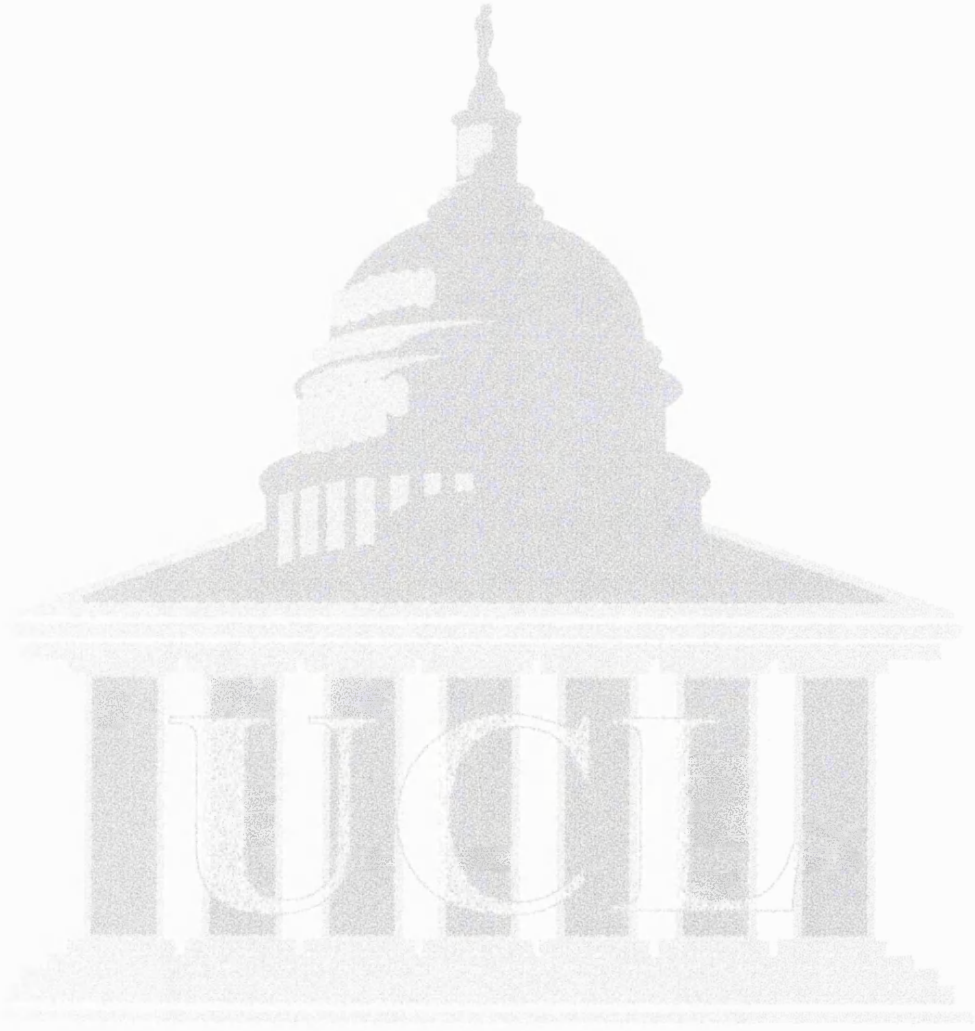
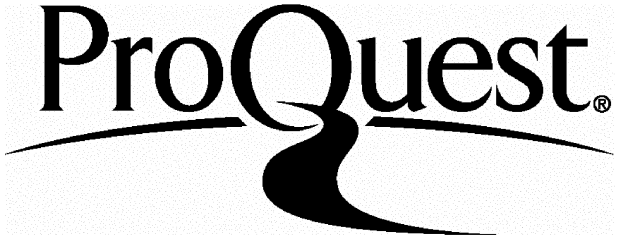
Power Amplifier Linearisation Through Second-Order Bias Injection

- PhD Thesis -



William Jenkins

May 2001



Abstract

This thesis describes the development of a novel linearisation technique for use in high-frequency power amplifiers. The need for linear power amplifiers is identified, and existing linearisation schemes are outlined and appraised in terms of their complexity, cost and efficacy. It is shown that currently-available linearisation schemes tend to have an effectiveness that is proportional to their complexity, and hence their cost of implementation.

Analysis and simulation results are presented to illustrate the mechanism through which the new linearisation technique reduces in-band distortion. The theoretical work is then verified with experimental measurements, initially using two unmodulated carriers and a feedback topology, and progressing to become a feedforward or 'injection' technique using four unmodulated carriers. The agreement between the simulated and measured performance was found to be excellent throughout.

The application of the technique to modulated 'real-world' signals is then investigated, with theoretical analysis, simulations and measured results presented to demonstrate the applicability of the technique to both single and multiple modulated-carrier input signals. It is shown that Second-Order Bias Injection can typically provide 15-18dB of in-band distortion improvement, and that the technique has potential for use in next-generation ($2G^+$ and 3G) mobile telecommunication networks.

Acknowledgements

There are numerous people without whom this work would not have been successful. I'd like to begin by expressing sincere thanks to Dr Ahmad Khanifar for his support, guidance and advice, especially during those periods of difficulty with the work and in my personal life. I am also indebted to Dr David Haigh and Dr Danny Webster at UCL for their generous assistance in the nonlinear model extraction process, as well as for the advice and constructive comments they have offered throughout the course of this study.

Within Nokia, I would like to acknowledge the invaluable support and encouragement provided by Dr Mahmoud Zadeh, and I would also like to thank Martin Goss and Richard Kybett for lending me their time and experience.

Finally, in gratitude for their endless support and encouragement, this thesis is dedicated to my parents.

Table of Contents

Glossary

ACP(R)	-	Adjacent Channel Power (Ratio)
ADC	-	Analogue-to-Digital Converter
ADS	-	Advanced Design System
AGC	-	Automatic Gain-Control
ALCPR	-	Alternate Channel Power Ratio
AM	-	Amplitude Modulation
AM/AM	-	Amplitude-modulation to amplitude-modulation distortion
AM/PM	-	Amplitude-modulation to amplitude-modulation distortion
ARIB	-	Association of Radio Industries and Broadcasting (ARIB)
ASIC	-	Application-Specific Integrated Circuit
AWG	-	Arbitrary Waveform Generator
BER	-	Bit Error Rate
BJT	-	Bipolar Junction Transistor
BTS	-	Base Transceiver Station
CAD	-	Computer-Aided Design
CDMA	-	Code Division Multiple Access
DAC	-	Digital-to-Analogue Converter
DAPD	-	Digitally-Adaptive Digital Predistortion
DFT	-	Discrete Fourier Transform
DSO	-	Digital Sampling Oscilloscope
DSP	-	Digital Signal Processing/Processor
DUT	-	Device Under Test
EDGE	-	Enhanced Data Rates for GSM Evolution
ETSI	-	European Telecommunications Standards Institute
EVM	-	Error Vector Magnitude
FDMA	-	Frequency Division Multiple Access
FFT	-	Fast Fourier Transform
GCP	-	Gain Compression Point
GMSK	-	Gaussian Minimum-Shift Keying
GSM	-	Global System for Mobile communications
IF	-	Intermediate Frequency
IMD	-	Intermodulation Distortion
IMD3	-	Third-order Intermodulation Distortion
IP_2	-	Second-Order Intercept Point
IP_3	-	Third-Order Intercept Point
LDMOS	-	Laterally-Doped Metal-Oxide Semiconductor
LINC	-	Linear Amplification Through Nonlinear Components
LPA	-	Linear Power Amplifier

LPF	-	Low-Pass Filter
LTI	-	Linear Time Invariant
PA	-	Power Amplifier
PM	-	Phase Modulation
MDS	-	Microwave Design System
MESFET	-	Metal-Semiconductor Field-Effect Transistor
MMIC	-	Monolithic Microwave Integrated Circuit
MCPA	-	Multicarrier Power Amplifier
PAE	-	Power-Added Efficiency
PAR	-	Peak-to-Average Ratio
PSK	-	Phase-Shift Keying
QoS	-	Quality of Service
(O)-QPSK	-	(Offset) Quadrature Phase-Shift Keying
RF	-	Radio Frequency
SCPA	-	Single Carrier Power Amplifier
SOBI	-	Second-Order Bias Injection
TDMA	-	Time Division Multiple Access
TWT	-	Travelling-Wave Tube
UHF	-	Ultra High-Frequency
UMTS	-	Universal Mobile Telephone System
VCO	-	Voltage-Controlled Oscillator
VHF	-	Very High-Frequency
WCDMA	-	Wideband Code Division Multiple Access

Table of Figures

1. Introduction

The recent explosion in the use of mobile telephony is the result of many technological innovations that have allowed both handsets and transmitters to be produced at ever-reducing costs, and with improved performance and functionality. Advances in DSP and ASIC architectures, MMIC design methods, signal processing techniques and even the foundry processes themselves have all contributed to the communications revolution, with new innovations being quickly adopted by the competitive telecommunications industry. However, despite rapid progress, there are still fundamental problems in vital system components, and these are rapidly becoming the limiting factor in terms of both cost and performance.

Of these essential components, power amplifiers (PAs) have been researched and developed more thoroughly than almost any other, and it is their unsatisfactory linearity and efficiency that are now emerging as prime concerns in the design of present-day and next-generation transmitters. It is the aim of this study to develop a novel solution to the amplifier linearity problem, which will improve the performance of high-frequency power amplifiers in a cost-effective manner.

There has already been a great deal of study directed towards the amplifier linearity issue, resulting in many diverse and varied techniques. The range of available solutions is still expanding as new techniques are proposed, and new technologies enable established linearisers to evolve further. The reason that such intensive work is still concentrated on linearisation is due to the fact that the cost of existing systems tends to be directly proportional to their effectiveness, as will be discussed shortly. The goal of an effective

lineariser of moderate complexity and at low cost remains frustratingly elusive, and for this reason has been referred to as 'The Holy Grail' of amplifier design.

It will be shown in the course of this study that the simultaneous amplification of multiple carriers in a single amplifier, without linearising circuitry of some form, is impractical due to the unacceptable degradation of signal quality that occurs when the device is operated with any reasonable degree of efficiency. Although there are several established linearisation techniques already available, they are limited either in terms of their bandwidth or by their cost and complexity. Each also tends to be bounded by the linearity improvement produced, and as a result it is unlikely that any one system will be able to simultaneously satisfy the orthogonal criteria of both cost and performance. It is therefore vital that new 'complementary' linearisation techniques are developed in order that they may be applied in tandem with other more well-established methods, with the aim of improving linearity in a cost-effective and more easily mass-produced manner.

The majority of commercially-available base-station transceivers circumvent the linearity problem by using single-carrier-per-amplifier (SCPA) architectures, which essentially provide a separate modular transmitter stage for each of the carriers in a particular cell. Although adequate, the SCPA solution is by no means ideal and has several disadvantages. The large physical size of the numerous filtering and amplifying stages is inconvenient, and the transceiver must be housed and maintained by the network provider, in the vicinity of the transmitting antennae. In cities, where mobile traffic density and hence infrastructure requirements are high, this entails long-term rental or purchase of space within or upon a suitable building - an expensive necessity. The amplifiers used in these single-carrier-per-amplifier systems are also relatively inefficient, generally being operated in a 'backed-off' Class A or Class AB regime to reduce both AM-PM distortion and spectral re-growth around

the carriers, so as to maintain the desired quality of service (QoS) without infringing broadcast regulations. As a result, considerable power is wasted through heat-dissipation, often requiring air-conditioning units in order to maintain operating temperatures. This further compounds the power consumption problem and increases the physical size of the transceiver.

Single-carrier-per-amplifier systems do not readily allow frequency-planning and dynamic channel-allocation, and this goal has eluded the designers and operators of cellular systems since the concept of cellular mobile communications first became a reality. The ability to re-plan frequency allocation without dismantling the transceiver would require either the use of remotely-tuneable combiners, or a broad-band, highly linear amplifier. The former solution implies the use of mechanical servo-controlled systems, which are slow and potentially unreliable, or varactor diodes, which both deteriorate the intermodulation performance of the combination process and restrict the power handling capability. The latter solution, a highly linear amplifier, is already realisable in the form of a Feedforward system; however, these are expensive, complicated, inefficient, difficult to mass-produce and are therefore not an ideal solution to the problem.

Apart from the issues of size, power dissipation and frequency allocation inflexibility, the capacity of a cell using a SCPA BTS is limited by the number of transceiver stages that have been installed. If demand increases in a particular cell, new hardware must be installed, consuming more space and power, generating more heat and requiring the intervention of a technician.

The main advantages to this approach are the maturity of SCPA technology, the relative ease manufacture and the lack of expensive linearising circuitry. Until recently, it was more cost-

effective for hardware manufacturers to employ SCPA architectures, as these offered the optimum compromise between cost and performance. This situation is shortly to change with the arrival of the next generation of mobile radio systems, however, as these require more flexible solutions with higher linearity amplifiers in both base-stations and handsets. Thus, the issue of amplifier linearisation is more important now than it ever has been.

The widespread adoption of GSM in Europe, and more recently the rest of the world, is the result of many contributing factors. The most prominent virtue of GSM for PA designers is the constant-envelope modulation scheme employed for the air interface - Gaussian Minimum-Shift Keying (GMSK). Although system imperfections cause GMSK to have a non-zero peak-to-average power ratio in practise, it is sufficiently small to allow PA designers to employ higher-efficiency amplifier operating modes such as Class AB, B or even C. As the PA is typically the most power-hungry block in any transceiver, this greatly extends handset battery-life and minimises the power-consumption, heat-dissipation and size of base stations. GSM has now matured into a highly efficient vehicle for voice traffic, but it has become the victim of its own success. Unfortunately, the many merits of GSM can no longer outweigh its fundamental limitations – namely, the relatively bandwidth-inefficient TDMA-FDMA access scheme and the low bit-rates (9.6kbit/sec) that the standard can support. These place severe constraints on the services that network providers can offer, and the demand for new high-bandwidth services coupled to extraordinary growth in subscriber numbers (now over 135 million world-wide [1]) has lead to the development of a new access scheme for the third generation of cellular systems. The goal is to provide subscribers with desirable new services such as streaming video and internet access, achieved through the use of a new code-division multiple-access (CDMA) radio standard similar to that developed by Qualcomm for IS-95, which is already operational in North America. A new 2GHz

frequency band that has been set aside for the new wideband-CDMA (WCDMA) standard [2] supported and standardised by both the European Telecommunications Standards Institute (ETSI) and the Association of Radio Industries and Broadcasting (ARIB).

In order to smooth the path from GSM to WCDMA, and go someway towards satisfying consumer demand for new services in the short-term, an intermediate '2G+' standard has been developed, referred to as EDGE (Enhanced Data Rates for GSM Evolution). The EDGE air-interface has been chosen carefully because it allows higher bit-rates to be carried within the GSM spectral mask without affecting burst duration, so it can co-exist with the original GSM voice-traffic and be introduced gradually. EDGE will allow a natural evolution of the GSM network and will also be particularly attractive to operators who are not granted a UMTS license, as it will support bit-rates of the same order of magnitude as future third-generation solutions (up to 384kbps). The increase in gross bit-rate from 28.8kbit/s to 69.2kbit/s (per timeslot) is achieved with a new modulation scheme, eight-phase-shift-keying (8-PSK), which is more spectrally efficient than GMSK and also relatively easy to implement. This will effectively double the traffic-handling capacity of the network with respect to GSM, at the expense of an increased peak-to-average power ratio due to the non-constant-envelope nature of 8-PSK modulation. Although EDGE uses a modified 8-PSK mapping to reduce its envelope variation, it still has a peak-to-average power ratio in the region of 3.4dB and a peak-to-minimum ratio of 17dB [3]. In order to avoid significant distortion, the average output power of an amplifier in an EDGE transmitter must be reduced by at least the peak-to-average value (with respect to the GSM operating-level), as there must be sufficient 'head-room' to allow for the peaks in the input-signal envelope. It has been found that even this degree of back-off is often insufficient to satisfy the linearity requirements of EDGE, which are quantified in terms of Error Vector

Magnitude (EVM). Unnecessary power back-off (over and above that required by the peak-to-average ratio) is undesirable for most applications as it fails to utilise the full range of the amplifier's output voltage swing, and thus lowers efficiency. One obvious alternative is to bias the amplifier in a more linear operating regime, but this also reduces efficiency and increases power-dissipation, incurring the heat-generation penalties already described. Therefore, an apparently minor change in the modulation scheme for the air interface has very wide implications, and it is the knock-on effect of these changes that is currently concerning PA designers and system engineers.

In the same way, the design of the transceivers for third-generation WCDMA systems is currently posing many challenges for engineers of all disciplines, including PA designers. Again, a non-constant-envelope modulation scheme (QPSK for the uplink and O-QPSK for the downlink) means that efficiency must be traded off against linearity and ultimately, cost. Spectral- and power-efficiency are mutually exclusive, and both the EDGE and WCDMA air-interfaces have a significant peak-to-average power variation with respect to GMSK, as shown in Table 1 below.

Air Interface	Modulation Scheme	Relative Spectral Efficiency	Peak-to-Average Power Ratio
GSM	GMSK	1	~0dB
EDGE	3π/8 8-PSK	3	~3.4dB
WCDMA	QPSK	2	~10dB

Table 1: Comparison of GSM, EDGE and WCDMA [4]

Therefore, even SCPA transceivers will be considerably less efficient when amplifying EDGE or WCDMA, and multi-carrier PAs with sufficient linearity will be even harder to realise than they are already.

In summary, there are two separate, but related, linearity issues that must be addressed. Firstly, highly linear multi-carrier PAs have been desirable for many years, primarily because they significantly reduce the amount of additional hardware required for a given traffic handling capacity. Any technique that can improve linearity, increase efficiency or achieve both of these goals at the same time in a cost-effective manner would be highly sought-after. Secondary to the general desire to develop new linearisation techniques is the problem caused by the imminent evolution of the mobile radio standards from GSM to CDMA via EDGE, as outlined above. It has become apparent that some form of linearisation is needed if the performance and/or cost of both handsets and base-stations is not to be adversely affected.

The aim of this study is to develop a novel linearisation technique that has the potential to improve amplifier linearity and/or efficiency in a cost-effective manner with only a moderate increase in circuit complexity. By doing so, it is hoped that the compromise between efficiency and linearity facing next-generation system designers can be made less arduous by adding a low-cost, low-power alternative to the existing range of distortion techniques.

2. Background

To appreciate the causes and effects of amplifier nonlinearity, a summary of the relevant background theory is first presented. The concepts developed in the first part of this chapter will be used to illustrate the function of the various linearisation techniques that will be discussed, and allow their merits and limitations to be assessed.

2.1 Nonlinearity

It is a fundamental truth of electronic engineering that all circuits are nonlinear. In the majority of cases, linearity is assumed by circuit designers in order to simplify analysis, and often this is a valid and useful assumption. For example, a resistor will only begin to behave in a measurably nonlinear fashion if driven to the extremes of its operating range, when thermal and other effects come into play - a regime not normally encountered. It has even been observed that RF connectors can produce small amounts of distortion at high power levels, due to the nonlinear resistance created at the junction between two dissimilar metals [6]. However, these effects are small enough for the idealising assumption of linearity to be justified in the vast majority of cases.

A circuit is said to be linear if the principle of *superposition* applies. In essence, superposition implies that if signals $x_1(t)$ and $x_2(t)$ are applied separately to the input of a circuit, producing outputs $y_1(t)$ and $y_2(t)$ respectively, an input consisting of $ax_1(t) + bx_2(t)$ will produce an output of the form $ay_1(t) + by_2(t)$. Further to this, the output spectrum of a linear circuit will contain no frequency components that were not present in the input signal. An amplifier is an inherently nonlinear device and will thus introduce distortion, reducing the fidelity of the output signal. The most well-known form of amplifier distortion, referred to as AM-AM (amplitude modulation-to-amplitude modulation), appears as extraneous

frequency components in the output signal. The other, less problematic type of amplifier nonlinearity is AM-PM (amplitude-modulation-to-phase-modulation), whereby changes in the instantaneous input signal amplitude are translated by the device to become phase variations in the output waveform. In systems where the phase of signals is important, these undesirable fluctuations are problematic. In modern communication systems using phase-modulated digital signals, these envelope-dependent distortions can alter the signal trajectory and thus lead to an increase in bit error rate (BER). However, AM-PM distortion is a secondary consideration in most systems as it tends to only appear when amplifiers are driven into saturation.

2.1.1 Strong and Weak Nonlinearity

There is no precise definition of the distinction between these two terms that is generally accepted, but it has been suggested that a weakly nonlinear circuit is one that may be described with adequate accuracy by a power or Volterra series expansion [7]. This implies that the characteristic is continuous, that it has continuous derivatives and that for most cases it may be described by the first few terms of the series. Strongly nonlinear behaviour, in comparison, cannot be described by a simple series expansion, is in general not continuous, and will not have continuous derivatives.

A device such as a MESFET or BJT amplifier exhibits both weak and strong nonlinear behaviour, depending on how hard it is driven. Figure 2.1 overleaf shows a typical idealised transfer function for such a device, showing both regimes plotted separately to highlight the differences between the two. The solid red line represents the behaviour of the device at the extremes of operation, between which it is assumed to be linear. Beyond the region bounded by $\pm V_{max}$, the output no longer varies with the input level, representing the limiting condition imposed by power supply and device constraints.

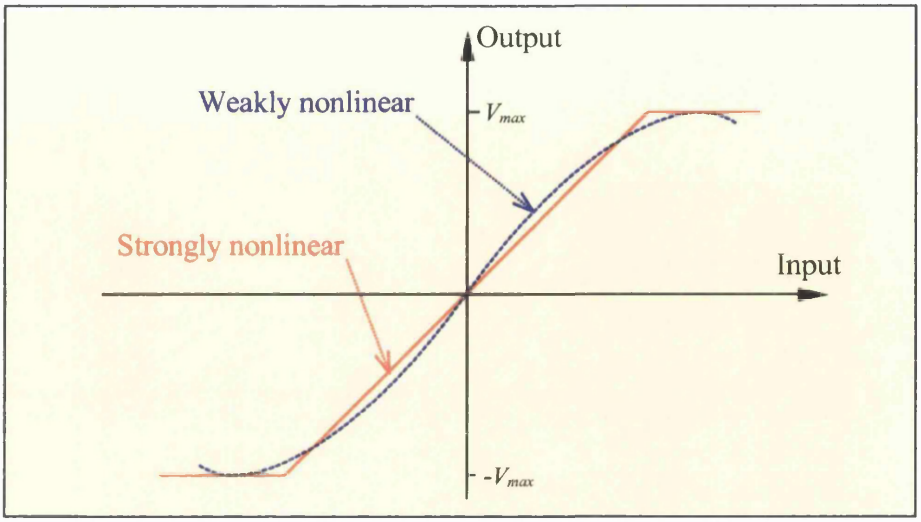


Figure 2.1: Typical amplifier transfer function

The dashed blue line, exaggerated here for clarity, represents the weakly nonlinear part of the amplifier characteristic which can be approximated by a power series expansion, and hence remains continuous around $\pm V_{max}$ as indicated in the figure. As the level of the input excitation varies up and down the weakly nonlinear characteristic, AM-AM distortion is generated in the form of spurious harmonic and intermodulation components at the output.

The weakly non-linear characteristic alone does not accurately depict the behaviour of the practical device at the extremes of the characteristic, where the strongly nonlinear behaviour applies. In order to accurately model the FET over the whole range of signal levels, a model consisting of a combination of both types of behaviour is required, with the strongly nonlinear characteristic superimposed upon the weak. Unfortunately, models of this type are complex to implement and are not required in the majority of applications; for example, an amplifier biased for Class A operation that is not driven into saturation remains within the weakly nonlinear regime at all times.

2.1.2 Amplitude Distortion

One of the most common types of nonlinearity observed in weakly nonlinear two-port systems is amplitude distortion, resulting in the type of transfer characteristic shown below in Figure 2.2.

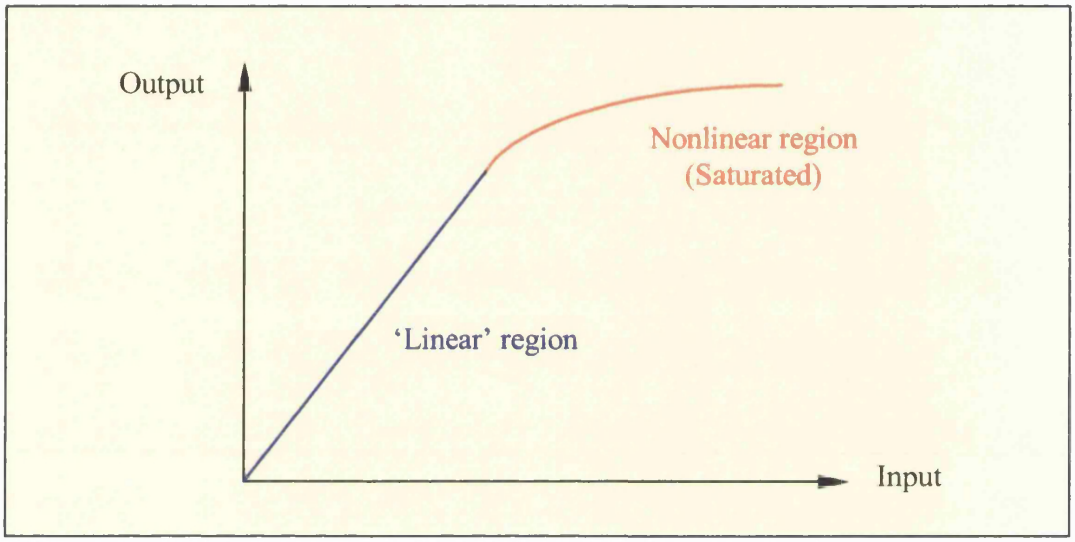


Figure 2.2: Input versus Output curve for a typical 2-port system

This type of behaviour is exhibited by all circuits, as available output power is always finite. If a two-port device such as this is treated as a 'memoryless' system - that is, the output depends only upon the instantaneous input - and its nonlinearity is weak, it may be described with reasonable accuracy by a power series expansion. Let us consider the case of a voltage-controlled voltage source having a weakly nonlinear characteristic such as that shown in Figure 2.1, with input V_{in} and output V_{out} related by a power series expansion as follows:

$$V_{out} = G_1 V_{in} + G_2 V_{in}^2 + G_3 V_{in}^3 + G_4 V_{in}^4 + \ldots. \tag{2.1}$$

where G_1 represents the linear gain and $G_1 >> G_2 >> G_3 >> G_4 >> \ldots..$

Most weakly-nonlinear device behaviour can be adequately modelled by only the first three terms of the above expansion, so it will be truncated beyond the third-order term in the following analysis. By inspection of Equation 2.1, and recalling that in most cases the

coefficients G_n are related as indicated, it can be observed that the amplitude distortion becomes more severe as the magnitude of V_{in} increases. As the input level increases further, successively higher terms of the expansion begin to affect, and then to dominate, the overall characteristic.

If the input voltage to such a system is a sinusoidal tone of the form $V_{in}=A\cos(\omega t)$, the output spectrum will contain harmonic distortion produced by the non-linear terms of the expression. The relative magnitude of these spurious components is given by the constants G_n and the magnitude of the driving input voltage, A, as can be seen in the following expansion:

$$V_{out}(t) = G_1 A\cos(\omega t) + G_2 A^2\cos^2(\omega t) + G_3 A^3\cos^3(\omega t)+\ldots$$

$$=\frac{G_2 A^2}{2} + (G_1 A + \frac{3}{4}G_3 A^3)\cos(\omega t) + \frac{1}{2}G_2 A^2\cos(2\omega t) + \frac{1}{4}G_3 A^3\cos(3\omega t)+\ldots \quad (2.2)$$

Equation 2.2 shows that the output voltage now contains a DC offset and spurious second and third-order harmonic components as well as the original fundamental frequency. Fortunately, the range of frequencies used in most communication systems allows these higher-order harmonics to be easily removed with bandpass filtering.

Referring again to Equation 2.2, the linear gain is now $G_1A + \frac{3}{4}G_3A^3$ and not simply G_1A as would be the case for a device with an ideal transfer characteristic. Thus, if the sign of coefficient G_3 is negative as is the case in almost all amplifiers, the linear gain is reduced as the input voltage increases, resulting in *Gain Compression*. This is the most common situation, and gives rise to the downward-sloping characteristic shown in Figure 2.2. In the cases where G_3 is positive (such as an amplifier biased for Class AB operation) the inverse applies, producing *Gain Expansion*.

If an input signal has a time-varying envelope, such as $V_{in}(t) = x_I(t)\cos(\omega_c t) - x_Q(t)\sin(\omega_c t)$ where $x_I(t)$ and $x_Q(t)$ are the in-phase and quadrature components of the baseband signal, the third-order term of Equation 2.1 can be re-written as:

$$G_3 V_{in}^3(t) = G_3 x_I^3(t)\frac{\cos(3\omega_c t)+3\cos(\omega_c t)}{4} - G_3 x_Q^3(t)\frac{-\cos(3\omega_c t)+3\sin(\omega_c t)}{4}+\ldots \qquad (2.3)$$

Thus, the output signal contains the spectra of $x_I^3(t)$ and $x_Q^3(t)$ centred around the main carrier frequency, ω_c. These third-order components have a bandwidth that is three times the width of the original carrier, so the spectrum 'grows' with the distortion appearing either side (and on top) of the main signal. The amount of spectral regrowth caused by an amplifier is quantified by the ratio between the total power in the main channel with respect to that in the adjacent channels, and is known as the 'Adjacent Channel Power Ratio' (ACPR). It should be noted that if the magnitude of the phasor represented by $V_{in}(t)$ were to remain constant, as is the case with a GMSK-modulated signal, spectral regrowth would not appear and the only distortion occurring would be in the form of higher-order harmonics. A more complete measure of an amplifier's linearity is given by 'Error Vector Magnitude' (EVM), which quantifies the degree to which the trajectory of the modulated output signal departs from its ideal path, and as such it accounts for both AM-AM and AM-PM distortion. However, the underlying mechanisms that give rise to both ACPR and EVM are the same, and as such either can be used as a measure of amplifier linearity.

Spectral-regrowth is a type of intermodulation distortion, which arises when nonlinearity causes signal components to interact or 'mix' with each other, a side-effect exploited in mixers. Again, this type of distortion can be most simply illustrated with the simple power series approximation of Equation 2.1, this time by applying a two-tone input signal which may be written as:

$$V_{in}(t) = A\cos(\omega_1 t) + B\cos(\omega_2 t) \tag{2.4}$$

The first-order term $G_1 V_{in}$ generates the linearly-amplified version of the input carriers. The second order term $G_2 V_{in}^2$ generates second-order harmonics and second-order mixing frequencies, as well as a spurious DC term as follows:

$$\begin{aligned} G_2 V_{in}^2(t) = {} & G_2\left(\frac{A^2 + B^2}{2}\right) && \text{- DC term} \\ & + G_2\left(\frac{A^2\cos(2\omega_1 t)}{2} + \frac{B^2\cos(2\omega_2 t)}{2}\right) && \text{- Second-order harmonics} \\ & + G_2 AB\cos\left([\omega_1 + \omega_2]t\right) + G_2 AB\cos\left([\omega_2 - \omega_1]t\right) && \text{- Second-order IM products} \end{aligned} \tag{2.5}$$

The third-order term of the power series, $G_3 V_{in}^3$, produces third-order harmonic and intermodulation distortion:

$$\begin{aligned} G_3 V_{in}^3(t) = {} & \left(\frac{3G_3\left(A^3 + 2AB^2\right)}{4}\right)\cos(\omega_1 t) + \left(\frac{3G_3\left(B^3 + 2BA^2\right)}{4}\right)\cos(\omega_2 t) && \text{- Fundamental components} \\ & + \frac{G_3 A^3}{4}\cos(3\omega_1 t) + \frac{G_3 B^3}{4}\cos(3\omega_2 t) && \text{- Third-order harmonics} \\ & + \frac{3G_3 A^2 B}{4}\left(\cos(2\omega_1 + \omega_2)t + \cos(2\omega_1 - \omega_2)t\right) \\ & + \frac{3G_3 B^2 A}{4}\left(\cos(2\omega_2 + \omega_1)t + \cos(2\omega_2 - \omega_1)t\right) && \text{- Third-order IM products} \end{aligned} \tag{2.6}$$

The carrier frequencies used in modern communication systems and the spacing between them (known as the 'delta-frequency') are typically such that the third-order products at $2\omega_1$-ω_2 and $2\omega_2$-ω_1 are produced very close to the carriers, or 'in-band'. Also as a consequence of the carrier frequencies used, the higher-order harmonics and other distortion components all appear far enough away from the carriers to be easily removed with careful bandpass filtering, with no adverse effects on the output signal. This is most clearly

illustrated by a power spectrum representation; Figure 2.3 shows a two-tone input, along with the relative frequencies of the distortion observed in the output of a weakly nonlinear device, having input carrier frequencies spaced such that $\omega_2-\omega_1 << \omega_1,\omega_2$.

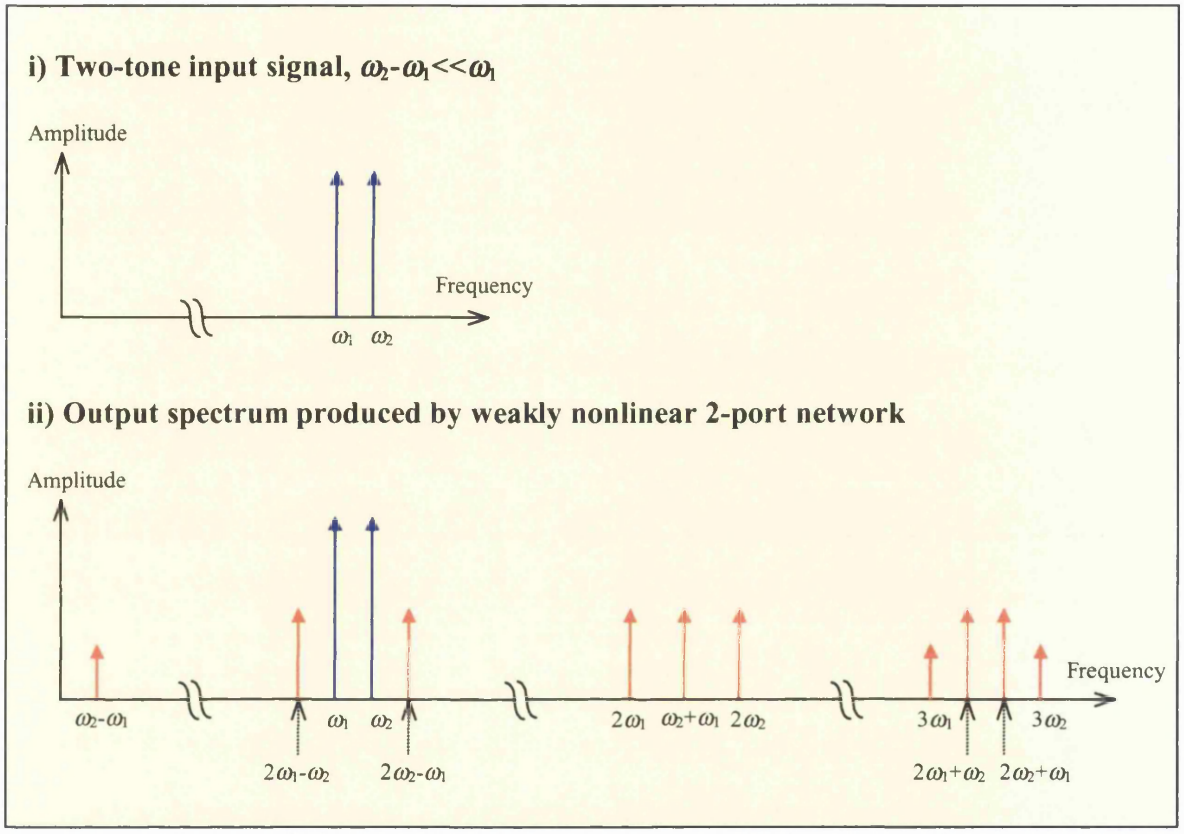


Figure 2.3: Input and output power spectrum for a 2-tone input

This case examines the distortion produced by only two input carriers. The complexity of the distortion increases rapidly with a larger number of carriers; for example, three carriers produce nine in-band third-order intermodulation products and a correspondingly greater number of out-of-band higher-order harmonic and intermodulation components. In fact, if we arbitrarily increase the number of carriers and reduce the delta-frequency to a vanishingly small value, the spectrum of the multi-carrier input resembles a single modulated carrier, and the collection of in-band third-order intermodulation components appearing either side correspond to the spectral regrowth already discussed.

The origins of AM-PM distortion can be similarly explained. Recall from Equation 2.2 that a single input sinusoid produces an output fundamental whose amplitude is a given by a combination of first and third-order terms:

$$V_{out}\big|_{\omega_c} = (G_1 A + \frac{3}{4} G_3 A^3)\cos(\omega_c t) \tag{2.7}$$

If the two components G_1A and $¾G_3A^3$ remain in phase, as in the above equation, then the distortion produced at the fundamental frequency is purely AM-AM. However, this very simple analysis using a memoryless nonlinearity takes no account of the phases of these two products; if the analysis were extended to include the effects of capacitive nonlinearities, the response at the fundamental becomes the sum of two vectors with some phase difference between them. Even if the value of this phase-shift is not itself dependent upon amplitude, the combined phasor will exhibit phase-deviations when the amplitude fluctuates, due to the fact that the first-order component varies linearly with amplitude whilst the third-order varies with the cube. These deviations will only become significant when the magnitude of the third-order component is comparable with the magnitude of the fundamental, and so AM-PM conversion only really becomes of concern when an amplifier is pushed into compression.

