t)$ are dependent on α.

Two criteria that may be used to solve for the optimal value of α are zero correlation between the error signal $v_e(t)$ and the input signal $v_m(t)$ or minimum power of $v_e(t)$. For α in the reference branch, the two criteria are equivalent and lead to the same optimal value of α [16]. In contrast, for α in the main branch, the two criteria, in general, lead to different optimal values as shown in Section 2.1. If α is optimized using the criterion of zero correlation between $v_e(t)$ and $v_m(t)$, then, because $v_e(t) = (\gamma_o\text{-}1)v_m(t) + v_d(t)$, it is clear that for the optimal value of α, γ_o must be unity since $v_m(t)$ and $v_d(t)$ are uncorrelated. Consequently, $v_e(t) = v_d(t)$, i.e. the error signal is composed solely of the intermodulation distortion generated by the amplifier with the reference signal completely canceled. On the other hand, for α optimized using the criterion of minimum power of $v_e(t)$, the linear gain is not unity in general; thus, $v_e(t)$ contains a component of the reference signal as well as the amplifier IMD even though the power of $v_e(t)$ is minimized.

Since gradient adaptation is the focus of the current work, the appropriate criterion for the optimization of α is zero correlation between $v_e(t)$ and $v_m(t)$. Unless otherwise specified, the optimal value of α is denoted α_{opt}. Using (3), the optimal value of α is found from a numerical solution of

$$E\left[v_e(t)v_m^*(t)\right] = \alpha_{opt} E\left[x_m G\left(\left|\alpha_{opt}\right|^2 x_m\right)\right] - P_m = 0 \tag{9}$$

for α_{opt} by noting that $v_e(t) = v_d(t)$ for $\alpha = \alpha_{opt}$.

The optimal value of β may be determined by posing the operation of the error cancellation circuit in linear estimation terms whereby the amplifier output signal is expressed as

$$v_a(t) = \beta_{opt} v_d(t) + v_o(t) \tag{10}$$

The basis of the estimate is the distortion signal $v_d(t)$ provided by the lower branch of the error cancellation circuit; the estimation error corresponds to the linearizer output signal $v_o(t)$. Again, two criteria that may be used to solve for β_{opt} are zero correlation between $v_o(t)$ and $v_d(t)$ or minimum power of $v_o(t)$. In contrast to the optimization of α, these two criteria are equivalent, since, as shown below, the error surface is quadratic. The latter criterion is used here.

Using (3), the linearizer output signal for $\alpha = \alpha_{opt}$ is

$$\begin{aligned} v_o(t) &= v_a(t) - \beta v_d(t) \\ &= \alpha_{opt} v_m(t) G\left[\left|\alpha_{opt}\right|^2 x_m(t)\right] - \beta v_d(t) \end{aligned} \tag{11}$$

and the average power of $v_o(t)$ as a function of β is

$$\begin{aligned} P_o(\beta) &= E\left[\left|\alpha_{opt} v_m G\left(\left|\alpha_{opt}\right|^2 x_m\right) - \beta v_d\right|^2\right] \\ &= P_a\left(\alpha_{opt}\right) - 2\,\mathrm{Re}\left\{\beta^* \alpha_{opt} E\left[v_m G\left(\left|\alpha_{opt}\right|^2 x_m\right) v_d^*\right]\right\} + |\beta|^2 P_d \end{aligned} \tag{12}$$

which is quadratic in β. If plotted, the error surface appears as a familiar paraboloid with its minimum value at $\beta = \beta_{opt}$.

Computing the gradient of the error surface gives

$$\frac{dP_o(\beta)}{d\beta} = -2E\left\{\alpha_{opt}\left[v_m G\left(\left|\alpha_{opt}\right|^2 x_m\right)\right] v_d^*\right\} + 2\beta P_d \tag{13}$$

Setting the derivative to zero and solving for β gives

$$\beta_{opt} = \frac{1}{P_d} \alpha_{opt} E\left[v_m G\left(\left|\alpha_{opt}\right|^2 x_m\right) v_d^*\right] \tag{14}$$

The numerator of this expression is equivalent to $E\left[v_a(t)v_d^*(t)\right]$, i.e. the covariance of the amplifier output and distortion signals for $\alpha = \alpha_{opt}$; thus,

$$\beta_{opt} = \frac{1}{P_d} E\left[\left(v_m(t) + v_d(t)\right)v_d^*(t)\right] = 1 \tag{15}$$

since $v_m(t)$ and $v_d(t)$ are uncorrelated. This result is expected and can be obtained simply by inspection of Figure 4.

Several useful expressions may be derived by first introducing expressions for the relative errors in the coefficients α and β defined, respectively, as

$$\varepsilon_\alpha = \left(\alpha - \alpha_{opt}\right)\Big/\alpha_{opt} \tag{16}$$

and

$$\varepsilon_\beta = \left(\beta - \beta_{opt}\right)\Big/\beta_{opt} = \beta - 1 \tag{17}$$

Since the linear gain γ_o is unity only for $\alpha = \alpha_{opt}$, the relative difference in γ_o (from unity) is defined as

$$\begin{aligned}\varepsilon_{\gamma_o} &= \gamma_o - 1 \\ &= \frac{1}{P_m}\alpha_{opt}\left(1+\varepsilon_\alpha\right)E\left[x_m G\left(\left|\alpha_{opt}\right|^2\left|1+\varepsilon_\alpha\right|^2 x_m\right)\right] - 1\end{aligned} \tag{18}$$

which shows that the relative difference in γ_o is a nonlinear function of the relative error in α and is level and pdf dependent. Table 1 shows the magnitude of ε_{γ_o} for a 1% relative error in α for several different backoff values. The modulation format used in calculating these values is $\pi/4$-DQPSK with 35% rolloff root raised cosine filtering resulting in a peak-to-average power ratio of approximately 2.5 dB. The complex gain function used is the one presented in Section 1.1 for the class AB amplifier. Note that a histogram

approximating the pdf of $x_m(t)$ is used to calculate the expectations. As can be seen, the relative difference in γ_o is on the same order as the relative error in α, but decreases somewhat for input levels approaching saturation.

Table 1. Relative difference in γ_o as a function of backoff for a 1% relative error in α

Backoff (dB)	$\lvert\varepsilon_{\gamma_o}\rvert$ (%)
20	1.3
10	1.2
9	1.1
8	1.1
7	1.0
6	1.0
5	1.0
4	0.9
3	0.9
2	0.8
1	0.5

Using the above expressions for ε_α, ε_β, and ε_{γ_o}, the amplifier output may be written as

$$v_a(t) = \left(1+\varepsilon_{\gamma_o}\right)v_m(t) + v_d(t) \tag{19}$$

the error signal as

$$v_e(t) = \varepsilon_{\gamma_o} v_m(t) + v_d(t) \tag{20}$$

and the linearizer output signal as

$$\begin{aligned} v_o(t) &= \left(1-\varepsilon_{\gamma_o}\varepsilon_\beta\right)v_m(t) - \varepsilon_\beta v_d(t) \\ &\approx v_m(t) - \varepsilon_\beta v_d(t) \end{aligned} \tag{21}$$

for $\varepsilon_{\gamma_o}\varepsilon_\beta << 1$. Note that the only signals appearing in above expressions are now $v_m(t)$ and $v_d(t)$ which are uncorrelated. Quantities useful in specifying the degree of distortion cancellation achieved by the feedforward linearizer are the ratios of IM power to desired signal power (*IMSR*) in both the amplifier output signal and in the linearizer output signal defined, respectively, as

$$IMSR_a = \frac{P_d}{\left|1+\varepsilon_{\gamma_o}\right|^2 P_m} \tag{22}$$

and

$$IMSR_o = \frac{\left|\varepsilon_\beta\right|^2 P_d}{P_m} \tag{23}$$

Combining (22) and (23) gives

$$\begin{aligned} IMSR_o &= \left|\varepsilon_\beta\right|^2 \left|1+\varepsilon_{\gamma_o}\right|^2 IMSR_a \\ &\approx \left|\varepsilon_\beta\right|^2 IMSR_a \end{aligned} \tag{24}$$

for $\left|\varepsilon_{\gamma_o}\right| << 1$, which gives the IM-to-signal ratio at the linearizer output as a function of the relative difference in γ_o, the relative error in β, and the observed IM-to-signal ratio at the amplifier output. Note that the degree of IMD suppression is largely determined by the relative error in β; a small relative difference in γ_o causes only a small change in the desired component of the linearizer output signal thus increasing $IMSR_o$ only slightly. For example, given $IMSR_a$ = -20 dB, a 1% error β, and a small relative difference in γ_o, the IMD generated by the power amplifier is suppressed by approximately 40 dB to give $IMSR_o \approx$ -60 dB.

2.1 Comparison of Optimization Criteria for Signal Cancellation Coefficient

Adaptation of α can be designed to either minimize the power of $v_e(t)$ or to decorrelate $v_e(t)$ and $v_m(t)$. For α in the reference branch of the signal cancellation circuit, as analyzed in Cavers' work [16], the two optimization criteria result in the same optimal value of α. The reason for this is that when the operation of the signal cancellation circuit is posed in linear estimation terms, $v_m(t)$ corresponds to the basis of the estimate of $v_a(t)$, $v_e(t)$ corresponds directly to the estimation error, and α directly to the variable estimation parameter. Consequently, the error surface is quadratic, and when the estimation error is minimized, $v_e(t)$ and $v_m(t)$ are automatically uncorrelated implying complete signal cancellation, i.e. $v_e(t) = v_d(t)$. In contrast, for α in the main branch, the two criteria are not equivalent, leading to different optimal values of α. This new result, demonstrated here, has practical consequences that are discussed later in this section.

The optimal value of α corresponding to the minimum value of P_e may be found analytically in a familiar fashion by writing the expression for P_e in terms of α, differentiating, setting the result to zero, and solving for $\alpha = \alpha_{minpwr}$. Note that the subscript *minpwr* is used to distinguish the optimal value of α corresponding to minimum P_e from α_{opt}.

Observing Figure 4, the error signal is

$$v_e(t) = \alpha v_m(t) G\left[|\alpha|^2 x_m(t)\right] - v_m(t) \tag{25}$$

Therefore, the average power of $v_e(t)$ as a function of α is

$$P_e(\alpha) = E\left[\left|\alpha v_m G\left(|\alpha|^2 x_m\right) - v_m\right|^2\right]$$
$$= |\alpha|^2 E\left[x_m \left|G\left(|\alpha|^2 x_m\right)\right|^2\right] - 2\,\mathrm{Re}\left\{\alpha E\left[x_m G\left(|\alpha|^2 x_m\right)\right]\right\} + P_m \qquad (26)$$

Clearly, $P_e(\alpha)$ is not quadratic because of its nonlinear dependence on α. Figure 5 shows the error surface for a $\pi/4$-DQPSK input with 1 dB backoff. Evidently, the surface is smooth with no local minima.

Fig. 5

Figure 5. Power of $v_e(t)$ as a function of α for a $\pi/4$-DQPSK input signal and 1 dB backoff

α_{minpwr} is found by first differentiating (26) with respect to α which gives the gradient of the error surface. To evaluate the derivative, we let $\alpha = \alpha_I + j\alpha_Q$ where $\alpha_I = \mathrm{Re}[\alpha]$ and $\alpha_Q = \mathrm{Im}[\alpha]$ and then write $P_e(\alpha)$ as $P_e(\alpha_I,\alpha_Q)$. The gradient of the error surface is then defined as

$$\frac{dP_e(\alpha)}{d\alpha} \equiv \frac{\partial P_e\left(\alpha_I,\alpha_Q\right)}{\partial \alpha_I} + j\frac{\partial P_e\left(\alpha_I,\alpha_Q\right)}{\partial \alpha_Q} \qquad (27)$$

Evaluating the partial derivatives in (27) and rewriting the result in terms of α instead of α_I and α_Q gives the gradient expression

$$\frac{dP_e(\alpha)}{d\alpha} = 2\alpha E\left[x_m\left|G(x)\right|^2\right] + 2\alpha|\alpha|^2 E\left\{x_m^2\left[\frac{dG(x)}{dx}G^*(x) + G(x)\frac{dG^*(x)}{dx}\right]\right\}$$
$$-2E\left[x_m G^*(x)\right] - 2\alpha^2 E\left[x_m^2\frac{dG(x)}{dx}\right] - 2|\alpha|^2 E\left[x_m^2\frac{dG^*(x)}{dx}\right] \qquad (28)$$

where $G(x)$ and its derivatives are evaluated at $x = |\alpha|^2 x_m$. In order to determine α_{minpwr}, it is necessary to equate this expression to zero and solve for α. Clearly, though, the gradient of the error surface is a highly nonlinear function of α; hence, only a numerical solution rather than an analytical one is possible. Figure 6 shows the negative of the gradient expression represented as arrows overlaid on a surface contour plot for the same surface plotted in Figure 5. The magnitude of the gradient is proportional to the length of the arrows, and the direction of the gradient is represented by the direction of the arrows. The negative of the gradient is plotted in order to indicate the direction of steepest descent. Note that the gradient is large for values of α leading to the optimal value, and is indeed zero at the global minimum of the surface. In this case, a numerical solution of (28) for the optimal value of α yields $\alpha_{minpwr} = 0.726 - j0.471$.

Fig. 6

Figure 6. Contours and negative gradient of $P_e(\alpha)$ surface for a $\pi/4$-DQPSK input signal and 1 dB backoff

The optimal value of α corresponding to the decorrelation of $v_e(t)$ and $v_m(t)$ is found, numerically, by solving equation (9) for α_{opt}. For a $\pi/4$-DQPSK input with 1 dB backoff, the result is $\alpha_{opt} = 0.838 - j0.501$. Evidently, the optimal values of α corresponding to the two different optimization criteria are different. Using (26), $P_e(\alpha_{opt})$ is 1.42 dB higher than $P_e(\alpha_{\text{minpwr}})$ and using (8), $\gamma_o(\alpha_{opt}) = 1$ and $\gamma_o(\alpha_{minpwr}) = 0.939$. Hence, using the criterion of minimum P_e does not result in the decorrelation of $v_e(t)$ and $v_m(t)$; thus, complete signal cancellation is not achieved, even though the power of $v_e(t)$ is minimized. In contrast, using the criterion of zero correlation between $v_e(t)$ and $v_m(t)$ does result in complete signal cancellation, but the power of $v_e(t)$ is higher than that for the other optimization criterion.

The above results are calculated for a backoff of 1 dB to emphasize the differences between the two optimization criteria. Further calculations show that the difference between the two optimal values decreases as the backoff increases. For example, for

backoff values higher than 3 dB, the two criteria lead to equivalent optimal values of α and virtually no difference in the power of $v_e(t)$.

The two different optimization criteria have practical consequences when considering the adaptation of β. If gradient adaptation of β is implemented, the criterion of zero correlation must be used since, as discussed in Section 2.3, if $v_e(t)$ and $v_m(t)$ are correlated, the resultant bias in β degrades the final distortion suppression. The error amplifier must be able to handle the higher power in the error signal though. If the minimum power criterion is used, the error amplifier does not have to handle as high a power, but, because of incomplete signal cancellation in the error signal, the desired component of the linearizer output signal is changed by a small amount. Since the two optimization criteria result in different optimal values of α only for the amplifier biased near saturation, the above issues may not have practical consequences for amplifiers operating with realistic backoff.

2.2 Adaptation of Signal Cancellation Coefficient

In a practical adaptation structure an iterative procedure for adjusting α towards its optimal value is necessary. Probably the simplest iterative procedure is the method of steepest descent. In the context of quadratic error surfaces, one begins by choosing an arbitrary initial value of α which defines some point on the error surface. The gradient of the error surface at that point is then calculated, and α is adjusted by a small increment in the direction opposite to the computed gradient. This ensures that α is adjusted in a direction towards the minimum point of the error surface. This process is repeated until,

eventually, the minimum is reached at which point the computed gradient is zero and no further change in α occurs. Well known in linear estimation theory is that, for quadratic error surfaces, the covariance of the basis $v_m(t)$ and estimation error $v_e(t)$ is identical to the gradient of the error surface and thus may be used to drive the adaptation of α. Since the basis and error are uncorrelated at the minimum point of the error surface, the covariance, and thus the gradient, is zero as required.

By analogy with the above situation, the adaptation of α for the non-quadratic error surface considered in this thesis proceeds in a parallel fashion. Since the criterion for the adaptation of α is the decorrelation of $v_e(t)$ and $v_m(t)$, an appropriate "gradient" signal is again, the covariance of $v_e(t)$ and $v_m(t)$:

$$\sigma_{em}^2 = E\left[v_e(t)v_m^*(t)\right] = \varepsilon_{\gamma_o} P_m \tag{29}$$

Clearly, for $\alpha = \alpha_{opt}$ implying $\varepsilon_{\gamma_o} = 0$, the gradient signal is zero as required. Although the covariance of $v_m(t)$ and $v_e(t)$ is not actually the true gradient of the error surface (the true gradient is given by (28) and plotted in Figure 6), the above expression is nonetheless treated as the gradient by analogy with quadratic error surfaces.

Figure 7 shows the above gradient represented as arrows overlaid on the contour plot of the error surface for a $\pi/4$-DQPSK input with 1 dB backoff. As can be seen, the gradient is a very good approximation to the true gradient; it is large for values of α leading to the optimal value, and is indeed zero at $\alpha = \alpha_{opt}$. The only real difference is that the gradient does not give the true path of steepest descent, but instead a spiral-like path towards the optimal value. Consequently, the adaptation time may be slightly longer. Note again, that α_{opt} in Figure 7 is not, in general, the minimum point of the error surface;

rather, the minimum point is shown in Figure 6. For larger backoff, though, α_{opt} and α_{minpwr} are coincident.

Fig. 7

Figure 7. Contours of $P_e(\alpha)$ surface and σ^2_{em} for a $\pi/4$-DQPSK input signal and 1 dB backoff

In a practical adaptation structure, an estimate of the above gradient is required to avoid the expectation operator. A noisy, but unbiased estimate is

$$D_\alpha(t) = v_e(t)v_m^*(t) \tag{30}$$

which is referred to as a stochastic gradient signal. The method of steepest descent coupled with the stochastic gradient suggests the following algorithm for the adjustment of α implemented as a first order adaptation loop

$$\alpha(t) = K_\alpha \int_0^t D_\alpha(\tau)d\tau \tag{31}$$

in which the integrator provides averaging to remove some of the self noise in the gradient estimate, and K_a controls the time constant of adaptation. When α_{opt} is achieved, $D_a(t)$ is zero on average since $v_m(t)$ and $v_e(t)$ are uncorrelated; thus, $\alpha(t)$ is held at its final integrated value. Of course any change in operating conditions, such as a sudden change

in input power level or a gradual change in temperature, causes $D_a(t)$ to become nonzero on average, and α is adjusted to a new optimal value.

2.3 Adaptation of Error Cancellation Coefficient

The concepts involved in the adaptation of β in the error cancellation circuit are similar to those for α in the signal cancellation circuit with two notable differences. Firstly, the convergence of β is coupled to that of α which leads to a bias effect that imposes high accuracy requirements on α for reasonable distortion cancellation. This is in contrast to a system using pilot tones in which the convergence of α and β are independent. Secondly, the weak IM component in $v_o(t)$ is masked by the much stronger signal component which leads to slow convergence of β. A novel use of DSP that overcomes the problems of bias and masking is discussed in Section 3.1.

An iterative procedure for achieving the optimal value of β amenable to implementation, is the method of steepest descent described previously. Again, the covariance of the basis $v_d(t)$ and the estimation error $v_o(t)$ forms a gradient signal for the adaptation of β. Thus, using (21) for $\varepsilon_{\gamma_o} = 0$, the desired gradient signal is

$$\sigma_{od}^2 = E\left[v_o(t)v_d^*(t)\right] = -\varepsilon_\beta P_d \tag{32}$$

Clearly, for $\beta = \beta_{opt}$, the gradient is zero as required implying $v_o(t)$ and $v_d(t)$ are uncorrelated. In contrast to the gradient used in the signal cancellation circuit, σ_{od}^2 is identical to the gradient of the error surface given by (13) since $P_o(\beta)$ is quadratic.

The problem with computing the above gradient signal in a practical adaptation structure, aside from calculating the expectation, is that the desired basis signal $v_d(t)$ is not

available unless the relative error in α is exactly zero and $v_e(t) = v_d(t)$. Of course $v_e(t)$ is a good approximation to $v_d(t)$; hence, it may be used to compute the gradient signal, but the bias effect mentioned previously must be considered. The fact that $v_e(t)$ instead of $v_d(t)$ is used as the basis of the estimate of $v_a(t)$ performed in the error cancellation circuit implies that the convergence of α and β is coupled since $v_e(t)$ is a function of α.

The bias effect is demonstrated by computing the covariance of $v_o(t)$ and $v_e(t)$ using (20) and (21)

$$\begin{aligned}\sigma_{oe}^2 &= E\left[v_o(t)v_e^*(t)\right] \\ &= \varepsilon_{\gamma_o}^*\left(1-\varepsilon_{\gamma_o}\varepsilon_\beta\right)P_m - \varepsilon_\beta P_d\end{aligned} \tag{33}$$

For α perfectly adjusted implying $\varepsilon_{\gamma_o} = 0$, the first term vanishes, and the gradient is proportional to ε_β as in (32). For α not adjusted perfectly, the first term is nonzero which causes the gradient to become zero for a nonzero error in β; that is, β does not converge to $\beta_{opt} = 1$. In effect, the bias is due to the strong correlation of any residual signal component in the error signal $v_e(t)$ with the desired signal component in the linearizer output signal $v_o(t)$.

The converged value of β including the bias may be found by setting equation (33) to zero and solving for β. The result is

$$\beta = \frac{\varepsilon_{\gamma_o}^*\left(1+\varepsilon_{\gamma_o}\right)P_m + P_d}{\left|\varepsilon_{\gamma_o}\right|^2 P_m + P_d} \tag{34}$$

More insight may be gained by rewriting this expression with respect to $IMSR_a$. Multiplying numerator and denominator by $\left(\left|1+\varepsilon_{\gamma_o}\right|^2 P_m\right)^{-1}$ gives

$$\beta = \frac{IMSR_a + \left(\frac{\varepsilon_{\gamma_o}}{1+\varepsilon_{\gamma_o}}\right)^*}{IMSR_a + \left|\frac{\varepsilon_{\gamma_o}}{1+\varepsilon_{\gamma_o}}\right|^2}$$

$$\approx 1 + \frac{\varepsilon^*_{\gamma_o}}{IMSR_a} \tag{35}$$

for $\varepsilon_{\gamma_o} << 1$ and $IMSR_a >> \left|\varepsilon_{\gamma_o}\right|^2$. This clearly shows the bias in β resulting from a small error in α. For example, for $IMSR_a$ = -20 dB and only a 0.1% relative difference in γ_o (corresponding to approximately the same relative error in α), β converges not to $\beta_{opt} = 1$, but to $\beta \approx 1.1$, i.e. a relative error of approximately 10%. Using (24), $IMSR_o \approx$ -40 dB indicating only a 20 dB improvement in IMD-to-signal ratio.

Combining (24) and (35) results in

$$\left|\varepsilon_{\gamma_o}\right| = \sqrt{IMSR_a\, IMSR_o} \tag{36}$$

which indicates, through ε_{γ_o}, the required accuracy of α for a desired $IMSR_o$ and an observed $IMSR_a$. This result is identical to the result for α in the reference branch [16] as long as ε_{γ_o} is replaced by ε_α. For example, for a desired $IMSR_o$ of -60 dB and an observed $IMSR_a$ of -20 dB (i.e. a desired 40 dB IM-to-signal ratio improvement), the required accuracy of α is approximately 0.01%—a difficult value to achieve in practice.

The stochastic gradient corresponding to (33) is

$$D_\beta(t) = v_o(t)v_e^*(t) \tag{37}$$

A problem with this gradient estimate, though, is excessive self noise due to the fact that the weak IM component in $v_o(t)$ is masked by the desired signal component which is

several orders of magnitude larger. The excessive self noise leads to a very long convergence time for β.

A solution to the masking problem, discussed in [16], is to suppress the desired signal component in $v_o(t)$ with a filter in the adaptation path for β. The mitigating effects of such a filter include faster convergence time for β and reduced bias, thus relaxed accuracy requirements for α. In [16], Cavers performs an analysis of the adaptation of β assuming a rectangular bandstop filter is used for signal suppression and the amplifier distortion is dominated by third order IM products. He shows, for example, that for 40 dB signal suppression, the time constant for adaptation may be reduced by a factor of 1000 to 100 reciprocal bandwidths—an acceptable operating range for single or multicarrier operation. He also shows that the required accuracy of α is reduced from that shown in (36) to

$$\left|\varepsilon_{\gamma_o}\right| = \frac{1}{3A_m}\sqrt{IMSR_a\, IMSR_o} \tag{38}$$

where A_m is the stopband attenuation of the signal suppression filter, and ε_α has been replaced with ε_{γ_o}. For example, assuming 40 dB stopband attenuation, the required accuracy from the previous example is reduced from 0.01% to approximately 0.3%, and for 60 dB stopband attenuation is reduced to approximately 3%—obtainable values in practice.

With signal suppression, the gradient estimate for the adaptation of β becomes

$$D_\beta^{'}(t) = v_o^{'}(t)v_e^{*}(t) \tag{39}$$

where $v'_o(t)$ is the linearizer output with the desired signal component suppressed. The algorithm for the adaptation of β then becomes

$$\beta(t) = K_\beta \int_0^t D'_\beta(\tau) d\tau \tag{40}$$

where K_β is analogous to K_α.

2.4 Effect of Delay Mismatches

Accurate delay matching is important for wide bandwidth operation as shown in this section. Even if α and β are adjusted to their optimal values, incomplete distortion cancellation can occur, depending on the bandwidth, if the delay matching in the signal and error cancellation circuits is inaccurate. With reference to Figure 4, assume a small delay mismatch τ_α in the reference branch of the signal cancellation circuit, and a small delay mismatch τ_β in the main branch of the error cancellation circuit. For $\alpha = \alpha_{opt}$ and $\beta = \beta_{opt}$, the linearizer output signal is thus

$$\begin{aligned} v_o(t) &= v_a(t-\tau_\beta) - v_e(t) \\ &= v_m(t-\tau_\beta) + v_d(t-\tau_\beta) - \left[v_m(t) + v_d(t) - v_m(t-\tau_\alpha)\right] \end{aligned} \tag{41}$$

Taking the Fourier transform gives

$$\begin{aligned} V_o(f) &= \left[e^{-j2\pi f\tau_\beta} + e^{-j2\pi f\tau_\alpha} - 1\right] V_m(f) + \left[e^{-j2\pi f\tau_\beta} - 1\right] V_d(f) \\ &\approx V_m(f) - \left(j2\pi f\tau_\beta\right) V_d(f) \end{aligned} \tag{42}$$

where the approximation holds for $f\tau_\alpha$ and $f\tau_\alpha << 1$. Evidently, the linearizer output signal is composed of the desired signal component plus some residual distortion, even though the coefficients α and β are adjusted to their optimal values. At band centre ($f = 0$), the

distortion cancellation is complete, but decreases towards the band edges. For example, if a minimum of 30 dB distortion suppression is desired, the product $2\pi f\tau_\beta$ must be less than or equal to 0.032 across the full band of the distortion signal. Assuming that the amplifier distortion is dominated by third-order IM products, $f \approx \frac{1}{2}(3B)$ at the band edge of the distortion signal where B is the bandwidth of the input signal. Therefore, the bandwidth-delay mismatch product $B\tau_\beta$ cannot exceed about 0.3%. Thus, for B = 1 MHz, τ_β cannot exceed about 3 ns, and for 10 MHz, cannot exceed 0.3 ns. Typical amplifier measurements performed in the current work, showed group delays on the order of 10 ns; thus, a mismatch of 0.3 ns implies that group delay measurements of both the amplifier and a compensating delay line must be accurate to within about 3%. Clearly, the accuracy requirements become quite stringent as the bandwidth exceeds a few tens of MHz.

2.5 Effect of Downconversion Errors

In a practical sense, the gradient signals in equations (30) and (39) may be calculated by way of either a bandpass correlation or baseband correlation method. As discussed in Section 3, the method of bandpass correlation is rejected in the current work in favour of a novel use of DSP to perform baseband correlation. Baseband correlation requires the recovery of the complex envelopes of the RF signals $\tilde{v}_m(t)$, $\tilde{v}_e(t)$, and $\tilde{v}_o(t)$, though. Consequently, an analysis of the effects of the downconversion of RF signals on the correlations is necessary.

A new analysis shown here, demonstrates that the following errors in downconversion and subsequent complex envelope recovery do not bias the correlation of either $v_e(t)$ and $v_m(t)$ or $v_o(t)$ and $v_e(t)$:

- frequency and phase offsets in the recovered complex envelopes
- linear distortion due to the filters in the downconversion chains, i.e. amplitude ripple, nonlinear phase response
- filter mismatches between two downconversion chains
- incomplete image suppression

Mixer DC offsets do bias the correlations though, but, as described in Section 3.1, these are avoided by downconverting the RF signals to a low enough intermediate frequency (IF) to be sampled, yet high enough to ensure no spectral occupancy at DC, performing the final downconversion in DSP, and filtering out the DC offsets along with the images at twice the IF.

The effect of the first three downconversion errors on the adaptation of α may be analyzed by replacing the downconversion chain for each of $v_e(t)$ and $v_m(t)$ by its complex baseband equivalent as illustrated in Figure 8. Note that mixer DC offsets are neglected and, initially, complete suppression of the images at twice the IF is assumed. Incomplete image suppression is considered later in this section. The frequency offset Δf is due to oscillator inaccuracies and/or drift, and the phase offset ϕ is due to possible component mismatches between the two downconversion chains. Note that the frequency offset is the same for each signal due to the assumption that the same oscillators are shared. The lowpass filters $H_e(f)$ and $H_m(f)$ with impulse responses $h_e(t)$ and $h_m(t)$, respectively, are the complex baseband equivalents of all filtering performed in downconversion of $v_e(t)$ and

$v_m(t)$. Each filter has arbitrary amplitude ripple and phase response; moreover, the two filters are not necessarily matched. Note that the same analysis applies to the adaptation of β except that the filter for $v_o(t)$ includes a notch to achieve suppression of the desired signal component.

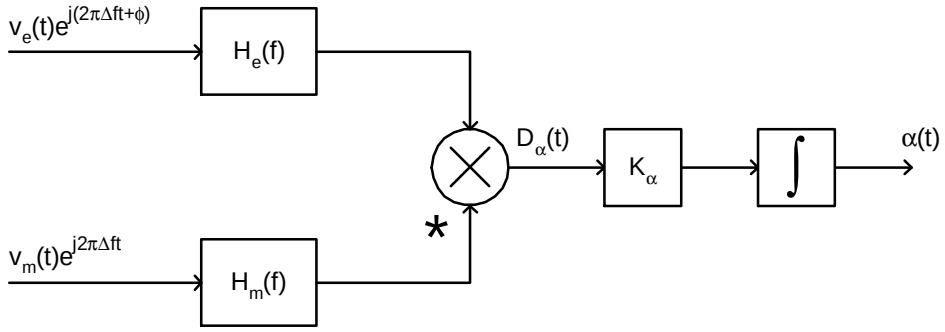


Figure 8. Complex baseband equivalent of adaptation loop for α

For the correlation of $v_e(t)$ and $v_m(t)$ to be unbiased, the expected value of the gradient estimate $D_\alpha(t)$ must be zero for α adjusted to its optimal value resulting in zero correlation between $v_m(t)$ and $v_e(t)$. In contrast, a biased correlation forces α to converge to a non-optimal value resulting in a certain degree of correlation between $v_e(t)$ and $v_m(t)$. In order to check for potential bias introduced by the filters, the expected value of the gradient estimate must be computed. Observing Figure 8, the gradient estimate is

$$\begin{aligned} D_\alpha(t) &= \left[v_e(t)e^{j(2\pi\Delta ft+\phi)} \otimes h_e(t)\right]\left[v_m(t)e^{j2\pi\Delta ft} \otimes h_m(t)\right]^* \\ &= e^{j\phi}\int v_e(t-\tau_1)e^{j2\pi\Delta f(t-\tau_1)}h_e(\tau_1)d\tau_1 \int v_m^*(t-\tau_2)e^{-j2\pi\Delta f(t-\tau_2)}h_m^*(\tau_2)d\tau_2 \end{aligned} \quad (43)$$

where ⊗ is defined as the convolution operator. Notice that the $e^{j2\pi\Delta ft}$ term in the first integral cancels its complex conjugate in the second integral. Hence, the rotation of the $v_e(t)$ and $v_m(t)$ signal constellations due to the frequency offset is canceled by the complex conjugate multiplication in the calculation of $D_\alpha(t)$. Therefore, (43) can be rewritten as

$$D_\alpha(t) = e^{j\phi}\int v_e(t-\tau_1)\tilde{h}_e(\tau_1)d\tau_1 \int v_m^*(t-\tau_2)\tilde{h}_m^*(\tau_2)d\tau_2$$
$$= e^{j\phi}\left[v_e(t)\otimes\tilde{h}_e(t)\right]\left[v_m(t)\otimes\tilde{h}_m(t)\right]^* \quad (44)$$

where $\tilde{h}_e(t) = h_e(t)e^{-j2\pi\Delta ft}$ and $\tilde{h}_m(t) = h_m(t)e^{-j2\pi\Delta ft}$ are the impulse responses of the filters $H_e(f)$ and $H_m(f)$ centred about $f = -\Delta f$. By a change of variables and combination of integrals, the expectation of (44) can be rewritten as

$$E\left[D_\alpha(t)\right] = e^{j\phi}\iint E\left[v_e(\tau_1)v_m^*(\tau_2)\right]\tilde{h}_e(t-\tau_1)\tilde{h}_m^*(t-\tau_2)d\tau_1 d\tau_2 \quad (45)$$

Clearly, for $v_e(t)$ and $v_m(t)$ uncorrelated, $E[D_\alpha(t)] = 0$, implying unbiased correlation regardless of the response of each filter, mismatches between the two filters, and the phase and frequency offsets. This shows that filter properties such as passband ripple and nonlinear phase response, both of which distort the complex envelopes, do not bias the correlation.

For the above analysis it is assumed that in recovering the complex envelopes, the images at twice the IF are completely suppressed by the lowpass filters in DSP. To investigate the effect of incomplete image suppression, assume that $H_e(f)$ and $H_m(f)$ are both rectangular lowpass filters with the following response

$$H(f) = \begin{cases} 1, & |f| \le W \\ a << 1, & \text{elsewhere} \end{cases} \quad (46)$$

where a is real, and W is large enough to ensure that $v_e(t)$ and $v_m(t)$, including frequency offset, fall completely within the passband of the filter. Consequently, the images at twice the IF are attenuated by the factor a. Including the images, the gradient signal is

$$D_\alpha(t) = \left[v_e(t)e^{j(2\pi\Delta ft+\phi)} + v_e^*(t)e^{-j\left[2\pi(2f_{IF}+\Delta f)t+\phi\right]}\right]\left[v_m(t)e^{j2\pi\Delta ft} + v_m^*(t)e^{-j\left[2\pi(2f_{IF}+\Delta f)t\right]}\right]^*$$
$$= v_e(t)v_m^*(t)e^{j\phi} + a^2 v_e^*(t)v_m(t)e^{-j\phi} + 2a\,\mathrm{Re}\left[v_e(t)v_m(t)e^{j\left[2\pi(4f_{IF})t+\phi\right]}\right] \quad (47)$$

where the third term with components at $f = \pm 4f_{IF}$ appears as self noise in the gradient estimate but is easily filtered out by the integrator in the adaptation loop. Taking the expectation of the first two terms gives

$$E\left[D_\alpha(t)\right] = e^{j\phi}E\left[v_e(t)v_m^*(t)\right] + a^2e^{-j\phi}E\left[v_e^*(t)v_m(t)\right] \quad (48)$$

which is proportional to the desired gradient plus its complex conjugate scaled by a^2. Clearly though, for $v_e(t)$ and $v_m(t)$ uncorrelated, $E[D_\alpha(t)] = 0$. Hence, incomplete image suppression does not result in a bias.

2.6 Effect of Vector Modulator Errors

The attenuation and phase shift required in each of the signal and error cancellation circuits may be achieved by the use of a vector modulator (VM). A practical vector modulator structure is discussed in Section 3.2.1; analyzed here are the effects on the adaptation of α and β of amplitude and phase imbalances as well as DC offsets in the VMs. The complex baseband equivalent of a practical VM is shown in Figure 9, in which the signals v_I and v_Q correspond to the DC control voltages, with offsets δv_I and δv_Q, and $v_{in}(t)$ and $v_{out}(t)$ correspond to the RF input and output voltages. The DC offsets are included since, in general, a nonzero control voltage in each branch of the VM may be necessary to achieve maximum attenuation. As discussed in Section 3.2.1, the attenuation in each branch is a nonlinear function of the control voltage which generally decreases monotonically with increasing voltage. In order to simply this analysis though, a linear

control characteristic is assumed. The amplitude factor *a* models the amplitude imbalance between the in-phase and quadrature branches of the VM, and the phase offset ϕ corresponds to a phase imbalance, e.g. due to an imperfect quadrature split of the input signal.

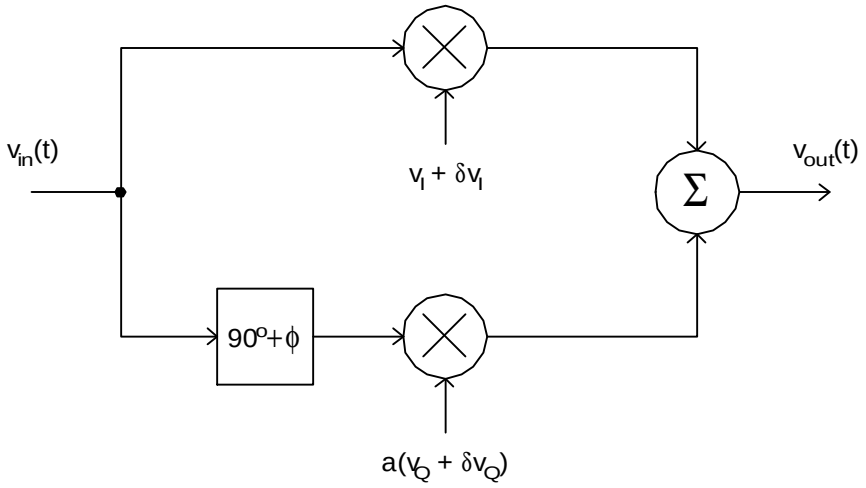


Figure 9. Complex baseband equivalent of vector modulator

Observing the complex baseband model, the output signal is expressed as

$$\begin{aligned} v_{out}(t) &= \left[\left(v_I + \delta v_I\right) + ja\left(v_Q + \delta v_Q\right)e^{j\phi}\right]v_{in}(t) \\ &= \left\{\left[\left(v_I + \delta v_I\right) - a\left(v_Q + \delta v_Q\right)\sin\phi\right] + ja\left(v_Q + \delta v_Q\right)\cos\phi\right\}v_{in}(t) \end{aligned} \tag{49}$$

The attenuation offered by the VM is thus

$$A = \sqrt{\left(v_I + \delta v_I\right)^2 + a^2\left(v_Q + \delta v_Q\right)^2 - 2a\left(v_I + \delta v_I\right)\left(v_Q + \delta v_Q\right)\sin\phi} \tag{50}$$

and the phase shift is

$$\Phi = \tan^{-1}\left[\frac{a\left(v_Q + \delta v_Q\right)\cos\phi}{\left(v_I + \delta v_I\right) - a\left(v_Q + \delta v_Q\right)\sin\phi}\right] \tag{51}$$

which shows that with suitable adjustment of the control voltages to overcome the imbalances and offsets, the range of A and Φ is [0,1] and [0°,360°] respectively. Note

that in an ideal vector modulator with no imbalances or offsets, i.e. $a = 1$, $\phi = 0$, and $\delta v_I = \delta v_I = 0$, the VM simply transforms the two control voltages in rectangular co-ordinates to attenuation and phase in polar co-ordinates. The effect of the imbalances and offsets is a distortion of this transformation.