The three types of distortion discussed above explain the dominance of SCPA transceiver architectures in today's mobile networks. The linearity requirements of a multi-carrier GSM/EDGE BTS (as specified by ETSI in GSM 05.05 [5]) are –75dBc, a very stringent requirement and one that is more applicable to passively-combined SCPA-architectures than MCPAs. Very few commercially-available MCPAs are capable of meeting this specification, and such amplifiers are notoriously difficult to manufacture in commercially-viable volumes. This, along with the constant-envelope nature of GMSK, is the reason why SCPA

architectures are used almost exclusively to carry GSM traffic. However, as discussed in the introduction, GSM is not the air interface of the future and is to be superseded by EDGE and eventually WCDMA. This evolution will require greater transceiver flexibility than is currently available, and multiple standards may need to be supported in a single BTS cabinet. Although this will be possible with SCPA-architectures, the greatest degree of flexibility will be provided by a single, highly-linear amplifier.

2.1.3 Linear Distortion

Linear distortion refers to the nonideal gain and phase variations that any practical RF amplifier displays across its band of operation. The most common manifestations of this type of nonlinearity are in the form of amplitude and phase ripple across the bandwidth of the amplifier, with amplitude roll-off and phase flattening occurring at the edges. Amplitude ripple can lead to the generation of additional scaled and delayed 'echoes' of the input signal [6], whilst a nonlinear phase-shift versus frequency characteristic results in the different frequency components of the input signal experiencing different time delays, resulting in signal distortion. A useful measure of phase distortion is given by 'group delay', defined as the negative of the derivative of phase shift versus frequency. If group delay is constant, a signal will pass through an amplifier without distortion.

Although these effects are undesirable, for most applications they are considered acceptable and are generally less detrimental than the nonlinear distortions discussed in the previous section, as they do not generate spurious in-band frequency components. In some test and measurement equipment, a highly linear response may be necessary and in these cases expensive hardware-intensive linearisation techniques such as Feedforward may be used (see Section 2.2).

2.1.4 Power Amplifier Intercept Points

In the preceding section it was shown that, for small-signal operation in the approximately linear regime, the output power of a device at the fundamental frequency with a single-tone input is linearly proportional to the amplitude of the input signal. When a two-tone input signal is applied, the second-order distortion produced is proportional to the square of the amplitude of the input signals, the third-order to the cube of input amplitude and so on. If a single-tone power sweep test is performed on a device, and the power of the fundamental plotted on the same logarithmic axis as the second and third-order distortion powers produced by a two-tone power sweep, the three traces produced are related in an approximately fixed ratio. Well below saturation, the slope of the fundamental is 1:1, the slope of the second-order power is 2:1 and the third-order is 3:1.

In the small-signal regime, the power of the fundamental and all the distortion products vary linearly with input power; toward the regions of compression and into saturation, the behaviour of the higher-order distortion products change more erratically, with peaks and troughs, the characteristics of which are dependent upon both the device and chosen operating-point.

By extrapolating the linear regions of the fundamental, second and third-order distortion powers, the so-called 'intercept points' are found. The second- and third-order intercept points, IP_2 and IP_3, are given by the intersections of the linear extrapolation of the fundamental and second-order, and fundamental and third-order distortion powers respectively. These conventions are summarised below in Figure 2.4, along with typical values of the intercepts with respect to the 1dB GCP.

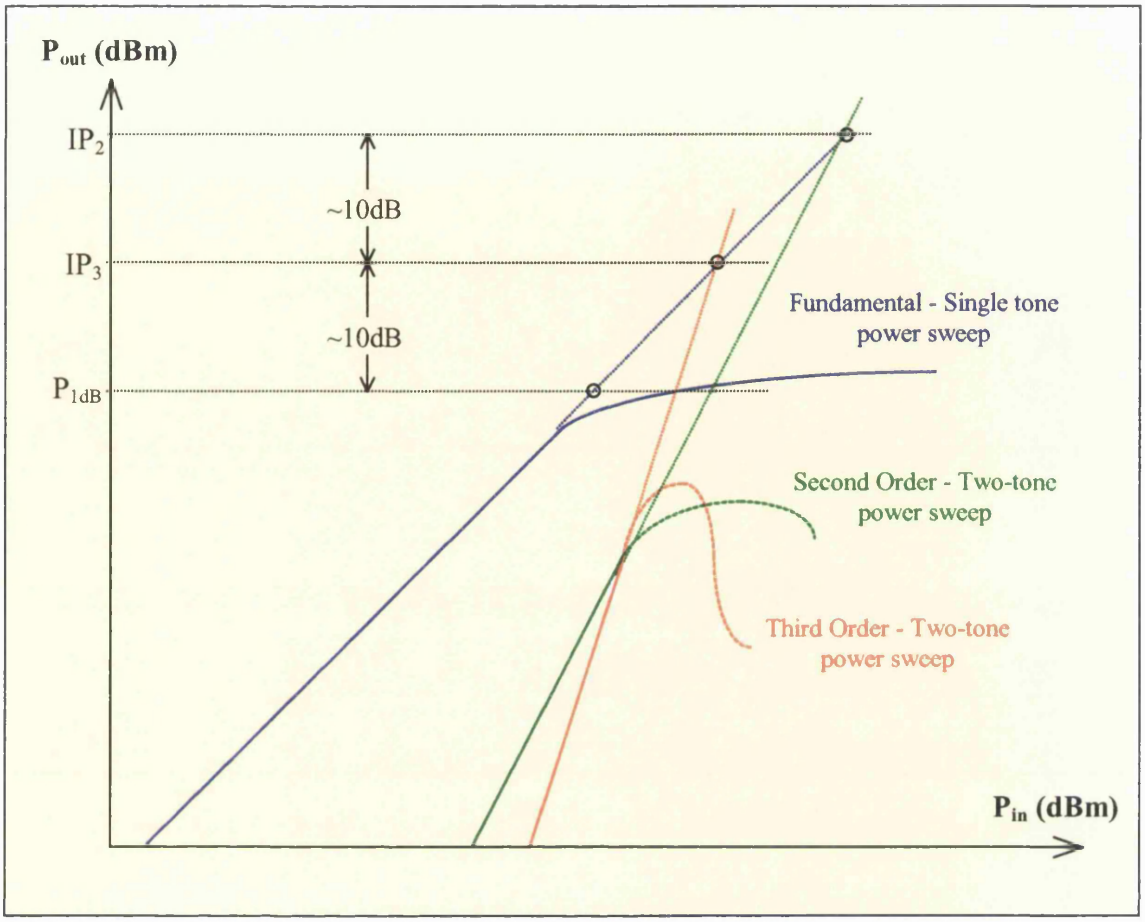


Figure 2.4: Intercept Points

The dashed sections of second and third-order distortion traces are arbitrary representations of the type of erratic behaviour that occurs due the complex interaction of the many distortion components.

IP_2 and IP_3 provide a measure of a device's linearity – the higher the values with respect to the 1dB GCP, the better the linearity will be. Occasionally this information will be supplied by manufacturers, but on the whole, experimental measurement is required to yield this data and it is only really useful as a rule-of thumb for designers. The process of extrapolation leads to some uncertainty, as the linear regions of the second- and third-order characteristics are often well below the intercept points produced, and can be near the noise floor of test-equipment. The greater this distance, the more that possible measurement errors are magnified by the extrapolation, and for these reasons IP_2 and IP_3 are only useful as

approximate guides. Despite these uncertainties, however, they are a useful benchmark of performance and are often used.

2.1.5 Peak to Average Ratio, Back-Off and Efficiency

Peak-to-Average Ratio (PAR) is a significant issue in multicarrier systems, as it has a great impact on efficiency. General literature indicates that WCDMA needs 10-13dB and multicarrier EDGE/GSM (>6 carriers) requires about 10-12dB of Peak-to-Average 'headroom'. This refers to the minimum amount of back-off required to prevent the amplifier output saturating, regardless of any other considerations. Figure 2.5 shows this graphically for the case of an input signal with a PAR of 10dB, assuming a two-tone linearity requirement of -45dBc.

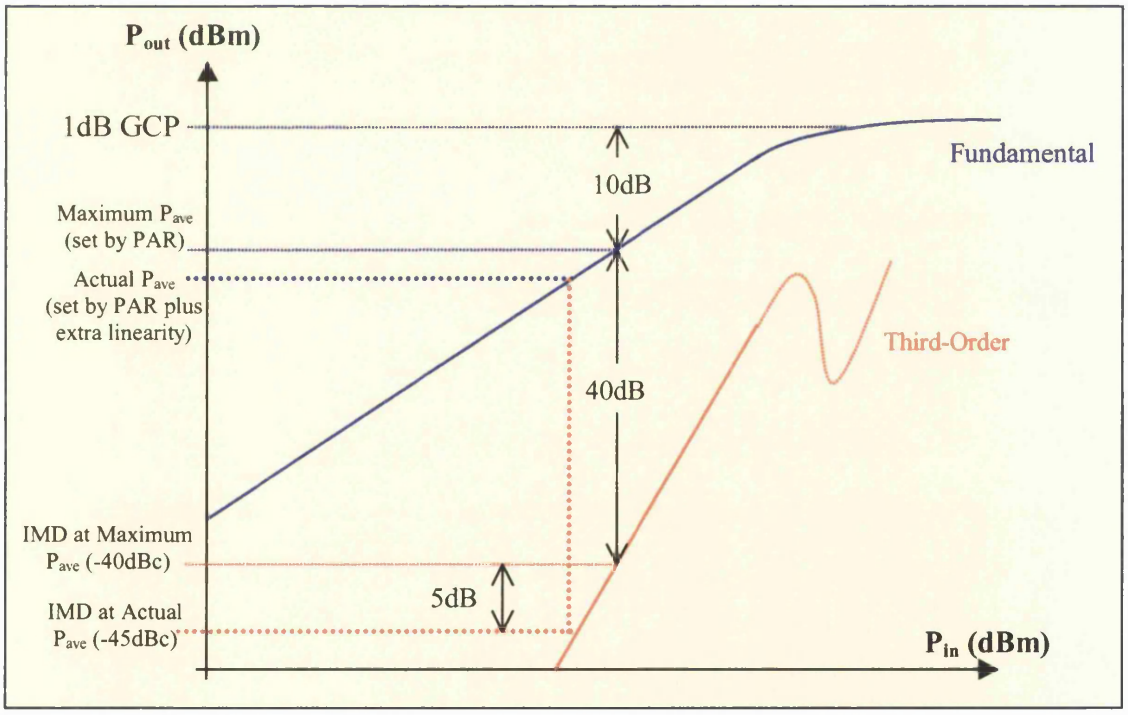


Figure 2.5: PAR, Linearity and Average Output Power

Referring to the above figure, the *theoretical* maximum average output power is less than the 1dB Gain-Compression Point power by an amount equal to the signal PAR. In this example, the two-tone linearity at this level of back-off is only 40dBc, so the extra 5dB of linearity

has to be gained by backing-off the output even further, giving the actual *achievable* average output power and efficiency. The maximum average output power for a given device is therefore governed firstly by input signal PAR and secondly by linearity requirements.

Although backing-off an amplifier reduces efficiency, this cannot be avoided when using non-constant-envelope signals (with PAR>1). The purpose of linearisation is therefore to ensure that the required linearity is achieved at the highest average output power and efficiency possible.

2.2 Linearisation Techniques

As discussed, linear PAs are now highly desirable. The choice is between using a linear Class A output stage, achieving 10-30% efficiency depending on the modulation scheme, or using a more efficient nonlinear amplifier with one or more linearisation techniques applied to it. Many such linearisation methods have been developed and evaluated at length in the literature, and these will be now be summarised in the following section. These methods are occasionally utilised in complex, expensive RF and microwave systems. As yet, they are not widely used in either mobile terminals or base-stations because they complicate the design process, are not suited to mass production and tend to become less effective as device characteristics fluctuate with temperature and output power. They can also consume a relatively large amount of power, so often the overall efficiency gains are marginal.

2.2.1 Power Back-Off

As described, the slopes of the fundamental output and IMD3 power versus input power for an amplifier are typically related in a 3:1 ratio, so if the drive level is reduced, or 'backed-off' by 1dB, it may be assumed that the third-order intermodulation distortion will be

reduced by 3dB. Therefore, the most straightforward way of achieving high linearity is to use a Class A amplifier at a reduced level of output power.

Although this method reliably produces predictable performance, it results in amplifiers with very low efficiencies. Gain and power at RF frequencies are valuable commodities, and as a result this solution to the problem is impractical and almost never used in multicarrier applications. For example, a 50W device backed off by 10dB would only produce 5W of output power - this low level of efficiency is simply unacceptable in nearly all instances.

2.2.2 Feedforward

Feedforward is perhaps the oldest approach to the linearisation problem, dating back to the original patent by Black [8]. It is also perhaps the most straightforward to understand, in that it is conceptually very simple. The configuration and basic operation of a typical Feedforward circuit is shown below in Figure 2.6 [9].

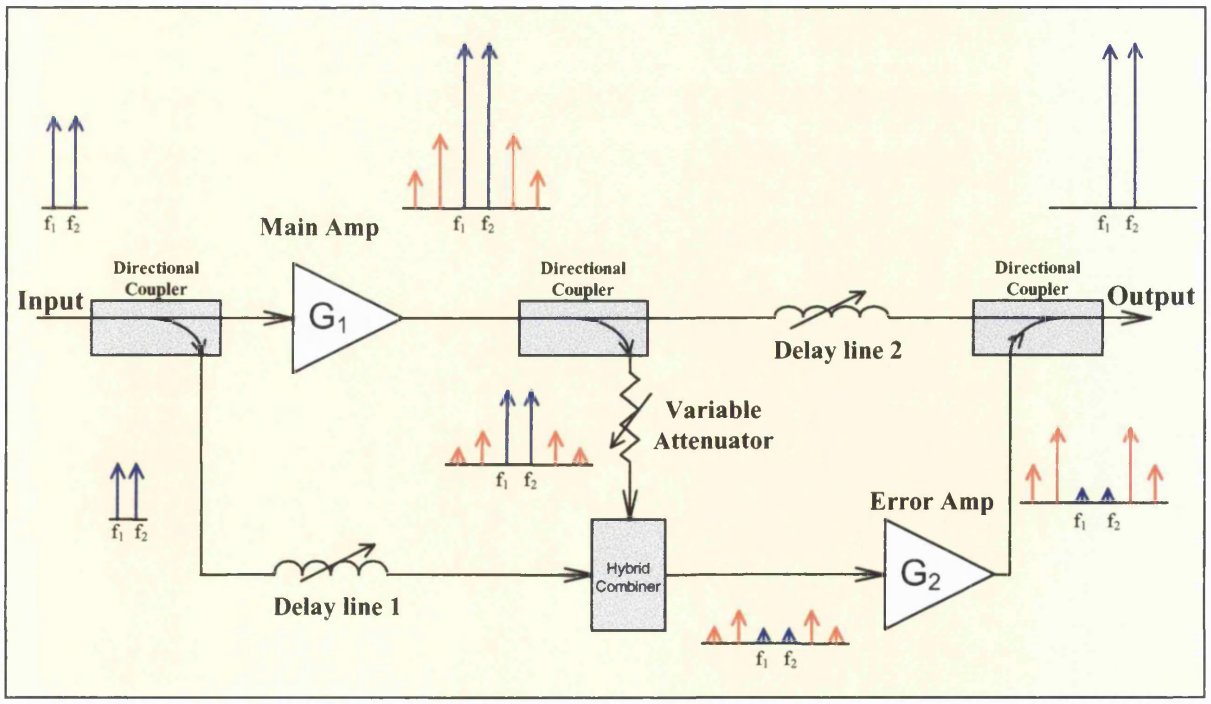


Figure 2.6: Feedforward topology

Referring to the above diagram, an input consisting of two closely-spaced tones of equal power is shown entering the circuit. A fraction of the clean, undistorted input signal is sampled by a directional coupler, before being amplified (and distorted) by the main PA. The output of the main amplifier is then sampled by a second directional coupler, and attenuated before being combined with the previously sampled undistorted input signals.

The phase shift introduced by Delay Line 1 and the attenuation introduced by the variable attenuator are chosen to be such that the two signals combine in anti-phase, and with equal magnitude, leaving only the distortion products. The error amplifier then amplifies these extraneous signals before they are combined with the output of the main amplifier, again in anti-phase and with equal magnitude. The resultant output signal is, theoretically at least, free of both intermodulation and harmonic distortion. Feedforward has inherent stability advantages over feedback topologies, despite the substantial phase shifts involved. This is of particular importance in RF and microwave circuits as inevitable poles and resonances at frequencies near the band of interest make it difficult for stable feedback to be achieved.

Despite these advantages, Feedforward is rarely employed in commercial BTSs as it is notoriously difficult to realise in practice. The system is open-loop, so variations in the characteristics of all the circuit components with time, temperature and output power are not automatically compensated for as they are in a feedback topology. The basic system shown in Figure 2.6 has been improved with adaptive cancellation control circuitry, employing microprocessors and algorithms to monitor the distortion cancellation and adjust the delay lines, attenuators and error amplifier gain to maintain performance [10, 11]. Although the results are impressive, with more than 30dB distortion improvement reported, the increase in complexity is considerable and as such, the use of such advanced techniques is even more expensive and unattractive than the 'basic' system shown here.

Further to these problems, the degree of linearisation is dependent upon the accurate cancellation of waveforms at two distinct points. It has been shown [12, 13] that if the two paths from the main input to the inputs of the first subtractor have a phase mismatch of $\Delta\phi$ and a relative gain mismatch of $\Delta A/A$, then the suppression of the magnitude of the IM products in the output is given by:

$$E = \sqrt{1 - 2\left(1 + \frac{\Delta A}{A}\right)\cos\Delta\phi + \left(1 + \frac{\Delta A}{A}\right)^2} \tag{2.8}$$

The relationship between distortion suppression and the phase and amplitude imbalance is shown plotted in Figure 2.7, were it can be seen that at least 5dB of cancellation is obtained across the whole swept range (±20°, ±20%). However, at least 20dB of cancellation requires a phase and amplitude accuracy of ±5° and ±10%, respectively.

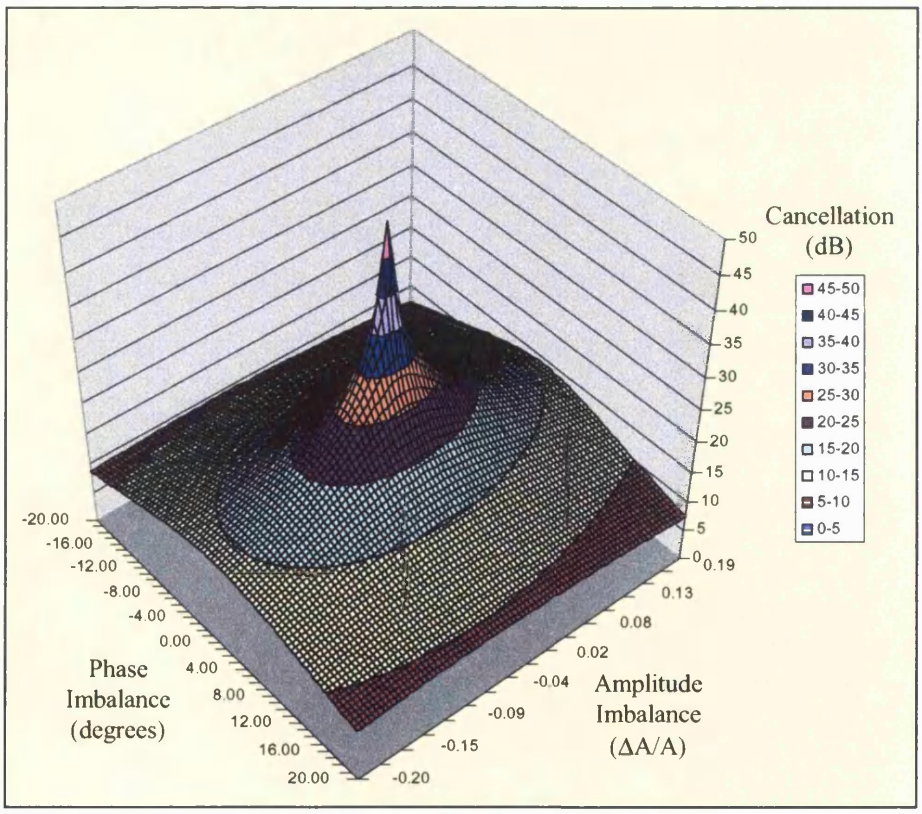


Figure 2.7: Cancellation vs. Phase and Amplitude Balance

The plot of Figure 2.7 represents the phase and amplitude tolerance of only a single loop, so the performance will be further degraded by any imbalance in the second loop. This tight

error tolerance implies that manual adjustment will usually be required to optimise performance, and as such, Feedforward circuits are unsuited to mass-production.

The signals that are processed by the error amplifier are by their very nature much smaller than those in the main PA, typically in the region of 25dB less. As a result, the intermodulation distortion introduced by this stage is much less severe than that in the main loop, and is not significantly detrimental to performance. However, the power handling capability of the error amplifier must be comparable with that of the main amplifier to achieve linear amplification, and this, coupled with the inherent losses in the system, results in low overall efficiencies, typically between 5% and 10% [14].

Feedforward systems may also be used to correct linear distortion (the nonideal gain and phase variations with frequency - i.e. frequency response ripple - that any practical RF amplifier exhibits) [15]. As discussed in Section 2.1.3, this type of distortion is not as detrimental to amplifier linearity as amplitude distortion, so is not considered for most applications.

The other advantages of Feedforward are as follows:

- Unconditional stability is assured.
- Gain is not substantially reduced as with feedback topologies, and the gain-bandwidth product is preserved within the band of interest.
- Distortion improvement is independent of the magnitude or shape of the amplifier delay and as the error amplifier is of lower power and lower noise, a lower overall noise figure results.
- Multiple loops maybe nested to increase linearity still further, though this increases the complexity and decreases the efficiency of the system.

2.2.3 Predistortion

Predistortion was originally developed for satellite TWT amplifiers, where it was used extensively, and has experienced a revival in recent years as a possible solution for solid-state applications as well. All predistortion techniques - closed or open-loop, active or passive, digital or analogue – work to the same underlying principle. That is, they deliberately distort the input signal prior to amplification in a manner contrary to the distortion caused by the PA, so a 'clean' output signal is produced. The concept is illustrated below in Figure 2.8.

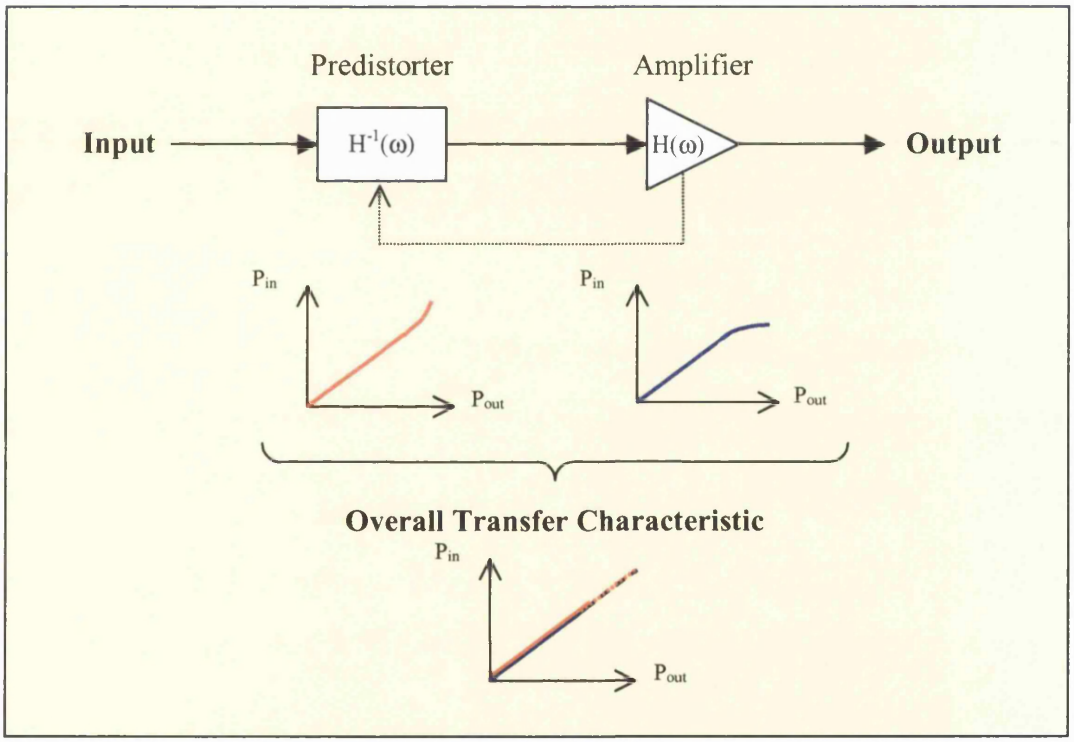


Figure 2.8: Predistortion Concept

Over the years, numerous predistorters have been proposed. As in most areas of engineering, the choice between different implementations is a compromise between cost and effectiveness. A simple predistorter that can oppose both AM-AM and AM-PM distortion has been realised with an RF level-dependent resistor combined with a fixed capacitor, achieving up to 10dB reduction in ACPR [16]. Unfortunately, in practical applications it is

very difficult for a simple lineariser such as this to achieve this improvement consistently, although even a reduction of only 1-2dB would still allow the output power of the PA to be increased for a given level of ACPR, increasing efficiency accordingly. For simple handset applications, this valuable improvement can be obtained relatively cheaply, with moderate changes to circuit complexity. However, when linearity requirements are more stringent – such as in multicarrier applications – simple analogue techniques such as this cannot adequately correct for both amplitude and phase distortion, and are incapable of achieving the required fidelity.

The majority of recent lineariser developments have focussed on exploiting the versatility and adaptability of DSP to apply controlled predistortion to the input signal, as shown in Figure 2.9. For this to be achieved, the previously-measured nonlinear amplitude and phase characteristic of the PA is stored in an array, which is accessed to give the required phase and amplitude correction according to the magnitude or instantaneous power of the input envelope [17].

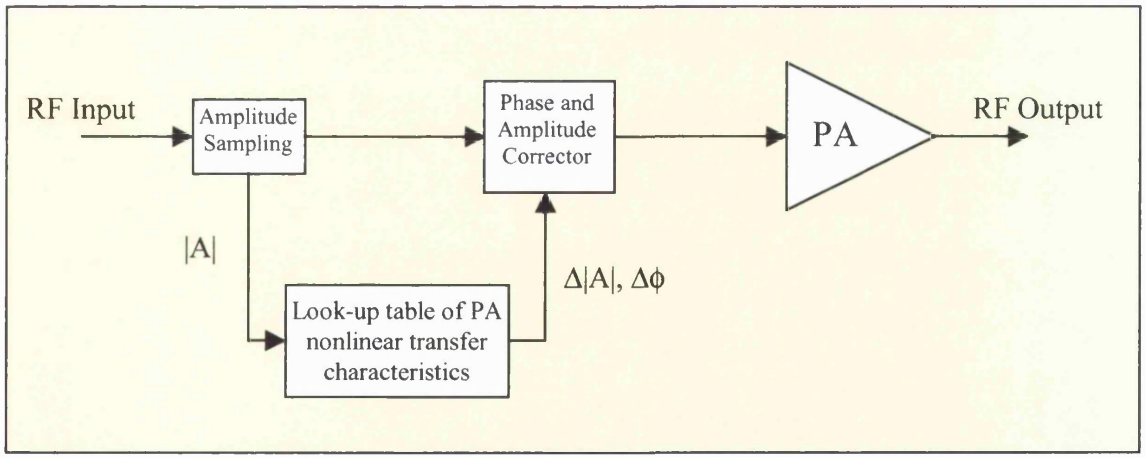


Figure 2.9: Digital Predistortion Schematic

The speed, bandwidth and dynamic range of the digital circuitry is the limiting factor in systems of this type, and multi-carrier linearisers with bandwidths greater than 20MHz or so are not currently viable. However, as the technology improves and gets faster, DSP-based

linearisers will become increasingly attractive. Enhancements to the basic topology of Figure 2.9 have also been proposed, with adaptive digital predistortion showing the most promise in terms of performance. Distortion corrections in the order of 10-20dB have been reported [18].

2.2.4 Negative Feedback

Feedback has been used as a means of prescribing the gain of an amplifier with external components for a long time. A 'classical' feedback topology is shown below in Figure 2.10.

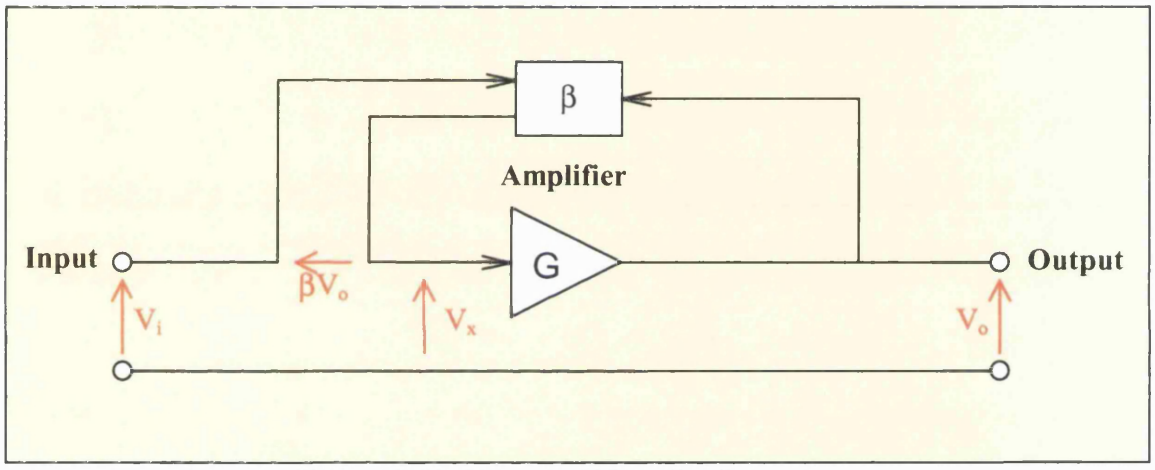


Figure 2.10: 'Classical' Feedback topology

Referring to Figure 2.10, the overall gain can be expressed as follows:

$$G_{tot} = \frac{V_o}{V_i} = \frac{GV_x}{V_x + \beta V_o} = \frac{G}{1+\beta\left(\frac{V_o}{V_x}\right)} = \frac{G}{1+\beta G} \tag{2.9}$$

Where G is the intrinsic gain of the amplifier. For systems with very high gain, $G_{tot} \approx 1/\beta$, so it can be seen that the overall transfer function of the system becomes less dependent upon the characteristics of the amplifier itself (non-linear or otherwise) as the gain increases.

For a non-compensated weakly nonlinear amplifier with sinusoidal input signal V_x, the output will be of the following form:

$$V_o = G_1 V_x + G_2 V_x^2 + G_3 V_x^3 + G_4 V_x^4 + \ldots \tag{2.10}$$

If the output of the amplifier with feedback applied is written as:

$$V_o = G_1' V_x + G_2' V_x^2 + G_3' V_x^3 + G_4' V_x^4 + \ldots + G_n' V_n^n \tag{2.11}$$

it may be shown that both the linear and non-linear products of the two cases are also related by:

$$G_n' = \frac{G_n}{1 - G_1 \beta} \tag{2.12}$$

Equation 2.12 shows that for this feedback configuration, the non-linear products are reduced by the same ratio as the gain, *for the same level of power output*. This property has great benefits at audio frequencies, where gain is plentiful and can be sacrificed in order to improve linearity. However, at RF there are severe disadvantages:

i) The smaller open-loop gains of high frequency devices require the use of several cascaded stages.

ii) Rapid phase rotation of the gain characteristic can quickly turn negative feedback into positive, resulting in instability and oscillation. The use of cascaded stages compounds this problem.

Although classical feedback has the potential for use in some microwave applications, the linearity benefits are too limited for multicarrier applications, and will remain so unless a new technology with enough RF voltage gain arrives.

2.2.5 Active Feedback

An improved feedback technique utilises a small-signal amplifier in the feedback path [19, 20], which works to generate distortion products from the fed-back output signal. The extra IMD3 products then pass through the main amplifier, along with the carriers, and the phase

shift and attenuation introduced in the feedback path are such that the products produced by the main amplifier and those that are fed back arrive at the output in antiphase and with equal magnitude. The result is distortion reduction, with the same degree of tolerance as other cancellation schemes (see Figure 2.7, page 34).

Again, this is a negative feedback loop so gain reduction results; however, this is not as severe as for 'lossy' feedback methods, and improves on the stability problems discussed in the previous section.

The system has similarities to Feedforward, in that a second amplifier is required and destructive interference is used to reduce the IMD3 products; however, there are advantages in that the structure is much simpler and the power handling of the auxiliary amplifier need not be close to that of the main amplifier.

2.2.6 Envelope Feedback

Envelope correction is a particular form of closed loop envelope predistortion that has been used in VHF and UHF solid-state amplifiers for many years [21]. This simple technique can be best illustrated by considering a two-tone signal in the time-domain, which appears as a single carrier, double sideband modulated at the difference frequency, as shown overleaf in Figure 2.11.

The peaks of the envelope of the RF carriers will be compressed as shown when this is amplified, due to AM-AM distortion and output power saturation as discussed in Section 2.1.2. This compression can be detected as it takes place at the difference frequency (typically MHz), and can therefore be removed by inclusion of an AGC loop, provided that the amplifier is well below saturation. An example implementation for an envelope correction scheme is shown overleaf in Figure 2.12.

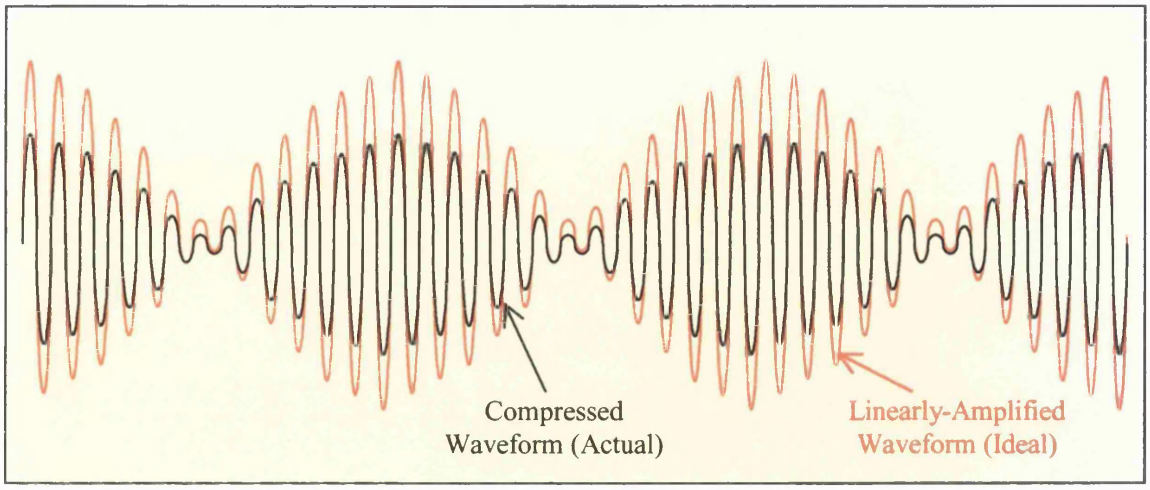


Figure 2.11: Time domain waveform of two closely-spaced RF carriers

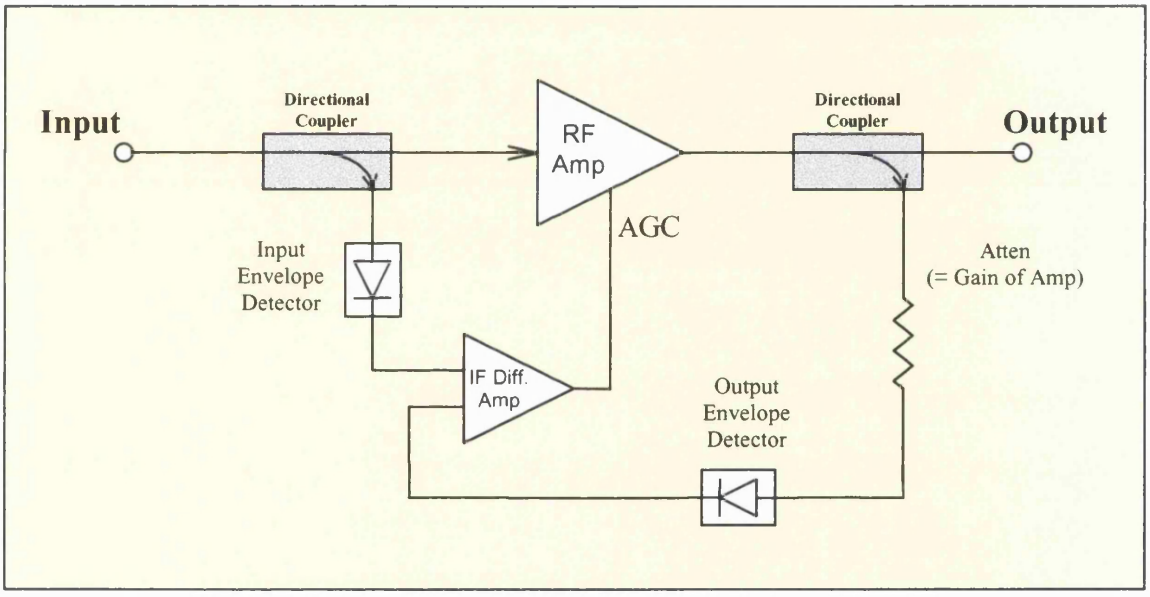


Figure 2.12: Schematic of Envelope Correction technique

The envelopes of the input and output signals are compared, and the difference - i.e. the amount of compression - is used to control the gain of the amplifier as shown in Figure 2.12. Although this technique can give useful improvement, it has several limitations. Firstly, the amplitude correction is bounded by the inherent power saturation of the amplifier, so it rapidly becomes ineffective as compression is approached. At much lower signal levels, the gain required from the video amplifier for any useful benefit increases rapidly, leading ultimately to bandwidth and stability problems. This technique also makes no attempt to

correct for AM-PM distortion, and can even create more AM-PM if the delays in the detection and video signal processing result in a video phase-difference between the AM and PM processes. Although this effect can be minimised with much faster video circuitry, this creates problems for higher data-rate signals, and as such this technique is unsuited to wideband and multicarrier applications.