Using (49), the necessary control voltages to achieve the optimal value of α, for example, are

$$v_I = \mathrm{Re}\left[\alpha_{opt}\right] + \mathrm{Im}\left[\alpha_{opt}\right]\tan\phi - \delta v_I \tag{52}$$

and

$$v_Q = \frac{\mathrm{Im}\left[\alpha_{opt}\right]}{a\cos\phi} - \delta v_Q \tag{53}$$

which indicates that the adaptation loop must adjust the control voltages to compensate for the imbalances and offsets. Note that the correlation performed in the adaptation of α is not biased though. The optimal value of α is still achieved implying $v_e(t)$ and $v_m(t)$ are uncorrelated resulting in no further change in the control voltages. This is in contrast to the case of mixer DC offsets which do bias the correlation. The difference is that mixer DC offsets appear before the integrator in Figure 8 whereas the VM imbalances and offsets occur after the integrator. Thus, even when α is adjusted properly to ensure $v_e(t)$ and $v_m(t)$ are uncorrelated, the control voltages still adapt due to the integration of the DC offset.

3. IMPLEMENTATION

Previously proposed gradient-driven, adaptive feedforward linearizers, e.g. the U.S. Patent [14], employ a bandpass correlation method in which the required gradient signal for the adaptation of α, given by (30), is computed using analog components in a circuit similar to that shown in Figure 10. Note that the signals $\tilde{v}_e(t)$ and $\tilde{v}_m(t)$ in this diagram are real RF signals as opposed to complex baseband signals. Ideally, the upper branch of this circuit performs the following computation

$$\begin{aligned}\mathrm{Re}\left[v_e(t)e^{j2\pi f_c t}\right]\mathrm{Re}\left[v_m(t)e^{j2\pi f_c t}\right] &= \frac{1}{2}\mathrm{Re}\left[v_e(t)v_m^*(t)+v_e(t)v_m(t)e^{j4\pi f_c t}\right] \\ &= \frac{1}{2}\mathrm{Re}\left[D_\alpha(t)\right]+\frac{1}{2}\mathrm{Re}\left[v_e(t)v_m(t)e^{j4\pi f_c t}\right] \end{aligned} \tag{54}$$

and the lower branch performs

$$\begin{aligned}\mathrm{Re}\left[v_e(t)e^{j2\pi f_c t}\right]\mathrm{Re}\left[v_m(t)e^{j\left(2\pi f_c t+\frac{\pi}{2}\right)}\right] &= \frac{1}{2}\mathrm{Re}\left[v_e(t)v_m^*(t)e^{-j\frac{\pi}{2}}+v_e(t)v_m(t)e^{j\left(4\pi f_c t+\frac{\pi}{2}\right)}\right] \\ &= \frac{1}{2}\mathrm{Im}\left[v_e(t)v_m^*(t)-v_e(t)v_m(t)e^{j4\pi f_c t}\right] \\ &= \frac{1}{2}\mathrm{Im}\left[D_\alpha(t)\right]-\frac{1}{2}\mathrm{Im}\left[v_e(t)v_m(t)e^{j4\pi f_c t}\right] \end{aligned} \tag{55}$$

The lowpass filters remove the components centred at $2f_c$ leaving the real and imaginary components of the lowpass gradient signal as desired. Two problems compromise the accuracy of this method though. The first is that DC offsets in the mixers bias the correlation; consequently, α converges to a non-optimal value. The second problem arises in mixing two modulated RF signals in each of the mixers: undesirable intermodulation distortion is generated in the process.

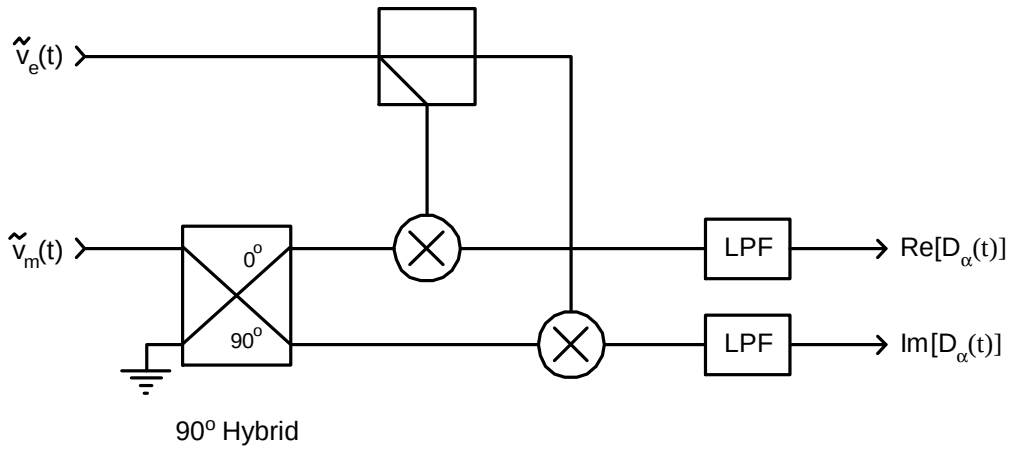


Figure 10. Analog circuit for bandpass computation of gradient signal for adaptation of α

For the adaptation of β a significant drawback to the bandpass correlation method, in addition to the DC offset and IMD problems, is excessive self noise due to the masking of the IM component by the desired signal component in $\tilde{v}_o(t)$. As stated previously, this leads to a long convergence time for β. The insertion of a filter in the adaptation path for β necessary to achieve signal suppression may not be feasible, though, since the design of an appropriate RF bandstop filter is difficult because of the stringent specifications of high stopband attenuation and narrow transition bandwidth.

In contrast, the current work adopts a novel use of DSP to perform baseband correlation, even for signals whose bandwidth far exceeds the available DSP sampling rate. The DSP method facilitates high accuracy gradient calculations and a simple approach to signal suppression, both of which overcome the inherent problems associated with bandpass correlation. In order to perform baseband correlation in DSP, though, the complex envelopes of the RF signals $\tilde{v}_m(t)$ and $\tilde{v}_e(t)$ and $\tilde{v}_o(t)$ must first be recovered as described next.

3.1 Baseband Correlation in DSP

Figure 11 indicates the points in the feedforward circuit where the RF signals $\tilde{v}_m(t)$, $\tilde{v}_e(t)$, and $\tilde{v}_o(t)$ are split to enable downconversion and sampling prior to baseband correlation in DSP. The same oscillators are used to downconvert each of the three RF signals; consequently, any frequency offset is the same amongst all three signals and is thus canceled as analyzed previously. The real and imaginary components of $\alpha(t)$ and $\beta(t)$, calculated in DSP, are used directly as the control signals for the vector modulators which realize the required attenuation and phase shift in the signal and error cancellation circuits.

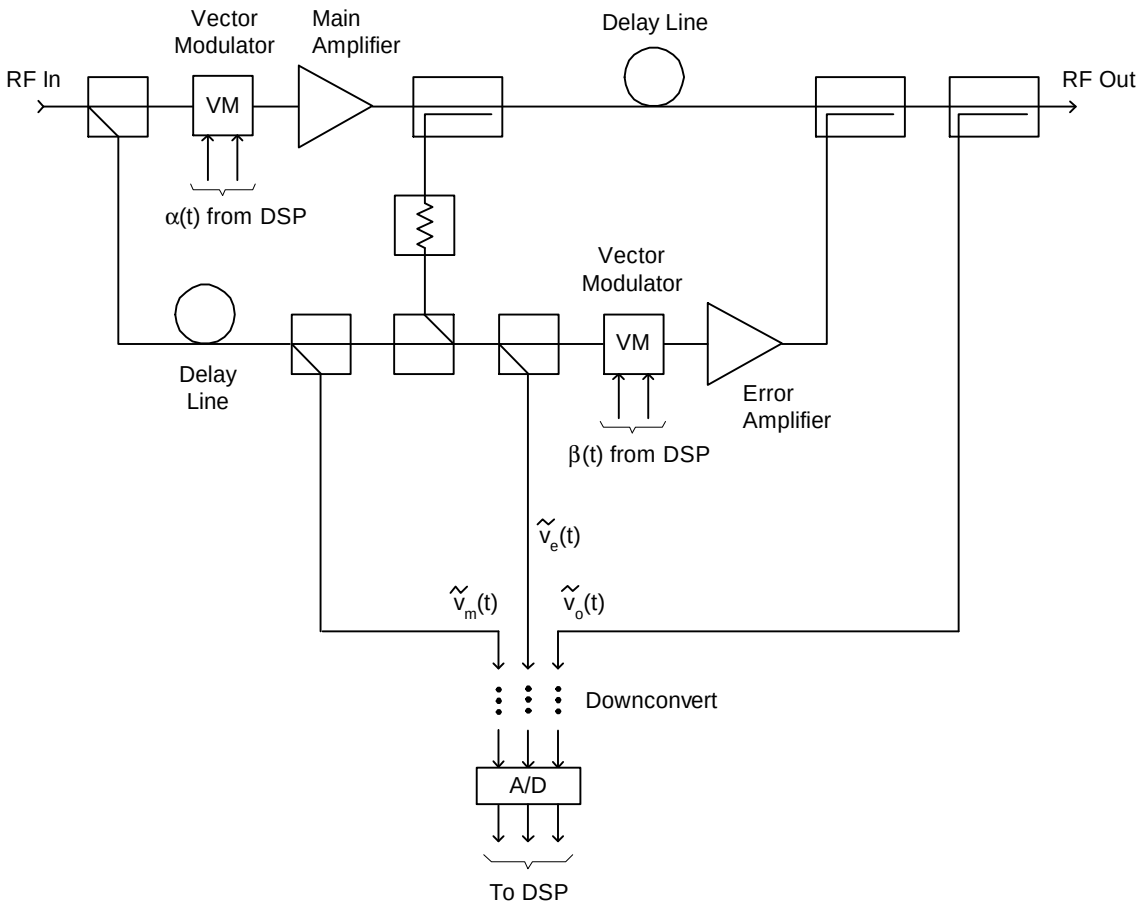


Figure 11. Block diagram of feedforward circuit showing downconversion of signals necessary for baseband correlation

Using Parseval's theorem, the correlation operations performed in the adaptation of α and β can be written as

$$\alpha = K_\alpha \int_{-\infty}^{\infty} v_e(t) v_m^*(t) dt = K_\alpha \int_{-\infty}^{\infty} V_e(f) V_m^*(f) df \tag{56}$$

and

$$\beta = K_\beta \int_{-\infty}^{\infty} v_o'(t) v_e^*(t) dt = K_\beta \int_{-\infty}^{\infty} V_o'(f) V_e^*(f) df \tag{57}$$

where $V_m(f)$ is the Fourier transform of $v_m(t)$—likewise for $V_e(f)$ and $V_o(f)$. Evident from these expressions is that the frequency band of interest for the correlation of $v_e(t)$ and $v_m(t)$ is that occupied by the reference signal only, since $V_e(f)V_m^*(f) = 0$ outside this band. In contrast, for the correlation of $v_o(t)$ and $v_e(t)$, the bands of interest are those occupied by the distortion on either side of the band occupied by the desired signal, since $V_o(f)V_e^*(f) \approx 0$ inside those bands due to signal suppression.

Observing the frequency domain versions of equations (56) and (57), each expression may be rewritten as a sum of integrals each over a narrow subband with the set of subbands spanning the full band of interest; subbands outside the band of interest may be excluded since the integrand is zero. This suggests that correlation of wideband signals may be accomplished in DSP by summing the results of partial correlations performed in the narrow subbands, where the width of the subbands is chosen to suit the available sampling rate. Signal suppression, necessary to speed the convergence of β, may be achieved simply by excluding the band occupied by the desired signal component in $v_o(t)$.

Various subbands of the RF signals may be selectively downconverted to baseband by employing a scanning receiver type architecture as shown in Figure 12, in which the first oscillator LO_1 selects the desired subband of width B, indexed by n. The width of each subband is determined by the bandwidth of the bandpass filter centred at the fixed intermediate frequency f_{IF_1}. The bandwidth of this filter is constrained by the DSP sampling rate f_s: in order to avoid aliasing, B must be chosen such that $B < f_s/2$. In order to avoid mixer DC offsets as well as aliasing, the second intermediate frequency IF_2 must be chosen such that $B/2 < f_{IF_2} < f_s/2 - B/2$; a convenient choice is $f_{IF_2} = f_s/4$.

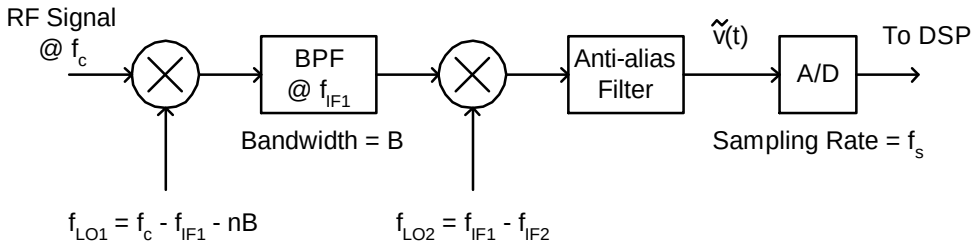


Figure 12. Two-step downconversion process for selection of subbands

The spectrum of a final downconverted signal subband prior to sampling is shown in Figure 13. Note that potential frequency offsets must also be considered in the selection of B and f_{IF_2}. The DC offset is easily eliminated by performing the final downconversion in DSP: the sampled signal subband plus DC offset first undergoes a left spectral shift by multiplying by the complex exponential $e^{-j2\pi f_{IF_2} t}$; the subsequent image at $f = -(2f_{IF_2} + \Delta f)$ plus the tone at $f = -f_{IF_2}$ is then filtered out with a lowpass filter. This suggests that in choosing B and f_{IF_2}, a notch of sufficient width must be left around DC to ensure that the spectrally shifted DC offset at $f = -f_{IF_2}$ may be attenuated by a sufficient degree by the lowpass image suppression filter in DSP.

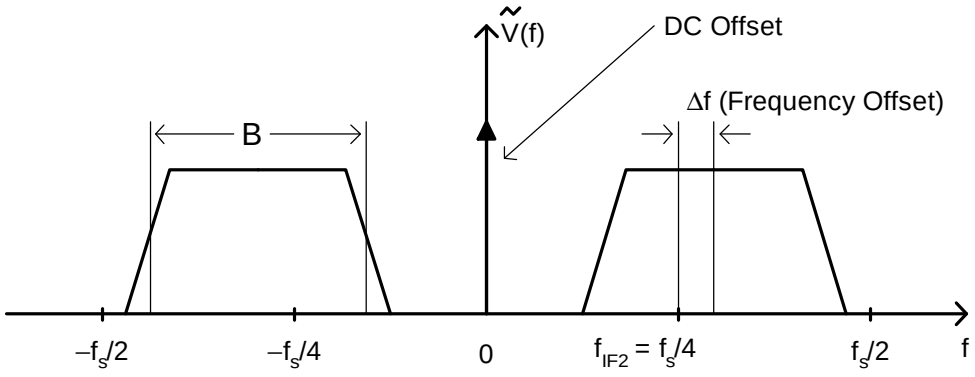


Figure 13. Conceptual diagram showing spectrum of downconverted signal subband

The ability to perform partial correlations in selectable frequency subbands for a wideband input signal is not implemented in the current work due to time and equipment constraints. Instead, correlation is implemented for a sufficiently narrowband input data signal such that the full bandwidth of each of the RF signals $\tilde{v}_m(t)$, $\tilde{v}_e(t)$, and $\tilde{v}_o(t)$ is less than or equal to *B*; hence, *LO1* remains fixed at $f_{LO_1} = f_c - f_{IF_1}$. Since subbands are not selectable in the current narrowband implementation, suppression of the desired signal component of $v_o(t)$ is achieved, instead, by use of a bandstop filter in DSP which selectively attenuates the desired signal component and passes the distortion component. Because the bandstop filter is implemented in DSP rather than at RF as would be necessary, but difficult, in a bandpass correlation scheme, good signal suppression may be achieved with relative ease with a digital filter having high stopband attenuation and sharp transition bands. Future work on adaptive feedforward linearization—a generalization of the current work—will concentrate on the implementation of partial correlation in selectable subbands for a wide bandwidth input signal.

For the narrowband implementation, equation (56) implemented in DSP results in the familiar LMS algorithm

$$\alpha(n) = \alpha(n-1) + \delta_\alpha v_e(n) v_m^*(n) \tag{58}$$

where δ_α is a small step size which determines the convergence rate. Note that the initial value of α is arbitrary, but a reasonable choice is $\alpha(0) = 0$ as can be seen by following the path of descent from the origin to α_{opt} defined by the gradient arrows shown in Figure 7. The adaptation of β, according to equation (57), is implemented in DSP in an analogous fashion:

$$\beta(n) = \beta(n-1) + \delta_\beta v_o'(n) v_e^*(n) \tag{59}$$

where δ_β is the step size parameter. The initial value of β is again arbitrary, but $\beta(0) = 0$ is a reasonable choice.

Shown in Figure 14, is a schematic diagram of the DSP algorithm for the adaptation of α in which the inputs are the sampled versions of the real signals $\tilde{v}_e(t)$ and $\tilde{v}_m(t)$ centred about the frequency $f = f_{IF_2} + \Delta f$ where $f_{IF_2} = f_s/4$. The first operation performed is quadrature demodulation of $\tilde{v}_e(n)$ and $\tilde{v}_m(n)$ followed by lowpass filtering to remove the images at $f = -\left(2 f_{IF_2} + \Delta f\right)$ and the DC offset at $f = -f_{IF_2}$. The next operation is complex conjugate multiplication of the recovered complex envelopes $v_e(n)$ and $v_m(n)$ to produce the gradient estimate $D_\alpha(n)$. Both the real and imaginary components of the gradient estimate are then multiplied by the step size δ_α and accumulated to produce the next update for the real and imaginary components of $\alpha(n)$. The only difference in the schematic for the adaptation of β is that, in order to achieve signal suppression, the lowpass filters used to recover the real and imaginary components of $v_o(n)$ include a stopband around DC equal to the bandwidth of the desired signal component.