2.2.7 Polar-Loop Feedback

The polar-loop lineariser is essentially an extension of the simple envelope-correction lineariser described in Section 2.2.6, with both amplitude and phase correction employed. The implementation shown in Figure 2.13 adds a phase-locked loop to maintain a constant phase transfer characteristic through the PA, the gain of which is also manipulated to reduce the amplitude compression as before [14].

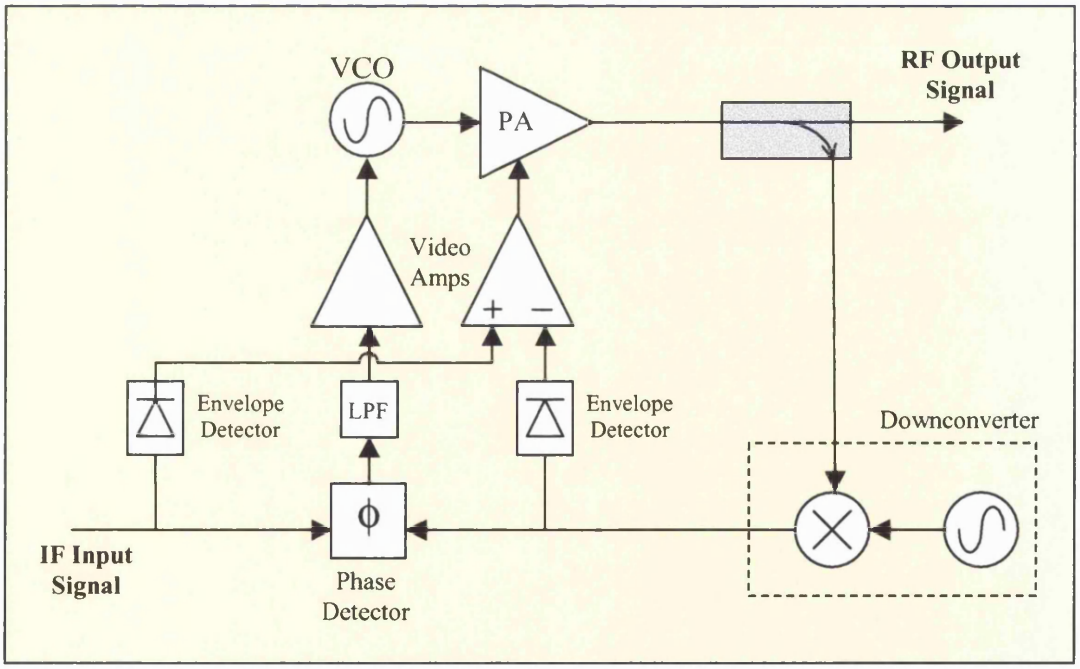


Figure 2.13: Polar Loop Lineariser

Again, the bandwidth requirements of the video circuitry are critical to the performance of the circuit and, as before, they limit the use of polar linearisers to single-carrier applications only.

2.2.8 Cartesian Loop Feedback

Cartesian Loop correction was developed more recently, and has certain advantages over the polar loop shown in the previous section [22-24]. It can only be employed in transmitters using quadrature-modulation, and particularly lends itself to digital architectures where the I and Q baseband waveforms are often directly available. An example of a Cartesian feedback lineariser is shown in Figure 2.14.

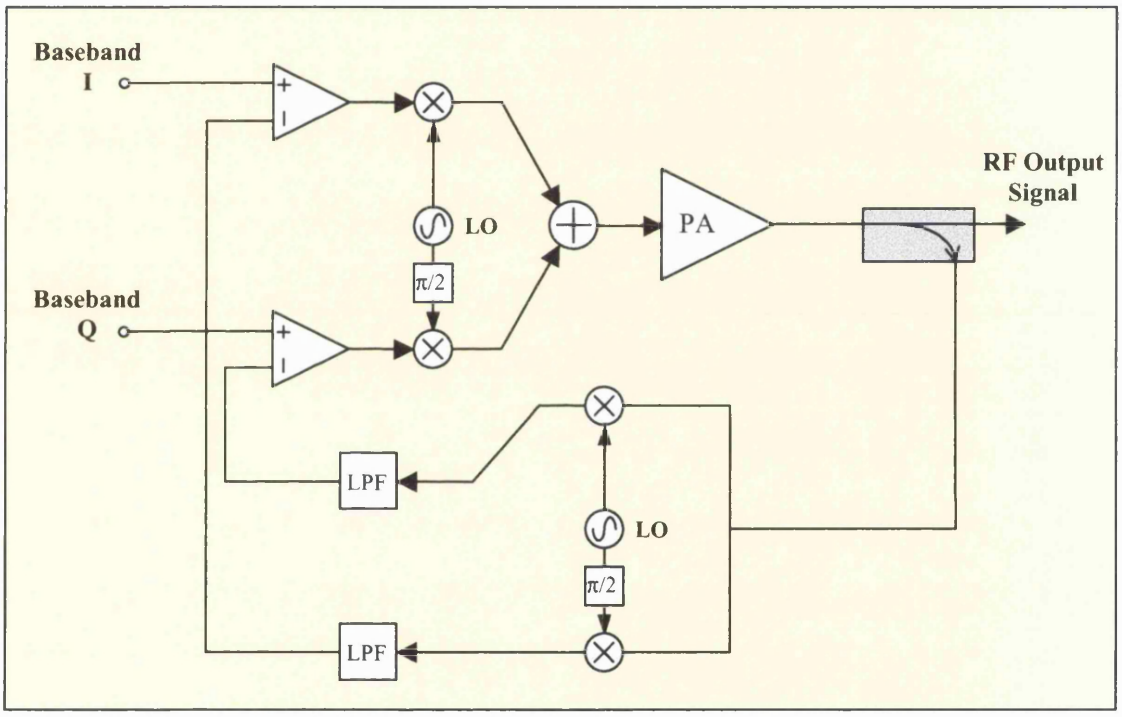


Figure 2.14: Cartesian Loop Linearisation

As the modulating waveform is split into two quadrature channels, it is possible to track and adjust for both amplitude and phase distortion. The I and Q channels can both be processed in well-matched paths, so it does not have the polar-loop problems of differing bandwidth and signal processing requirements for the magnitude and phase paths, thus reducing the introduction of phase-shifts between the AM-AM and AM-PM correction processes. Again, the ability to handle multicarrier signals is limited by video bandwidth and stability

limitations, but it does represent a neat and attractive transmitter architecture that has been reported to achieve up to 45dB of linearity improvement in narrowband applications.

2.2.9 'Linear Amplification Through Nonlinear Components' (LINC)

This interesting approach increases efficiency by avoiding the inherent problems of non-constant-envelope modulated signals by exploiting the fact that a signal such as $v_{in}(t)=a(t)\cos[\omega_c t+\phi(t)]$ can be expressed as the sum of two constant-amplitude phase-modulated signals, $v_1(t)=0.5V_o\sin[\omega_c t+\phi(t)+\theta(t)]$ and $v_2(t)=-0.5V_o\sin[\omega_c t+\phi(t)-\theta(t)]$, where $\theta(t)=\sin^{-1}[a(t)/V_o]$. These two components can therefore be formed from $v_{in}(t)$, amplified in separate high-efficiency nonlinear PAs and then recombined, with the resultant signal being a linearly-amplified version of $v_{in}(t)$. However, realisation of the component signals $v_1(t)$ and $v_2(t)$ is complex, mainly due to the fact that their phase must be modulated by $\theta(t)$, a nonlinear function of $a(t)$. Nonlinear frequency-translating loops have been suggested as possible implementations [25, 26], but loop stability is often a problem. This problem can be bypassed by expressing $v_1(t)$ and $v_2(t)$ differently [27], in such a way that the nonlinear operation required by the mapping can be implemented far more easily.

Aside from problems of implementation complexity, there are two other issues. Firstly, the gain and phase characteristics of both nonlinear amplifiers must be well-matched to avoid residual distortion in the output signal, and secondly, the output combiner must provide high-isolation between the two PAs, and will consequently introduce significant loss.

2.2.10 Second-Harmonic Feedback and 'Interstage Second Harmonic Enhancement'

These two techniques have been recently reported in the literature [28, 29], but are yet to establish themselves as viable linearisation solutions. Although, to the author's knowledge,

they have so far only been applied in experimental 'one-off' prototypes, there are similarities to the lineariser developed in the course of this thesis that merit their inclusion here.

Second-harmonic feedback can be most clearly illustrated with simple power-series analysis of the type developed in Section 2.12. We will begin by recalling the nonlinear products produced when the two-tone signal of Equation 2.4 is amplified by a device having a third-order power-series transfer function as shown in Equation 2.1. The second and third-order distortion components may be written as:

$$G_2V_{in}^2 = \underbrace{G_2A^2}_{\text{DC term}} + \underbrace{\frac{1}{2}G_2A^2\left(\cos(2\omega_1 t)+\cos(2\omega_2 t)\right)}_{\text{Second-Order Harmonics}} + \underbrace{\frac{1}{2}G_2A^2\left(\cos(\omega_2-\omega_1)t+\cos(\omega_1+\omega_2)t\right)}_{\text{Second-Order IM Products}} \quad (2.13)$$

$$G_3V_{in}^3 = \underbrace{\frac{9}{4}G_3A^3\left(\cos(\omega_1 t)+\cos(\omega_2 t)\right)}_{\text{Fundamentals}} + \underbrace{\frac{1}{4}G_3A^3\left(\cos(3\omega_1 t)+\cos(3\omega_2 t)\right)}_{\text{Third-Order Harmonics}}$$

$$+\underbrace{\frac{3}{4}G_3A^3\left(\cos(2\omega_1+\omega_2)t+\cos(2\omega_2+\omega_1)t\right)}_{\text{Out-of-Band Third-Order IM Products}} + \underbrace{\frac{3}{4}G_3A^3\left(\cos(2\omega_1-\omega_2)t+\cos(2\omega_2-\omega_1)t\right)}_{\text{In-Band Third-Order IM Products}} \quad (2.14)$$

The important terms of the above expansions are the in-band IMD3 products at $2\omega_1$-ω_2 and $2\omega_2$-ω_1 and the second-order harmonic components at $2\omega_1$ and $2\omega_2$. If the latter two components are 'selectively' fed back to the input of the amplifier, the input signal is modified to become:

$$V_{in}'(t) = A\cos(\omega_1 t)+B\cos(\omega_2 t)+C\cos(2\omega_1 t+\phi_1)+D\cos(2\omega_2 t+\phi_2) \quad (2.15)$$

Where C, D and ϕ represent magnitudes and a phase-shift that are dependent upon the transfer characteristics of the feedback network. When this input signal interacts with the amplifier transfer function, new output signal components are produced. The two IMD3

products are now phasors consisting of the vector sum of three signals. For example, the third-order component at $2\omega_2$-ω_1 is given by:

$$ADG_2\cos(2\omega_2 t-\omega_1 t-\phi_2)+\frac{3}{4}AB^2G_3\cos(2\omega_2 t-\omega_1 t)+\frac{3}{2}CDG_3\cos(2\omega_2 t-\omega_1 t+\phi_2-\phi_1) \quad (2.16)$$

The middle term of Equation 2.16 represents the third-order intermodulation due to the amplifier as before, whereas the other two components are produced by the addition of the second-order harmonic. The amplitude and phase of the first of these signals are determined by the transfer characteristics of the feedback loop, so these can be manipulated in order to cancel the existing IMD3. This condition is satisfied when:

$$D=\frac{3B^2G_3}{4G_2} \qquad \text{and} \qquad |\phi_2|=180° \quad (2.17)$$

This analysis may be repeated to show that a similar condition exists for the cancellation of the other IMD3 product. Achievable distortion reduction with a two-tone input signal has been reported to be as high as 30dB, and can be maintained across a wide dynamic range.

However, in this form, the technique has not been applied to a practical system. The difficulties lie in the design of the feedback loop, which must have a transfer function that gives the correct amplitude and phase for both the newly-produced cancellation signals. This is not something that can be accurately predicted in simulation, and although it is possible that a fixed feedback network could be realised, this would be difficult. This is borne out by the fact that the results reported in the literature [28] were obtained by using externally-generated, not fed-back, second-order products. There is also an issue of causality, in that the analysis presented here assumes instantaneous feedback, with no time delay around the loop. This will not be the case, and although one or two RF cycle 'slips' can be tolerated (if the rate of change of a carrier's phase and/or amplitude is small in comparison to its frequency), this will cause problems when applied to modulated carriers.

A variation of second-harmonic feedback that overcomes some of these limitations is known as 'Second-Order Interstage Enhancement' [29]. This requires two amplifiers, using a pre-amplifying stage to generate the required second-order harmonic signal that is then separated from the amplified carrier, phase-shifted, amplitude-adjusted and added back in to the RF path at the input of the second amplifier. As such, this is really a feedforward technique, and thus can overcome some of the causality issues that are present in a feedback topology. It has been shown experimentally that this technique is capable of reducing ACPR by up to 15dB.

Again, the effectiveness of this technique is dependent on maintaining accurate phase and amplitude balance. The characteristics of both amplifiers and the interstage loop will all vary independently with operating conditions and temperature, so it is unlikely that this performance can be achieved reliably in a practical application without some kind of closed-loop control. Further to this, the need for an additional amplifier reduces overall efficiency, and increases both cost and physical size.

2.2.11 Low-Frequency Feedback

In a similar way to second-harmonic feedback, this technique also produces additional anti-phase third-order components by the selective feeding back of second-order distortion components to the input [30]. Referring again to Equation 2.13, it can be seen that in addition to the harmonics at $2\omega_1$ and $2\omega_2$ there are also second-order intermodulation products at $\omega_2+\omega_1$ and $\omega_2-\omega_1$. If the difference-frequency, $\omega_2-\omega_1$ is selectively fed-back to the input of the amplifier, the input signal will become:

$$V_{in}(t) = A\cos(\omega_1 t) + B\cos(\omega_2 t) + C\cos([\omega_2 - \omega_1]t + \phi) \qquad (2.18)$$

Where C and ϕ are both determined by the transfer characteristic of the feedback loop. The second-order term of the amplifier transfer function of Equation 2.1 will now produce the following distortion components:

$$G_2V_{in}^2(t) = G_2\left(\frac{A^2+B^2+C^2}{2}\right) + G_2BC\cos(\omega_1 t-\phi) + G_2AC\cos(\omega_2 t+\phi) \quad \} \text{ New Fundamentals}$$

$$+G_2\left(\frac{A^2}{2}\cos(2\omega_1 t) + \frac{B^2}{2}\cos(2\omega_2 t)\right) \quad \} \text{ Second-order harmonics}$$

$$+G_2AB[\cos(\omega_1+\omega_2)t + \cos(\omega_2-\omega_1)t] \quad \} \text{ Second-order IM products}$$

$$+ G_2AC\cos[(2\omega_1-\omega_2)t-\phi] + G_2BC[(2\omega_2-\omega_1)t+\phi] \quad \} \text{ New third-order IM products} \qquad (2.19)$$

From the above expression, it can be seen that two new products are generated by the second-order term of the transfer function, that they appear at the IMD3 frequencies, and that their phase and amplitude are quantities that are determined by the transfer characteristic of the feedback network. It should also be noted that there are new products generated at the fundamental frequencies ω_1 and ω_2, but their amplitude is so small in comparison to the fundamentals that they can be ignored. In order for distortion cancellation to occur, the amplitudes of the two sets of IMD3 components must be equal and their phases opposite. By comparing Equations 2.19 and 2.14 it may be seen that these conditions are satisfied if:

$$G_2AC = \frac{3G_3A^2B}{4}, \quad G_2BC = \frac{3G_3AB^2}{4} \quad \text{and} \quad |\phi| = 180° \qquad (2.20)$$

Recalling that C is the amplitude of the fed-back second-order difference component, and that this is equal to ABH, where H represents the amplitude response of the feedback loop at ω_2-ω_1, it may be easily shown that both the upper and lower IMD3 components will be cancelled if:

$$H = \frac{3G_3}{4G_2^2} \quad \text{and} \quad |\phi| = 180° \qquad (2.21)$$

Thus, if the coefficients G_2 and G_3 are constant, the amplitude and phase response of the feedback loop can be fixed – the cancellation condition is independent of the amplitudes of the carriers themselves. Although in a real device the nonlinear coefficients G_n will not

remain constant as bias and loading conditions vary, the assumption is valid if their rate of change around the operating point is small.

Figure 2.15 shows the topology of a low-frequency feedback lineariser, which has been reported in the literature as giving a reduction in IMD3 of 12dB in a two-tone test [30].

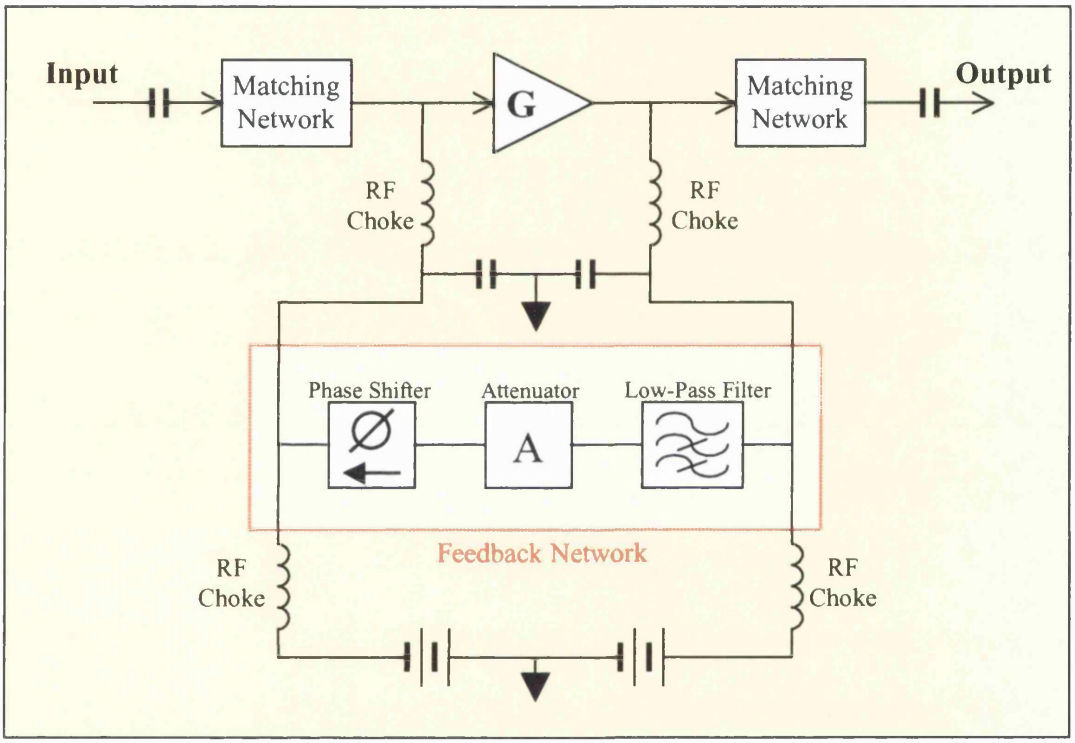


Figure 2.15: Schematic of 'low-frequency feedback' linearisation

Unfortunately, there are problems with this technique if applied in the manner shown above, the most obvious being the realisation of the feedback loop. Aside from the feedback causality issues mentioned previously, the transfer characteristic of the feedback network must be satisfy the cancellation condition of Equation 2.21 *across the whole range of carrier spacings*. This is not an issue for two sinusoidal carriers separated by a fixed difference frequency – with some adjustment it is relatively simple to obtain the required H and ϕ at a single frequency. However, maintaining the required phase and amplitude response across any useful range of carrier separation would pose a formidable synthesis problem, regardless of the choice of implementation. Referring again to Figure 2.15, the

low-frequency ω_2-ω_1 component appears at the end of the output RF choke, and is fed back to the input by the same means. The frequency responses of these low-pass networks cannot be ignored, and thus a practical feedback loop would need to equalise their transfer functions as well as providing the correct overall H and ϕ. It will be shown later that this is compounded when modulation is applied to the carriers.

The technique that is developed in the course of this thesis also exploits the second-order nonlinearity of the amplifier to reduce third-order distortion, but does so in a manner that neatly avoids the problems outlined above. In a similar way to 'interstage second-harmonic enhancement', the solution is to generate the second-order signal by some other means, prior to the amplifier itself to avoid bandwidth and feedback stability issues. However, now the required linearising signal is at IF – not at twice the carrier frequency – and as a result it can be formed far more easily. This will be explored in more detail in Chapter 4 and Chapter 5.

2.3 Summary

This chapter began with a summary of the basic concepts concerning amplifier nonlinearity, showing how the theory can be applied to analyse and explain power amplifier behaviour, under certain conditions. This background theory was followed by a summary of exisiting linearisation techniques, with discussion highlighting the benefits and limitations of each. It was shown that in general, all linearisation techniques are derived from Feedforward, Feedback or Predistortion, but there are some that are best categorised as a hybrid of two or even all three of these methods. The technique developed in this study falls into the latter category, and it will be shown that it compares well with the more well-established methods of linearisation that have already been discussed.

3. Simulated and Practical Proof-of-Concept

The discussion of nonlinearity in this study has so far centred on nonlinear systems that can be characterised by a simple power series expression involving only one variable. In many instances, if the device is not pushed into gain compression, this is a valid approximation and as demonstrated, it may be used to illustrate several important aspects of power amplifier behaviour. Unfortunately, real devices are not so well behaved and have distortion characteristics that are dependent upon bias, loading conditions and temperature. A comprehensive PA model would also need to describe both strong and weak nonlinear behaviour, as defined in Section 2.1.1. This poses a big challenge to those wishing to accurately predict the performance of such devices in CAD simulations, and although an enormous amount of effort has been devoted to producing amplifier models, a definitive solution remains elusive. Models are generally divided into two types; 'behavioural' (or 'top-down'), which employ curve-fitting techniques to produce equations that approximate measured behaviour; and 'physical' (or 'bottom-up') which model performance from device physics and geometry.

For the simulation of PAs, behavioural models are used almost exclusively. Although physical models can give great accuracy for low-power small-size devices, they do not model the secondary effects that occur when they are scaled up by tens or even hundreds of times to the geometry of power amplifiers. These limitations are well known to PA designers, and although many models attempting to describe both large and small-signal behaviour have been proposed in the literature and incorporated into commercial CAD packages, in general they serve only to provide a useful starting point for designs. In nearly all cases, a prototype circuit will require a period of manual optimisation after construction.

This Doctorate thesis is a continuation of an MSc project in which the basic principles of low-frequency linearisation were established and several important findings were made. One of the first conclusions was that the well-established 'Curtice-Cubic' FET model is inappropriate for simulating this linearisation technique, leading to the development of a behavioural nonlinear amplifier model that has been employed throughout the majority the subsequent work. As this was developed before the start of this Doctorate thesis [31], the details of the model extraction process and the design of the completed amplifier are not provided here. However, a summary of the model is provided in the next section.

3.1 Nonlinear Amplifier Model

3.1.1 2-D Maclaurin Series Description

The drain current of a common-source FET amplifier is not simply dependent upon the excitation voltage – it varies with the output drain voltage as well as well as the voltage at the gate. I_{ds} is then dependent upon two control voltages, and as such it can be expressed as a Taylor Series expansion of the form shown in Equation 3.1 below, where the expansion is truncated beyond the third-order terms as usual.

$$I_{ds}(V_{gs},V_{ds}) = I_{DS} + \frac{\delta I_{ds}}{\delta V_{gs}} v_{gs} + \frac{\delta I_{ds}}{\delta V_{ds}} v_{ds} + \frac{1}{2}\frac{\delta^2 I_{ds}}{\delta V_{gs}^2} v_{gs}^2 + \frac{\delta^2 I_{ds}}{\delta V_{gs}\partial V_{ds}} v_{gs}.v_{ds} + \frac{1}{2}\frac{\delta^2 I_{ds}}{\delta V_{ds}^2} v_{ds}^2 + \frac{1}{6}\frac{\delta^3 I_{ds}}{\delta V_{gs}^3} v_{gs}^3 + \frac{1}{2}\frac{\delta^3 I_{ds}}{\delta V_{gs}^2 \partial V_{ds}} v_{gs}^2.v_{ds} + \frac{1}{2}\frac{\delta^3 I_{ds}}{\delta V_{gs}\delta V_{ds}^2} v_{gs}.v_{ds}^2 + \frac{1}{6}\frac{\delta^3 I_{ds}}{\partial V_{ds}^3} v_{ds}^3 \quad (3.1)$$

I_{DS} represents the quiescent bias current, $I_{ds}(V_{GS}, V_{DS})$; v_{gs} and v_{ds} are the deviations of V_{gs} and V_{ds} away from the bias point such that $v_{gs}=V_{gs}-V_{GS}$, $v_{ds}=V_{ds}-V_{DS}$; and all the derivatives are evaluated at $V_{gs}=V_{GS}$, $V_{ds}=V_{DS}$. If the deviations away from V_{GS} and V_{DS} are not too severe, Equation 3.1 may be rewritten in terms of incremental voltages and currents as follows:

$$i_{ds}(v_{gs}, v_{ds}) = g_{m1}v_{gs} + g_{m2}v_{gs}^2 + g_{m3}v_{gs}^3 + g_{d1}v_{ds} + g_{d2}v_{ds}^2 + g_{d3}v_{ds}^3 + m_{11}v_{ds}v_{gs} + m_{12}v_{ds}v_{gs}^2 + m_{21}v_{ds}^2v_{gs} \quad (3.2)$$

The first six terms of the above expansion describe the dependence of i_{ds} on the input and output voltages, v_{gs} and v_{ds}; g_{m1} and g_{d1} correspond to the FET's linear transconductance and output conductance respectively, g_{m2} and g_{m3} describe the variation of transconductance with v_{gs}, and g_{d2} and g_{d3} the variation of output conductance with v_{ds}.

The three terms with coefficients m_{11} m_{12} and m_{21} are known as 'mixing terms'; they represent the physical interaction that occurs between the input and the output of a MESFET, an effect that can be observed experimentally. The coefficients m_{11} and m_{12} describe the first and second-order nonlinear dependence of g_{ds} on v_{gs}, whilst m_{11} and m_{21} represent the nonlinear dependence of g_m on v_{ds}. It was shown in a previous study [32] that these mixing terms can have a significant effect on the nonlinear behaviour of a FET, and that their contribution is often greater than that due to the nonlinear output conductance.

Several other sources of FET nonlinearity, such as the nonlinear gate-channel capacitances C_g and C_d, can also be included, and are discussed in detail in text-books and in the literature. However, both these effects are minimal in comparison to the current-source nonlinearity discussed above and can be treated as constant, provided the device is biased, and remains, within its weakly nonlinear regime.

The drain-current expression of Equation 3.2, in conjunction with the simplifying assumptions, can only describe weakly nonlinear behaviour, losing validity when strongly nonlinear effects such as gain compression and saturation come into play. However, this type of model is adequate for predicting the behaviour of a Class A amplifier not being driven

beyond its 'linear' output power limits, conditions which are assumed throughout the majority of the analytical and simulation work presented here.

A distortion measurement set-up developed separately at UCL was used to characterise the chosen device (Fujitsu FLL351ME) across a range of bias and output loading conditions [33]. This data was then used to de-embed the nonlinear current source and provide the nine coefficients of Equation 3.2, that would be valid at one bias point, chosen as V_{DS} = 10V, V_{GS} = -1.2V, ensuring Class A operation. The nine coefficients are shown below in Table 2.

g_{m1}	0.976679	g_{d1}	0.0301944	m_{11}	8.5153×10^{-3}
g_{m2}	0.2244	g_{d2}	-4.587×10^{-4}	m_{21}	-6.183×10^{-3}
g_{m3}	-0.255912	g_{d3}	-2.8475×10^{-4}	m_{12}	-4.6481×10^{-2}

Table 2: Nonlinear Coefficients

The nonlinear voltage-current relationship was then realised in the form of a 2-port Symbolically-Defined Device (SDD) in *MDS*, and a small-signal FET equivalent circuit model to represent the FET at the chosen bias point was assembled around it. The use of a Class A model allowed the manufacturer's small-signal S-parameter data to be used in the model optimisation process, as this had been measured under the same quiescent conditions. Various standard-type topologies given in the literature were investigated, with the final design selected on the basis that it gave the best fit to the manufacturer's quoted S-parameter data.

The complete small-signal equivalent circuit with the parasitic component values found to show closest agreement with the measured S-parameters is shown in Figure 3.1, with the *MDS* implementation included as Appendix A.

Referring to Figure 3.1, the intrinsic FET has been represented in its classical form with linear capacitances. Although gate and drain capacitances also contribute to FET distortion, this is a secondary effect, to the extent that it may be neglected [34].

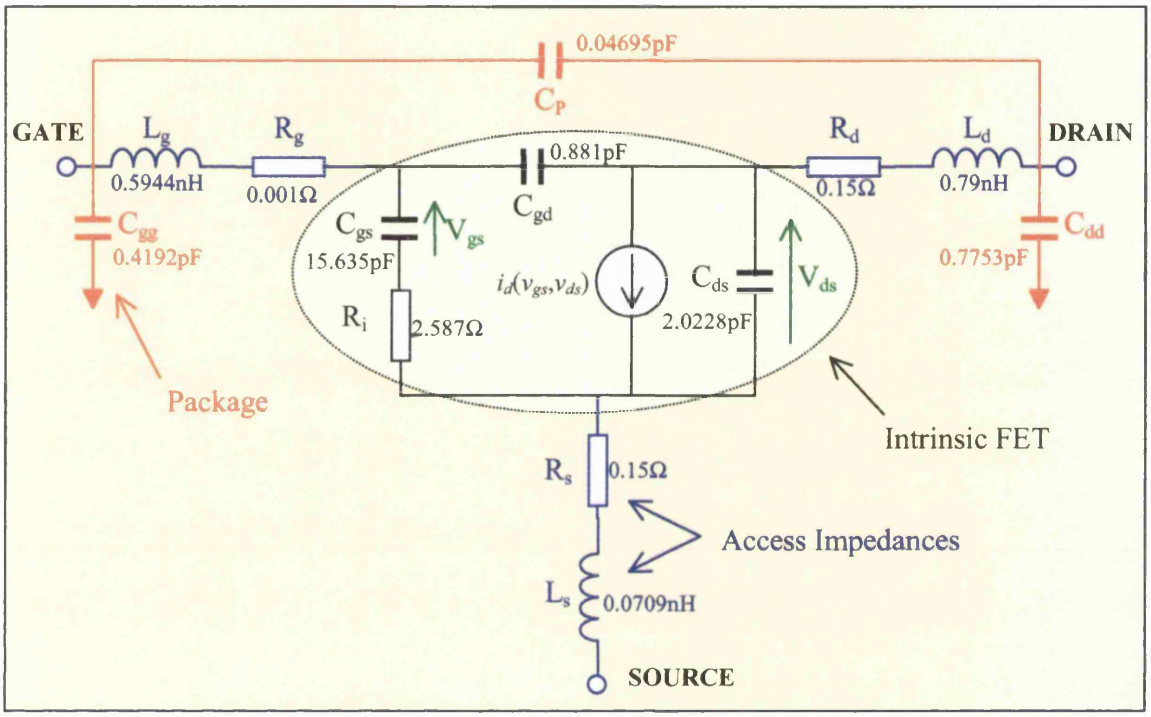


Figure 3.1: Equivalent circuit of selected FET model

The bias-independent access impedances are represented by series resistors and inductors as shown, whilst the effect of the package is modelled by the inclusion of C_{gg}, C_{dd} and C_p as indicated. As the effect of the FET output conductance has been incorporated into the nonlinear current expression, $i_{ds}(v_{gs},v_{ds})$, g_{ds} is not included as a separate lumped component as would be the case in a classical linear model.

The S-parameters that were produced by the optimised model were plotted on the same axis as the measured data, as well as that produced by the Curtice Cubic FET model for the same device. The traces of Figure 3.2 and Figure 3.3 show the result of this comparison, over a frequency span of 0.5-4GHz. The red traces were produced by the equivalent circuit model shown in Figure 3.1, the blue traces by measured library data and the green by the Curtice

Cubic model of the FLL351ME biased at V_{gs}=-1.2V, V_{ds}=10V. It can be seen that the S-parameters produced by the new model are almost indistinguishable from both the Curtice-Cubic and also the measured data.

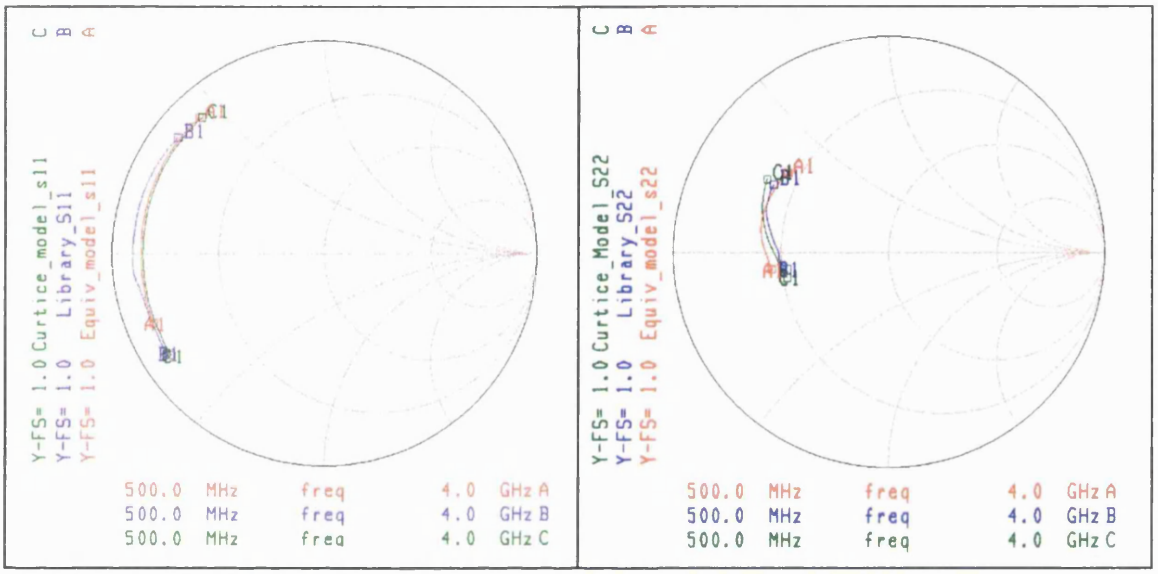


Figure 3.2: Comparison of measured S_{11} and S_{22}, new and Curtice Cubic models

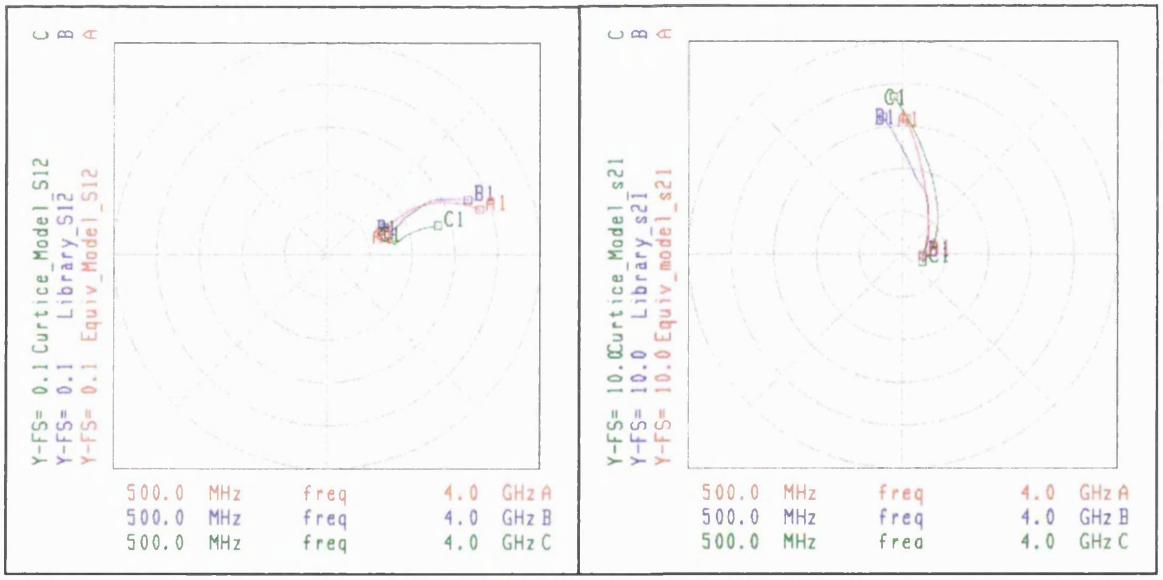


Figure 3.3: Comparison of measured S_{12} and S_{21}, new and Curtice Cubic models

3.2 Model Evaluation

Input and output matching and bias circuits for the amplifier model were designed in *MDS* and the completed circuit was used to perform simulations, enabling the measured and modelled and behaviour to be compared, as shown on the following pages.

3.2.1 S-Parameters

The S-parameters of the real amplifier were measured with a Network Analyser (HP-8510) and plotted on the same axis as those produced by the *MDS* simulation of the same circuit. These comparisons are shown in Figure 3.4, Figure 3.5 and Figure 3.6.

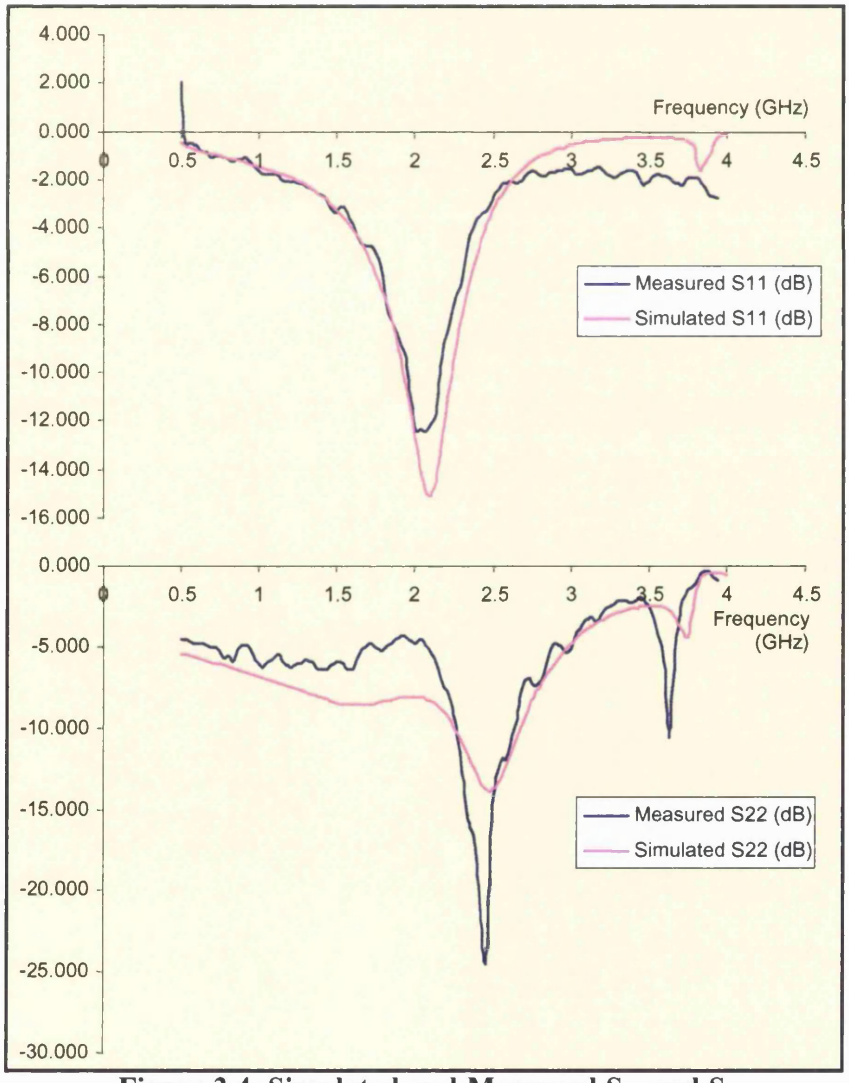


Figure 3.4: Simulated and Measured S_{11} and S_{22}

Referring to Figure 3.4, excellent agreement between the measured and simulated S_{11} was observed. The results for S_{22} are not quite so accurate and, although similar in shape, they show that the output matching is by no means optimal, with a centre frequency somewhat above that which was intended. Despite this anomaly, the impedance matching was thought to be adequate for the purposes of the investigation at this stage, as it is the amplifier's distortion performance and not gain or power performance that is under investigation.

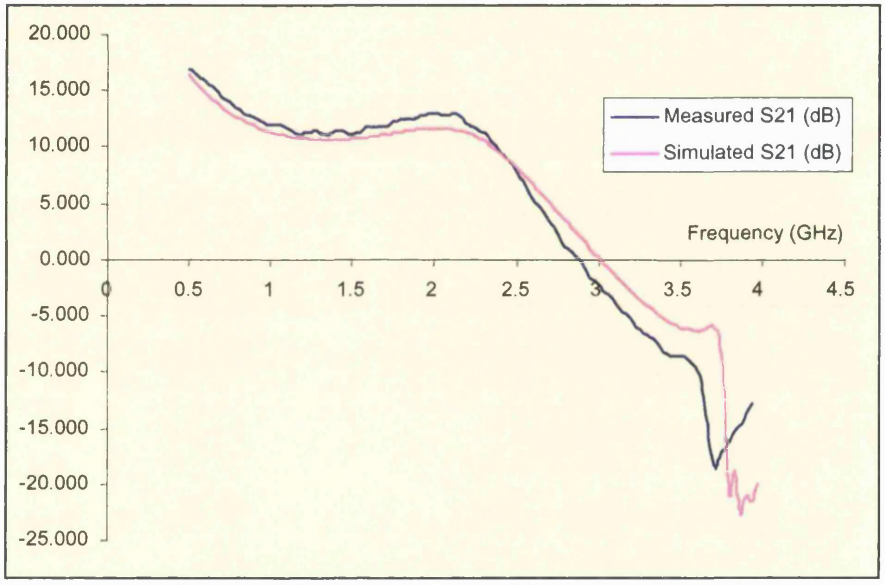


Figure 3.5: Simulated and Measured S_{21}

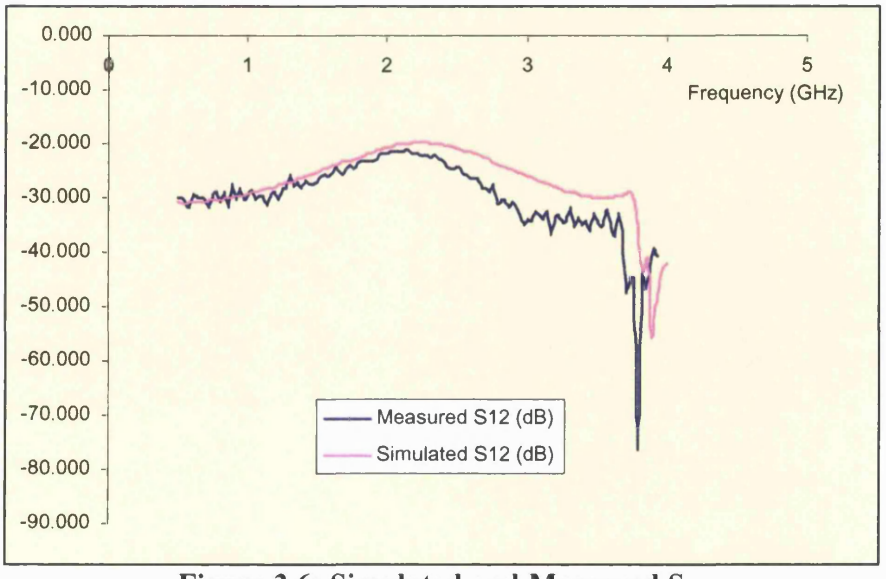


Figure 3.6: Simulated and Measured S_{12}

From the plot of S_{21} shown in Figure 3.5, it can be seen that the measured gain of the amplifier was found to be higher than that predicted by the model, despite the nonideal output impedance matching.

3.2.2 Input Power Sweep and Two-Tone Distortion Measurement

The amplifier's distortion performance was then measured, and is shown below in Figure 3.7. It can be seen that at 1.81GHz, the 1dB Gain Compression Point (GCP) is 31dBm, whilst the second and third-order intercept points are 54dBm and 45dBm respectively.

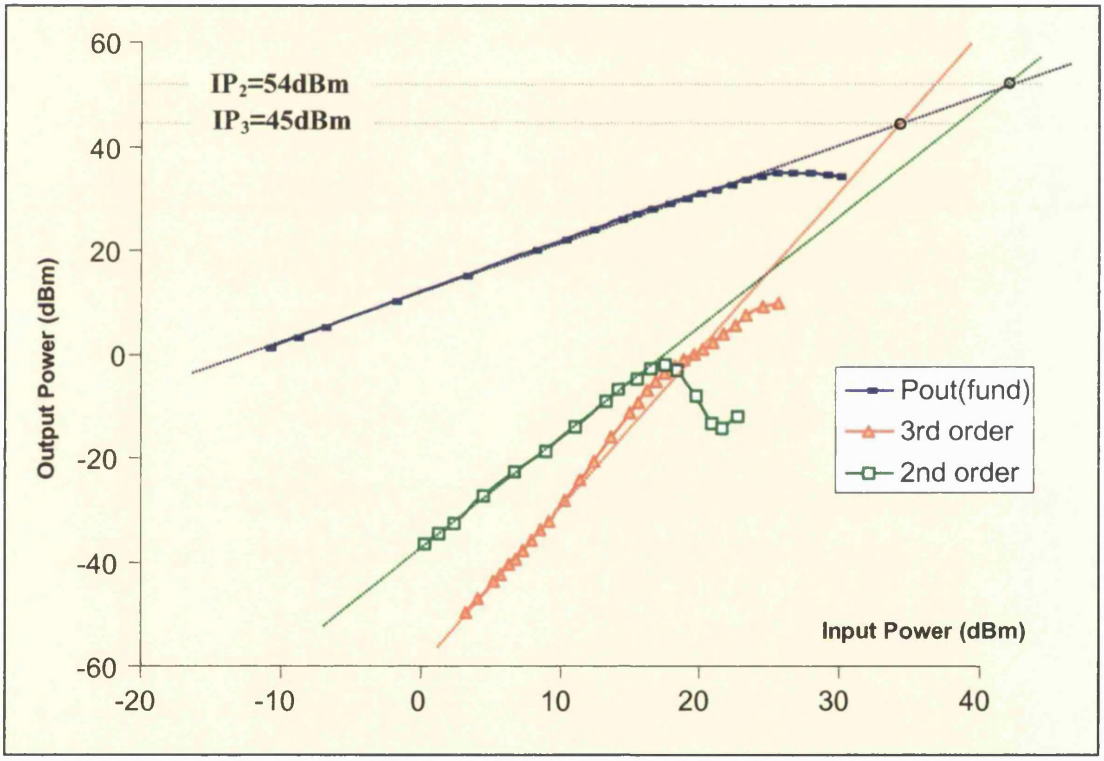


Figure 3.7: Measured Distortion Performance of FLL351ME Amplifier at 1.81GHz

The extrapolations that have been superimposed onto the above figure illustrate the difficulties of making accurate intercept-point measurements – the linear regions of the second- and third-order traces are almost down into the noise floor of the measurement apparatus, and the large distances between these linear regions and the intercept points

greatly magnify any errors. As such, the IP_2 and IP_3 shown here are to be treated as only an estimate.

The same test was then performed on the nonlinear model and circuit shown in Appendix A by simulating a two-tone test at the same frequency and over the same range of power levels. The results of this measured and simulated data were plotted on the same axis for easy comparison, and can be seen in Figure 3.8, Figure 3.9 and Figure 3.10 overleaf.

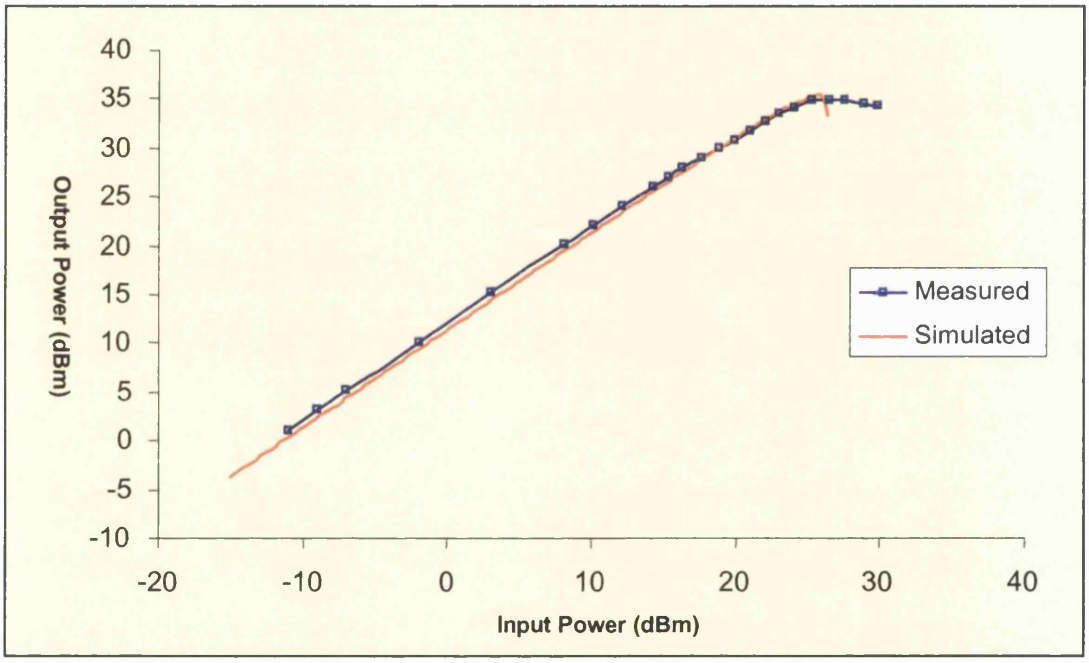


Figure 3.8: Comparison of measured and modelled fundamental power-sweep

From the distortion plots, it can be seen that the agreement between the modelled and measured data for both the fundamental and third-order characteristics is close, under small-signal conditions. Although the second-order distortion comparison shows a discrepancy of 5dB across a wide range of power levels, it was found during subsequent work that the behaviour of the model was close enough to the actual device to make further adjustment of the model unnecessary at this stage. Further justification for this is provided later.

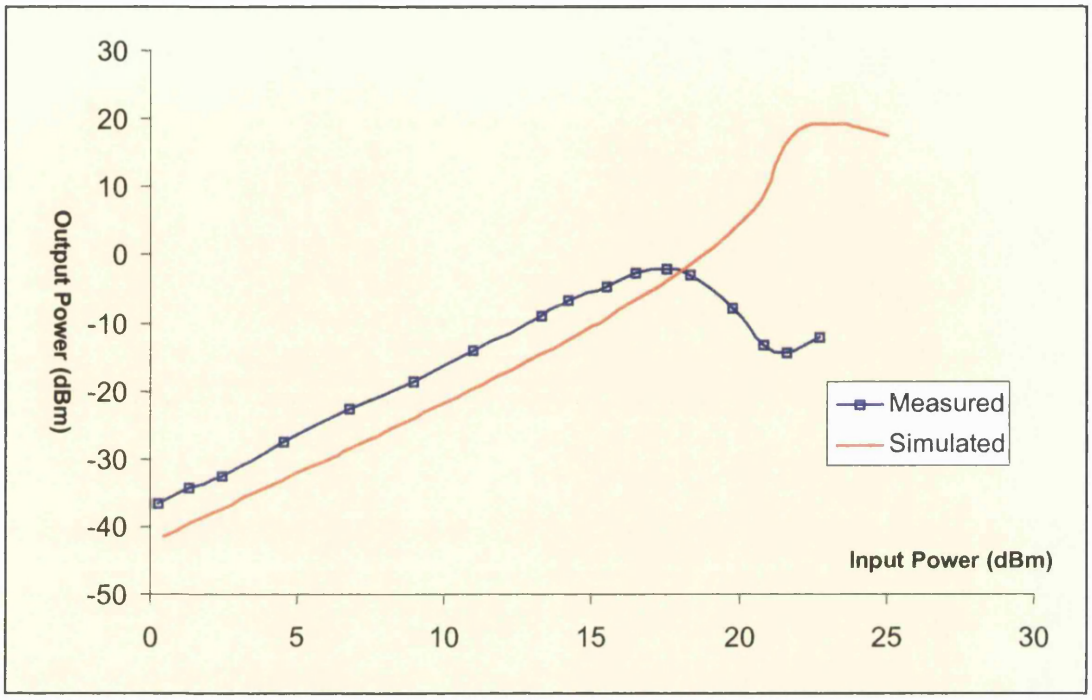


Figure 3.9: Comparison of measured and modelled second-order distortion

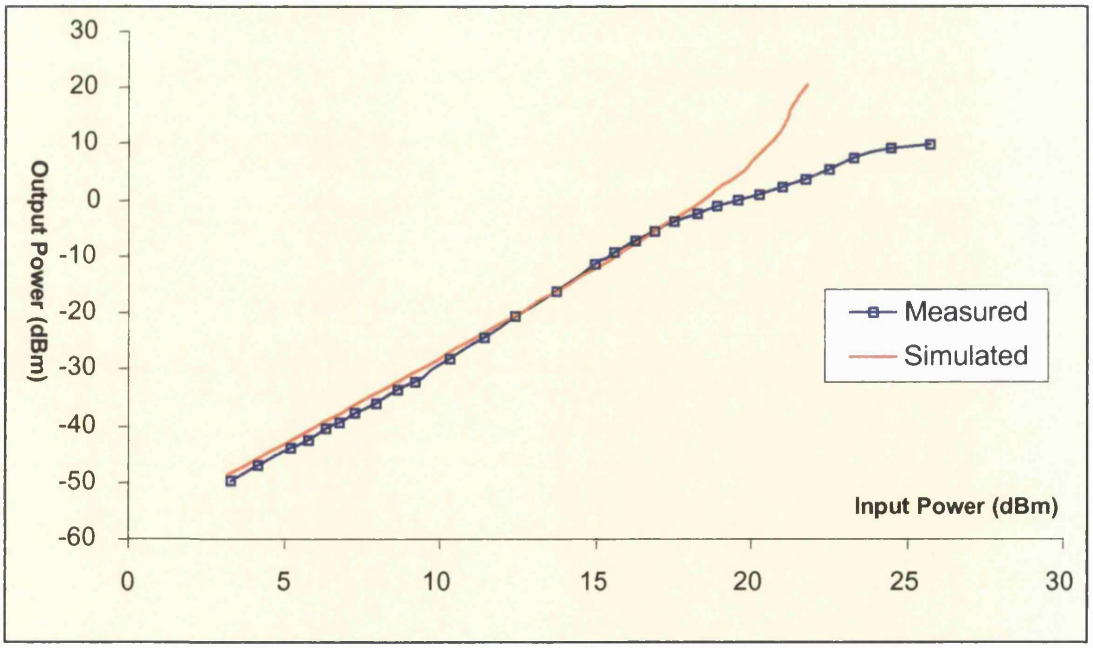


Figure 3.10: Comparison of measured and modelled third-order distortion

3.3 2-Tone Proof of Concept

3.3.1 Simulation of Low-Frequency Feedback

The practical work began by applying second-order difference-frequency feedback to the nonlinear PA model, via the quarter-wavelength bias-feeding lines in the same manner as a previous investigation [see Section 2.2.11, page 47]. In order to achieve this, a simple narrowband feedback network consisting of an adjustable 'all-pass' phase-shifter and attenuator was designed, the basic topology of which is shown in Figure 3.11 below.

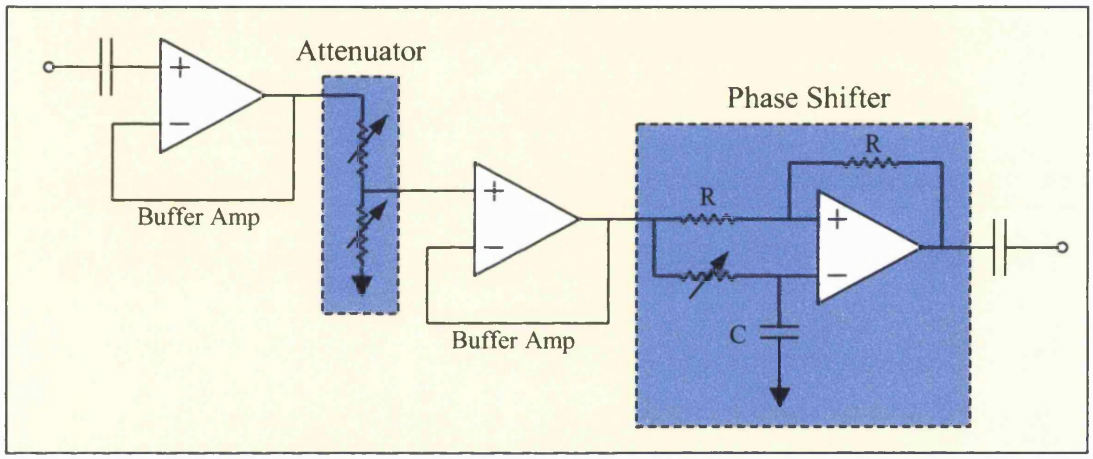


Figure 3.11: Schematic of narrowband feedback network

Referring to the above circuit, a potential divider was used to enable amplitude adjustment, and an adjustable resistor was used in the RC network controlling the phase-shift introduced by the all-pass network. Voltage-followers were used to isolate the two stages as shown.

This network was then applied to the amplifier, connected at the ends of the inductive bias-feed lines as shown in Figure 3.12 overleaf. A two-tone test (f_1=1.805GHz, f_2=1.815GHz) was then applied to the input of the circuit, using optimisation to minimise the power of the two IMD3 components, with the values of the potential divider and phase-shifter resistor chosen as the optimisable variables.

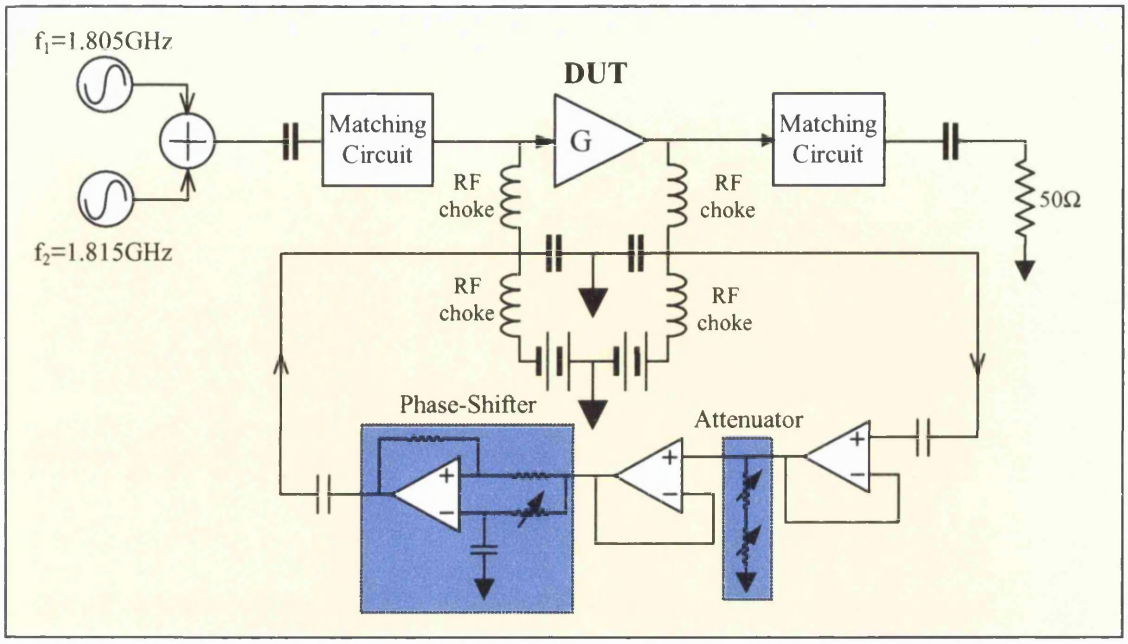


Figure 3.12: Amplifier with Low-Frequency Feedback

The output spectrum around the carriers with no feedback applied as well as the results of the optimisation are shown in Figure 3.13, with the power of the carriers and the relative IMD3 level included to highlight the reduction.

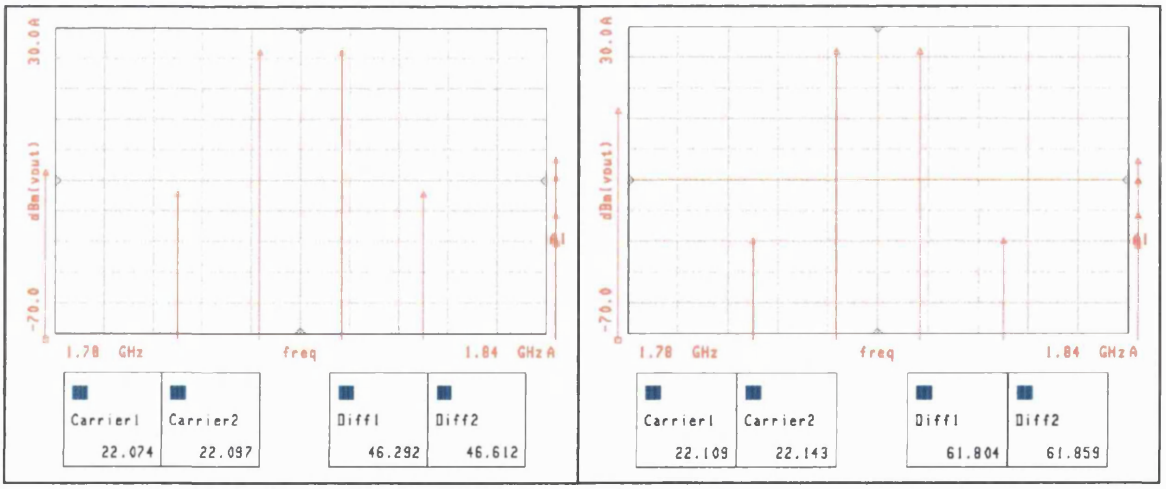


Figure 3.13: Simulation results for low-frequency feedback

Referring to the results, it may be seen that application of low-frequency feedback reduced both the upper and lower in-band intermodulation distortion by over 15dB, and also fractionally increased the power of both sinusoidal carriers. As well as validating the

analytical theory presented earlier, this simulation allowed other aspects to be investigated, as will now be shown.

3.3.2 Feedback Phase and Amplitude Sweeping

It is interesting to note that according to the simple power-series analysis, low-frequency feedback should have the capability to completely eliminate IMD3 if the optimum amplitude and phase is applied to the feed-back signal. However, the results of the simulation suggest that this is not the case, as the distortion reduction was found to be limited to approximately 15dB. In order to explore this further, the sensitivity of the distortion reduction to phase and amplitude imperfections was then assessed, by replacing the phase-shifter and attenuator in the feedback path with ideal components, the values of which were swept. The plots of Figure 3.14 show the results of this simulation, with the reduction in the lower IMD3 product plotted as '*Improvement1*' and the reduction in the upper IMD3 product plotted as '*Improvement2*'. The horizontal axis shows an attenuation deviation span of ±0.5dB, with successive traces representing phase increments of ±0.1°, ±0.5°, ±1° and ±2°.

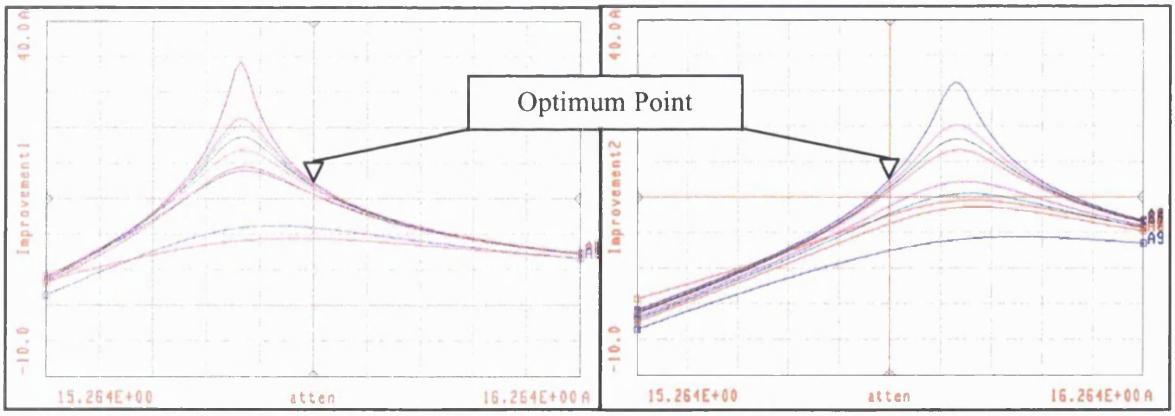


Figure 3.14: Simulated sensitivity to phase and attenuation deviation

The above plots show that there *is* a value of phase and attenuation that will yield very high IMD3 reduction, but that the optimum values are slightly different for the two components. The values of amplitude and phase-shift that gave complete cancellation of the upper and

lower IMD3 components were found to be separated by 3.7° and 0.15dB respectively. The 'optimum point' shown represents the best compromise between the two optimums, giving 15dB reduction in both, as in the two-tone simulation shown in Figure 3.13.

This behaviour can be explained by the fact that the model used in these simulations includes both transconductance (input-output) and output conductance (output-output) nonlinearities, as described in Section 3.1.1. This causes the amplifier to produce many more nonlinear components than those given by simple third-order power-series analysis employed in Section 2.1.2. For example, the low-frequency second-order component appearing at the output of the amplifier will mix with the carriers via the *output* second-order nonlinearity (given by g_{d2} in Equation 3.2), giving rise to new signal components at the same frequencies as those produced with low-frequency feedback. In fact, as reported recently this effect can be exploited as to improve linearity, in the form of 'low-frequency feedforward' [35]. This can be illustrated by considering two carriers producing a low-frequency second-order component, giving an amplifier output identical to the input signal described by Equation 2.18, as follows:

$$V_{out}(t) = A\cos(\omega_1 t) + B\cos(\omega_2 t) + C\cos([\omega_2 - \omega_1]t + \phi) \tag{3.3}$$

The components arising from the second-order output nonlinearity are identical to those given in Equation 2.19, which are re-written as:

$$\begin{aligned} G_{d2}V_{out}^2(t) = {} & G_{d2}\left(\frac{A^2+B^2+C^2}{2}\right) + G_{d2}BC\cos(\omega_1 t - \phi) + G_{d2}AC\cos(\omega_2 t + \phi) \\ & + G_{d2}\left(\frac{A^2}{2}\cos(2\omega_1 t) + \frac{B^2}{2}\cos(2\omega_2 t)\right) \\ & + G_{d2}AB[\cos(\omega_1+\omega_2)t + \cos(\omega_2-\omega_1)t] \\ & + G_{d2}AC\cos[(2\omega_1-\omega_2)t - \phi] + G_{d2}BC[(2\omega_2-\omega_1)t + \phi] \end{aligned} \tag{3.4}$$

The last two terms in Equation 3.4 show the components that fall on the IMD3 frequencies, and as before, their amplitude and phase is governed by the amplitude and phase of the second-order component, which in this case are determined by the impedance presented at the output of the amplifier at the difference (or envelope) frequency. Typically, third-order transconductance nonlinearity (given by g_{m3}) dominates the in-band distortion, generating larger components than either second- or third-order output conductance nonlinearity (g_{d2} and g_{d3}, respectively). For example, in this investigation the measured g_{m3}, g_{d2} and g_{d3} were found to be -0.255912, -4.587×10^{-4} and -2.8475×10^{-4} (see Table 2 on Page 54). Despite the large difference in the magnitudes of the nonlinear coefficients, if the impedance presented to the amplifier output at the difference frequency is not a perfect short-circuit, the second-order component will have a non-zero amplitude, and as shown in Equation 3.4, any phase offset is added to one of the new components and subtracted from the other. This is illustrated below in Figure 3.15, where the left-hand panel depicts the output and third-order intermodulation components produced by a third-order transconductance nonlinearity and the right-hand panel shows the same output with the effect of nonlinear output conductance included.

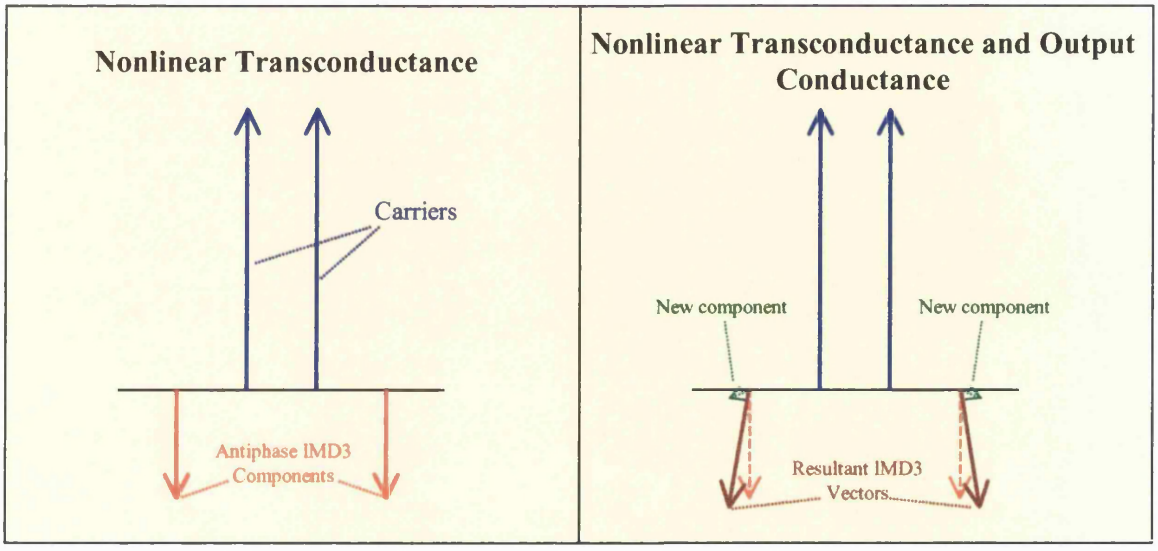


Figure 3.15: Third-order IMD skewing due to output conductance nonlinearity

The additional components produced at the IMD3 frequencies alter the amplitude and phase of the resultant vectors, tending to skew them in the manner shown. This is commonly referred to as 'memory effect' [36], a generic term to describe the frequency-dependence of in-band nonlinearities arising from both bias-circuit interactions and dynamic temperature fluctuations within the device.

Although it is possible that the additional distortion will be generated in antiphase to the dominant components, resulting in reduced distortion overall, this requires very careful control of the envelope impedance with 'intelligent' bias-circuit design. Studies to investigate the use of 'Envelope Load-Pull' [37] have shown that this is indeed possible, but difficult to exploit over large envelope bandwidths such as those required in multicarrier amplifiers. Far more commonly, these secondary effects are removed with a 'brute force' approach, ensuring a near short-circuit across the band of envelope frequencies with the use of large-value capacitors in the output bias-circuit.

However, if the effects of second-order *harmonic* components and mixing terms were also included, or if the analysis were extended to higher-orders, it would be found that there are numerous interactions that result in components at the IMD3 frequencies, so the distortion will always consist of the sums of the multiple nonlinear products that fall upon them. Hence, the upper and lower IMD3 products or ACPR bands will almost always have a slightly different amplitude and phase, and it is this fact that limits the performance of linearisation schemes such as analogue tion, or indeed low-frequency feedback. Despite this, it will be shown that useful distortion improvements are still possible, as long as the effects of the undesirable signal interactions are minimised.

3.3.3 Simulated Power Dependence

By sweeping the input power, it was possible to assess whether a fixed feedback network could achieve distortion reduction across a useful dynamic range. Figure 3.16 shows the result of this simulation, with the power of the IMD3 components plotted against input power level, both for the amplifier alone and also with feedback applied for comparison.