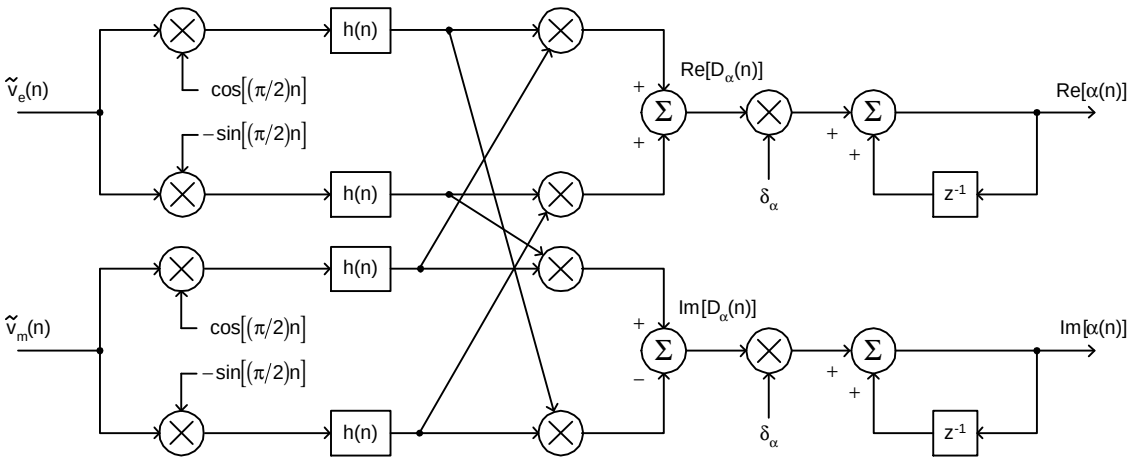


Figure 14. DSP algorithm for adaptation of α

In a wideband implementation, the algorithm for the adaptation of α and β is exactly the same as above, except that the input signals correspond to particular subbands of the original RF signals. The outputs for each subband must then be summed before outputting the next update for the real and imaginary components of α. In contrast to the narrowband case, the DSP lowpass filters are the same for both the adaptation of α and β. Signal suppression is achieved simply by ignoring subbands of $\tilde{v}_o(t)$ that contain the desired signal component.

3.2 Hardware Design

Figure 15 and Figure 16 show schematic diagrams of the RF circuitry in the signal and error cancellation circuits respectively. Primarily, Mini-Circuits™ SMA connectorized components are used in the feedforward linearizer prototype to enable quick construction and modification. The modulation format of the input signal is π/4-DQPSK at a symbol rate of 20 ksym/sec using 35% rolloff root raised cosine filtering. It is generated by an HP8782B Vector Signal Generator at an IF of 185 MHz and

subsequently upconverted to a centre frequency of 815 MHz using a 1 GHz LO. A resonant cavity bandpass filter is used to suppress the image at 1185 MHz.

Note that since the goal of the current work is the demonstration of DSP-based adaptation, the prototype circuit is not designed to optimize power efficiency. Efficiency is determined primarily by the efficiency of the main and error amplifiers, losses in the main branch following the PA, and the value of the output coupler [18]. Efficiency is optimized by minimizing any losses in the main branch as well as selecting the value of the output coupler based on the efficiency of the main and error amplifiers. These design criteria are not followed in the development of the prototype.

Fig. 15

Figure 15. Schematic diagram of signal cancellation circuit

Fig. 16

Figure 16. Schematic diagram of error cancellation circuit

In the signal cancellation circuit, the class AB power amplifier has a gain of 24 dB and output 1-dB compression point of +37 dB_m. The class A preamplifiers each have a gain of 10 dB and a 1-dB compression point of +17 dB_m. The attenuation provided by the vector modulator is chosen to be nominally 15 dB. As shown in Section 3.2.1, this value falls safely within the 7 – 37 dB range of attenuation that the VM is capable of providing. The value of the fixed attenuator in path 2-3 is chosen such that the signal levels at points 3 and 5 are approximately matched. This ensures that when the attenuation and phase of the VM are adjusted precisely under DSP control, complete signal cancellation is achieved. Using a 28-dB fixed attenuator and assuming the VM operates at a nominal attenuation of 15 dB, the gain in path 1-2-3 is approximately -5 dB. The loss in the reference branch between points 4 and 5 is approximately 4 dB (~1 dB due to the 2.8 m coaxial delay line and ~3 dB due to the 3-dB splitter). Thus, the levels at points 3 and 5 are, indeed, approximately matched before the VM is adjusted precisely.

Delay matching in the signal cancellation circuit, accurate to within approximately 1 - 2%, is accomplished by measurement of the group delay in path 1-2-3 using a network analyzer. The delay line is then trimmed such that the delay in reference branch between points 4 and 5 matches this measurement; the necessary delay is 13.1 ns.

In order to demonstrate the distortion suppression capability of the feedforward linearizer, the output power of the PA is chosen to be close to the 1-dB compression point so that a significant level of IMD appears in the PA output signal. The IMD is generated primarily by the PA, although some distortion is also generated by the last stage of the preamplifier chain. Assuming P_a = +35 dB_m, the required linearizer input power is approximately -5 dB_m. Although any distortion generated by the vector modulator in

addition to that generated by the PA may be canceled in the linearizer output, a 6-dB attenuator is inserted in the main branch to ensure that the input power to the vector modulator is at a low enough level to avoid the generation of additional IMD. In this way, distortion is only generated by the amplifiers.

The downconversion chains for $\tilde{v}_m(t)$, $\tilde{v}_e(t)$, and $\tilde{v}_o(t)$ all share the same 745 MHz and 69.9625 MHz oscillators for the first and second downconversions. The first IF is chosen to be the common value of 70 MHz, and the second IF is chosen to be 37.5 kHz—approximately one quarter the DSP sampling rate. Operational amplifiers boost the signal levels of the downconverted signals prior to sampling to ensure that the full ±3 V range of the A/D converters is used. The attenuator prior to the first mixer in each of the downconversion chains is a temporary solution to the problem of various mixing products of the two oscillators leaking back into the signal paths and falling in-band. A proper solution may be to choose a lower frequency for the first IF, e.g. 455 kHz, to allow for a bandpass filter with sharp transition bands to suppress the backwards leakage of the second LO.

In the error cancellation circuit, the class A error amplifier has a gain of 40 dB and output 1-dB compression point of +38 dB_m—approximately the same as the power amplifier. The two class A preamplifiers each have a gain of 20 dB and a 1-dB compression point of +20 dB_m. Such a large error amplifier is found to be necessary because high backoff (output power approximately 30 dB lower than the 1-dB compression point) is required to ensure linear operation. Even with this large an error amplifier, the linearizer input power is limited to approximately -5 dB_m giving $P_a \approx$ +35 dBm. Any higher results in an error signal of sufficiently high power that IMD

generated by the error amplifier shows up in the linearizer output signal. This demonstrates that the linearity requirements of the error amplifier are quite stringent.

The attenuation provided by the vector modulator is chosen, again, to be nominally 15 dB. The total gain in the error cancellation circuit from points 6 to 7 is chosen such that the amplifier IMD levels at points 7 and 8 (determined by $IMSR_a$) are approximately matched. This ensures that when the attenuation and phase of the VM are adjusted precisely under DSP control, complete distortion cancellation is achieved. Assuming the VM operates at a nominal attenuation of 15 dB, the gain in path 6-7 is approximately 49 dB. The loss in path 2-3-6 (~51 dB) minus the loss in the main branch between points 2 and 8 (~1 dB due to the 2.6 m coaxial delay line) is approximately 50 dB. Thus, the levels at points 7 and 8 are, indeed, approximately matched before the VM is adjusted precisely.

Delay matching in the error cancellation circuit, accurate to within approximately 3 - 4%, is accomplished by measurement of the group delay in path 2-3-6-7. The delay line is then trimmed such that the delay in the main branch between points 2 and 8 matches this measurement; the necessary delay is 12.3 ns.

In contrast to the main branch of the signal cancellation circuit, the error cancellation circuit is intolerant to frequency variations across the band of interest. Linear distortion introduced in the main branch of the signal cancellation circuit is ultimately canceled in the linearizer output signal along with nonlinear distortion, but any form of distortion introduced in the error cancellation circuit appears unchanged at the linearizer output. For this reason, components in the error cancellation circuit must have flat amplitude response across the frequency band of interest. This is a concern primarily for

the amplifiers and the vector modulator. For example, to ensure 40 dB distortion cancellation, the accuracy required of the coefficient β is 1% as shown previously. Thus, amplitude ripple across the band of interest must be held to within approximately ±0.1 dB. This accuracy requirement was kept in mind in the selection of components for the error cancellation circuit.

3.2.1 Vector Modulator Design

Figure 17 shows a schematic diagram of the vector modulators used in the signal and error cancellation loops. The VMs were constructed at SFU using surface-mount components mounted on regular double-sided copper clad board. One side of the board serves as the ground plane; the other side is used for traces of appropriate width to ensure a characteristic impedance of 50 Ω. The DC voltages v_I and v_Q are the control inputs for the bi-phase voltage controlled attenuators (VCAs) in each branch of the VM which provide a phase shift of 0 or 180° depending on the polarity of the control voltage. The fact that the two branches of the VM are in phase quadrature and that the VCAs are capable of bi-phase operation, means that the VM can achieve phase shifts anywhere in the range [0,360°]. The structure of the VCAs consists of a diode ring, similar to that of a double-balanced mixer, except that an effort is made towards precise diode matching. Thus, the RF input-output characteristic of each VCA is essentially linear implying the generation of only very low levels of IMD. Measurements show that for a suitably low input power ($\leq$ -15 dB_m), the IMD generated by the vector modulator is at most -60 dB_c.

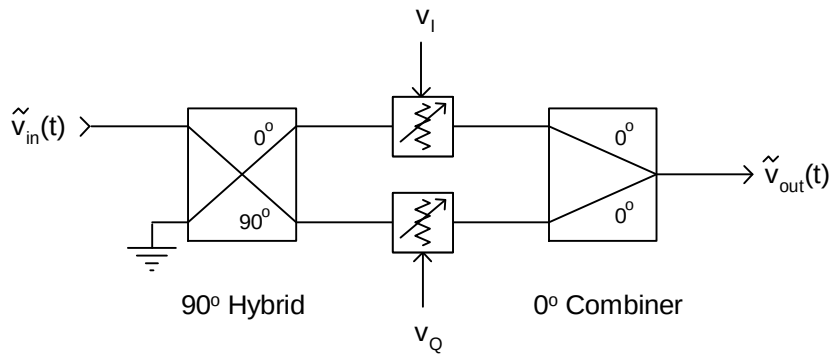


Figure 17. Schematic diagram of vector modulator

The attenuation provided by each VCA is a nonlinear function of the control voltage. The attenuation is maximum for zero voltage and decreases monotonically for increasing control voltage. Measured values of attenuation and phase shift provided by the vector modulator are shown in Figure 18 as a function of $r = \sqrt{v_I^2 + v_Q^2}$ and $\phi = \tan^{-1}\left(v_Q / v_I\right)$ respectively. Clearly, the attenuation provided by the VM behaves in the same fashion as for each individual VCA, namely a monotonic decrease with increasing control voltage. For increasing ϕ, the phase shift provided by the VM varies monotonically as well, changing fastest in mid-quadrant, i.e. around 45°, 135°, 225°, and 315°. The nominal attenuation provided by the vector modulator in the signal and error cancellation circuits is chosen to be 15 dB, since this value is high enough to be on the steep part of the attenuation vs. *r* curve—desirable for quick convergence of the adaptation algorithms for α and β. A nominal value any higher would necessitate additional preamplifiers ahead of both the main and error amplifiers.

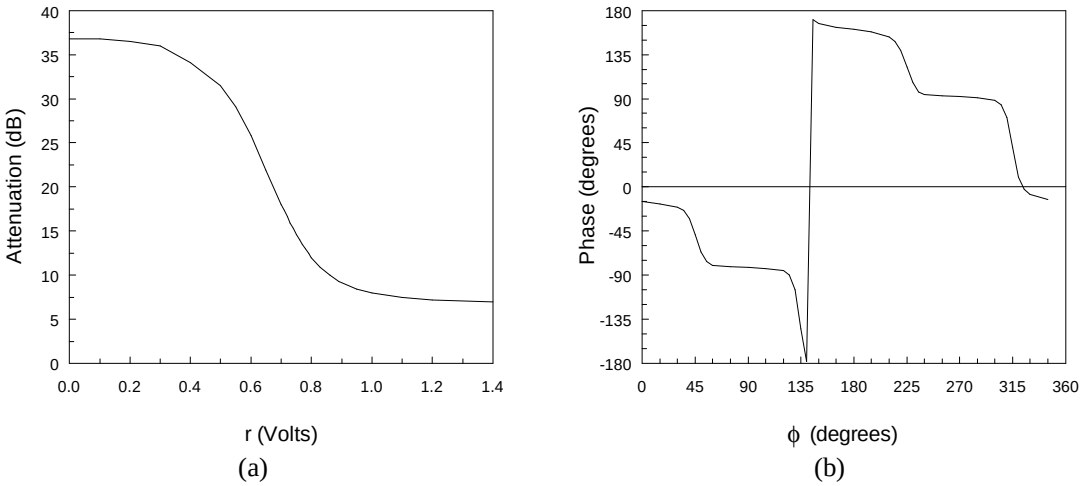


Figure 18. Measured (a) attenuation for $\phi = 45°$ and (b) phase for r = 0.7 V provided by vector modulator

3.3 Software Design

The algorithms for the adaptation of α and β are implemented using the TMS320C30 floating-point digital signal processor which runs at 33 MHz and has an instruction cycle time of 60 ns. A floating-point processor, rather than fixed-point, is used in the development of the adaptive feedforward prototype to speed development time. The processor is mounted on a card manufactured by Spectrum Signal Processing that is housed inside an Intel 80386 based PC. The PC is used to write, assemble, and link the DSP code, and a program written in C is used to download and run the resultant program. The DSP card has two banks of external RAM, each 64 kwords long, and two channels of analog I/O. Each channel consists of 16-bit A/D and D/A converters, each with a ±3V range, along with programmable anti-alias and reconstruction filters; the maximum sampling rate is 151.5 kHz. Since only two channels of analog I/O are available per board, two DSP boards are used for the feedforward prototype: one is for the adaptation

of α with input signals $\tilde{v}_e(t)$ and $\tilde{v}_m(t)$ and output signals Re[$\alpha(t)$] and Im[$\alpha(t)$]; the other is for the adaptation of β, with inputs $\tilde{v}_o(t)$ and $\tilde{v}_e(t)$, and outputs Re[$\beta(t)$] and Im[$\beta(t)$]. In a production version of the feedforward linearizer one fixed-point processor, rather than two floating-point processors, would be used to reduce cost and power consumption; all analog I/O would be handled by this single processor.

Figure 14 introduced in Section 3.1 illustrates the LMS algorithm used for the adaptation of α; the algorithm for β is identical except for the filtering. The first step required is quadrature demodulation of both of the real signals $\tilde{v}_e(n)$ and $\tilde{v}_m(n)$ which is accomplished by multiplication of each signal by $\cos[(\pi/2)n]$ and $-\sin[(\pi/2)n]$. Note that each function alternates between zero and plus or minus one, thus only two multiplications per sample are required for this operation. The next step consists of lowpass filtering the four resultant signals to remove the double frequency images. This requires approximately $4L$ multiplications and additions per sample for FIR filters of length L. The following step is calculation of the real and imaginary components of the gradient estimate which requires four multiplications and two additions per sample. The last step is scaling by the LMS step size parameter and accumulation which requires two multiplications and two additions per sample. Additions may be performed in parallel with multiplications on the C30 processor; thus, the adaptation algorithm requires approximately $8 + 4L$ floating-point instructions per sample.

Since the instruction cycle for the C30 processor is 60 ns and the sampling rate used is 151.5 kHz, the number of available instruction cycles between samples is 110. Neglecting overhead such as memory accesses, context saves and restores, 2's

complement-to-floating point conversions of the input samples and the inverse for the output samples, the maximum FIR filter length is approximately $L = 24$ for sample-by-sample update of α and β. If the above overhead operations are accounted for and/or a longer filter is required, sample-by-sample updates are not possible.

A solution to this processing power constraint is decimation of the filter outputs, i.e. the filter outputs are calculated only every K samples leaving more time to perform the necessary calculations. Decimation is possible because α and β vary much more slowly than the input signals due to the fact that the gradient estimates are continuously accumulated—a lowpass operation which significantly restricts the bandwidth of $\alpha(t)$ and $\beta(t)$. The effect of decimation is that, since updates to α and β are not generated as often, convergence times are greater by a factor of K than if no decimation is required.

For the implementation of adaptation for the feedforward prototype, the decimation factor used is $K = L$. Although a value this large is not actually necessary, it is nonetheless used in order to simplify the assembly code. As shown later in this section, the required filters do not have to be too long; sufficient image suppression may be obtained for a lowpass filter as short as $L = 16$. Thus, optimized code for the C30 processor and a sampling rate of 151.5 kHz might require a decimation factor as low as $K = 2$. Better yet, use of a current generation processor might avoid decimation altogether because of enhanced processing power.

In the actual implementation of adaptation for the feedforward prototype, a simplification is made in Figure 14 by combining the quadrature demodulation operation with the lowpass filters $h(n)$. The result, shown here, is a bandpass filter $\tilde{h}(n)$ centred

about $f = f_s/4$ with complex-valued coefficients. Quadrature demodulation of the real signal $\tilde{v}_e(n)$ and subsequent image suppression is written compactly as

$$y_e(n) = \tilde{v}_e(n) e^{-j\frac{\pi}{2}n} \otimes h(n) \tag{60}$$

where the real and imaginary components of $y_e(n)$ give the outputs of the lowpass filters and $\otimes$ denotes the convolution operation. Expanding and rearranging this expression gives,

$$\begin{aligned} y_e(n) &= \sum_{k=0}^{\infty} h(k)\tilde{v}_e(n-k) e^{-j\frac{\pi}{2}(n-k)} \\ &= e^{-j\frac{\pi}{2}n} \sum_{k=0}^{\infty} \tilde{h}(k)\tilde{v}_e(n-k) \\ &= e^{-j\frac{\pi}{2}n} \left[\tilde{v}_e(n) \otimes \tilde{h}(n)\right] \end{aligned} \tag{61}$$

where $\tilde{h}(n) = e^{j\frac{\pi}{2}n} h(n)$ is the complex bandpass filter derived by spectrally shifting the real lowpass filter $h(n)$ to a centre frequency of $f = f_s/4$. The output of the lowpass filters for $\tilde{v}_m(n)$ is written in an analogous fashion. The gradient estimate is thus

$$\begin{aligned} D_\alpha(n) &= y_e(n) y_m^*(n) \\ &= \left[\tilde{v}_e(n) \otimes \tilde{h}(n)\right]\left[\tilde{v}_m(n) \otimes \tilde{h}(n)\right]^* \end{aligned} \tag{62}$$

Evidently, the exponential factor in (61) is canceled due to the complex conjugate multiplication. Thus, as shown in Figure 19, for each real-valued signal $\tilde{v}_e(n)$ and $\tilde{v}_m(n)$, the quadrature demodulation operation can be neglected, and the real lowpass filters $h(n)$ replaced by the real and imaginary components of the complex bandpass filters $\tilde{h}(n)$. Also shown in this diagram is the decimation of the filter outputs by the factor K. In effect, when $\tilde{v}_e(n)$ and $\tilde{v}_m(n)$ are filtered with $\tilde{h}(n)$, the resultant rotating signal

constellations of the filter outputs are derotated by the complex conjugate multiplication performed in calculating the gradient estimate.

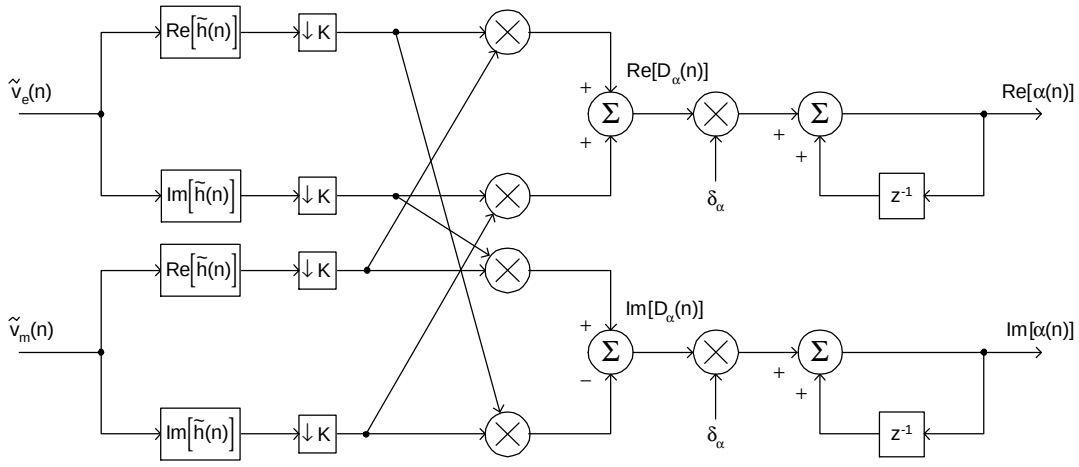


Figure 19. Simplified DSP algorithm for adaptation of α

For the adaptation of β, the appropriate gradient estimate, analogous to (62), is

$$D_{\beta}(n) = \left[\tilde{v}_o(n) \otimes \tilde{h}_o(n)\right]\left[\tilde{v}_e(n) \otimes \tilde{h}_e(n)\right]^* \tag{63}$$

Note that in contrast to the algorithm for α, the complex bandpass filters for the two signals $\tilde{v}_o(n)$ and $\tilde{v}_e(n)$ are different; $\tilde{h}_e(n)$ is similar to $\tilde{h}(n)$ above, but $\tilde{h}_o(n)$ has a notch at band centre to suppress the desired signal component thereby reducing self noise in the gradient estimate (see Section 2.3).

3.3.1 Filter Design

As mentioned previously, the modulation format of the linearizer input signal is π/4-DQPSK with 35% rolloff root raised cosine filtering at a symbol rate of 20 ksym/sec. Use of this symbol rate ensures that aliasing of $\tilde{v}_e(t)$—the widest band signal to be

sampled—is avoided. Remember that $\tilde{v}_m(t)$, $\tilde{v}_e(t)$, and $\tilde{v}_o(t)$ are each downconverted to an IF of 37.5 kHz which corresponds to approximately one quarter the sampling rate.

The magnitude response of the complex bandpass filter $\tilde{h}(n)$ used for both $\tilde{v}_m(n)$ and $\tilde{v}_e(n)$ in the adaptation of α is shown in Figure 20. Note that all filters shown in this section are FIR with linear phase response and are designed in MATLAB using the REMEZ exchange algorithm. In the design of this filter, the passband is chosen to have a width of 25 kHz—wide enough to pass $\tilde{v}_m(n)$ unattenuated. The band edges of $\tilde{v}_e(n)$ are attenuated, but this does not affect the correlation, since the band of interest is that occupied by the reference signal only. The stopband attenuation is chosen, somewhat arbitrarily, to be 50 dB. The filter length *L*, is then minimized to achieve these criteria; the result is the relatively short length of $L = 16$.

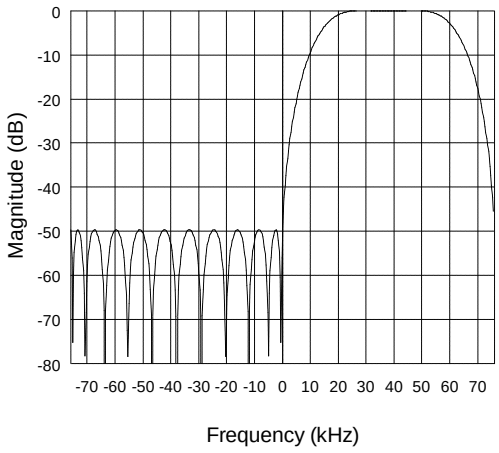


Figure 20. Complex bandpass filter $\tilde{h}(n)$ used for $\tilde{v}_m(n)$ and $\tilde{v}_e(n)$ in adaptation of α

Shown in Figure 21 are power spectra of the filtered reference signal and the filtered error signal for $\alpha = \alpha_{opt}$, i.e. for the reference signal completely canceled. These spectra are estimated by first sampling the signals $\tilde{v}_m(t)$ and $\tilde{v}_e(t)$ and then storing the

samples in a 64 kword memory bank on one of the DSP boards. The data is then uploaded to a file and processed off-line in MATLAB using the same filter shown Figure 20. The ability to process actual signals off-line in simulation studies demonstrates a hidden advantage of performing baseband correlation. Evident from these spectra is that the filtered signals contain no DC offsets due to the high attenuation provided by the filters at DC. This is one of the fundamental, and novel, aspects of using DSP to perform baseband correlation. Remember, in viewing these spectra, that incomplete image suppression does not bias the correlation of the two signals as shown in Section 2.5

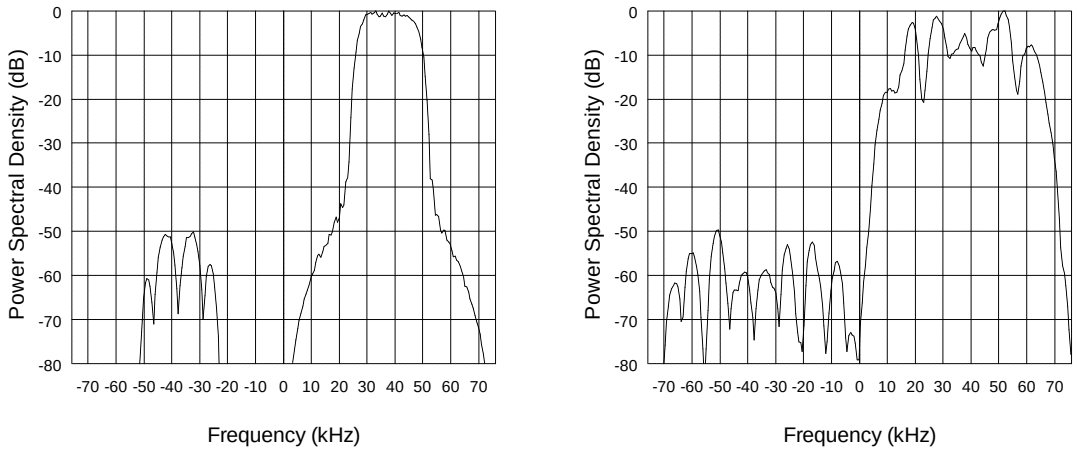


Figure 21. (a) Reference signal and (b) error signal for $\alpha = \alpha_{opt}$ both filtered using $\tilde{h}(n)$

The complex bandpass filter $\tilde{h}_e(n)$ used for $\tilde{v}_e(n)$ and $\tilde{h}_o(n)$ used for $\tilde{v}_o(n)$ in the adaptation of β are shown in Figure 22. In the design of both filters, the transition bands are chosen to be significantly narrower than for $\tilde{h}(n)$ so that the amplifier distortion in $\tilde{v}_e(n)$ and $\tilde{v}_o(n)$ is passed unattenuated. The reason for this is that the band of interest for the correlation of $\tilde{v}_e(n)$ and $\tilde{v}_o(n)$ is that occupied by the amplifier distortion outside the band of the desired signal. For $\tilde{h}_o(n)$ specifically, a notch in the bandpass response is

designed in order to suppress the desired signal component of $\tilde{v}_o(n)$. The minimum width of the notch is chosen to be 25 kHz to cover virtually the whole band of the desired signal. The stopband attenuation is chosen to be high (60 dB) in order to realize the advantages of good signal suppression, namely faster convergence and reduced bias in β, hence relaxed accuracy requirements for α. This is the other fundamental, and novel, aspect of using DSP to perform baseband correlation. Using the criteria of narrow transition bands and high stopband attenuation, the length of $\tilde{h}_o(n)$ is minimized to achieve these criteria; the result is $L = 53$—roughly three times longer than the filters used in the adaptation of α. The length of $\tilde{h}_e(n)$ is also $L = 53$ so that both signals experience equal group delay. Note that in a wideband implementation in which subbands of the RF signals are selectable, the filters used in the adaptation of β are identical to those used for α—simply complex bandpass filters with no notch. Suppression of the desired signal component is achieved, instead, by neglecting the subbands of $\tilde{v}_o(t)$ occupied by the desired signal component.

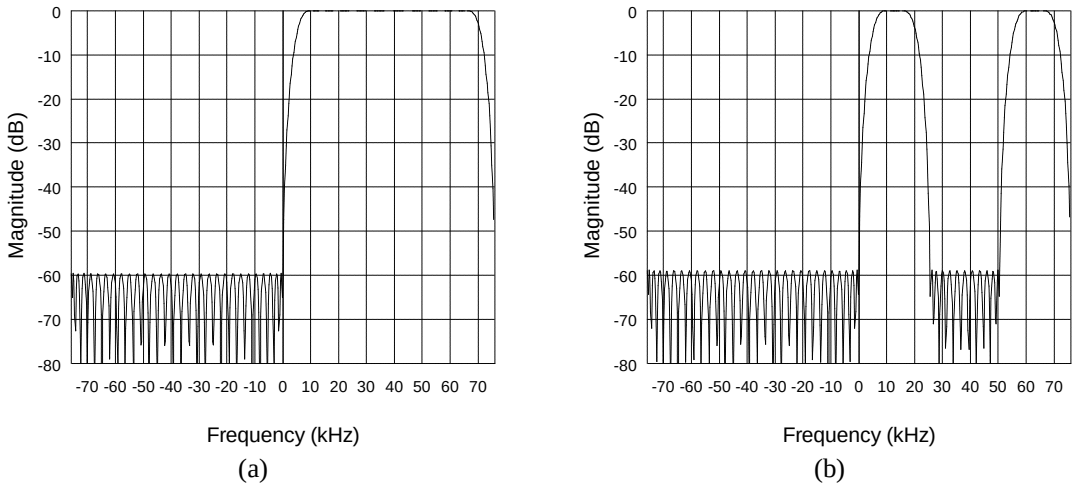


Figure 22. Complex bandpass filters (a) $\tilde{h}_e(n)$ used for $\tilde{v}_e(n)$ and (b) $\tilde{h}_o(n)$ used for $\tilde{v}_o(n)$ in adaptation of β

Shown in Figure 23 is the power spectrum of the filtered error signal, this time using $\tilde{h}_e(n)$, and the spectrum of the filtered linearizer output signal using $\tilde{h}_o(n)$ for $\beta = 0$, i.e. no distortion cancellation. Clearly, the desired signal component of $\tilde{v}_o(n)$ has been suppressed by a sufficient degree so as not to mask the distortion component. Also, the band edges of $\tilde{v}_e(n)$ are attenuated by a lesser degree than $\tilde{v}_e(n)$ in Figure 21 due to the sharper filter transition bands. Note again that no DC offsets are present in either of the filtered signals.

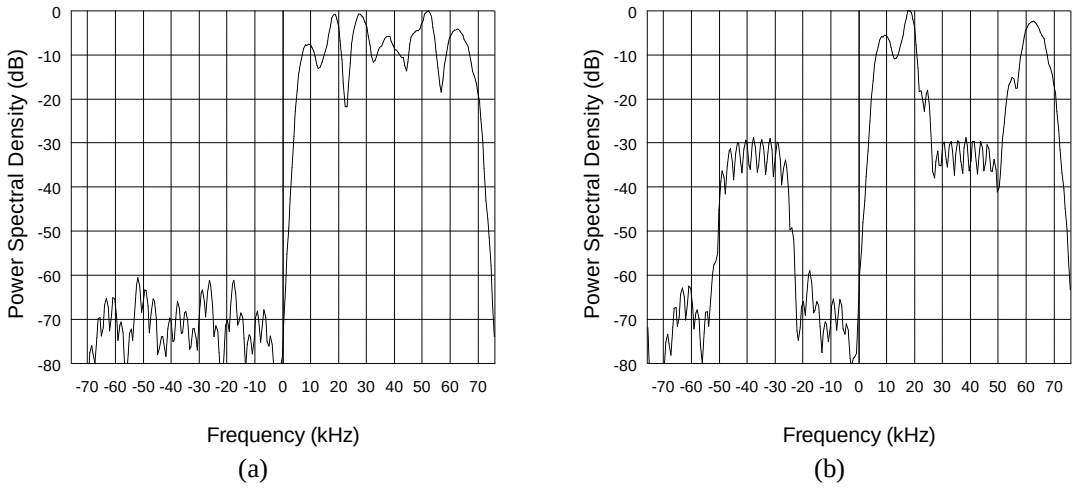


Figure 23. (a) Error signal for $\alpha = \alpha_{opt}$ filtered using $\tilde{h}_e(n)$ and (b) linearizer output signal for $\beta = 0$ (no distortion cancellation) filtered using $\tilde{h}_o(n)$

3.3.2 TMS320C30 Assembly Code Design

Since all of the operations in the algorithm for the adaptation of α and β are time-critical, the DSP software implementing adaptation is written in TMS320C30 assembly language. The assembly code for α is in Appendix A, and that for β is in Appendix B. An attempt is made at providing enough comments along with the code so that the flow may be followed effectively.

Signal sampling is handled by the use of an interrupt service routine. An on-chip counter/timer is configured to generate a clock signal at the sampling rate of 151.5 kHz. This corresponds to a clock period of 6.6 μs or 110 machine cycles. Each period, the clock triggers both the A/D and D/A to start a conversion; the A/D conversion takes roughly 5 μs. When the conversion is complete, the A/D triggers the INT1 interrupt pin on the DSP. When the interrupt is acknowledged, execution branches to the interrupt

service routine (ISR). Note that the DSP board is designed in such a way that input samples are read from and output samples are written to the same memory mapped register connected to both the A/D and D/A converters. Thus, input samples must be read from this register before the output samples are written. Approximately 1.6 μs (27 machine cycles) into the ISR, the next clock period begins; thus, immediately, the current sample is read, and the previous sample written before the next conversion begins to avoid conflicts in the single register connected to both the A/D and D/A converters. This also ensures a delay of only 1 sample from input to output.

As mentioned previously, the easiest decimation factor to use from a software standpoint is $K = L$. Thus, the filter outputs are only calculated every L samples. This reduction in processing means that $110L$, instead of 110, machine cycles are available for calculating the filter outputs and the next update to either α or β—more than adequate time. Of course the input sampling rate remains the same, so buffering of the input samples is necessary. For $K = L$, a double-buffering technique is used: input samples are stored in a buffer of length L; when this buffer is full, a second buffer is filled during which time the samples in the first buffer are filtered and the next update of α or β is generated. Filtering the buffers of length L simply involves calculating the inner product of the buffer contents and the filter coefficients. These calculations finish long before the second buffer is full. When the second buffer is full, the buffers are switched, and the process is repeated. Addressing of the double buffer is simple since the two buffers are arranged contiguously in memory and the top of the first buffer is aligned on an R-bit boundary, where $R = \log_2(2L)$ if L is a power of 2, or $R = \text{floor}[\log_2(2L)] + 1$ if not. The modulo

two addition of L to the starting address of one buffer gives the starting address of the other buffer.

Since processing of one buffer must take place while the other buffer is being filled, multi-tasking techniques are used. The ISR runs in the foreground and the filtering in the background; thus, context saves and restores are required when switching between foreground and background processes.

The primary functions of the ISR (foreground process) are the following:

- Save context of background process
- For the adaptation of α, read $\tilde{v}_m(n)$ and $\tilde{v}_e(n)$ input samples from input register
- For the adaptation of β, read $\tilde{v}_e(n)$ and $\tilde{v}_o(n)$ input samples from input register
- Write current α or β output samples to output register
- Convert input samples from 2's complement integer format to floating-point format
- Store each input sample in appropriate double buffer
- Check for end of buffer
- If at end, switch buffers and set flag indicating buffer full
- Restore context of background process

The background process continually polls the flag (set by the ISR) indicating whether or not the buffers are full. When the flag is set, the following primary functions are carried out:

- For the adaptation of α, filter both $\tilde{v}_m(n)$ and $\tilde{v}_e(n)$ buffers with coefficients of Re[$\tilde{h}(n)$] and Im[$\tilde{h}(n)$] using FIR subroutine
- For the adaptation of β, filter $\tilde{v}_e(n)$ buffer with coefficients of Re[$\tilde{h}_e(n)$] and Im[$\tilde{h}_e(n)$]; filter $\tilde{v}_o(n)$ buffer using coefficients of Re[$\tilde{h}_o(n)$] and Im[$\tilde{h}_o(n)$] using FIR subroutine
- Calculate gradient estimate
- Multiply by LMS step size parameter
- Accumulate scaled gradient estimate
- Convert update to α or β in floating-point format to 2's complement integer format
- Clear flag indicating buffer full
- Continue polling for next flag set

4. RESULTS

Shown in Figure 24 is the convergence behaviour of α and β via plots of the actual vector modulator control voltages. Note that the control voltage data is collected by storing the output samples of α and β in memory on the DSP boards and then uploading the data to a file for plotting. The convergence behaviour shows some differences from what one would expect from a first-order control system. Specifically, the curves for both α and β show regions of fast and slow convergence, and the curves for β show a degree of overshoot. The regions of fast and slow convergence are most likely due to the nonlinear control characteristics of the vector modulators and possibly due to the shape of the non-quadratic error surface in the case of α. Recall that the nominal operating point of the vector modulators is chosen to correspond to the steep portion of the attenuation vs. control voltage curve to allow for quick convergence near the optimal coefficient values. The overshoot may be due to the 53 sample delay between updates of β by virtue of the decimation required in the DSP algorithm, though a more thorough investigation is required to adequately explain this behaviour.