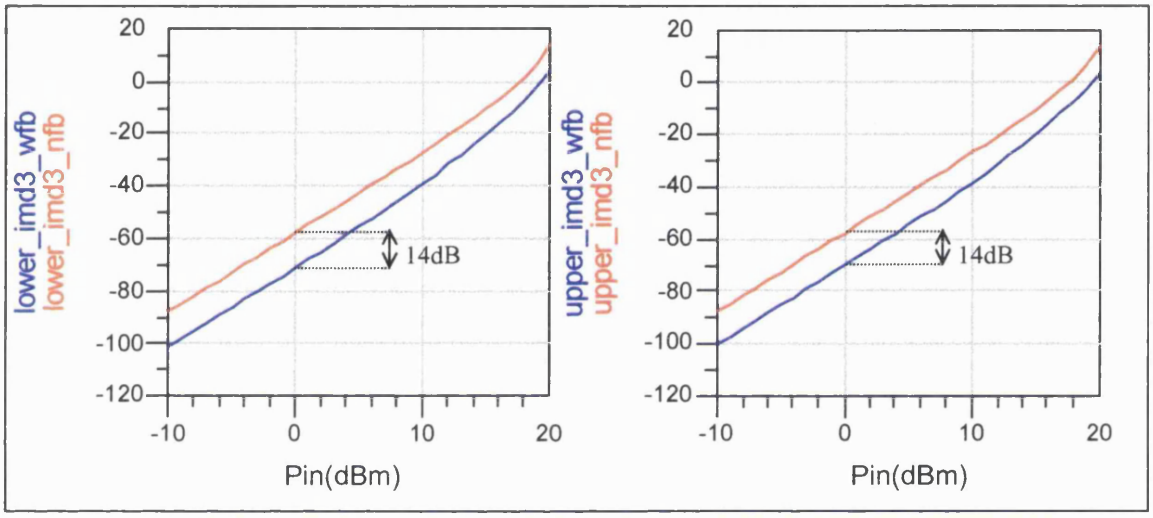


Figure 3.16: Simulated IMD3 Improvement vs. Input Power

It can be seen that the reduction of 14dB is achieved across a wide dynamic range, only becoming less effective as the device is pushed towards saturation, at which point the validity of model becomes questionable. The lack of small-signal power-dependence was expected as it is predicted by the power-series analysis presented in Section 2.2.11. However, the model used in these simulations was relatively simple, so it is probable that a practical device will not exhibit such well-behaved performance.

3.4 Experimental Verification

In order to complete the simple proof-of-concept, the same two-tone test was applied to the fabricated PA, with carrier frequencies f_1=1.805GHz and f_2=1.815GHz. The narrowband feedback network of Figure 3.11 was constructed using high-power op-amps and surface-mount passive components wherever possible. This was connected to the ends of the bias-feeding lines via DC-blocking capacitors as shown in Figure 3.12, and manually adjusted to give the greatest degree of IMD3 cancellation in both upper and lower products concurrently. The results of this test are shown in Figure 3.17 below.

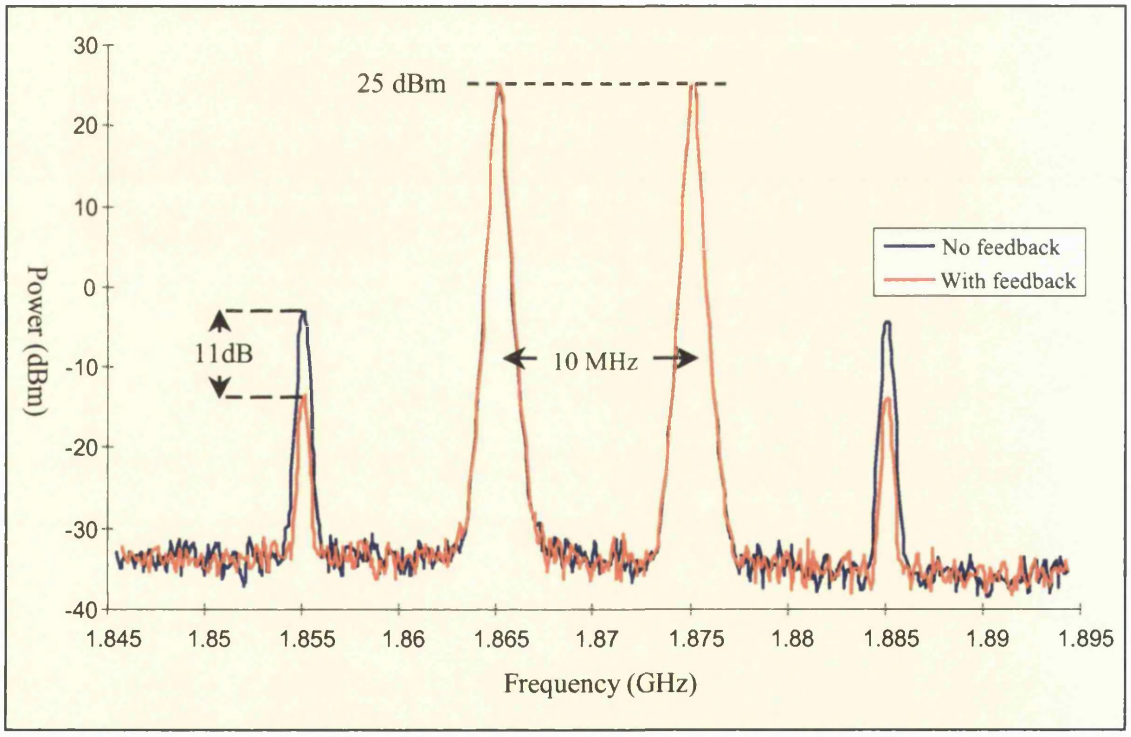


Figure 3.17: Measured Two-Tone Test Results with and without feedback applied

It can be seen that an improvement of 11dB in both IMD3 products was obtained. This is 4dB less than the 15dB that was obtained in the simulation, but still represents a useful level of distortion reduction.

Although the feedback network could be tuned with the variable resistors, the range of phase-shift adjustment was limited and the output op-amp was found to introduce a

considerable amount of harmonic distortion. Further investigation revealed that this was due to the low impedance at the end of the input bias-feed line, at the point where feedback was applied, which led to saturation of the output of the last op-amp. Although measures were taken to remedy this, the limited output current of the high-speed op-amps remained a problem. In order to explore the relationship between feedback phase and amplitude more fully, it was decided that a slightly different approach was required.

3.4.1 Linearisation Through 'Low-Frequency Injection'

As discussed above, it became necessary to provide a greater degree of tuning than that available with the simple feedback network. Although the narrowband circuit allowed the technique to be validated, it could only be tuned over a very limited range, so the range pf carrier spacings that could be tested was limited. If this technique were to be applied to a multicarrier amplifier, it would have to linearise carriers spaced by anywhere between 200kHz and 35MHz for a second-generation transmitter, or between 2MHz and 75MHz for future systems. It was therefore decided that the design of a feedback loop that would allow the required degree of tuning across bandwidths of this size would be a time-consuming (if not impossible) task and that far more insight could be gained by feeding the linearising signal *forward.*

Figure 3.18 illustrates the concept; a pair of signal generators were used to generate a two-tone test signal, with their outputs amplified separately as shown. The carriers were summed in a power-combiner, with ferrite isolators used to minimise residual intermodulation, before the composite signal was applied to the FLL351ME PA. A third signal generator was used to produce the second-order difference frequency at f_2-f_1, connected to the end of the bias-feed line at the input of the DUT as shown. The three signal generators used were all digital synthesisers, so they could be locked to a common external reference to maintain frequency

stability. The use of frequency synthesisers also allowed both the amplitude and phase of the 'injected' linearisation signal to be varied independently, effectively removing the tuning limitations of the narrowband feedback loop and allowing much greater flexibility.

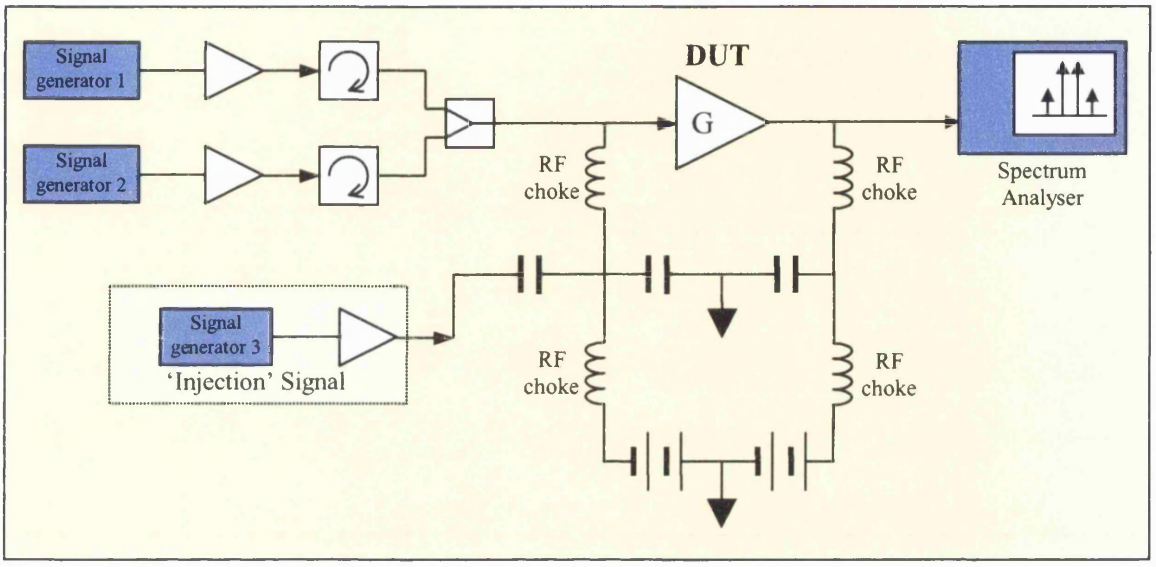


Figure 3.18: Schematic of new test-bench

The carrier frequencies were initially selected as before, with f_1=1.805GHz and f_2=1.815GHz, giving a spacing of f_2-f_1=10MHz for the third signal generator. The amplitude and phase of the injected signal were adjusted to give the greatest 'common' reduction in both upper and lower IMD3 components. The results of a typical two-tone test can be seen overleaf in Figure 3.19, where it can be seen that both IMD3 components are reduced by 16.5dB. The performance was therefore superior to that obtained with the simple feedback loop, the most likely reasons being the greater degree of phase and amplitude adjustment that was available and the lack of output power saturation. By comparison with the results shown in Figure 3.13, it can be seen that the reduction is IMD3 is also 1.5dB better than that achieved in simulation, suggesting that the nonlinear model was producing slightly pessimistic results.

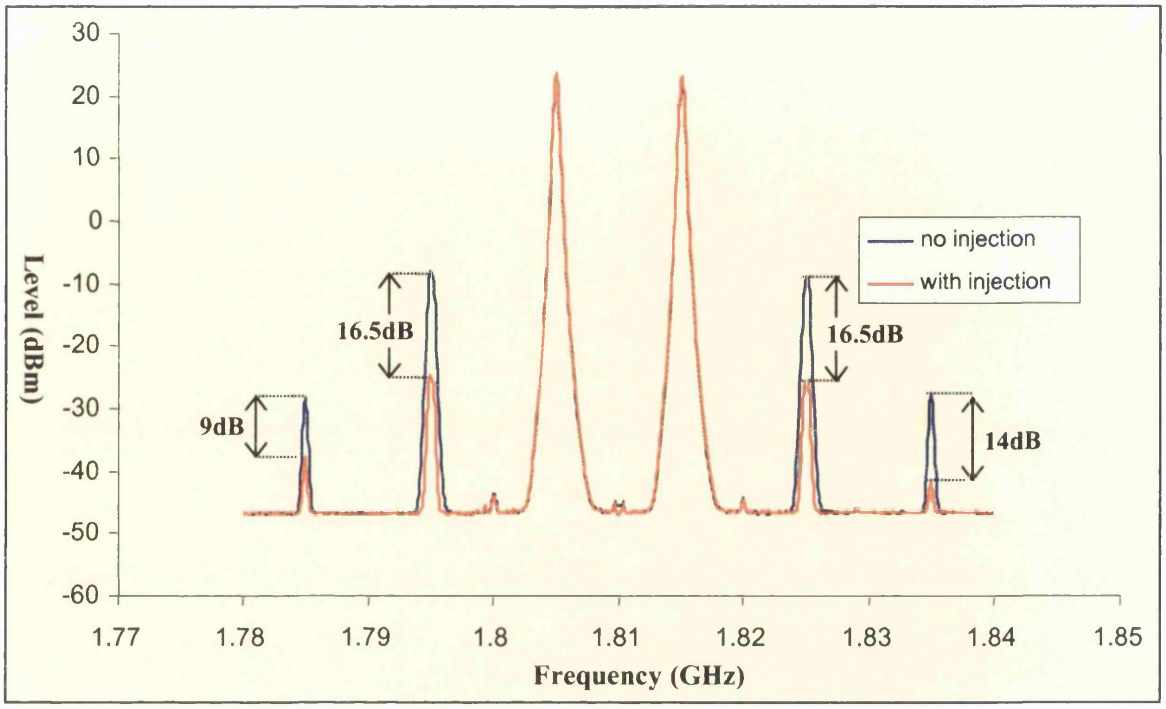


Figure 3.19: Two-Tone test results for low-frequency 'injection'

Interestingly, it can be seen that the fifth-order intermodulation components at $3f_1$-$2f_2$ and $3f_2$-$2f_1$ were also reduced appreciatively. As discussed, fifth-order intermodulation is also of concern in modern communication systems, especially in high-efficiency amplifiers, so this finding was of great encouragement. The likely cause is the generation of new fifth-order components through the interaction of the low-frequency component and the carriers via the *third-order* nonlinearity, i.e. if $A=f_1$, $B=f_2$ and $C=f_2-f_1$ then fifth-order components are generated by the third-order interactions $A-2C=f_1-2(f_2-f_1)+2f_1=3f_1-2f_2$ and $B+2C=f_2+2(f_2-f_1)=3f_2-2f_1$. The reduction in fifth-order distortion is therefore dependent on the relative magnitudes and signs of the second, third and fifth-order nonlinearities, and as these parameters are strongly bias-dependent it is thought that this effect will also vary with the quiescent operating point.

By varying the phase and amplitude around the optimum point it was also possible to observe the type of phase and amplitude imbalance 'tolerance' that was simulated and is

shown plotted in Figure 3.14. Any deviation in phase or amplitude from the mid-point caused the level of one IMD3 component to be reduced at the expense of an increase in the other. Thus, the simulation findings were qualitatively confirmed - infinite reduction is indeed possible for both the upper and lower IMD3 products, though not simultaneously.

3.4.2 Phase Tolerance

The relationship between the distortion reduction and the phase of the linearising signal was then investigated to evaluate the range over which useful improvement would result. The injected signal amplitude was fixed whilst the phase shift was varied, and the reduction in the power of the third-order intermodulation products was observed. These were then plotted on the same axis as those obtained from the nonlinear *MDS* simulations, and are shown in Figure 3.20.

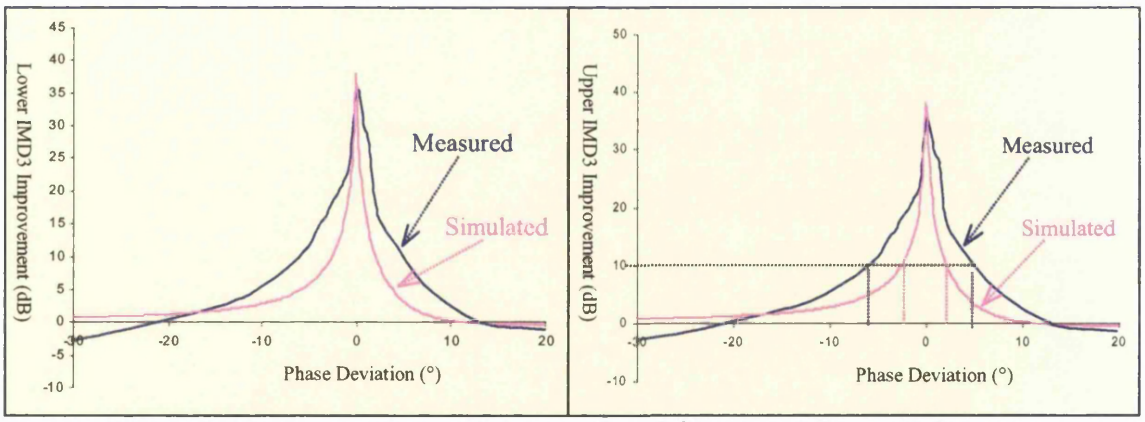


Figure 3.20: Comparison of Measured and Simulated Phase Tolerance

Referring to the two plots, it appears that the variation of distortion improvement with phase deviation predicted by the model is again pessimistic. In this example, over 10dB of reduction is achieved over a phase deviation range of ±7°, more relaxed than the ±3° suggested by the simulations. Despite this, the overall shapes of the characteristics are similar, and both show regions where the IMD3 levels are worse than they were before linearisation was applied.

The final point of note is the difference between the values of phase-shift that gave the maximum reduction in the two IMD3 components. In simulation, the two optimum points are separated by 3.7°, though in practise it was found to be around 5°. If the limitation in measurement accuracy (due to phase-jitter in the frequency synthesisers) is taken into account, these two figures may be considered to be in close agreement.

3.4.3 Two-Tone Power Sweep

With this arrangement, it was also possible to compare the simulations of IMD3 reduction against input power shown in Figure 3.16, which showed that a fixed feedback network should give intermodulation reduction over a wide dynamic range. An oscilloscope was used to measure the waveforms at the two points where feedback would be applied, and it was possible to track the phase and attenuation values that would give maximum distortion improvement as the output power was varied. Figure 3.21 and Figure 3.22 overleaf show that, as expected, the phase and attenuation requirement is fixed (within measurement limits) across a wide dynamic range, only deviating as the amplifier begins to saturate and other effects come into play. The change of optimum phase can be best explained by the fact that IMD3 components contain contributions from both AM-AM and AM-PM effects. At low levels, AM-AM distortion dominates, with AM-PM growing rapidly as compression is approached. Thus, as the drive level increases, it modifies the phase of both distortion components and requires a corresponding linearisation-signal phase-adjustment.

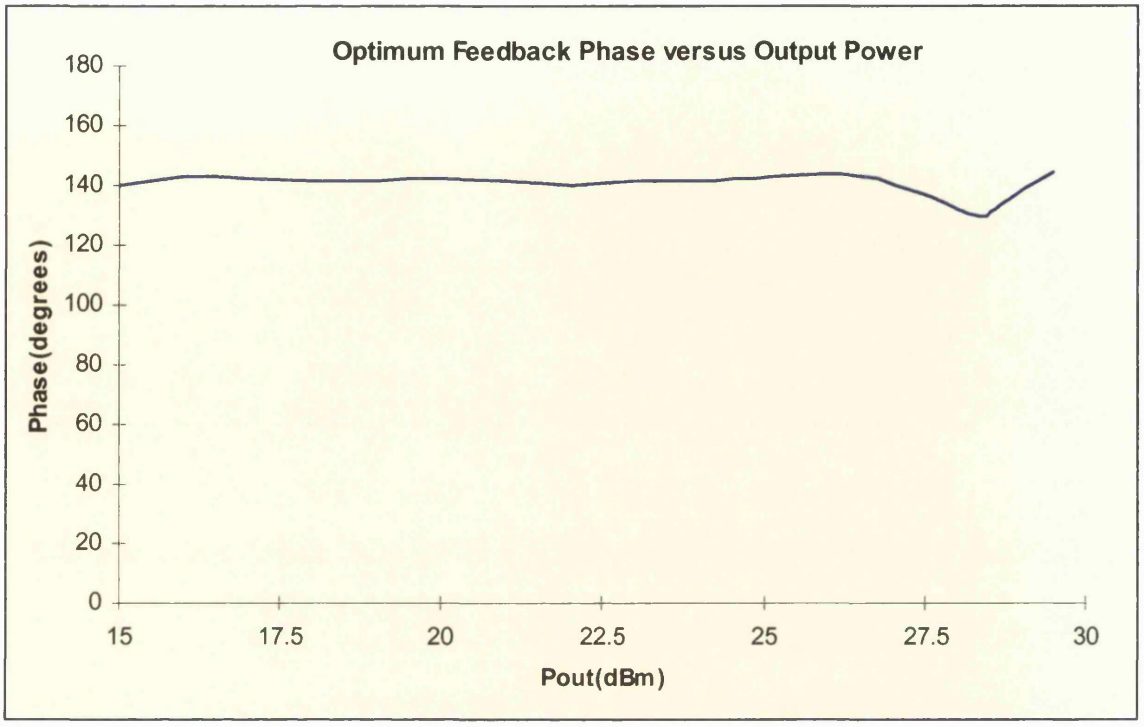


Figure 3.21: Optimum Feedback Phase versus Output Power

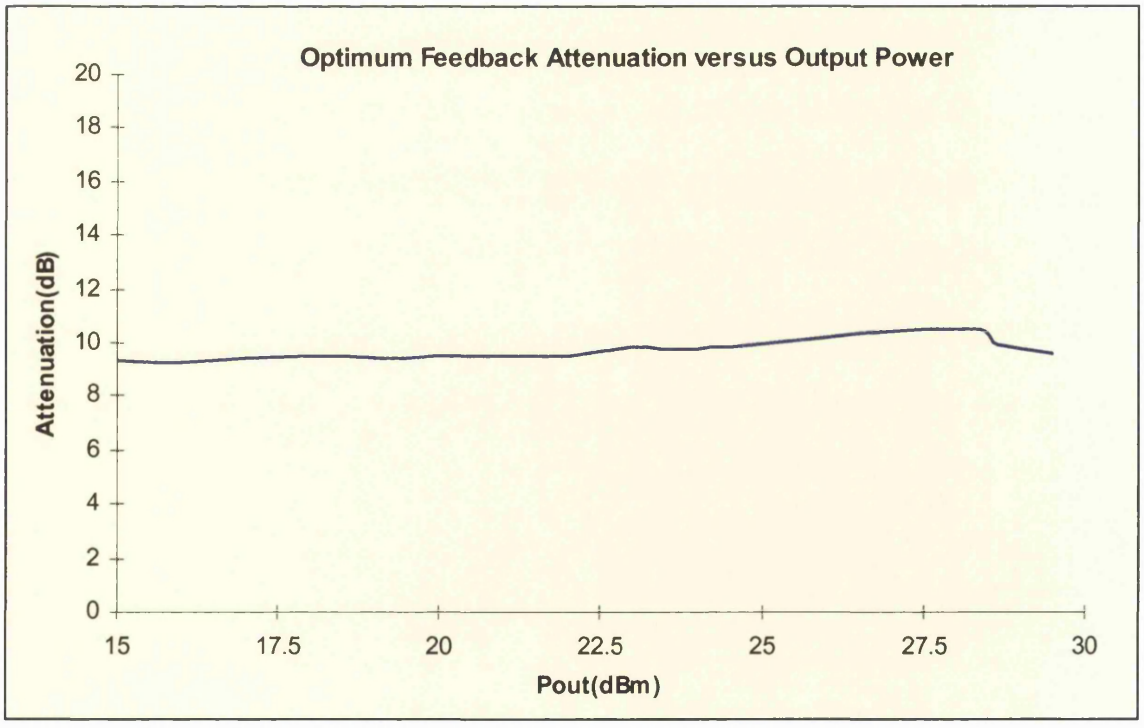


Figure 3.22: Optimum Feedback Attenuation versus Output Power

3.5 Summary

This chapter began by discussing some of the modelling issues that were encountered during the development of this technique. It was found that commercially-available device models, both physical and behavioural, have shortcomings that cause problems when low-frequency linearisation is simulated. With the experimentally-derived model, it was found that distortion reductions in the region of 15dB could be expected, and this was verified in practice. The shortcomings of a feedback-implementation were then highlighted, and the reasons for employing the modified difference-frequency-injection technique were outlined. Measurements of improvement, phase and amplitude tolerance and also power-dependence of the technique show excellent agreement with the simulation results, and verify the analysis presented.

4. Multi-Tone Input Signals

The preliminary work detailed in the previous chapters demonstrated conclusively that a narrowband feedback network could achieve distortion reduction with two-tone input signals, using a single feedback (or injection) component. The next phase of the work focussed on extending the theory and practical work to apply the linearisation technique to multi-tone input signals.

4.1 Multi-Tone Linearisation

As discussed in Chapter 2, multiple input carriers generate multiple low-frequency second-order components, each corresponding to one of the carrier spacings. It be may shown that the relationship between the number of carriers, N, and the number of low-frequency second-order difference products produced, M, is given by:

$$M = \sum_{n=0}^{N-1} N - n - 1 \tag{4.1}$$

Table 2 shows the number of second-order difference products produced by various numbers of carriers.

Number of Carriers, N	Low-frequency components, M
2	1
3	3
4	6
6	15
8	28

Table 3: Number of carriers vs. number of second-order difference products

4.1.1 Multi-Tone Analysis and Discussion

Figure 4.1 shows an example of a four-tone input signal with the second-order low-frequency distortion components that are generated shown explicitly.

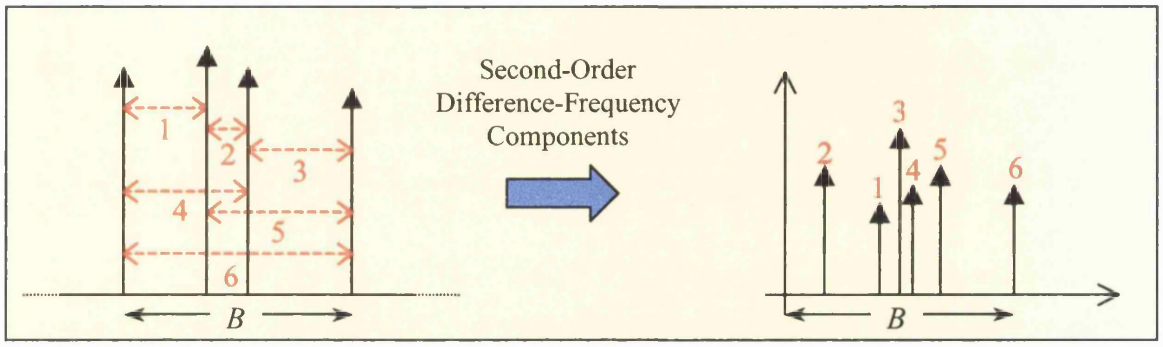


Figure 4.1: Four-tone input signal producing 6 second-order difference frequencies

The simple 2-tone analysis of low-frequency feedback presented earlier is the simplest case that can be considered; as the number of carriers increases, the analysis complexity increases rapidly and becomes time-consuming, yet some further investigation is still required. We will begin by considering a three-tone input signal:

$$V_{in}(t) = A\cos(\omega_1 t) + B\cos(\omega_2 t) + C\cos(\omega_3 t), \quad \omega_1 < \omega_2 < \omega_3 \tag{4.2}$$

And an amplifier transfer function that may be described by:

$$V_{out} = G_1 V_{in} + G_2 V_{in}^2 + G_3 V_{in}^3 + G_4 V_{in}^4 + \ldots. \tag{4.3}$$

The in-band third-order intermodulation components that are produced are shown tabulated in Table 4.

i) $\frac{3}{4}G_3AB^2\cos[(2\omega_2-\omega_1)t]$	iv) $\frac{3}{4}G_3AC^2\cos[(2\omega_3-\omega_1)t]$	vii) $\frac{3}{4}G_3A^2B\cos[(2\omega_1-\omega_2)t]$
ii) $\frac{3}{4}G_3A^2C\cos[(2\omega_1-\omega_3)t]$	v) $\frac{3}{4}G_3BC^2\cos[(2\omega_3-\omega_2)t]$	viii) $\frac{3}{4}G_3B^2C\cos[(2\omega_2-\omega_3)t]$
iii) $\frac{3}{2}G_3ABC\cos[(\omega_1+\omega_2-\omega_3)t]$	vi) $\frac{3}{2}G_3ABC\cos[(\omega_2+\omega_3-\omega_1)t]$	ix) $\frac{3}{2}G_3ABC\cos[(\omega_1+\omega_3-\omega_2)t]$

Table 4: IMD3 products produced by a three-tone input signal

In order to prove that cancellation of all nine IMD3 products can be achieved with low-frequency feedback, the action of the second-order term, $G_2V_{in}^2$, needs to be investigated. As well as the second-order harmonics and sum components, three second-order difference products are now produced, at ω_2-ω_1, ω_3-ω_2 and ω_3-ω_1. If it is assumed that these are fed back to the input in the same fashion as before, with the three components undergoing amplitude change H and phase-shift ϕ before being added to the original carriers, the input signal becomes:

$$V_{in}'(t) = A\cos(\omega_1 t) + B\cos(\omega_2 t) + C\cos(\omega_3 t) + D\cos(\delta_1 t + \phi_1) + E\cos(\delta_2 t + \phi_2) + F\cos(\delta_3 t + \phi_3) \quad (4.4)$$

Where δ_1=ω_2-ω_1, δ_2=ω_3-ω_2, δ_3=ω_3-ω_1, D=G_2ABH, E=G_2BCH, and F=G_2ACH.

The second-order term of the transfer function now produces a great many more components:

$$\begin{aligned}
G_2V_{in}^2 = G_2\Big\{&\frac{1}{2}\left(A^2 + B^2 + C^2 + D^2 + E^2 + F^2\right) \quad \text{- DC terms} \\
&+ \frac{A^2}{2}\cos(2\omega_1) + \frac{B^2}{2}\cos(2\omega_2) + \frac{C^2}{2}\cos(2\omega_3) + \frac{D^2}{2}\cos(2\delta_1 + 2\phi) + \frac{E^2}{2}\cos(2\delta_2 + 2\phi) + \frac{F^2}{2}\cos(2\delta_3 + 2\phi) \quad \text{2nd-order harmonics} \\
&+ AB[\cos(\omega_1 + \omega_2) + \cos(\omega_2 - \omega_1)] + AC[\cos(\omega_1 + \omega_3) + \cos(\omega_3 - \omega_1)] + BC[\cos(\omega_2 + \omega_3) + \cos(\omega_3 - \omega_2)] \quad \text{2nd-order IMD} \\
&+ AD[\cos(\omega_1 + \delta_1 + \phi) + \cos(\omega_1 - \delta_1 - \phi)] + AE[\cos(\omega_1 + \delta_2 + \phi) + \cos(\omega_1 - \delta_2 - \phi)] \\
&+ AF[\cos(\omega_1 + \delta_3 + \phi) + \cos(\omega_1 - \delta_3 - \phi)] + BD[\cos(\omega_2 + \delta_1 + \phi) + \cos(\omega_2 - \delta_1 - \phi)] \\
&+ BE[\cos(\omega_2 + \delta_2 + \phi) + \cos(\omega_2 - \delta_2 - \phi)] + BF[\cos(\omega_2 + \delta_3 + \phi) + \cos(\omega_2 - \delta_3 - \phi)] \\
&+ CD[\cos(\omega_3 + \delta_1 + \phi) + \cos(\omega_3 - \delta_1 - \phi)] + CE[\cos(\omega_3 + \delta_2 + \phi) + \cos(\omega_3 - \delta_2 - \phi)] \\
&+ CF[\cos(\omega_3 + \delta_3 + \phi) + \cos(\omega_3 - \delta_3 - \phi)] + DE[\cos(\delta_1 + \delta_2 + 2\phi) + \cos(\delta_1 - \delta_2)] \\
&+ DF[\cos(\delta_1 + \delta_3 + 2\phi) + \cos(\delta_3 - \delta_1)] + EF[\cos(\delta_2 + \delta_3 + 2\phi) + \cos(\delta_3 - \delta_2)]\Big\} \quad \text{3rd-order IMD}
\end{aligned} \quad (4.5)$$

The time variable t is omitted from this analysis for the purpose of clarity. Substituting δ_1=ω_2-ω_1, δ_2=ω_3-ω_2 and δ_3=ω_3-ω_1 into the above expression and ignoring out-of-band components gives:

$$\mathbf{G}_2\Big\{AD[\cos(\omega_2+\phi)+\underbrace{\cos(2\omega_1-\omega_2-\phi)}_{\text{(vii)}}]+AE[\underbrace{\cos(\omega_1+\omega_3-\omega_2+\phi)}_{\text{(ix)}}+\underbrace{\cos(\omega_1+\omega_2-\omega_3-\phi)}_{\text{(iii)}}]$$
$$+AF[\cos(\omega_3+\phi)+\underbrace{\cos(2\omega_1-\omega_3-\phi)}_{\text{(ii)}}]+BD[\underbrace{\cos(2\omega_2-\omega_1+\phi)}_{\text{(i)}}+\cos(\omega_1-\phi)]$$
$$+BE[\cos(\omega_3+\phi)+\underbrace{\cos(2\omega_2-\omega_3-\phi)}_{\text{(viii)}}]+BF[\underbrace{\cos(\omega_2+\omega_3-\omega_1+\phi)}_{\text{(vi)}}+\underbrace{\cos(\omega_1+\omega_2-\omega_3-\phi)}_{\text{(iii)}}]$$
$$+CD[\underbrace{\cos(\omega_2+\omega_3-\omega_1+\phi)}_{\text{(vi)}}+\underbrace{\cos(\omega_1+\omega_3-\omega_2-\phi)}_{\text{(ix)}}]+CE[\underbrace{\cos(2\omega_3-\omega_2+\phi)}_{\text{(v)}}+\cos(\omega_2-\phi)]$$
$$+CF[\underbrace{\cos(2\omega_3-\omega_1+\phi)}_{\text{(iv)}}+\cos(\omega_1-\phi)]+DE[\cos(\omega_3-\omega_1+2\phi)+\cos(2\omega_2+\omega_3-\omega_1)]$$
$$+DF[\cos(\omega_2+\omega_3-2\omega_1+2\phi)+\cos(\omega_3-\omega_2)]+EF[\cos(2\omega_3-\omega_1-\omega_2+2\phi)+\cos(\omega_2-\omega_1)]\Big\} \quad (4.6)$$

Table 5 below summarises the in-band IMD3 components in the above expansion, along with those produced by the third-order term of the amplifier transfer function with no linearisation applied.

Linearising Components	Pre-existing IMD3
i) $G_2^2HA^2BH\cos(2\omega_1-\omega_2-\phi)$	$\frac{3}{4}G_3A^2B\cos(2\omega_1-\omega_2)$
ii) $G_2^2HA^2C\cos(2\omega_1-\omega_3-\phi)$	$\frac{3}{4}G_3A^2C\cos(2\omega_1-\omega_3)$
iii) $G_2^2HAB^2\cos(2\omega_2-\omega_1+\phi)$	$\frac{3}{4}G_3AB^2\cos(2\omega_2-\omega_1)$
iv) $G_2^2HBC^2\cos(2\omega_3-\omega_2+\phi)$	$\frac{3}{4}G_3BC^2\cos(2\omega_3-\omega_2)$
v) $G_2^2HAC^2\cos(2\omega_3-\omega_1+\phi)$	$\frac{3}{4}G_3AC^2\cos(2\omega_3-\omega_1)$
vi) $G_2^2HB^2C\cos(2\omega_2-\omega_3-\phi)$	$\frac{3}{4}G_3B^2C\cos(2\omega_2-\omega_3)$
vii) $G_2^2H[ABC+ABC]\cos(\omega_1+\omega_2-\omega_3-\phi)$	$\frac{3}{2}G_3ABC\cos(\omega_1+\omega_2-\omega_3)$
viii) $G_2^2H[ABC+ABC]\cos(\omega_2+\omega_3-\omega_1+\phi)$	$\frac{3}{2}G_3ABC\cos(\omega_2+\omega_3-\omega_1)$
ix) $G_2^2H[ABC\cos(\omega_1+\omega_3-\omega_2-\phi)+ABC\cos(\omega_1+\omega_3-\omega_2+\phi)]$	$\frac{3}{2}G_3ABC\cos(\omega_1+\omega_3-\omega_2)$

Table 5: Second-order linearising components and IMD3 products produced by a three-tone input signal

By inspection of Table 5, it is clear that the two sets of signals can be made to cancel if:

$$H = \frac{3G_3}{4G_2^2} \qquad \text{and} \qquad |\phi| = 180° \tag{4.7}$$

This same result was obtained for the simple two-tone case shown earlier, and demonstrates that low-frequency feedback may be applied to multi-carrier input signals, a finding unpublished in the literature until very recently [38, 39].

4.1.2 Feedback Implementation Issues

The above analysis also highlights important issues regarding the realisation of the feedback-loop. Even this simple case, using an amplifier transfer function that is constant with frequency, shows that a flat amplitude response is required in the loop as each of the fed back components must undergo the same amplitude and phase shift. This may be put into perspective by considering a practical application; for example, a GSM multicarrier transmitter could have carrier spacings of between 200kHz and 35MHz, requiring a feedback loop with a flat amplitude and phase response *across this entire bandwidth.* This corresponds to a fractional bandwidth of 200%, an 'ultra-wideband' response. The synthesis problem is further compounded by the fact that the feedback is applied via low-pass networks, whose frequency responses need to be equalised by the feedback loop if a flat characteristic is to be obtained. Although it is conceivable that such a network could be realised with analogue components (say with a carefully-designed filter and all-pass phase-shaping network or other means), this would result in a fixed response that could not be adjusted easily. The simple analysis presented above shows that any change in the coefficients G_2 and G_3 will affect the required amplitude of the linearising signal(s), so it may be anticipated that a real device whose distortion characteristics are varying with loading, time, and temperature will require feedback that can be adjusted to cater for these fluctuations. Further to this, the distortion characteristics of amplifiers are also frequency-

dependent, so the required H and ϕ for each of the fed-back components will not, in reality, be equal. It may therefore be assumed, without further justification, that the feedback network cannot be realised as a passive analogue network.