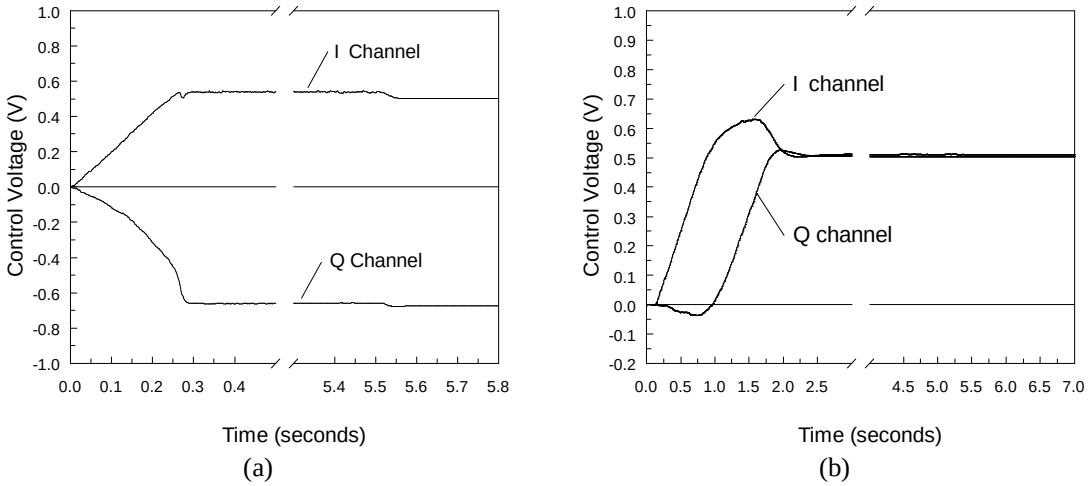


Figure 24. Convergence of (a) α and (b) β for narrowband $\pi/4$-DQPSK input signal (PA output power $\approx +35$ dB_m)

The initial convergence time for α is approximately 0.3 sec, and that for β is approximately 2.5 sec. These values are pessimistic by an order equal to the decimation factor K = 16 for α and K = 53 for β. Scaling the results by these factors gives an initial convergence time for α of approximately 20 msec, and for β approximately 50 msec. Due to the good signal suppression achieved by the filter in the adaptation path for β, the initial convergence time for β is of the same order of magnitude as that for α. This is in stark contrast to implementations relying on a bandpass correlation method without signal suppression in which the convergence time for β can be several orders of magnitude greater than for α.

Once initial convergence is achieved, the input power is dropped by a factor of 6 dB at about $t = 5.55$ sec in order to measure reconvergence time which is usually of more interest than the initial convergence time. A 6 dB decrease in input power might correspond, for example, to a drop in 3 carriers out of 4 in a multicarrier application, or a drop due to power control at a cellular basestation. The result is an approximate 3 msec

reconvergence time for α (when scaled by K = 16) and no change in β, implying no loss in IMD suppression—a very satisfying result. Notice that the jitter in the converged values of α and β decreases when the input power drops which may be explained simply by the fact that for a lower input power, the amplitude of the gradient signals is smaller. Thus, since the LMS step-size parameters δ_α and δ_β in equations (58) and (59) remain fixed, the incremental updates made to α and β are smaller.

Presently, the step-size parameters are selected empirically: δ_α and δ_β are chosen to minimize initial convergence time while maintaining the jitter in the converged values of α and β at a low enough level such that the degree of distortion suppression is not compromised. An automated technique for the selection of the appropriate step-size parameters might consist of estimating the average power of the reference signal in DSP and choosing the appropriate step-size from a set of pre-determined values stored in a look-up table. The step-size could then be adjusted if the input power changes. A "gear-shifting" technique could also be used in which the initial step-size is decreased after both α and β have converged such that jitter in the converged values is minimized, thus maximizing distortion suppression.

In order to measure the distortion improvement offered by the adaptive feedforward linearizer, all measurements of distortion cancellation, unless otherwise noted, are made for a PA output power of approximately +35 dB_m—about 2 dB below the amplifier's output 1-dB compression point. This value ensures that a significant level of IMD is generated by the PA. Measurements show that the corresponding output power of the class A error amplifier—with output 1-dB compression point of +38 dB_m—is approximately +12 dB_m. This is just low enough to ensure that no additional IMD is

generated by the error amplifier, which indicates that significant backoff is required to ensure linear operation. This may be due to the fact that the error signal has a high peak-to-average power ratio. A histogram of the instantaneous power of the error signal is shown in Figure 25, in which the average power is normalized to unity. Note that the error signal data is collected for a PA output power 3 dB lower than that stated above. The peak-to-average power ratio under these conditions is approximately 9 dB, whereas the peak-to-average power ratio for the input $\pi/4$-DQPSK signal with 35% rolloff root raised cosine filtering is approximately 2.5 dB. For a PA output power 3 dB higher, the peak-to-average power ratio of the error signal may be even higher than 9 dB.

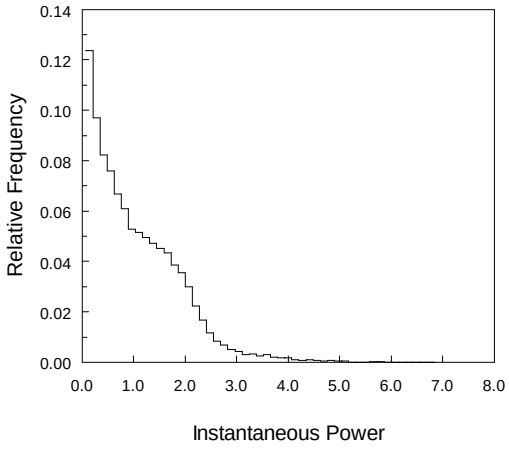


Figure 25. Histogram of instantaneous power of error signal normalized to unity average power for $\pi/4$-DQPSK input signal (PA output power $\approx$ +32 dB_m)

Spectra of both the narrowband input data signal and the error signal are shown in Figure 26. For the measurement of the error signal spectrum, α is allowed to converge to its optimal value. Clearly, the reference signal is canceled from the error signal to a high degree; as a result, the two signals are virtually uncorrelated.

Fig. 26

Figure 26. Spectra of π/4-DQPSK input signal and error signal with reference signal completely canceled

Illustrated in Figure 27 is the degree of distortion cancellation offered by the adaptive feedforward linearizer for a narrowband data signal. The peak distortion improvement is on the order of 30 dB and results in the linearizer output signal being virtually identical to the original input signal as shown in Figure 28. Note that the spectrum of the linearizer output signal before distortion cancellation is measured by allowing α to converge to its optimal value and disconnecting the error cancellation circuit from the output coupler.

Fig. 27

Figure 27. Spectrum of linearizer output signal before and after distortion cancellation

Fig. 28

Figure 28. Spectra of linearizer output signal after distortion cancellation and linearizer input signal

In order to assess the bandwidth of the feedforward circuit, a single tone, offset from band centre by 2.5 MHz, is injected at the linearizer input as shown in Figure 29. Since the tone is offset from band centre by a large degree compared to the bandwidth of the data signal, it is attenuated by the anti-alias filters at the DSP inputs. Thus, the tone is

effectively transparent to the adaptation of α and β. Note that the span of the spectrum analyzer is set at 20 MHz; thus, no detail can be seen in the narrowband data signal at 815 MHz.

Fig. 29

Figure 29. Spectrum of narrowband $\pi/4$-DQPSK input signal at 815 MHz plus tone at 812.5 MHz

The injected tone and the data signal form various intermodulation products (IMPs), as shown in Figure 30, when amplified by the nonlinear PA. Inspecting each IMP more closely, the products at odd multiples of 2.5 MHz contain a tone of varying levels plus some form of distortion of the data signal; the products at even multiples contain no tone.

Fig. 30

Figure 30. Spectrum of amplifier output signal for narrowband π/4-DQPSK input signal at 815 MHz plus tone at 812.5 MHz

Illustrated in Figure 31 is the degree of distortion cancellation achieved by the feedforward linearizer across a wide bandwidth. Cancellation of the IMP at 817.5 MHz is approximately 40 dB—a very good result. This indicates that the bandwidth of the circuit for 40 dB distortion suppression is at least twice the frequency offset of the injected tone, i.e. at least 5 MHz. In fact, additional measurements for varying tone offsets show that 40 dB distortion suppression is achievable for offsets up to 3.5 MHz indicating an effective bandwidth of 7 MHz. For bandwidths wider than 7 MHz, the distortion cancellation degrades. This is consistent with the theory presented previously. For 40 dB distortion cancellation across a maximum bandwidth of W, the product $2\pi(W/2)\tau_\beta$ cannot exceed 0.01 at the band edges where τ_β is the delay mismatch in the error cancellation circuit (see Section 2.4). Thus, for $W = 7$ MHz, τ_β must be less than approximately 0.45 ns. The delay line in the error cancellation circuit has a delay of approximately 12.3 ns, thus the relative error in τ_β cannot exceed approximately 3.5%. As mentioned

previously, the delay matching in the error cancellation circuit is accurate to within 3 to 4%; thus, the degradation in distortion cancellation outside a bandwidth of 7 MHz, observed in practice, agrees with that predicted in theory. If operation across a wider bandwidth is required, the delay matching in the error cancellation circuit must be made more accurate by trimming the delay line to give precisely the correct delay, although this becomes more difficult for increasing bandwidths. For bandwidths on the order of several tens of MHz or more, adaptive delay matching, as described in [19], may be necessary.

Fig. 31

Figure 31. Spectrum of linearizer output signal before and after distortion cancellation for narrowband $\pi/4$-DQPSK input signal at 815 MHz plus tone at 812.5 MHz

5. CONCLUSIONS

An extension to previous analysis of adaptive feedforward linearization has been made by considering placement of the signal cancellation coefficient (α) in the main branch of the signal cancellation circuit ahead of the power amplifier, rather than in the reference branch. New results show that for α in the main branch, the error surface, corresponding to the power in the error signal, is a non-quadratic function of α. As a result, the criteria of minimum power of the error signal and zero correlation between the error and reference signals (equivalent for quadratic error surfaces) generally lead to different optimal values of α. As a consequence of the non-quadratic error surface, formation of a gradient signal through the covariance of the error and reference signals does not give the true path of steepest descent towards the minimum of the error surface, but does give a path that leads towards the point corresponding to the optimal value of α defined using the criterion of zero correlation. In general, the optimal value of α corresponding to minimum power does not result in complete signal cancellation. In contrast, use of the zero correlation criterion—a necessity for gradient adaptation to avoid bias in the error cancellation coefficient—does result in complete signal cancellation, but the power of the error signal can be higher having possible implications for the selection of an appropriate error amplifier. Numerical analysis, though, shows that for the PA backed off by several dB, the different optimal coefficient values become coincident.

The effects of various implementation inaccuracies, such as frequency offsets, phase offsets, filter mismatches, and incomplete image suppression in the recovery of the complex envelopes were analyzed as well as amplitude imbalances, phase imbalances, and

DC offsets in the vector modulators. This new analysis shows that these inaccuracies do not affect the final converged coefficient values.

An adaptive feedforward amplifier linearizer for an 815 MHz, 5-Watt, class AB power amplifier has been implemented with a novel use of DSP for baseband calculation of the necessary gradient signals to drive the adaptation. Use of DSP has been shown to overcome the two main problems that compromise the accuracy and convergence time of previously proposed analog implementations that rely on bandpass gradient calculations. The problem of mixer DC offsets, causing incorrect coefficient convergence, has been solved by accurately recovering the complex envelopes of RF signals in DSP before performing the gradient calculations. The masking problem, causing slow convergence of the error cancellation coefficient, has been solved by easily suppressing the desired signal component of the linearizer output signal with a suitable filter in DSP.

The adaptive feedforward prototype shows 40 dB cancellation of intermodulation products across a bandwidth of at least 7 MHz with extremely fast tracking. Initial coefficient convergence occurs within approximately 50 msec of start-up, and following a sudden change in input signal level by 6 dB, reconvergence occurs in approximately 3 msec with no loss in IM suppression.

REFERENCES

[1] Private Communication, A.S. Wright, NovAtel, 1992.

[2] J.F. Wilson, "The TETRA system and its requirements for linear amplification," *IEE Colloquium on 'Linear RF Amplifiers and Transmitters,'* Digest no. 1994/089, 1994, pp. 4/1-7.

[3] Y. Akaiwa and Y. Nagata, "Highly efficient digital mobile communications with a linear modulation method," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 890-895, June 1987.

[4] Y. Nagata, "Linear amplification technique for digital mobile communications," *Proceedings of IEEE Vehicular Technology Conference*, 1989, pp. 159-164.

[5] J.K. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 4, pp. 374-382, November 1990.

[6] D. Hilborn, S.P. Stapleton, and J.K. Cavers, "An adaptive direct conversion transmitter," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 223-233, May 1994.

[7] H. Seidel, "A microwave feed-forward experiment," *Bell Systems Technical Journal*, vol. 50, no. 9, pp. 2879-2918, November 1971.

[8] S. Kumar and G. Wells, "Memory controlled feedforward lineariser suitable for MMIC implementation," *IEE Proceedings-H*, vol. 138, no. 1, pp. 9-12, February 1991.

[9] T.J. Bennet and R.F. Clements, "Feedforward—an alternative approach to amplifier linearization," *The Radio and Electronic Engineer*, vol. 44, no. 5, pp. 257-262, May 1974.

[10] H.S. Black, "Inventing the negative feedback amplifier," *IEEE Spectrum*, pp. 55-60, December 1977.

[11] P.B. Kenington and D.W. Bennett, "Linear distortion correction using a feedforward system," *IEEE Transactions on Vehicular Technology*, vol. 45, no. 1, pp. 74-81, February 1996.

[12] S. Narahashi et al, “Feed-forward amplifier,” U.S. Patent 5,166,634, November 24, 1992.

[13] M.G. Obermann and J.F. Long, “Feed-forward distortion minimization circuit,” U.S. Patent 5,077,532, December 31, 1991.

[14] R.H. Chapman and W.J. Turney, “Feedforward distortion cancellation circuit,” U.S. Patent 5,051,704, September 24, 1991.

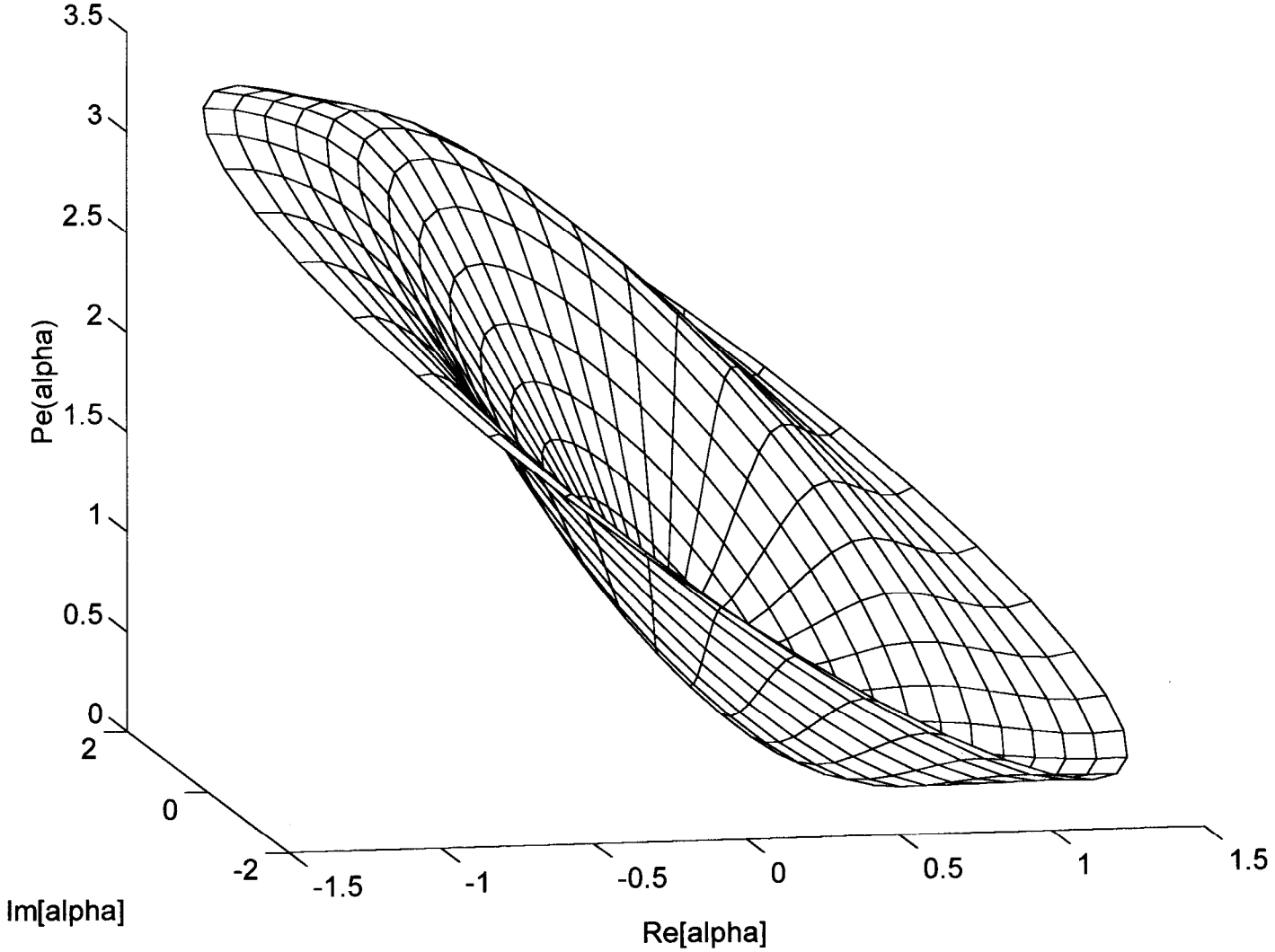
[15] R.M. Bauman, “Adaptive feed-forward system,” U.S. Patent 4,389,618, June 21, 1983.

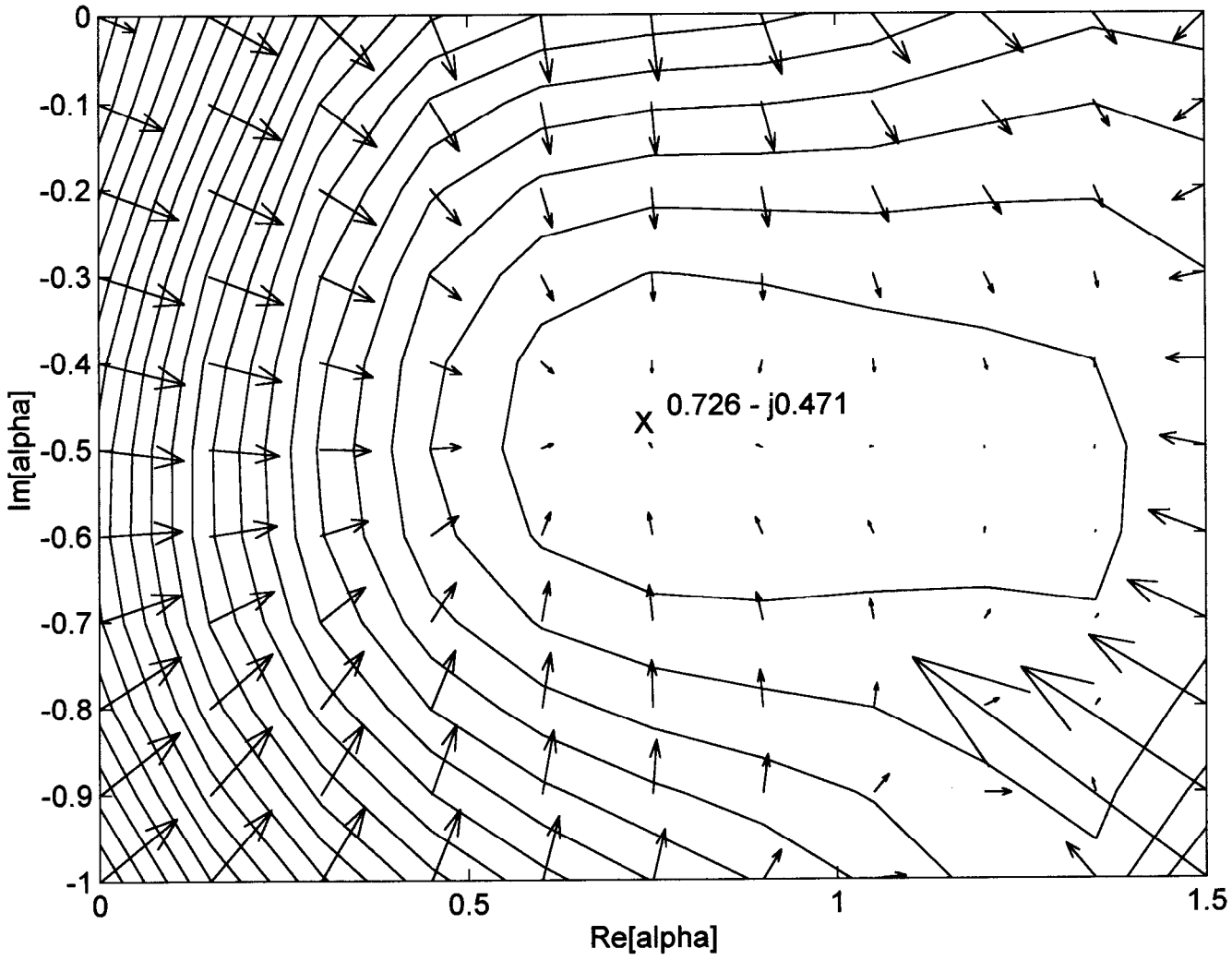
[16] J.K. Cavers, “Adaptation behavior of a feedforward amplifier linearizer,” *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 31-40, February 1995.

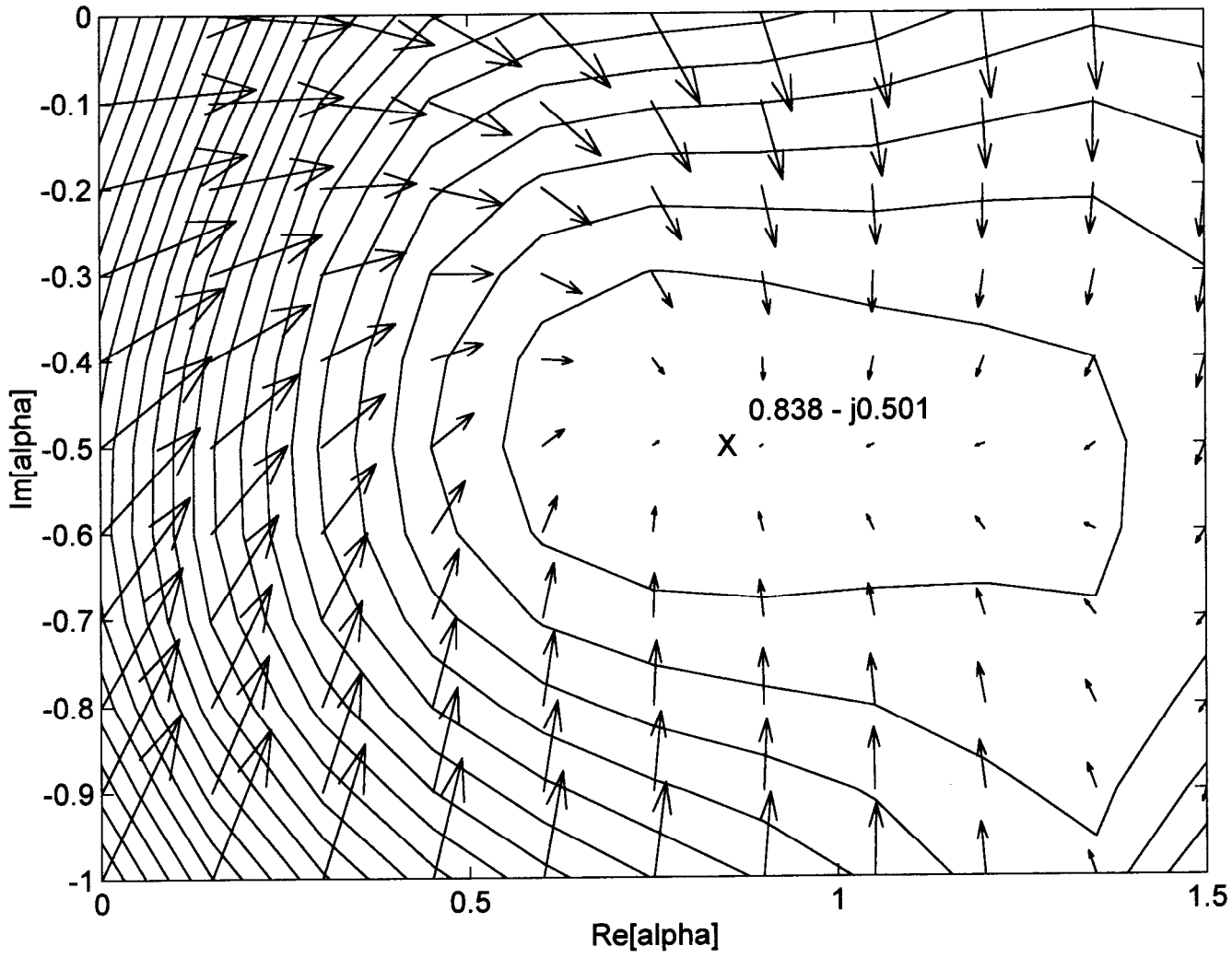
[17] S. Narahashi and T. Nojima, “Extremely low-distortion multi-carrier amplifier—Self-adjusting feed-forward (SAFF) amplifier,” *Proceedings of IEEE International Communications Conference*, 1991, pp. 1485-1490.

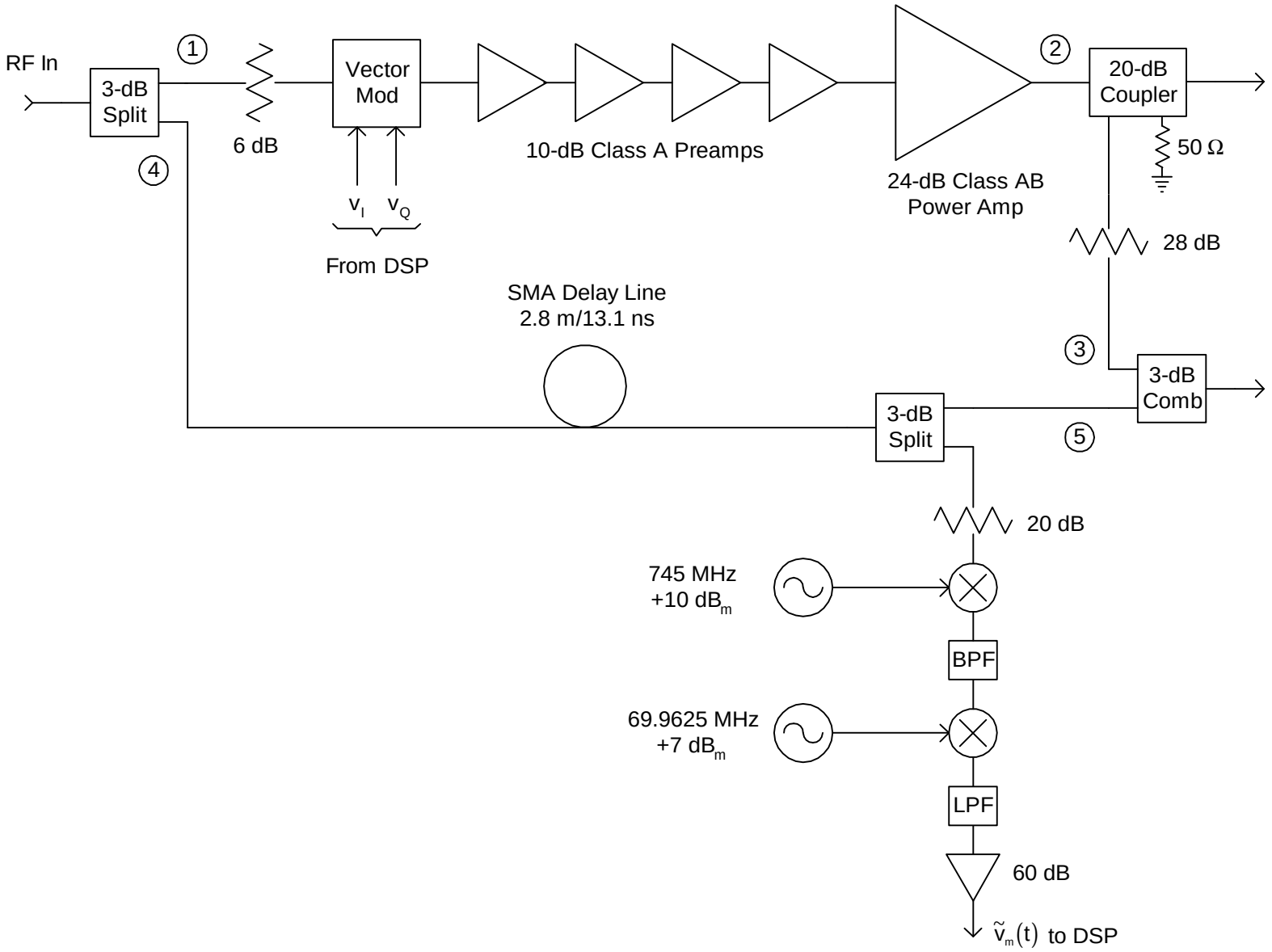
[18] K.J. Parsons and P.B. Kenington, “The efficiency of a feedforward amplifier with delay loss,” *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 407-412, May 1994.

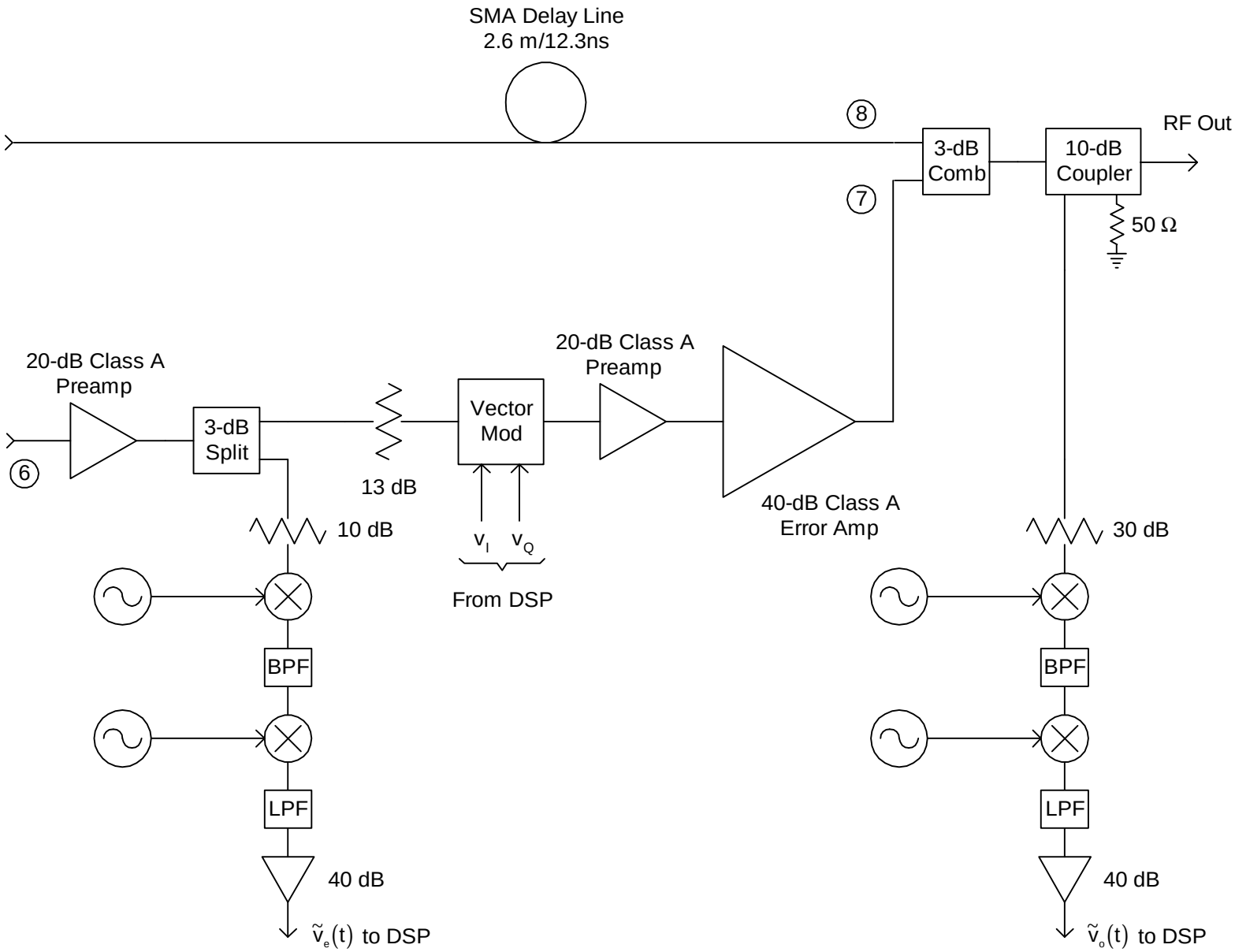
[19] J.K. Cavers, “Adaptive Feedforward Linearizer for RF Power Amplifiers,” U.S. Patent 5,489,875, February 6, 1996.

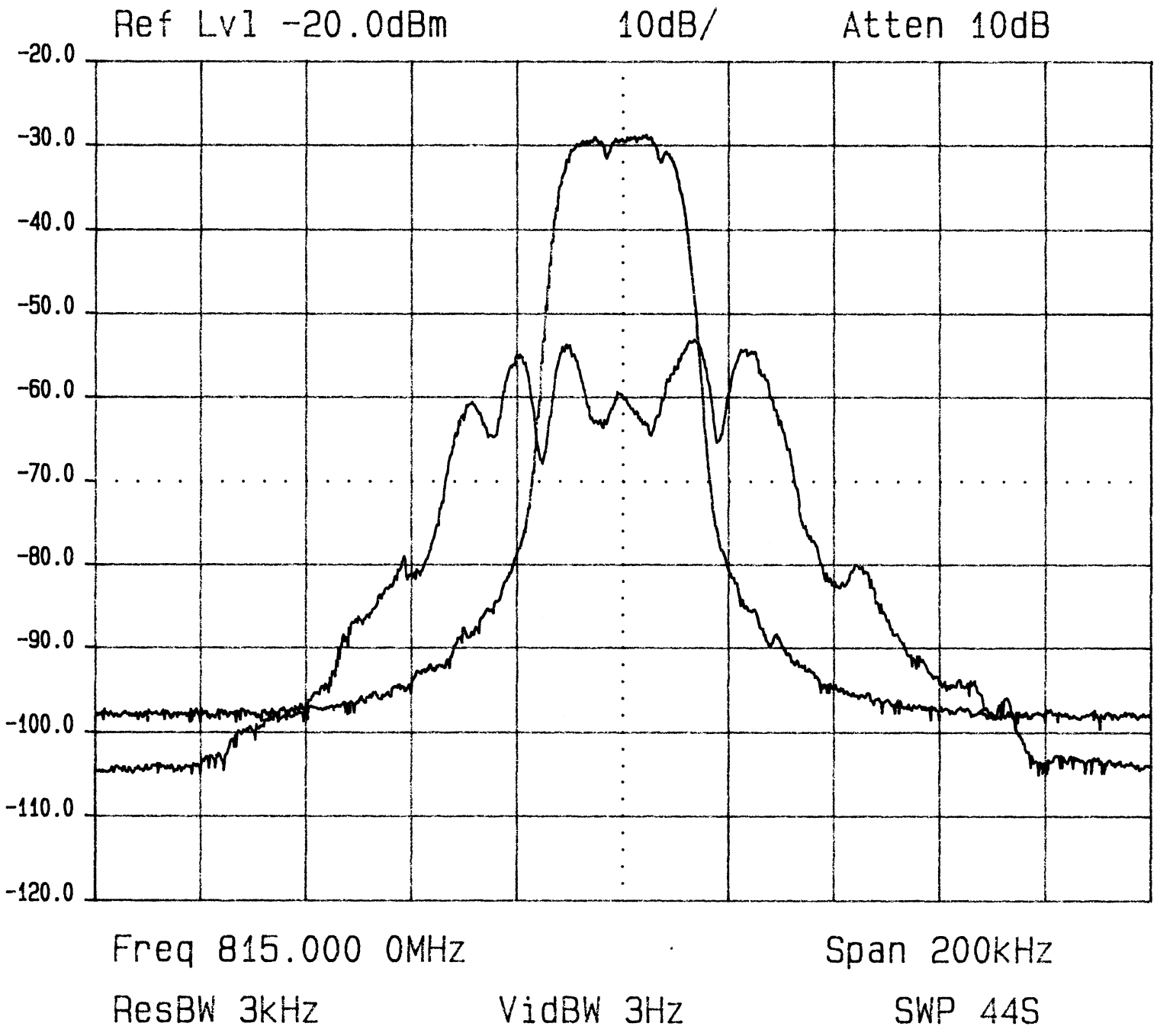

3.5
3
2.5
2
1.5
1
0.5
0
Pe(alpha)
2
0
-2
Im[alpha]
-1.5
-1
-0.5
0
0.5
1
1.5
Re[alpha]

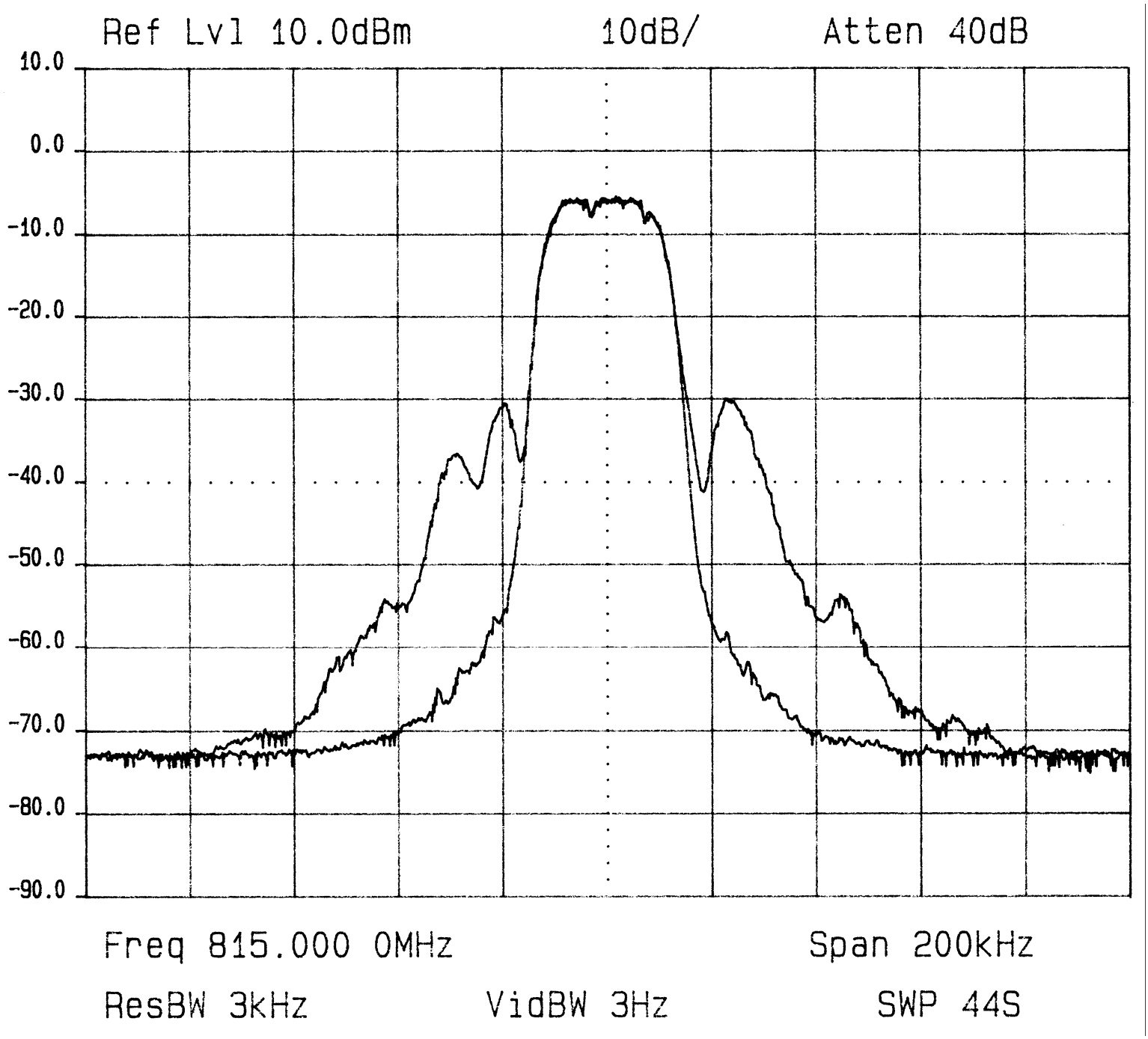

0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1
Im[alpha]
0
0.5
1
1.5
Re[alpha]
0.726 - j0.471