The problems outlined above might suggest a digital solution, and an implementation such as that shown in Figure 4.2 could, theoretically, solve many of these issues.

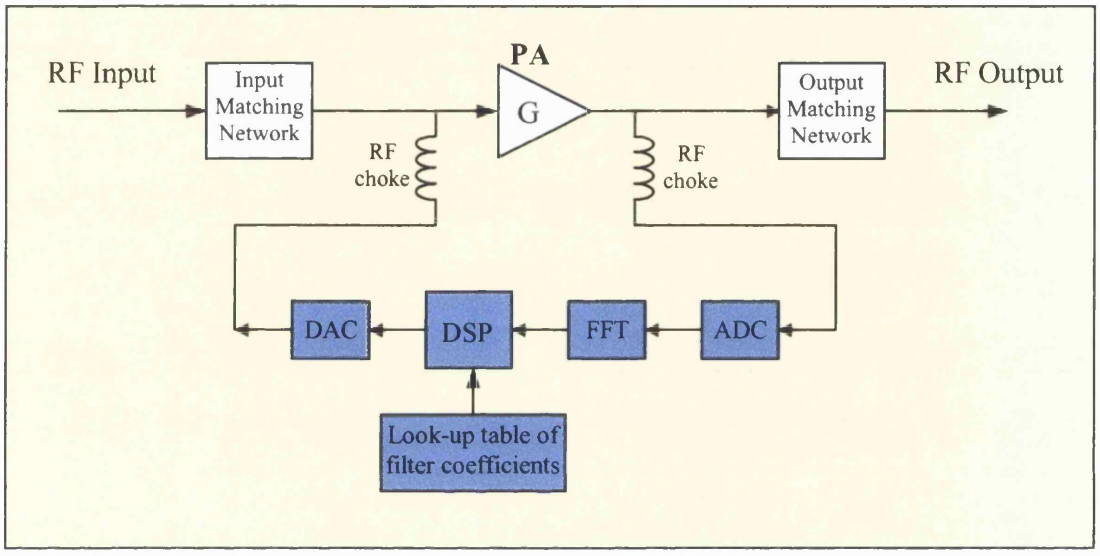


Figure 4.2: Digital Feedback-Loop

In the above implementation, the second-order difference signal is first sampled with an ADC at the end of the output RF choke, as shown. A hardware-implemented FFT is then used to transform the time-domain digital signal into the frequency-domain, after which a DSP block applies phase and frequency adjustment to the frequency-domain signal. The equalised linearising signal is then converted back into the time-domain, and passed through a DAC to re-produce the equalised analogue signal, before it is 'injected' back into the input of the amplifier.

The advantage of this topology is that the response-shaping is performed in the frequency-domain, so any effective transfer function can be implemented; this would allow any undesirable effects due to the low-pass RF chokes and amplifier non-idealities to be

removed. However, it will take a finite amount of time for the digitally-implemented feedback loop to process the signal, depending on the sampling frequency used. The significance of this is illustrated by considering the requirements of a future WCDMA system, whose carriers may be separated by up to 75MHz. If this is taken as being the upper cut-off frequency of the feedback loop, sampling at the Nyquist rate would result a bit-rate of f_s=150MHz, which is well within the limits of current technology. However, at this frequency, one sample period is also equivalent to 180° of phase. As even the most basic digital filtering operations entail tens of sample delays, the sampling rate needs to be increased to 'buy' extra processing time in the loop. For this very simple case, assuming no phase-shift occurs anywhere else, it is clear that f_s must be increased by a factor of at least the total number of sample delays for the loop phase not to exceed 180°. Thus, with present digital technology limited to sub-GHz frequencies, a loop having the complexity indicated by Figure 4.2 cannot be realised.

Fortunately, there is another approach with the potential to overcome this loop-delay limitation, and it will be described in the next section.

4.1.3 External Generation and Injection

A logical solution to the feedback implementation problems described above would be to generate the linearising signal externally to the amplifier by some other means, in a manner similar to that employed in 'Second-Order Interstage Enhancement' [29], as described earlier (see Section 2.2.10). However, in this case the linearising signal is at a much lower frequency, so it may be generated with simpler circuitry, and the distortion improvement doesn't depend upon on the nonlinearity of a second amplifier, which is likely to fluctuate over time. The nature of a feedforward topology allows the time delay penalties outlined in

the previous section to be compensated for elsewhere in the circuit, greatly reducing the causality limitations.

4.2 Practical Verification of Multi-Tone Linearisation

In order to test the theory outlined above, a multi-tone test-bench was required. This process began with the design and construction of a four-tone combining and pre-amplifying module, the details of which are given in the next section.

4.2.1 Design and Build of Four-Tone Combiner

Generating multi-tone signals from separate sources is by no means straightforward. If the outputs of two signal generators are connected to the inputs of a passive power combiner with no isolating components to separate them, the nonlinearity of the generator outputs causes cross-modulation to occur, resulting in intermodulation distortion.

A second consideration is the maximum output power of most high-frequency signal generators, which is usually in the order of 10-20dBm. If carriers from 4 sources are to be combined using passive components, at least 6dB of power, per carrier, will be wasted. Assuming that source power levels are limited to 20dBm and that no other losses in the combining process occur, each carrier will be only 14dBm (25mW), with a combined peak envelope power that is up to 6dB greater than this. For even moderately-sized power-amplifiers, such as the FLL351-ME employed in this study (GCP_{1dB}=35.5dBm), input powers in excess of 25dBm are required in order to push the amplifier into compression. Therefore, in order to generate multi-tone signals of any significant power level, some pre-amplification is usually required. This usually takes the form of a discrete pre-amplifying stage, consisting of a single heavily backed-off Class A amplifier, or preferably, a separate discrete amplifier for each carrier. The former of these two options is simpler to realise, but

requires a large, heavily backed-off pre-amplifier with high second- and third-order intercept points to minimise the residual distortion in the test signal. In order to perform accurate measurements, the signal degradation caused by such a pre-amplifier must be corrected for – a relatively simple procedure, but not ideal. For the purposes of this investigation where the accurate measurement of distortion (and its reduction) is of primary concern, this is an unacceptable solution.

It was therefore decided that a purpose-built pre-amplifying and combining module was required, as shown schematically in Figure 4.3 below.

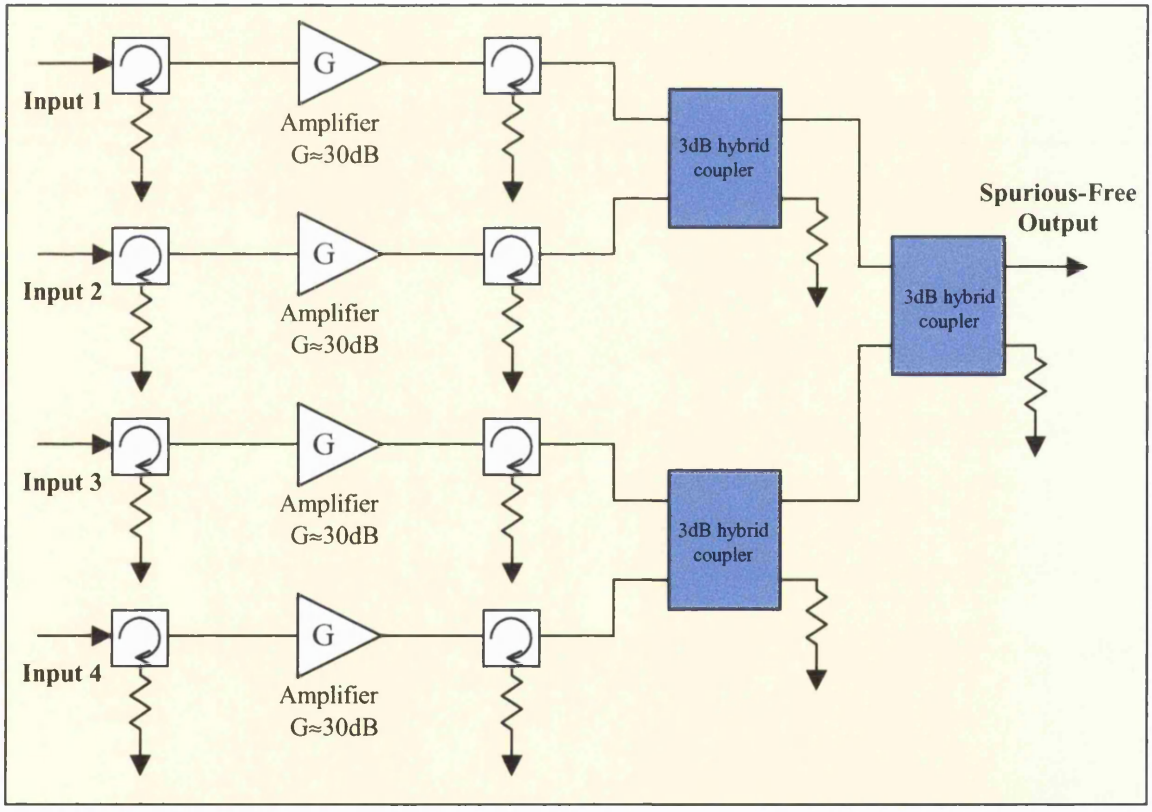


Figure 4.3: Schematic of 4-tone Power Combining/Amplifying Module

Four 20W amplifier modules (Philips BGY1816) with internal matching and de-coupling networks were supplied by the industrial sponsor, along with high-power hybrid circulators, low-power surface-mount circulators and three surface-mount hybrid couplers. The microstrip layout and locations of the various components were then finalised, taking into

consideration the bias-feed connections and ensuring cross-coupling between components was minimised. The CAD package '*Wavemaker*' was used to translate this layout into a *Gerber* file that was then used to control a milling-machine that removed the required areas from the upper surface of the microstrip substrate, chosen as Rogers Duriod 5870 (ε_r=2.33). A design for an aluminium jig to support and provide heat-sinking for the amplifiers and other components was then fabricated in-house and the various components were assembled. A top-down plan view of the layout is shown overleaf in Figure 4.3.

The efficiency of the pre-packaged 20W amplifiers was then investigated to ascertain whether additional heat-sinking would prove necessary, with the calculation shown below.

Amplifier Efficiency (from datasheet) = ≥30%

Maximum RF Output Power = 20W

DC Power Required @ η = 30% = 60W

DC Power dissipated as heat ~ 40W (minimum, assuming no other losses)

$\therefore$ Heat dissipation requirement for 4 devices = **>160W**

From this brief calculation, it was apparent that the aluminium jig used to mount the components would not form a sufficiently large heat-sink, and that additional measures were required. Suitable heat-sinks were selected, taking into account the above figure and the dimensions of the jig that had been designed, which had a specific heat capacity of 0.5°C/W. Two such heatsinks were used together, and combined with fans to force air across the fins and thus increase the rate of heat dissipation. Two suitable fans were provided by the industrial sponsor, and the heatsinks were modified to accommodate them.

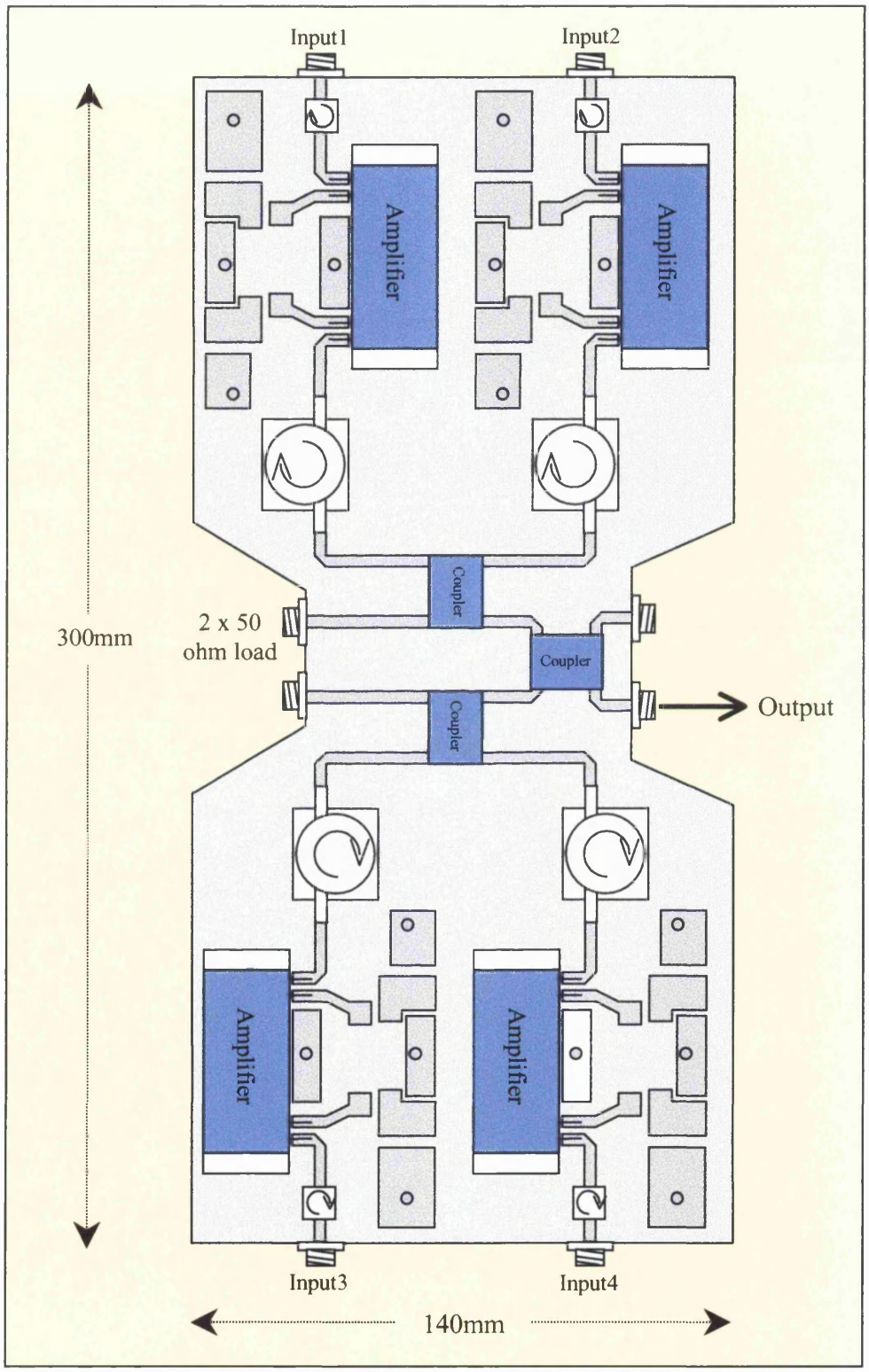


Figure 4.4: Layout of 4-tone combining module

Heatsink-compound was used to reduce the thermal resistance between the underside of each amplifier and the aluminium jig, and also between the underside of the jig and the two heatsinks. Additional bias de-coupling components were then affixed and the unit was tested.

The performance was found to be as expected, with each branch of the module providing in the region of 15dB of gain with a maximum available output-power per carrier of 33dBm (2W). The residual intermodulation appearing at the output was found to be undetectable for the power output levels required by the FLL351-ME test amplifier. A photograph of the completed unit may be seen below in Figure 4.5.

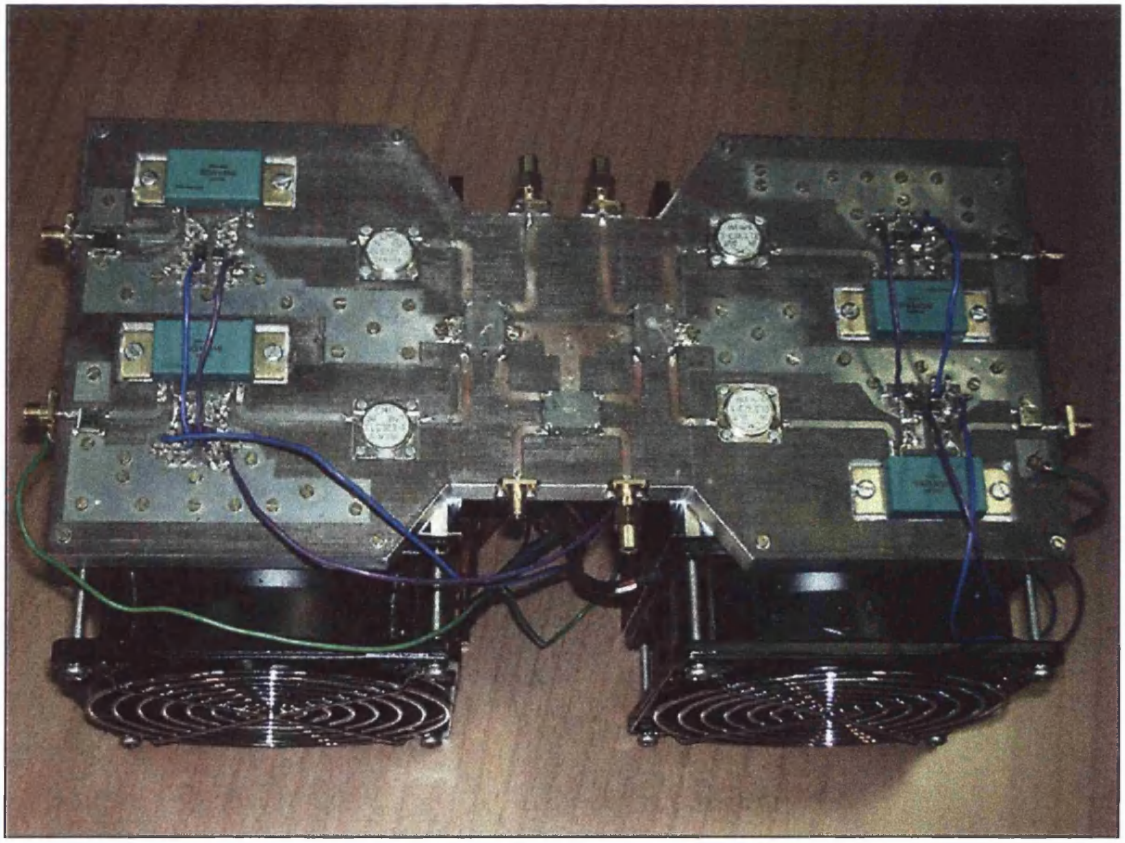


Figure 4.5: Photograph of 4-way amplifying/combining module

4.2.2 4-tone Test-Bench

The external generation of a low-frequency linearising signal was demonstrated in Chapter 3, where a third signal-generator was used to produce a single difference-frequency component that was added to the two RF carriers. Although this overcame the bandwidth and tuning limitations of narrowband feedback, it is very hardware-intensive when applied to an amplifier with more than two carriers. For example, four carriers would produce six low-frequency second-order components, requiring ten separate signal-generators in total. Further to this, the outputs of the six sources used to produce the linearising signal would also need to be combined in a similar fashion to the RF carriers, which as described above is not straightforward. Although this allows a wide degree of flexibility, it is extremely hardware-intensive, impractical and inconvenient to set up.

It was also shown earlier that a digital feedback implementation such as that shown in Figure 4.2 could not be applied to a practical system, due to excessive time-delays around the feedback loop. However, if the carriers are all sinusoidal, the low-frequency second-order difference signal will be periodic, allowing any amount of time delay in the loop to be tolerated, as long as the total phase is equal to $180°+n360°$. Therefore, under these conditions, the feedback loop *can* be implemented digitally, and it was decided to exploit this in the next phase of practical verification.

Figure 4.6 shows the schematic of the test-bench that was assembled to implement the digital feedback loop, comprised of the following components:

(i) A Sampling oscilloscope, used to capture the periodic second-order waveform at the end of the output bias-feed line, in effect an ADC.

(ii) A PC, controlling all the equipment over GP-IB as shown. This acts as a versatile (if slow) DSP block, performing equalisation and wave-shaping as will be described shortly.

(iii) An HP33120A Arbitrary Waveform Generator (AWG), used to generate the resulting linearising signal produced by the PC.

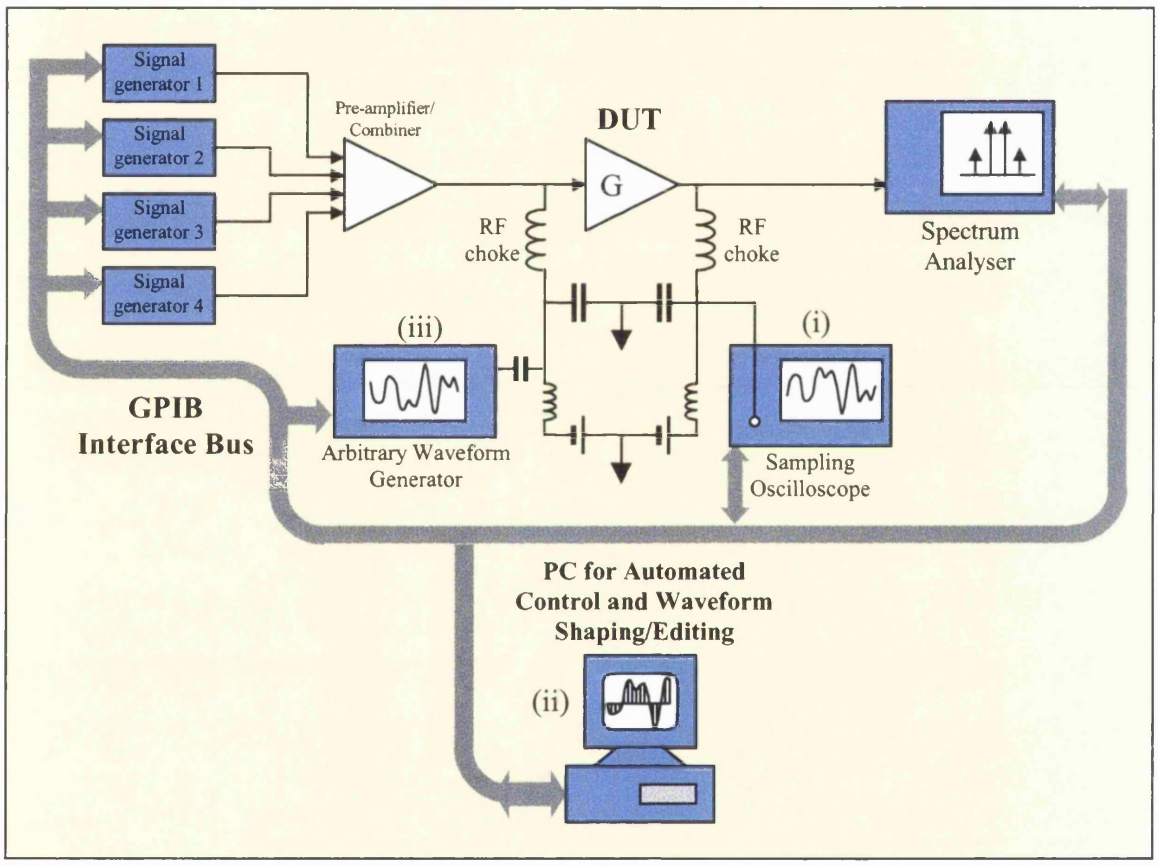


Figure 4.6: Schematic of four-tone test-bench

The four signal generators and the AWG were all locked to a common reference signal to maintain frequency stability between the sources over time. The main advantages of this arrangement over those that have been employed up until this point are as follows:

1. **Versatility.** As there is no need to design, build, test and optimise prototype feedback circuitry, and a wide bandwidth is available from the arbitrary waveform generator, this arrangement provides a large degree of flexibility. Any combination of difference-

frequency amplitudes and phases may be synthesised, up to the maximum output frequency of the AWG, 15MHz. Alterations can be made at the press of a key.

2. **Similarity to Intended Implementation.** As already discussed, a digital implementation will be required in order to make this linearisation technique viable, and also adaptable. The PC can be used to implement many DSP functions including filtering, convolution and Fast Fourier Transforms, thus enabling the processing power and speed required in a practical application to be estimated.
3. **Automation.** The GP-IB interface allows a large amount of data to be collected rapidly, and can also be used to automate the measurement process. This will enable various aspects of the linearisation technique to be accurately measured, such as its tolerance to feedback phase balance (Section 3.3.1) and the behaviour under different bias conditions.

A photograph of the assembled test-bench is shown below in Figure 4.7:

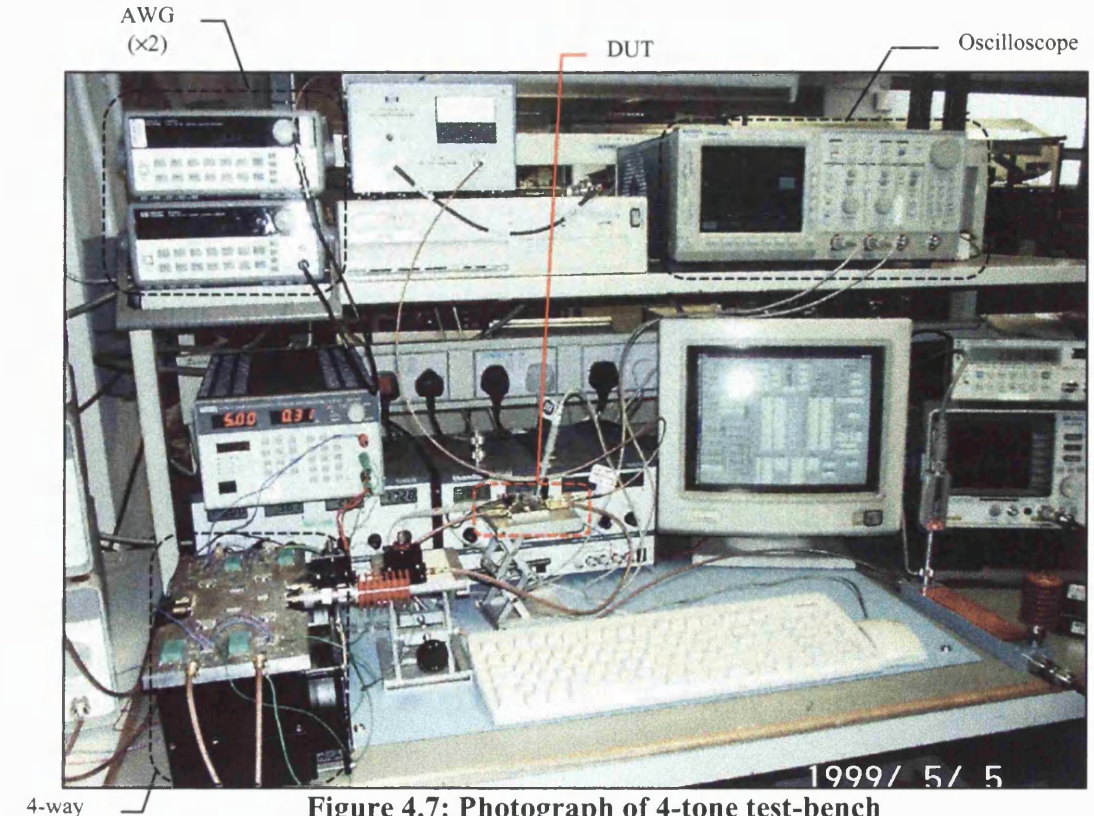


Figure 4.7: Photograph of 4-tone test-bench

4.3 Experimental Work

The testing began with the development of simple VisualBasic programs to send commands to, and read data from, the separate pieces of equipment. Once the necessary protocols were established, multi-tone tests could be performed. The aim was to develop a 'virtual' digital feedback loop that would generate a linearising signal from the low-frequency second-order difference signal appearing at the end of the output bias-feed line. This would be achieved with the following steps:

1. The composite second-order difference signal was captured at the end of the output bias-feed line with a digital sampling-oscilloscope.
2. One period of the captured waveform was downloaded to the PC across GB-IB as a sequence of samples with 16-bit resolution.
3. VisualBasic routines were used to apply phase and amplitude equalisation to the captured waveform.
4. The processed signal was downloaded across the interface bus to the AWG, which was set to continuously output the periodic waveform.
5. The time reference and amplitude of the AWG were then adjusted until the distortion reduction was optimised.

Steps 1, 2, 4 and 5 in the above sequence remained almost unchanged as the software was developed. However, the equalisation described by Step 3 was implemented in several different ways, and this evolution is described in the following sections.

4.3.1 Simple Inversion

Referring to Figure 4.6 it may be seen that the linearising signal is both captured (by sampling oscilloscope) and injected (by arbitrary waveform generator) via inductive bias-feeding lines. The frequency response of these and the other de-coupling components is therefore imposed upon the transfer function of the 'virtual' feedback loop, and needs to be taken into consideration. In order to assess the impact of this ancillary low-pass circuitry and the degree of correction that might be required, the first type of signal conditioning applied in the PC was simple inversion. This is the simplest means of achieving 180° of loop phase-shift, as required by the analysis presented earlier.

Three carriers at 1.858GHz, 1.86GHz and 1.868GHz were pre-amplified, combined, and applied to the DUT, with the equipment arranged as shown in Figure 4.7. The frequencies were initially chosen arbitrarily, with the condition that the second-order difference signal produced should be within the useful frequency range of the AWG (~12MHz). Accordingly, the composite second-order difference-frequency signal contained IF components at 2MHz, 8MHz and 10MHz, and this waveform was captured on a digital sampling oscilloscope (DSO) at the end of the output bias-feed line. One complete period was transferred via GP-IB to the PC, where it was inverted and subsequently downloaded to the output buffer of the AWG. This unit was set to output the waveform at a rate that would exactly reproduce the inverted version of the original signal, which was injected to the input of the DUT as shown, via a capacitive feed-through.

Time-alignment of the AWG was achieved by rapidly hopping the output frequency away from and back to its original value, essentially 'nudging' the waveform backwards or forwards in time, after which the amplitude of the injected signal was adjusted by hand until

the distortion around the three carriers was minimised. Figure 4.8 shows a typical result produced by the inverted second-order signal, where it may be seen that many of the distortion components are reduced, some considerably so.

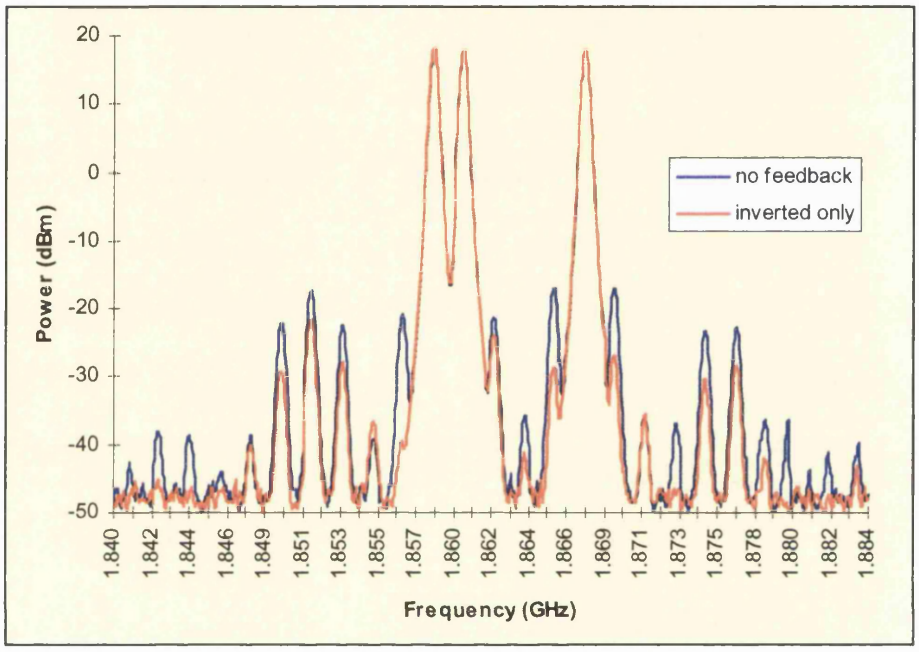


Figure 4.8: Early 3-tone test using inverted injection signal

However, the above figure also shows several distortion components that could not be reduced concurrently. As already discussed, the low-pass bias circuitry was expected to affect the phase and amplitude of the linearising signal, so this result was anticipated.

4.3.2 Equalisation with Convolution

After the success of these initial tests, the next phase of work concentrated on engineering the response of the virtual feedback loop by altering the signal conditioning applied by the PC. This process began with direct characterisation of the input and output bias-feeding networks. In order to obtain these measurements, the DUT was modified and connected to a standard Network Analyser as shown in Figure 4.9.

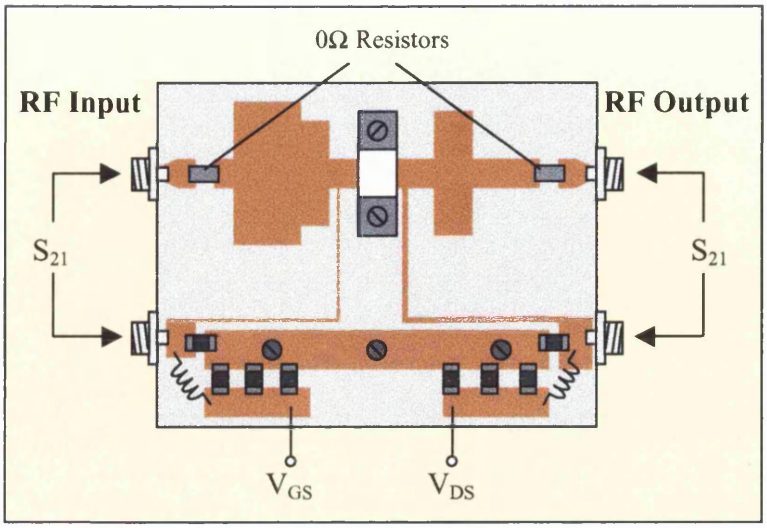


Figure 4.9: Determining the frequency responses of the bias networks

The above figure shows that input and output impedance matching circuits complicate the characterisation of the bias networks as they cannot be bypassed easily. However, by replacing the DC-blocking capacitors at the input and output with zero-ohm resistors, the effect of the matching networks was minimised and assumed to be insignificant over the IF frequency range of interest.

The transfer characteristics of the input and output bias networks over the range 1-50MHz are shown in Figure 4.10 and Figure 4.11.

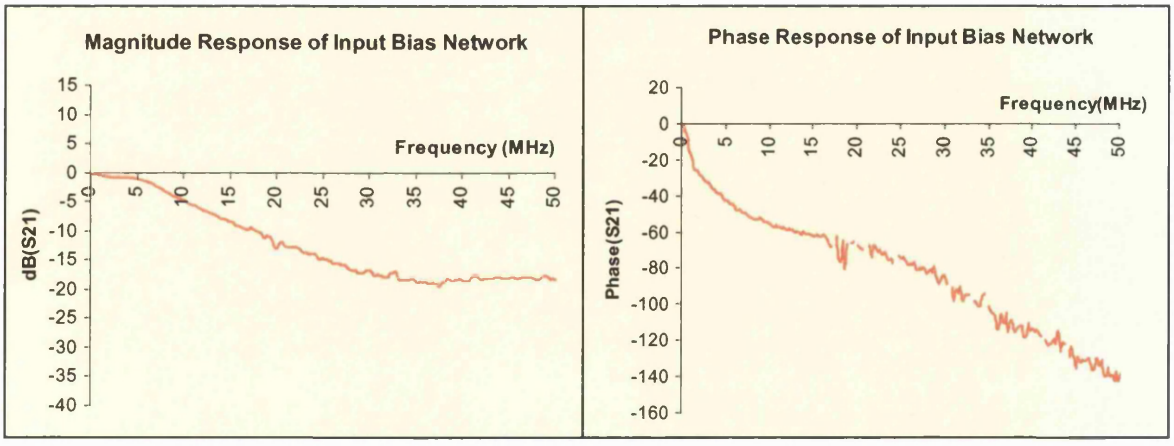


Figure 4.10: Measured frequency response, input bias-network

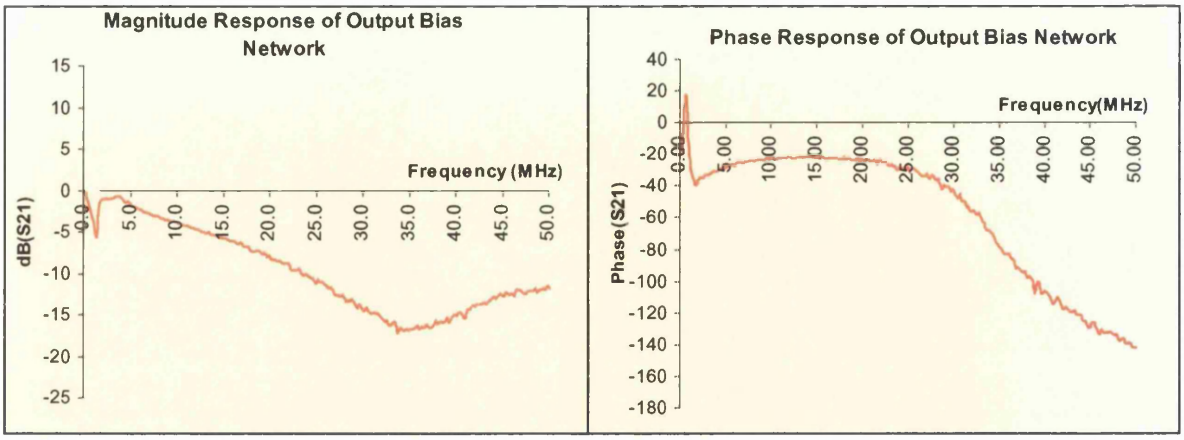


Figure 4.11: Measured frequency response, output bias-network

Finally, the frequency response of the capacitive feed-through was measured, as shown below.

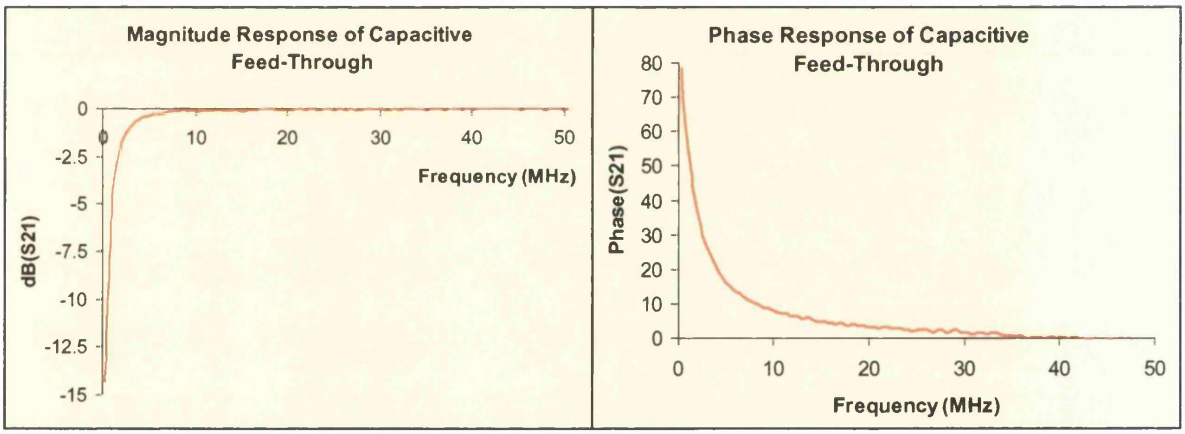


Figure 4.12: Measured frequency response, capacitive feed-through

The S-Parameters of the three separate networks were combined in *ADS* to simulate the overall transfer characteristic. The frequency response required to equalise the bias circuitry and produce a flat overall characteristic was then derived, and transferred to a different CAD package (*Systemview*) to produce digital filter approximations that could be implemented with the PC.

Several sets of FIR impulse-responses using differing numbers of numerator coefficients were obtained because, as expected, no digital filter could be found to match the required transfer function exactly.

A VisualBasic routine was written to convolve the captured waveforms with the digital impulse-responses that had been derived. The result of each convolution was normalised and downloaded to the arbitrary waveform generator over the GPIB interface, and then injected to the input of the amplifier via the input bias-feed line. Care had to be taken to ensure the effective sampling frequency of the oscilloscope capture matched that used in the digital filter synthesis tool. For convenience, sets of filter weights for 125MHz, 250MHz and 500MHz sampling rates were generated and stored in separate files.

The functionality of the VisualBasic code is illustrated below in Figure 4.13.

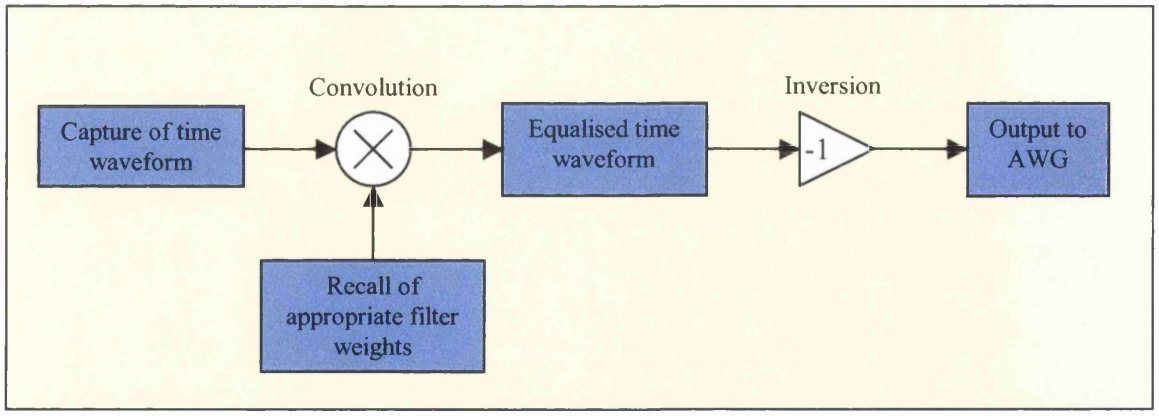


Figure 4.13: Functionality of 'Convolution-Equalisation'

Once the captured waveform has been equalised, inverted and downloaded to the arbitrary waveform generator, its bulk phase (or time reference) and amplitude were adjusted manually until the best performance was achieved. Many tests were carried out, with the traces shown in Figure 4.14 overleaf depicting a typical result with a three-tone input signal.

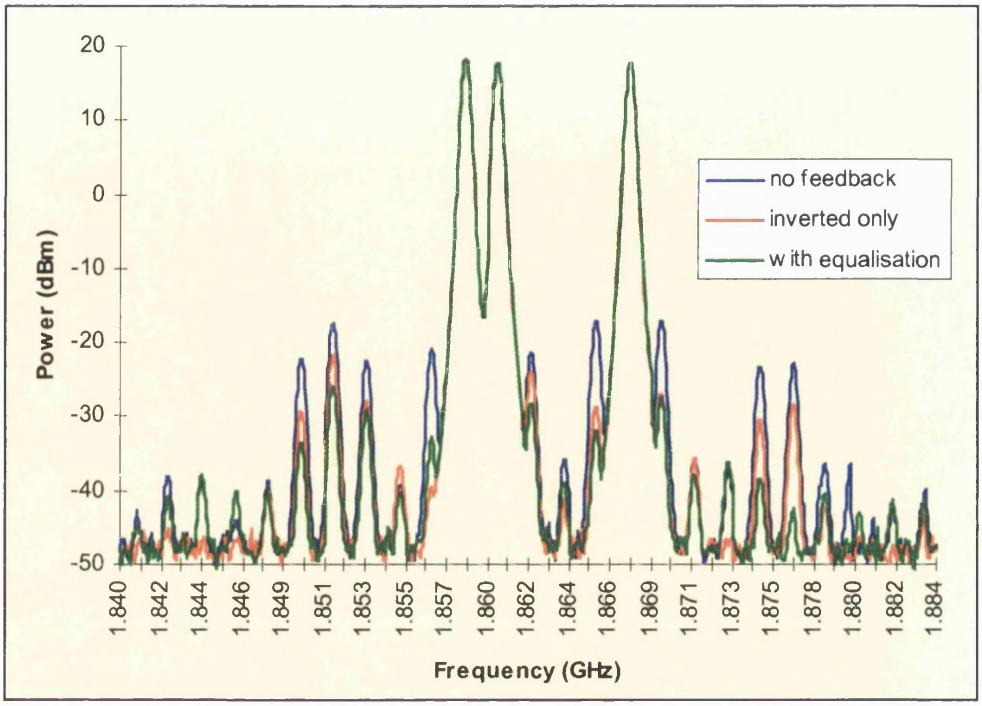


Figure 4.14: Performance of Equalisation by Convolution

The blue trace in the above figure represents the performance of the amplifier without any linearisation; the red trace shows the improvement achieved with only inversion of the captured waveform; the green trace represents the performance with inversion *and* equalisation applied to the captured waveform. It can be seen from these results that equalisation improves the performance, sometimes improving the linearity by 10dB over the simple inversion-only scenario. The individual improvements observed in the various IMD3 products were between 5dB and 20dB. Despite the overall improvement, however, several low-level IMD3 components actually increase when the equalisation is applied. Although there are several uncertainties, such as the manual AWG phase and amplitude adjustment used to optimise the performance, it was thought that the approximations used in deriving the digital impulse-responses were responsible for this poor performance. Other shortcomings of this system include:

- **Carrier spacing limitation.** The performance was seen to vary with different carrier spacings. As already mentioned, this was probably due to inaccuracies in the digital filter

synthesis process, and causing variations in performance due to the fact that the desired and actual responses were closer at some frequencies than at others. Attempts were made to alter the bias-circuitry in an effort to produce a more realisable response but this proved to be ineffective.

- **Processing Time.** The computational overhead for the convolution process used in this method proved to be very high, taking up to 30 seconds each time it was executed.
- **Lack of easy equalisation-adjustment.** Most RF circuits, especially linearisers, require tuning when the transition from theory to practise is made. The only adjustments that could be effected with this method were the selection of different impulse-responses for use in the convolution.

For the above reasons, it was decided that a different approach was required.

4.3.3 Equalisation by Discrete Fourier Transform

During the previous stage of the investigation, it was discovered that the AWG output response rolled off gradually above a few MHz (due to the anti-imaging filter in the output of the unit) and that this response was being imposed upon the 'feedback loop' of the system, along with those of the bias-circuitry. For this reason, it was apparent that some adjustment would probably be required in order to obtain the optimum performance from the technique, and it was decided that this could be achieved most easily in the frequency domain.

In the same way as before, the composite low-frequency difference signal was captured with the sampling oscilloscope. VisualBasic code was written to perform a Discrete Fourier Transform (DFT) on this signal at 'spot' frequencies (chosen automatically according to the carrier spacings that were selected), giving the relative amplitude and phase value of each

component present. The frequency-response data that had been collected previously was then loaded into look-up tables that were referenced to adjust the amplitude and phase of each component, thus equalising the effect of the bias circuitry. Unlike the time-domain convolution already described, this allowed the *precise* measured responses to be used, and the phase and amplitude adjustment applied to each component could be adjusted individually as and when required.

Figure 4.15 illustrates the functionality of the VisualBasic routine that was written. The process was carried out for each difference-frequency component in the captured signal, before all were re-combined to form a time-domain waveform that was then transferred to the Arbitrary Waveform Generator.

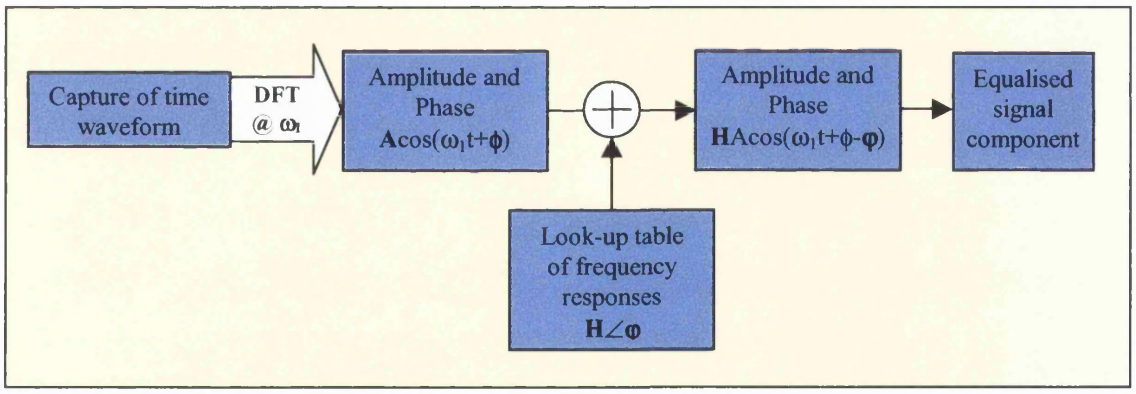


Figure 4.15: Functionality of 'DFT-Equalisation'

After the necessary code had been written and tested, it was applied to three-tone input signals similar to those shown earlier. The results were so impressive that the software was immediately extended to work with four tones, the objective of this phase of the work.

A look-up table of amplitude- and phase-adjustment coefficients was developed to fine-tune the performance of the system, allowing for correction of unknowns such as the roll-off of the AWG and also the amplifier nonlinearity. This was developed through a long process of

trial-and-error, and as such is not necessarily optimal. The final amplitude and phase coefficients resulting from this process are shown plotted below.

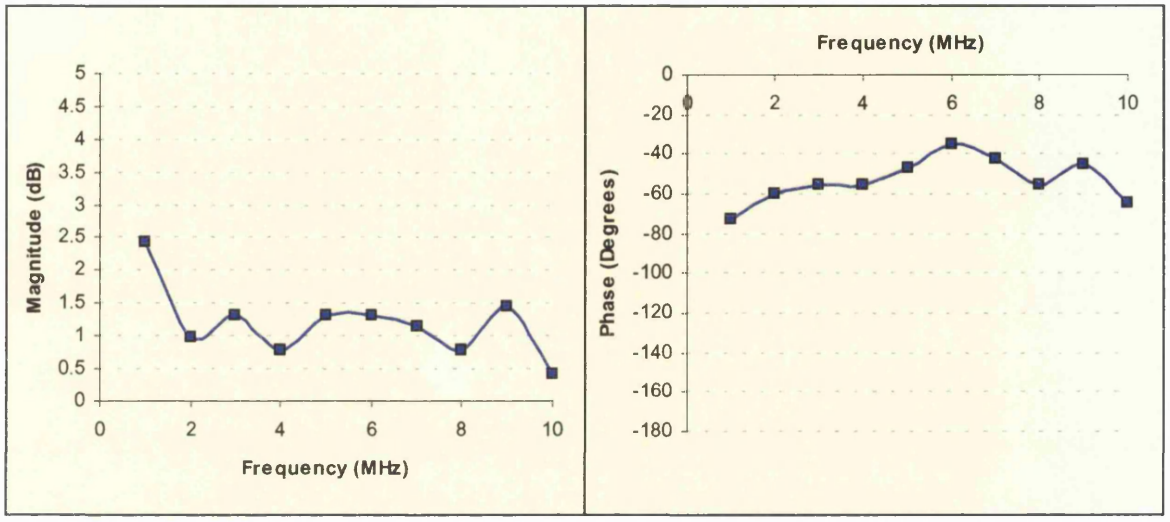


Figure 4.16: Amplitude and phase coefficients

Figure 4.16 shows the result of a typical 4-tone test, with equal-power carriers, and Figure 4.17 shows the same test with a narrower frequency sweep to enable resolution of the close-in products. The carrier spacings were selected as 1MHz, 4MHz and 3MHz, as this produces six distinct second-order difference frequency components (1MHz, 3MHz, 4MHz, 5MHz, 7MHz and 8MHz) and twenty of the twenty-four possible IMD3 components.

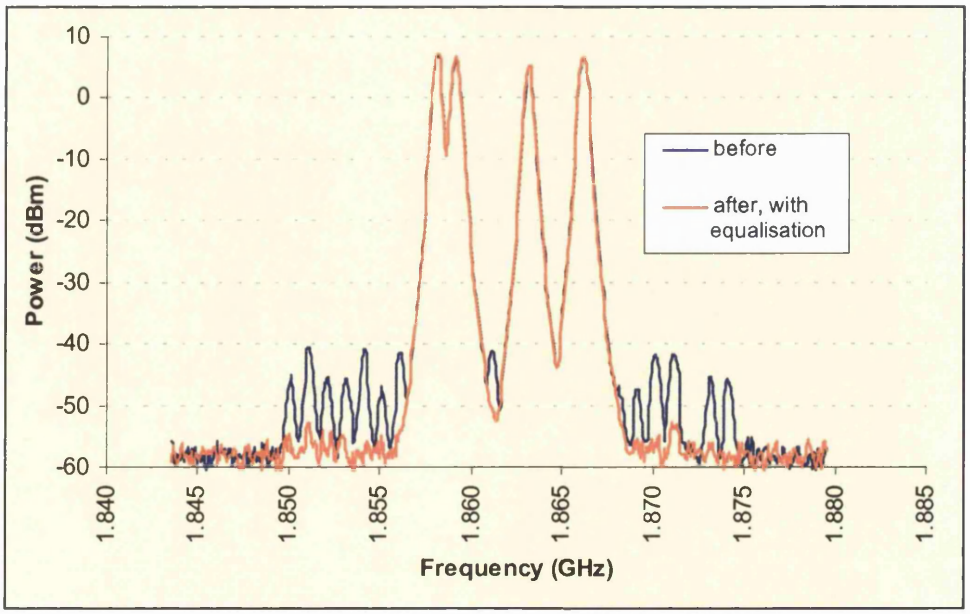


Figure 4.17: 4-tone test results using 'DFT-equalisation' – wide frequency span

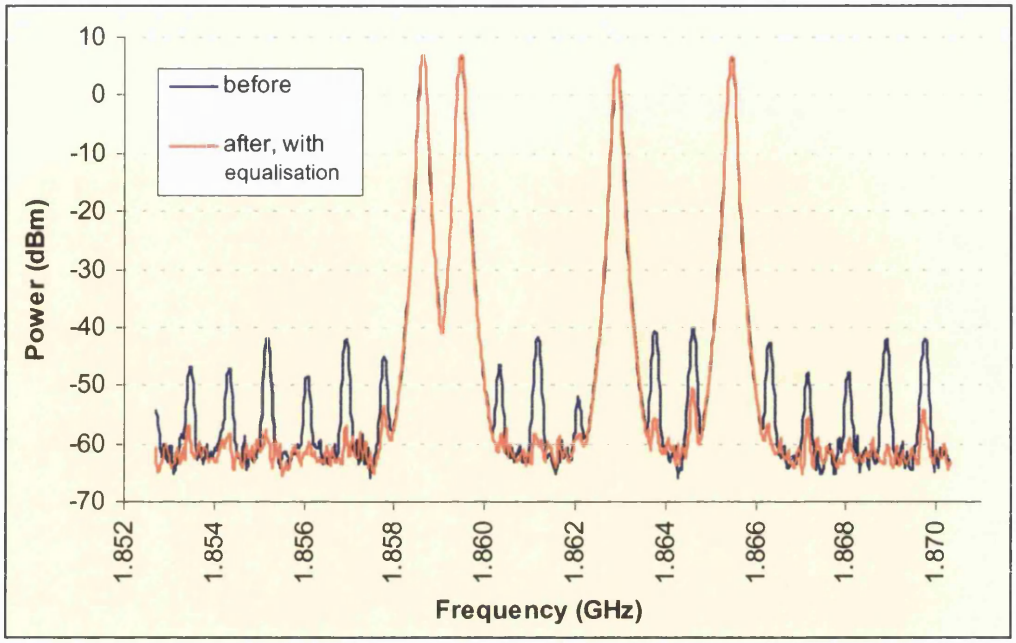


Figure 4.18: 4-tone test results using 'DFT-equalisation' – narrow frequency span

The details of the improvements observed in each of the 24 in-band IMD3 products are given in Table 6. Where no value is given the IMD3 product concerned appears at the same frequency as the one shown in parenthesis.

IMD3 Product	Improvement (dB)	IMD3 Product	Improvement (dB)
$2f_2-f_1$	18.2	$f_1+f_2-f_3$	23.4
$2f_3-f_1$	11.4	$f_2-f_1+f_3$	15
$2f_1-f_2$	10.9	$f_1-f_2+f_3$	9.5
$2f_1-f_3$	14.8	$f_1+f_2-f_4$	15.5
$2f_3-f_2$	16.1	$f_2-f_1+f_4$	$(2f_3-f_2)$
$2f_2-f_3$	17.8	$f_1-f_2+f_4$	12.2
$2f_1-f_4$	14.2	$f_2+f_3-f_4$	20.3
$2f_2-f_4$	18.3	$f_3-f_2+f_4$	20.5
$2f_3-f_4$	$(2f_2-f_1)$	$f_2-f_3+f_4$	$(f_1-f_2+f_3)$
$2f_4-f_1$	18.2	$f_1+f_3-f_4$	$(2f_2-f_3)$
$2f_4-f_2$	18.2	$f_3-f_1+f_4$	13.3
$2f_4-f_3$	17.7	$f_1-f_3+f_4$	21.6

Table 6: 4-tone test results giving details of IMD3 improvement in dB

It may be seen from Table 6 that at least 10dB of improvement was observed in all but one distortion component, and that the majority showed a reduction well in excess of this value. The average reduction observed in all the products was 16.4dB.

The power of all four carriers was then increased to 18dBm, with a corresponding increase in the power of the IMD3 components. Again, the linearising signal was generated in the same manner and phase- and amplitude-adjusted to optimise distortion reduction. The results are shown below, where it may be seen that the distortion improvement is considerable. It is likely that fifth-order intermodulation components are significant at this power level, so the fact that the overall average reduction is 17.7dB shows great promise.

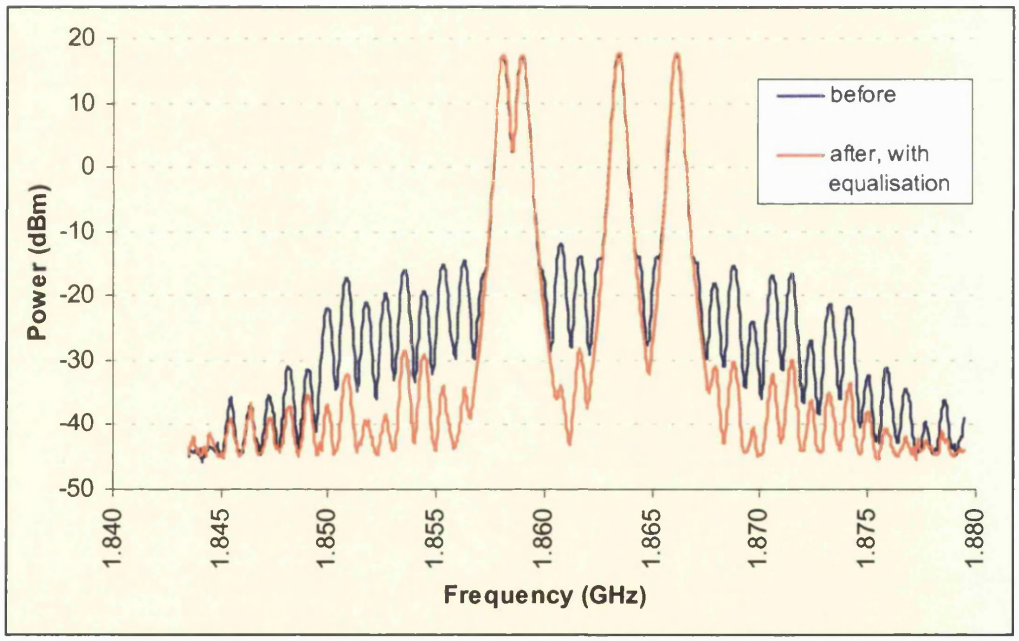


Figure 4.19: 4-tone test results using 'DFT-equalisation' – narrow frequency span

The next set of results concern the most widely-used benchmark for LPA performance, and involves equally-spaced carriers. Four carriers separated by 2MHz were applied to the system, and the linearising signal was captured, equalised and re-assembled in the manner already described. Figure 4.20 shows the output both with and without linearisation applied.

The reduction in individual components was found to be between 13dB and 20dB with an average of 16.7dB.

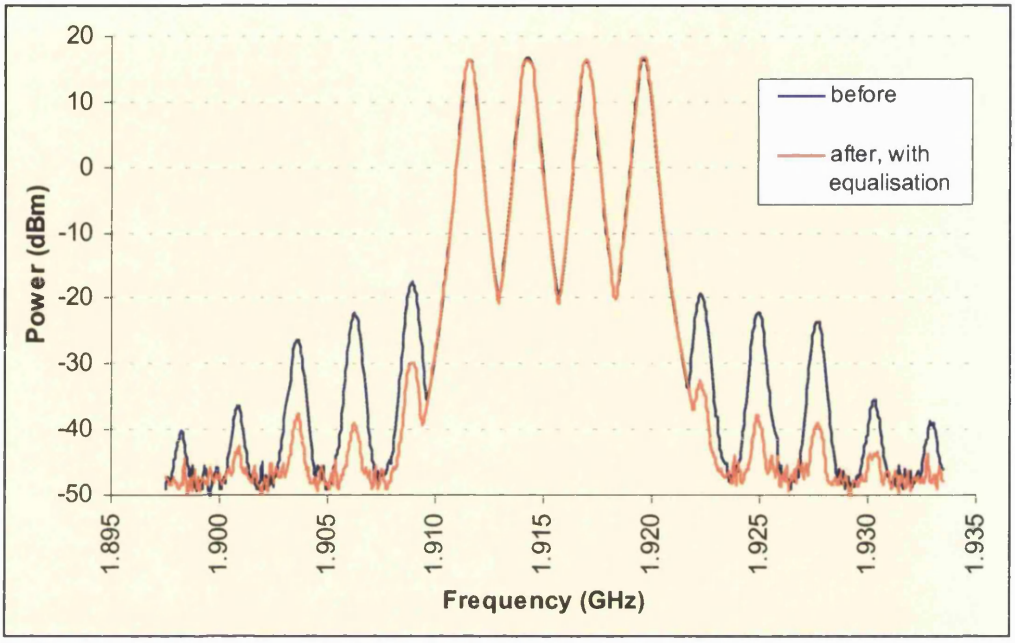


Figure 4.20: 4-tone test with equally-spaced carriers

A more realistic test was then performed, with four carriers of differing amplitudes and frequency spacings. Typical results using this type input signal are shown in Figure 4.21 and Figure 4.22, and are also tabulated in Table 7 overleaf.

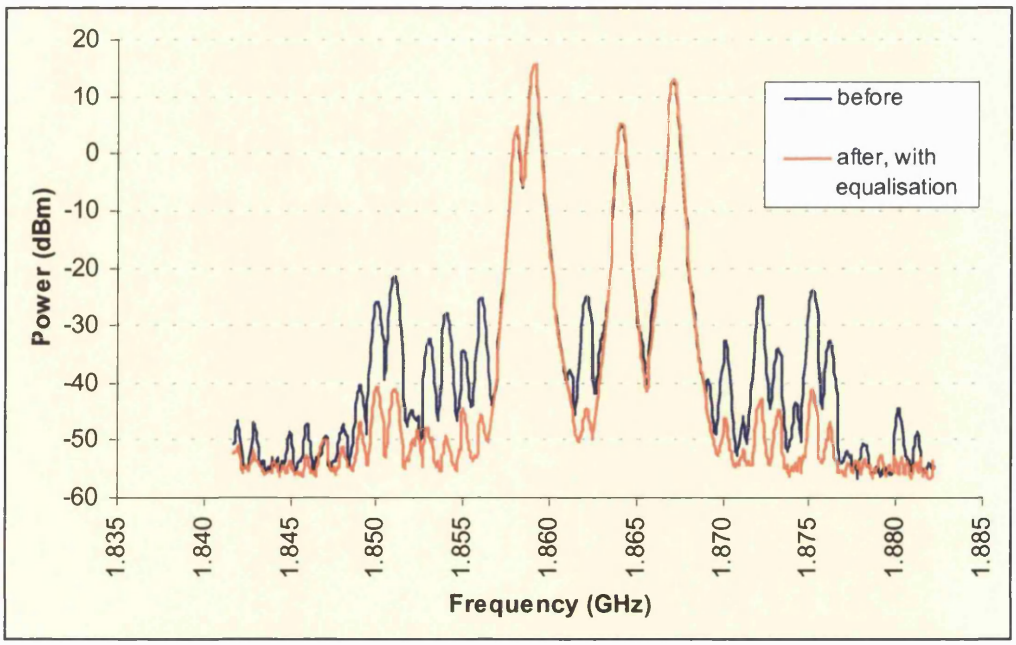


Figure 4.21: 4-tone test with staggered power levels – wide span

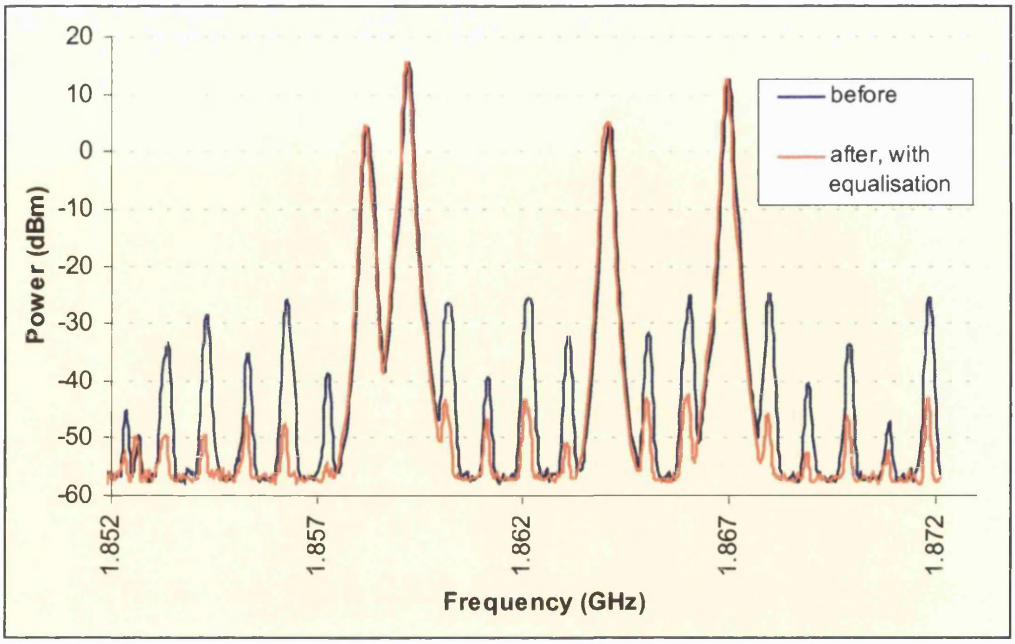


Figure 4.22: 4-tone test with staggered power levels – narrow span

IMD3 Product	Improvement (dB)	IMD3 Product	Improvement (dB)
$2f_2-f_1$	18	$f_1+f_2-f_3$	19.1
$2f_3-f_1$	20.5	$f_2-f_1+f_3$	13.1
$2f_1-f_2$	21.7	$f_1-f_2+f_3$	19.75
$2f_1-f_3$	9.85	$f_1+f_2-f_4$	9.9
$2f_3-f_2$	13.9	$f_2-f_1+f_4$	25.7
$2f_2-f_3$	26.25	$f_1-f_2+f_4$	16.1
$2f_1-f_4$	13.2	$f_2+f_3-f_4$	25.5
$2f_2-f_4$	20.6	$f_3-f_2+f_4$	18.85
$2f_3-f_4$	10	$f_2-f_3+f_4$	20.2
$2f_4-f_1$	18.2	$f_1+f_3-f_4$	13.1
$2f_4-f_2$	18.6	$f_3-f_1+f_4$	13.3
$2f_4-f_3$	$(2f_3-f_1)$	$f_1-f_3+f_4$	$(2f_3-f_4)$

Table 7: 4-tone test results giving exact improvement observed

Again, these results are impressive, with only two distortion products showing (marginally) less than 10dB of reduction, the majority being well in excess of this. The average value of improvement for all the products is 17.5dB.

The advantages of DFT-equalisation are numerous:

- **Reduced processing overhead**. As there is prior knowledge of the signal components present in the captured waveform, the DFT need only be performed at 'spot' frequencies, requiring far fewer multiply-and-add operations than either an FFT or a time-domain convolution.
- **Flexibility.** Accurate phase and amplitude equalisation can be applied to each frequency component individually, and these values may be easily adjusted if required. Thus, the 'virtual' feedback loop may have any transfer characteristic. This was not possible with convolution.
- **Overall performance.** As the results clearly demonstrate, the linearity improvement given by this method is far greater and more consistently successful than that produced previously.

This solution therefore provides the greatest degree of freedom with the smallest processing overhead and turnaround times, and produces superior results as shown.

4.4 Summary

The experiments described in this chapter demonstrate that the technique can be successfully applied to multi-carrier input signals, and that it can achieve significant reduction. The 'virtual' feedback loop that was developed was able to produce useful reduction with a wide variety of carrier spacings and power-levels with an essentially fixed transfer function. It may therefore be surmised that the linearising signal may be generated from the input signal by some other means, external to the amplifier, and that it will be sufficient for this generation process to have fixed characteristics also. As such, this technique is no longer best employed (or described) as low-frequency feedback, and this will be explored in more detail in the next chapter. However, the characteristics of any amplifier are likely to vary

over time and temperature, so it is possible that some kind of closed-loop control may be required.

5. Modulated Carriers

The analysis, simulation and practical results shown in the previous chapters have used sinusoidal input signals; this allowed the basic principles to be established, and allowed important conclusions to be drawn. However, the implications of applying modulation to the carriers are not immediately apparent from the simple analysis presented in the preceding chapters, so this aspect will now be investigated in more detail.

Third-order distortion is not only a problem in multi-carrier amplifiers – it affects any system where the waveform to be amplified has a time-varying envelope. As discussed in the introduction to this thesis, PA linearisation is currently a 'hot topic' in the both the telecommunications industry and also academia, as the air-interfaces of the next generation mobile systems will employ non-constant-envelope modulation schemes. This will increase the linearity requirements of PAs and unless suitable linearisers are developed, amplifier efficiencies will be reduced, BTS running costs will increase and handset talk-times will shorten. It is therefore essential that the application and effectiveness of the proposed linearisation technique with this type of input signal is now investigated.

5.1 Analysis of Single-Carrier Linearisation with Digital Modulation

It is convenient to begin this analysis in the frequency domain, as any arbitrary digitally-modulated signal may be represented as baseband spectrum convolved with a single sinusoidal carrier [29]:

$$x(t) \Rightarrow X(j\omega) = aB(j\omega) \otimes 1/2[\delta(\omega+\omega_o)+\delta(\omega-\omega_o)] \qquad (5.1)$$

where $B(j\omega)$ represents the spectrum of the baseband, a is the amplitude and ω_0 the carrier frequency that the modulated RF signal is centred at.

As before, we will consider an amplifier whose nonlinearity may be represented by a third-order power series, and proceed by examining the action of the second- and third-order distortion separately. The second-order term of the nonlinear transfer function, $g_2x^2(t)$, produces the following new output signal components:

$$g_2x^2(t) \Rightarrow g_2X(j\omega)\otimes X(j\omega) = g_2a^2B(j\omega)\otimes B(j\omega)\otimes 1/4[\delta(\omega\pm 2\omega_0)+2\delta(\omega)] \quad (5.2)$$

Equation 5.2 shows that second-order nonlinearity produces two 'self-convolved' or auto-correlated versions of the baseband spectrum, one appearing at $2\omega_0$, the second harmonic frequency, and the other at $\omega = 0$, or DC. It is the latter of these two components, the low-frequency portion, that is of interest here.

The third-order term of the transfer function, $g_3x^3(t)$, also produces new output signal components:

$$g_3x^3(t) \Rightarrow g_3X(j\omega)\otimes X(j\omega)\otimes X(j\omega) = g_3a^3B(j\omega)\otimes B(j\omega)\otimes B(j\omega)\otimes 1/8[\underbrace{\delta(\omega\pm 3\omega_0)}_{3^{rd}\text{-order harmonic}}+\overbrace{3\delta(\omega\pm\omega_0)}^{\text{Spectral Regrowth}}] \quad (5.3)$$

Equation 5.3 shows that, as expected, spectral regrowth is primarily due to third-order nonlinearity, and as such, it is logical to propose that this may be reduced by applying the new method of linearisation. (It is also logical to conclude that, if the analysis were extended to higher orders, that fifth-order nonlinearity would be another source of in-band spectral spreading, but with a bandwidth of five times that of the RF carrier.)

The application of the linearising technique will again be investigated by forming a second-order 'injection' signal from the low-frequency portion of the square of the input signal, which in this case may be written as:

$$I(j\omega) = \alpha e^{-j\theta}[B(j\omega)\otimes B(j\omega)] \quad (5.4)$$

where $\alpha e^{-j\theta}$ represents the bulk amplitude and phase-shift applied to the injection signal with respect to the original input. If this signal is examined, it can be seen that it consists of the convolution of the baseband spectrum with itself, in the frequency-domain. As the time-domain equivalent of convolution is multiplication, in this case the injection signal may be formed simply by squaring the time-domain baseband waveform and applying the required amplitude and phase-shift.