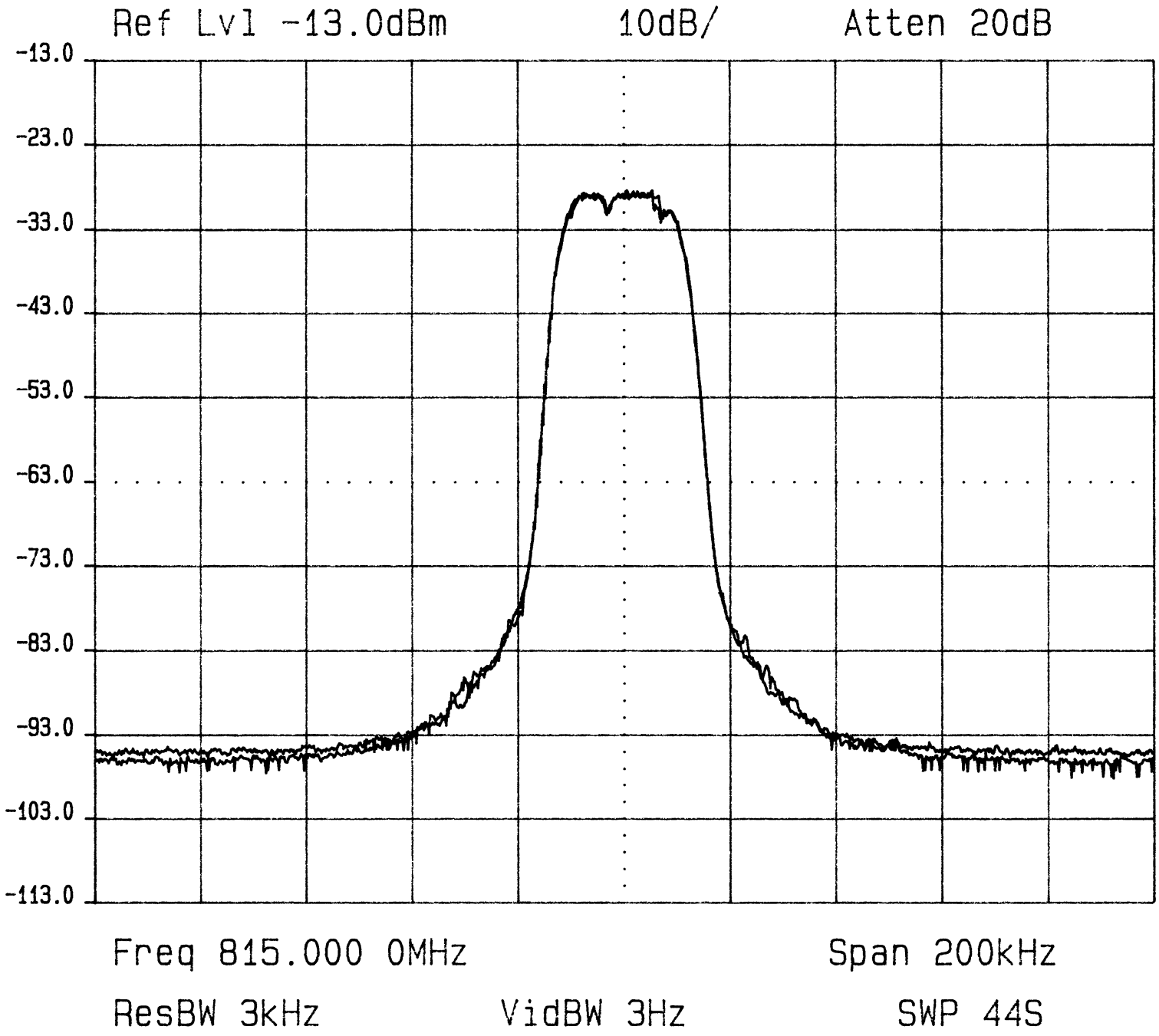

0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1
Im[alpha]
0
0.5
1
1.5
Re[alpha]
0.838 - j0.501
X

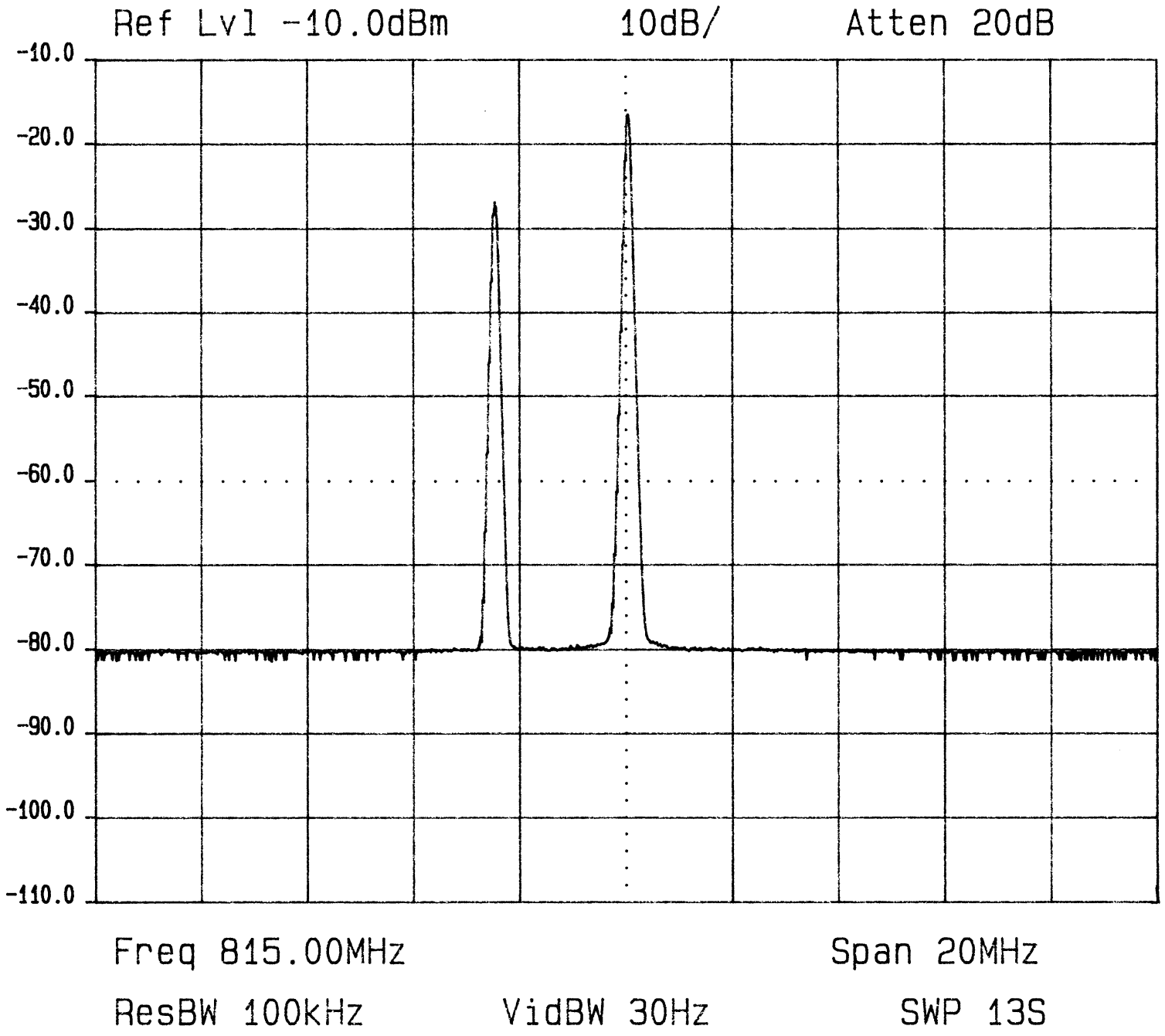

RF In
3-dB Split
1
6 dB
Vector Mod
v_I v_Q
From DSP
10-dB Class A Preamps
24-dB Class AB Power Amp
2
20-dB Coupler
50 Ω
28 dB
4
SMA Delay Line
2.8 m/13.1 ns
3
3-dB Comb
3-dB Split
5
20 dB
745 MHz
+10 dB_m
BPF
69.9625 MHz
+7 dB_m
LPF
60 dB
$\tilde{v}_m(t)$ to DSP

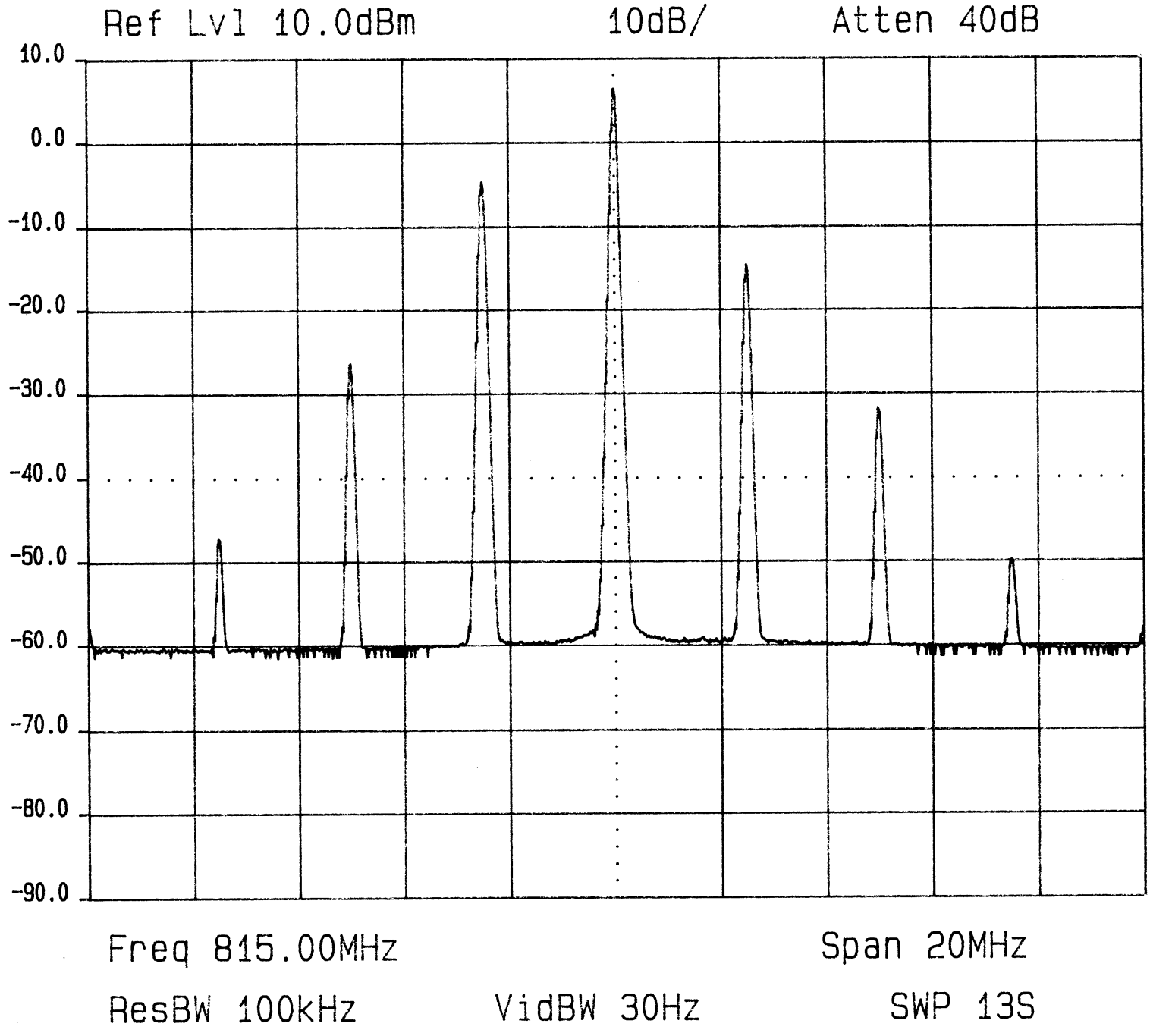

SMA Delay Line
2.6 m/12.3ns
8
7
3-dB
Comb
10-dB
Coupler
RF Out
50 Ω
20-dB Class A
Preamp
6
3-dB
Split
13 dB
Vector
Mod
20-dB Class A
Preamp
40-dB Class A
Error Amp
10 dB
30 dB
v_I v_Q
From DSP
BPF
LPF
40 dB
$\tilde{v}_e(t)$ to DSP
BPF
LPF
40 dB
$\tilde{v}_o(t)$ to DSP

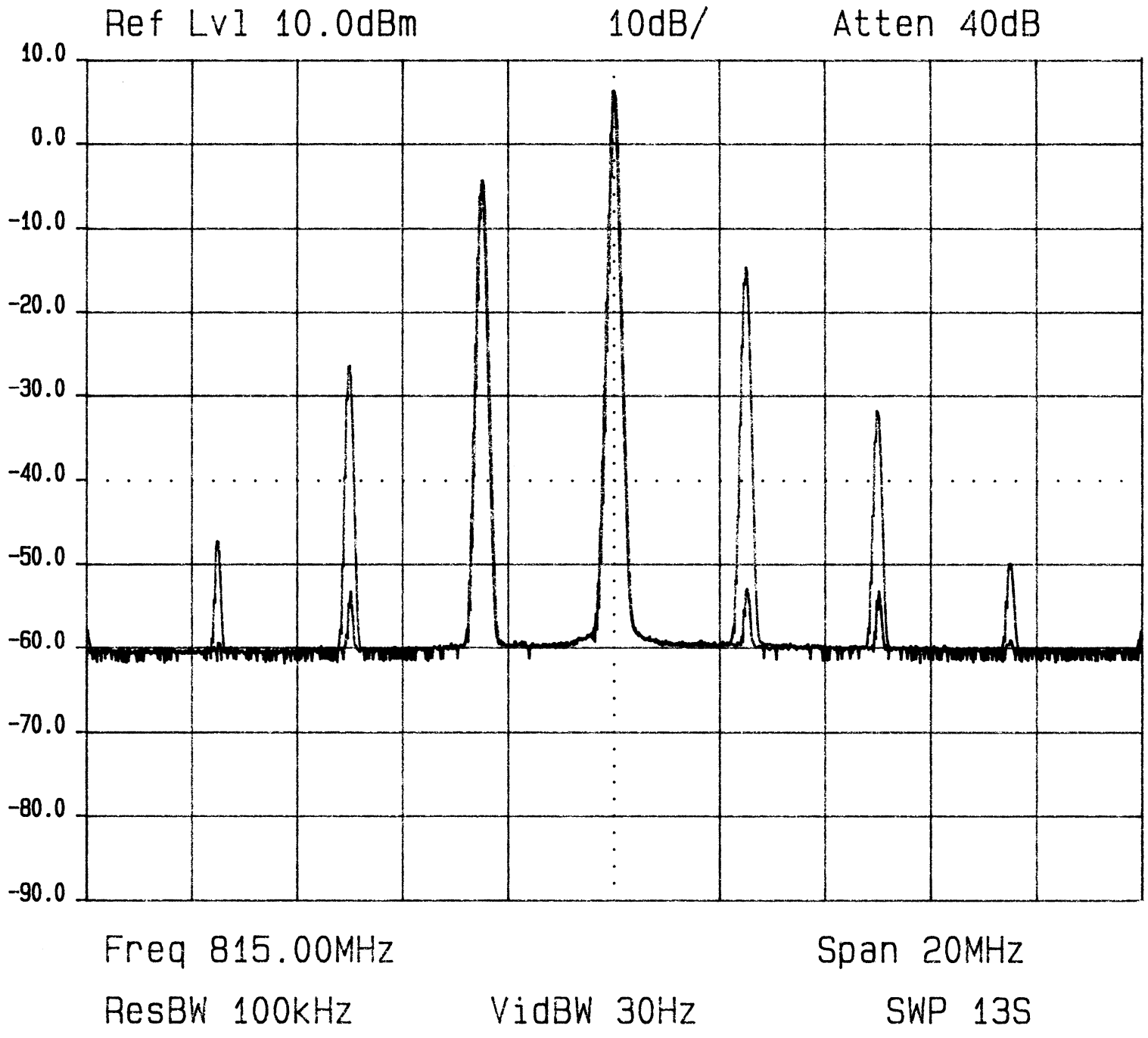

Ref Lvl -20.0dBm
10dB/
Atten 10dB
-20.0
-30.0
-40.0
-50.0
-60.0
-70.0
-80.0
-90.0
-100.0
-110.0
-120.0
Freq 815.000 0MHz
Span 200kHz
ResBW 3kHz
VidBW 3Hz
SWP 44S


Ref Lvl 10.0dBm
10dB/
Atten 40dB
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
-60.0
-70.0
-80.0
-90.0
Freq 815.000 0MHz
Span 200kHz
ResBW 3kHz
VidBW 3Hz
SWP 44S


Ref Lvl -13.0dBm
10dB/
Atten 20dB
-13.0
-23.0
-33.0
-43.0
-53.0
-63.0
-73.0
-83.0
-93.0
-103.0
-113.0
Freq 815.000 0MHz
Span 200kHz
ResBW 3kHz
VidBW 3Hz
SWP 44S


Ref Lvl -10.0dBm
10dB/
Atten 20dB
-10.0
-20.0
-30.0
-40.0
-50.0
-60.0
-70.0
-80.0
-90.0
-100.0
-110.0
Freq 815.00MHz
Span 20MHz
ResBW 100kHz
VidBW 30Hz
SWP 13S


Ref Lvl 10.0dBm
10dB/
Atten 40dB
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
-60.0
-70.0
-80.0
-90.0
Freq 815.00MHz
Span 20MHz
ResBW 100kHz
VidBW 30Hz
SWP 13S


Ref Lvl 10.0dBm
10dB/
Atten 40dB
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
-60.0
-70.0
-80.0
-90.0
Freq 815.00MHz
Span 20MHz
ResBW 100kHz
VidBW 30Hz
SWP 13S