If the injection signal is added to the original modulated carrier that is centred at frequency ω_o, the input signal is modified thus:

$$X'(j\omega) = X(j\omega) + I(j\omega) = aB(j\omega) \otimes 1/2[\delta(\omega \pm \omega_o)] + \alpha e^{-j\theta}[B(j\omega) \otimes B(j\omega) \otimes \delta(\omega)] \quad (5.5)$$

and the second-order components generated at the output of the amplifier are also changed accordingly:

$$g_2 X'(j\omega) \otimes X'(j\omega) =$$

$$g_2 a^2 B(j\omega) \otimes B(j\omega) \otimes 1/4[\delta(\omega \pm 2\omega_o) + 2\delta(\omega)] \quad \text{- Original 2nd-Order Components}$$

$$+ g_2 a\alpha e^{-j\theta} 1/2[B(j\omega) \otimes B(j\omega) \otimes B(j\omega) \otimes \delta(\omega \pm \omega_o)] \quad \textbf{- New ACPR around carrier at } \omega_o$$

$$+ g_2 \alpha^2 e^{-j2\theta}[B(j\omega) \otimes B(j\omega) \otimes B(j\omega) \otimes B(j\omega)] \quad \text{- New components at } \omega=0 \quad (5.6)$$

There are now new third-order distortion components produced by the second-order term of the nonlinear transfer-function, as shown in Equation 5.6. If this new ACPR is compared against the spectral regrowth highlighted in Equation 5.3, it can be seen that the two terms are identical except for their amplitudes and the phase-shift term $e^{-j\theta}$. The value of the injection signal amplitude α may be found my equating the two amplitudes as follows:

$$\frac{3a^3 g_3}{8} = \frac{a g_2 \alpha}{2} \Rightarrow \alpha = \frac{3a^2 g_3}{4 g_2} \quad (5.7)$$

Hence, if the value of α is as shown in Equation 5.7 with θ equal to 180°, the two sets of components in Equations 5.6 and 5.3 can be made to cancel. In fact, if the sign of the third-

order coefficient is negative, the phase relationship between the existing third-order distortion and the new linearising components should already be such that they are in anti-phase without the application of this 'bulk' phase-shift.

To further investigate the application of the technique to a digitally-modulated carrier, it is instructive to revert to time-domain analysis. Referring to Figure 5.1, filtered baseband inputs $b_1(t)$ and $b_2(t)$ are upconverted in a quadrature modulator to the RF carrier frequency, ω_{RF}. The two quadrature waveforms are then summed before being squared and low-pass filtered to produce the basic linearising signal.

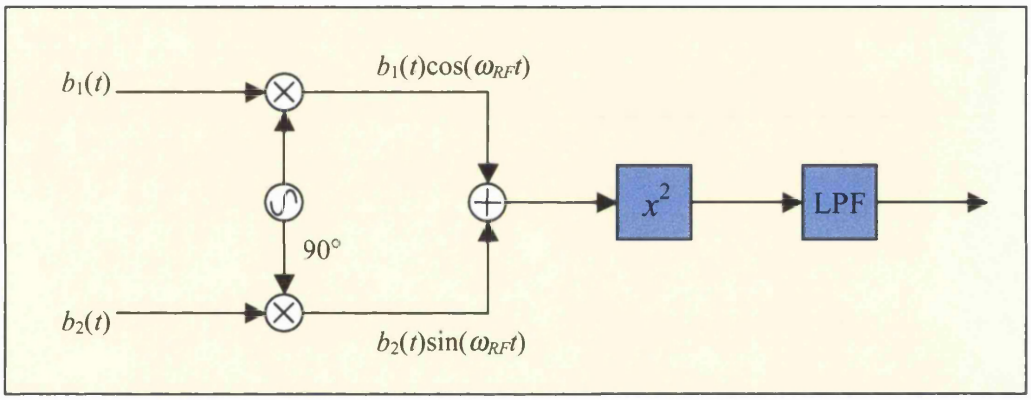


Figure 5.1: Linearisation signal generation with a digitally-modulated carrier

The square of the input signal may now be written as:

$$\begin{aligned} V_{RF}{}^2 &= [b_1(t)\cos(\omega_{RF}t) + b_2(t)\sin(\omega_{RF}t)]^2 \\ &= b_1^2(t)\cos^2(\omega_{RF}t) + b_2^2(t)\sin^2(\omega_{RF}t) + 2b_1(t)b_2(t)\cos(\omega_{RF}t)\sin(\omega_{RF}t) \\ &= \frac{b_1^2(t)}{2}(1+\cos(2\omega_{RF}t)) + \frac{b_2^2(t)}{2}(1-\cos(2\omega_{RF}t)) + b_1(t)b_2(t)\cos(2\omega_{RF}t) \end{aligned} \quad (5.8)$$

By examination of Equation 5.8, it is clear that the only components at a lower frequency than the RF carrier are those given by the square of the separate time-varying baseband waveforms, $b_1{}^2(t)$ and $b_2{}^2(t)$. Therefore, it may be postulated that the linearising signal for a single carrier may be formed from the sum of the squares of the separate baseband waveforms. This conclusion will be verified in simulation in the next section.

5.2 Simulation of Single-Carrier Linearisation with Digital Modulation

5.2.1 Idealised Envelope Simulation

In order to verify the above analytical findings, the circuit shown in Figure 5.2 was assembled in *Advanced Design System* (*ADS*). Although *ADS* has the capability to realistically 'co-simulate' digital baseband and RF circuits, it was decided to begin with an idealised 'proof-of-concept' scenario that, once optimised, could then be developed into a more authentic representation of a transmitter.

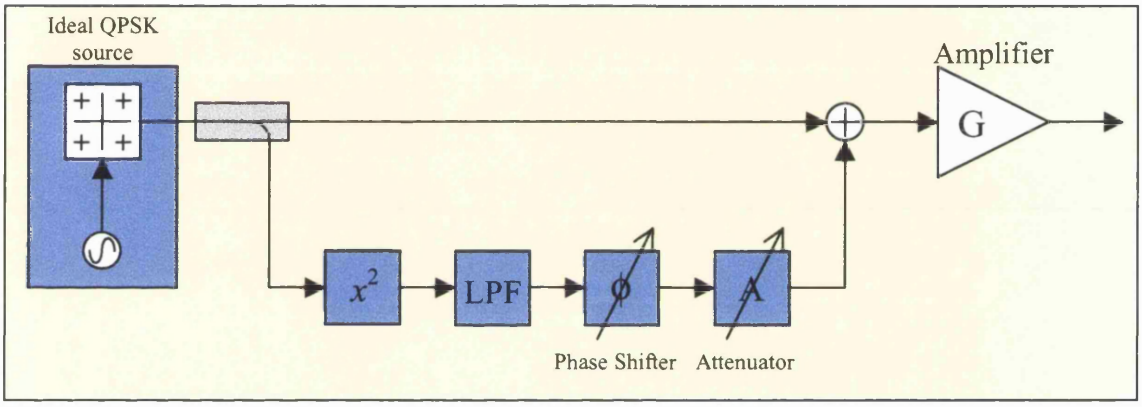


Figure 5.2: Idealised proof-of-concept lineariser schematic

As this initial circuit was to be used in an Envelope simulation, a Quadrature Phase-Shift Keying (QPSK) source from the library was employed to generate the upconverted RF carrier signal. Unfortunately, this component did not allow the I and Q baseband signals to be accessed directly. This problem was solved in the manner shown above, where a coupled-off sample of the RF input is first squared and low-pass filtered before being phase- and amplitude-adjusted, again with idealised components.

The nonlinear amplifier model used in these simulations was simplified to a one-dimensional power-series nonlinearity of the type used in earlier analytical work to reduce the number of unknown variables and make simulator convergence easier. The RF matching circuitry and parasitic components around the nonlinear current generator within the amplifier were also

removed for the same reasons. This simplifying assumption is often made in simulations of this type, and is valid for 'bandpass' nonlinearities of the type represented by a memoryless LTI system such as this. Although in a real amplifier, the impedances presented to the low-frequency and RF input signal components will be different as well as frequency dependent, this is ignored for the present. If it proves to be necessary, the second-order injection signal can be pre-equalised in a similar fashion to the practical four-tone proof-of-concept system described in Section 4.3.

The results in Figure 5.3 below are for a typical simulation, with a carrier centred at 900MHz and a modulation rate of 100kHz. Although the power level had to backed off by approximately 20dB to ensure convergence of the harmonic balance simulation, the amplifier produced appreciable distortion, as shown by the blue trace.

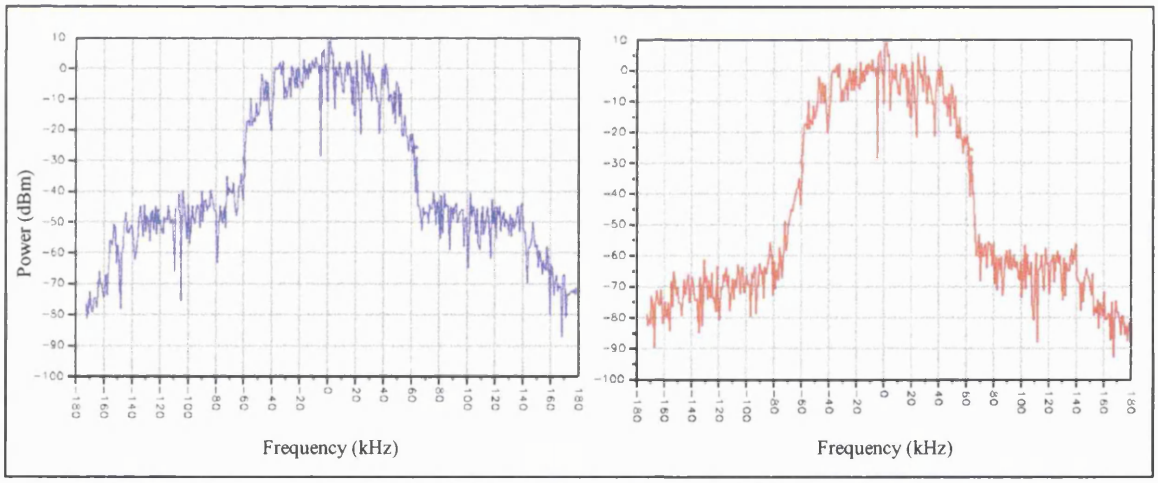


Figure 5.3: *ADS* Envelope-Simulation Results with 100kHz QPSK input signal

The upper and lower ACPR were found to be –57.8dBm and –57.6dBm respectively. The red trace shows the same circuit with the addition of the low-frequency injection signal, after optimisation of the attenuation and phase-shift values. The upper and lower ACPR were both reduced by approximately 18dB, with the level of the main carrier being increased by 0.1dB.

5.2.2 Idealised DSP/RF Co-Simulation

Having demonstrated that second-order bias injection could be applied to modulated signals, the next step entailed generating the linearising signal without squaring and low-pass filtering the RF carrier itself. In order to do this, it was necessary to employ the 'co-simulation' facilities of *ADS*, a new feature allowing analogue and digital circuits to be linked and analysed together. Figure 5.4 shows the schematic of the circuit that was assembled.

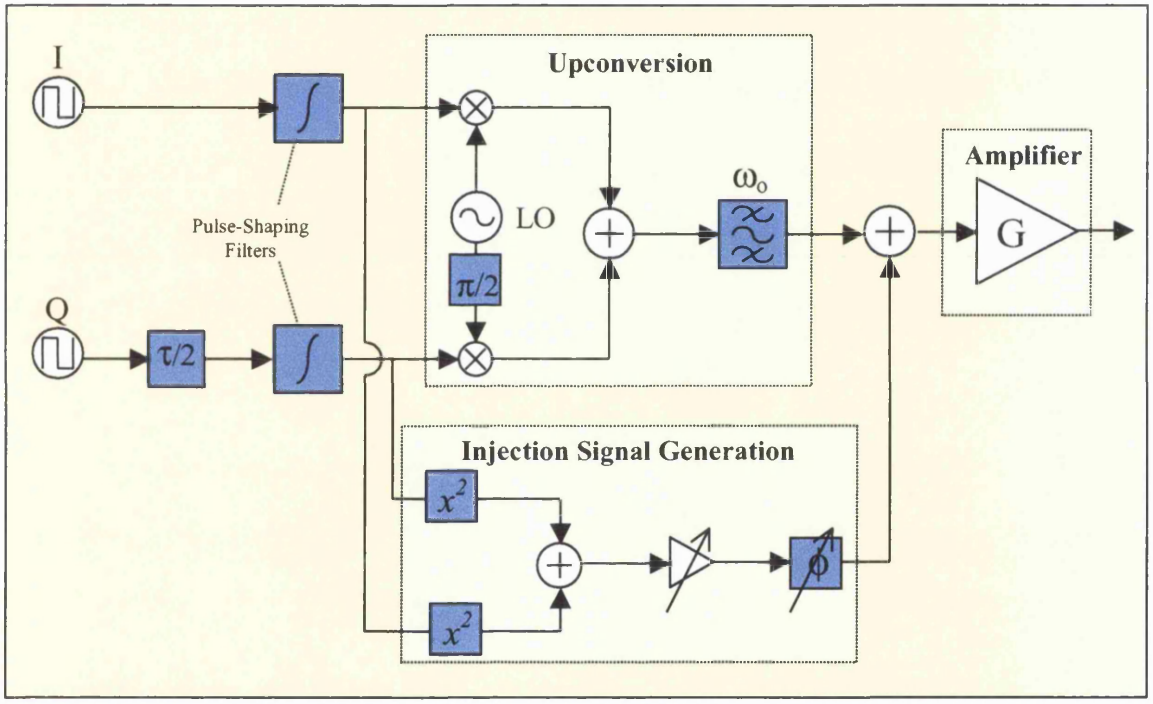


Figure 5.4: Schematic of *ADS* Co-simulation of second-order injection circuit

Referring to the above figure, the I and Q digital bitstreams at the left-hand side of the diagram are first pulse-shaped before being mixed onto quadrature sinusoids which are summed to form the RF input signal. The quadrature baseband waveform is delayed by half a bit-period with respect to the in-phase component, to generate offset-QPSK modulation, as used for the uplink of the North American CDMA system, IS-95.

As described earlier, the linearising signal can be generated directly from the baseband waveforms, so the I and Q bitstreams were passed separately to squaring networks, then summed, phase- and amplitude-adjusted, and re-injected before the amplifier to form a new composite input signal. Figure 5.5 shows a typical result, with a carrier centred at 2GHz, with a bandwidth of 2MHz.

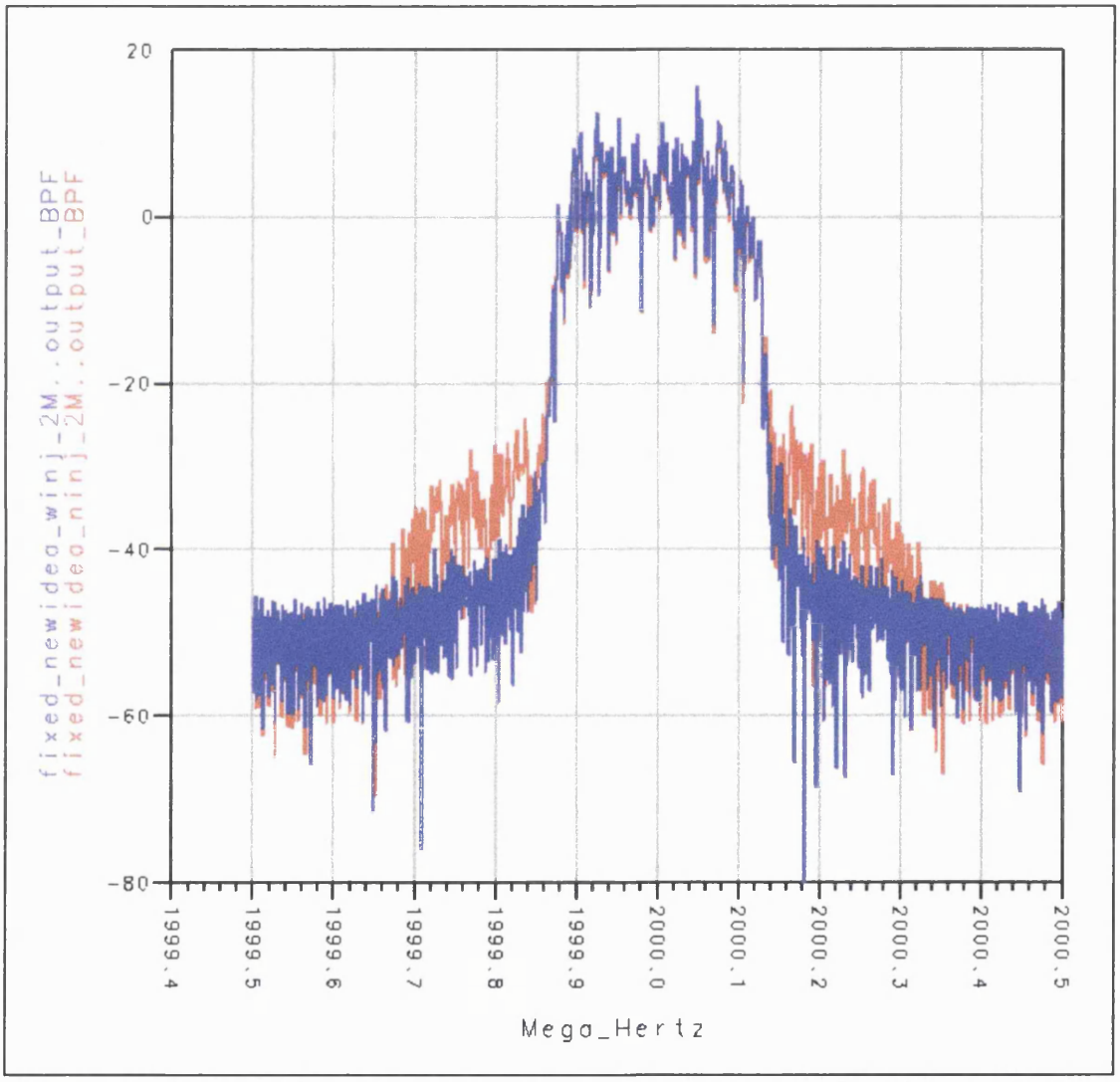


Figure 5.5: Results of *ADS* Co-Simulation, both with (blue) and without (red) the injection signal

The results confirm the analysis presented earlier, as the output spectrum shows considerably less spectral spreading around the carrier when the injection signal is added. The selection of

frequencies and modulation standard was made arbitrarily – the technique may be applied in this form to any scheme employing digital modulation.

5.2.3 Power Dependence

It was shown in Equation 5.7 that the optimum amplitude of the linearising signal is given by:

$$\alpha = \frac{3a^2 g_3}{4g_2} \tag{5.9}$$

Therefore, for distortion reduction to be maintained with changing power level, the amplitude of the linearising signal must vary with the square of the amplitude of the input signal. To verify this, the topology shown in Figure 5.4 was modified to incorporate this, and the amplitude of the carrier was swept. The results of this are shown in Figure 5.6 below.

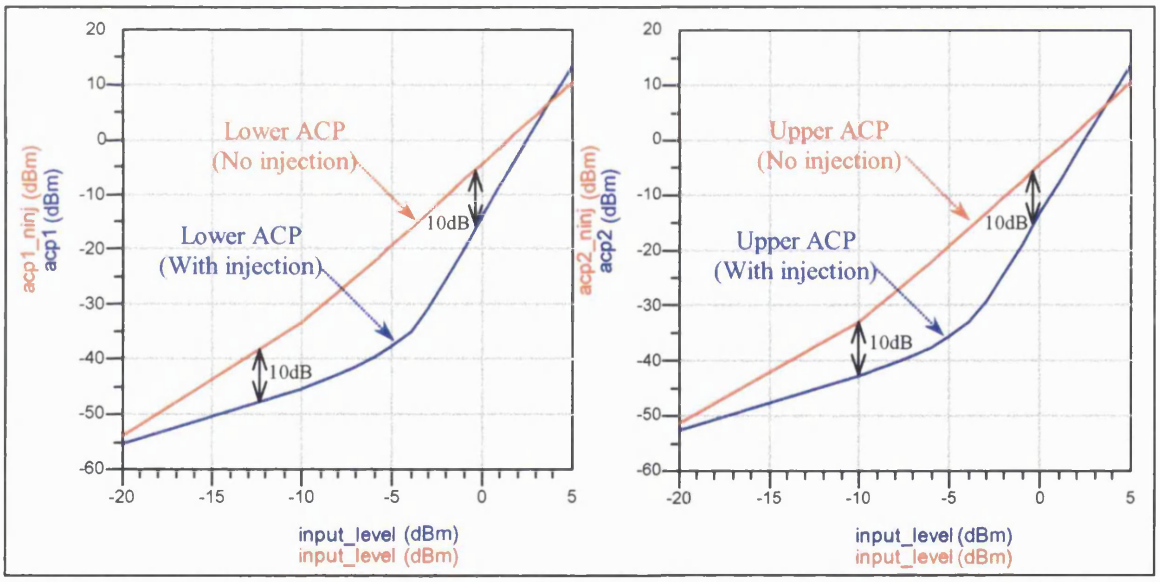


Figure 5.6: Upper and lower ACP versus input power level, with and without injection

Referring to Figure 5.6, the red traces show upper and lower ACP against input power level without linearisation, while the blue traces show the level of the distortion after linearisation has been applied. The plots show that the level of distortion reduction appears to be level-dependent, with 10dB of ACP improvement occurring over an power level range of approximately 10dB. The blue (linearised) curves have a distinctive shape; at the lower end

of the range, the distortion reduction is limited by the power of the residual sidebands around the carrier and the noise floor of the simulation. This dominates the lower half of the curves, until the level of the third-order ACP becomes significant, as shown by the increase in gradient. The main part of the plots show that distortion reduction is approximately constant at 12dB, while the upper end of the scale is dominated by another mechanism that will be discussed shortly.

5.2.4 Simulated Phase and Amplitude Tolerance

By adjusting the relative phase and amplitude of the injection signal, the tolerance of the distortion reduction was assessed. Figure 5.7 below shows the amount of ACPR reduction against phase deviation.

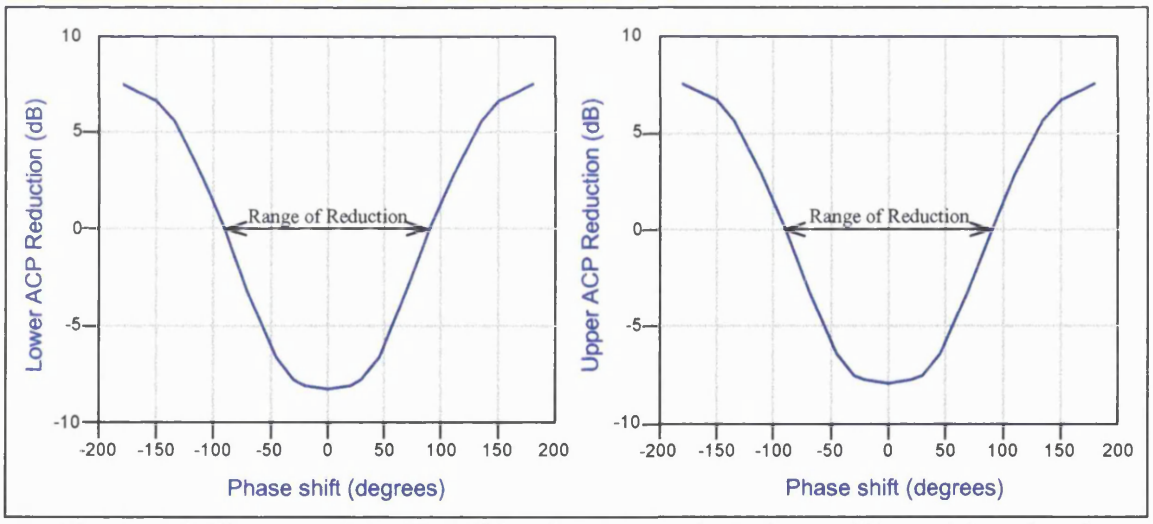


Figure 5.7: Upper and lower ACP reduction against phase of linearising signal

The above plots also show that distortion reduction occurs over a very wide range of phase-shift, with some improvement being obtained over a range spanning approximately ±90° around the optimum value. This is considerably wider than the tolerance of a scheme such as Feedforward, and is due to the fact that the linearising signal is now formed from the scalar product of the RF carrier signal and the low-frequency linearising component. As the level

of the RF carrier is considerably larger than that of the injection signal, the linearising signal makes less contribution to the resulting distortion-cancelling components than the carrier. Figure 5.8 attempts to illustrate this, with input and output signals shown as vectors whose thickness denotes the instantaneous (or modulation) bandwidth of the signal component represented.

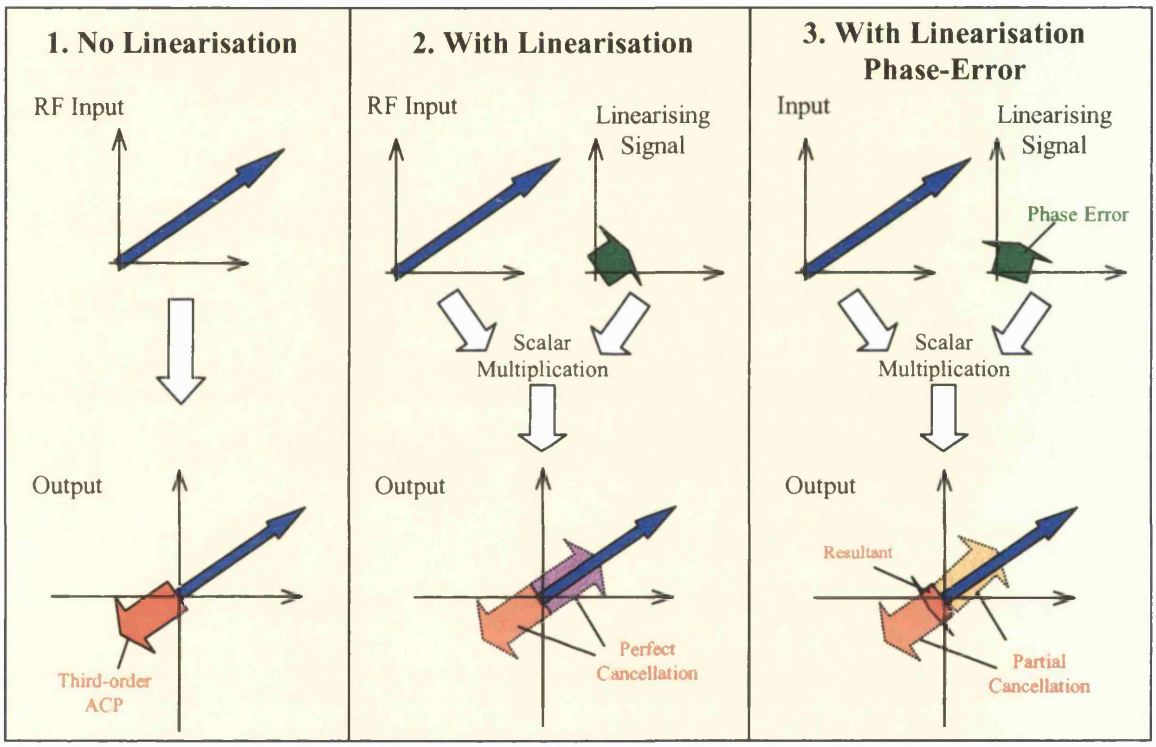


Figure 5.8: Vectorial-representation of linearising signal phase-imbalance

The first panel of Figure 5.8 shows the input and amplifier output signals with no linearisation signal, with the thicker red arrow in the output representing the third-order distortion, appearing in antiphase. The middle panel shows an optimised linearising signal being added to the input carrier, producing a new third-order component that perfectly cancels the in-band ACP. The right-hand panel shows a linearising signal whose phase has been slightly off-set, resulting in a distortion-cancellation component that is no longer in perfect antiphase with the in-band distortion. However, the resulting phase perturbation is

far less than that of the phase-error of the linearising component, so its effects are reduced significantly and some reduction of the distortion still occurs.

This suggests that deviation in the amplitude of the linearising signal will have a similarly reduced impact on the distortion reduction. Figure 5.9 shows the simulated variation in ACP reduction against the amplitude of the linearising signal, varied around its optimum value.

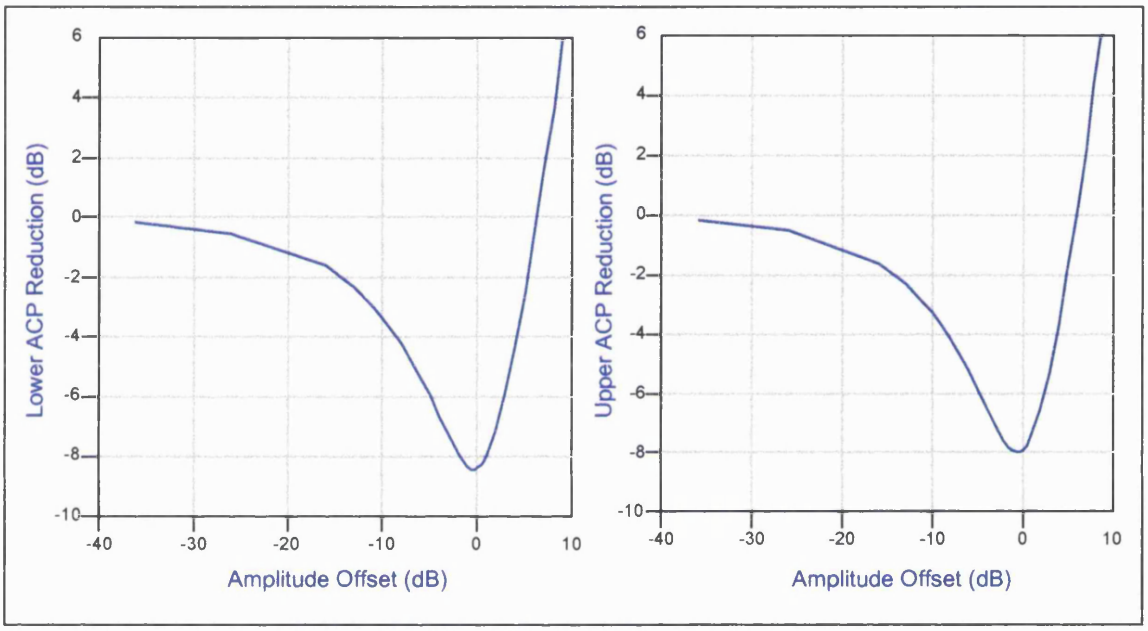


Figure 5.9: Simulated ACPR reduction against linearising signal amplitude offset

Figure 5.9 shows that the tolerance to amplitude-shift is also high, with distortion reduction now occurring over a range up to 35dB. The plots show that the performance rapidly degrades as the amplitude of the linearising signal exceeds the optimum value, however, as it has an increasingly large influence on the amplitude of the linearising component.

The shapes of the characteristics shown in Figure 5.8 and Figure 5.9 are determined by the relative levels of the second- and third-order nonlinearities, so in reality they are likely to be device- dependent and will vary with bias point and operating conditions. However, the tolerance to linearising-signal amplitude and phase error will always be larger than in a cancellation scheme such as Feedforward. This has already proven to be the case in practice,

where two-tone measurements revealed tolerances in the order of 10dB and 100°. This suggests that the effect of bias-circuit transfer-functions will be reduced, and that the gain and phase-flatness requirements of the injection-signal path are far more relaxed than for other schemes such as Feedforward or RF predistortion.

The results shown suggest that the maximum achievable reduction in ACPR at this power level is limited to about 8dB. However, as these results were obtained with idealised models and components, near-perfect cancellation should be possible and it is not immediately apparent why this is not so. The reason for this limit can be explained by referring to Figure 5.10 below, which shows linearised and un-linearised output spectrums obtained from an optimised circuit.

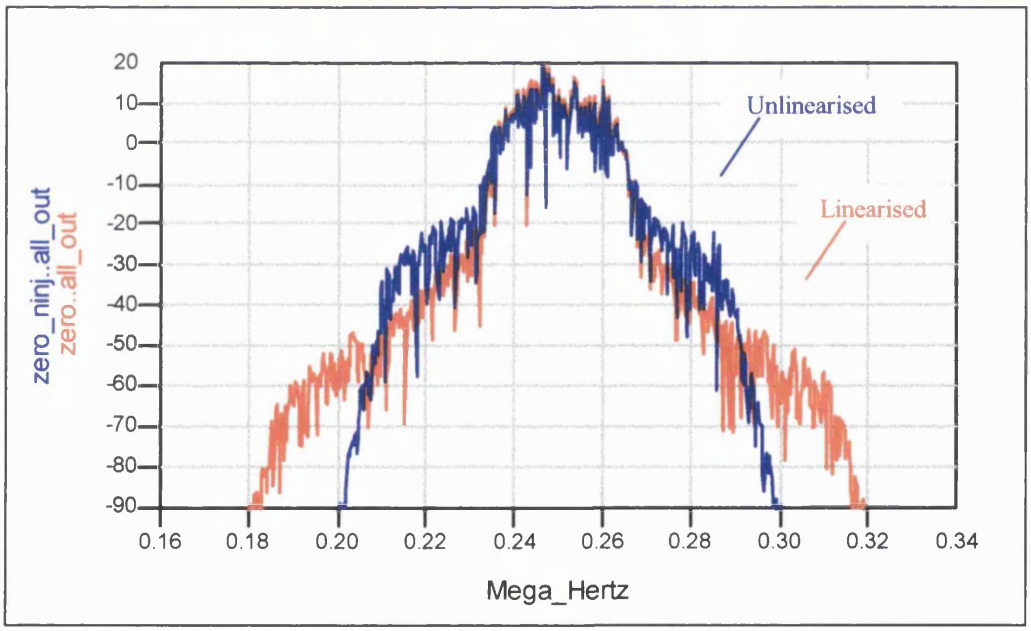


Figure 5.10: Output spectrum with (red) and without (blue) linearisation

The blue trace (representing the output of the amplifier with no linearisation applied) shows the carrier with third-order spectral spreading around it. The red trace shows the output of the amplifier after the application of linearisation, and it is here that the reason for the limited ACPR reduction lies. The linearised output contains what appears to be fifth-order

spectral-spreading that is in fact due to the interaction of the low-frequency linearising signal and the carrier via the *third-order* amplifier nonlinearity. The overall ACPR cannot be reduced beyond the level of this new distortion, so the maximum improvement that can be achieved is limited. Again, the level of these undesirable components is determined by the relative size of the second- and third-order nonlinearities, as well the relative power of the carrier and linearising signals.

As the model used in these simulations takes no account of fifth-order nonlinearity, there is no way of predicting at this stage whether the performance of a real device will be limited or improved by this mechanism. Fifth-order spectral-spreading components will already exist in the output of a real device, and the overall effect cannot be predicted until an experimental investigation is carried out. This will be further explored in the practical verification of the technique described in the next section.

5.3 Experimental Single-Carrier Linearisation with Digital Modulation

In order to verify the conclusions presented in the preceding sections, it was necessary to recreate the simulation scenario shown in Figure 5.4. In order to do this, a test-bench was assembled as shown overleaf in Figure 5.11 overleaf. Although less hardware is now required to produce the test-signals, the equipment itself and the software to drive it is more sophisticated than that used previously. The digital synthesiser shown in Figure 5.11 (ESG4433B from Agilent) was chosen as it incorporates an internal dual arbitrary waveform generator to which externally-generated waveforms can be downloaded. These stored waveforms can then be used to modulate a carrier that is upconverted within the unit before being output at RF, a highly desirable feature as it allows the digital I and Q waveforms to be generated in a software environment, before being transferred to the unit. This gave the

greatest flexibility in terms of the modulation types that could be tested, and it also allowed the raw I and Q data to be manipulated and processed to create the linearising signal.

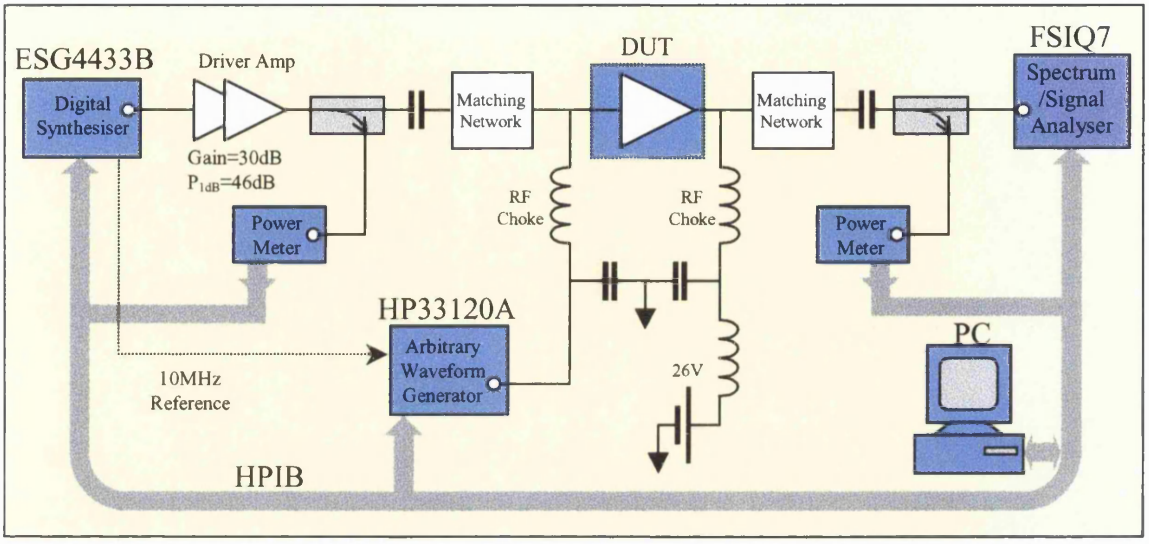


Figure 5.11: New single carrier testbench

Two 20dB directional couplers were used with power meters on the DUT input and output as shown. These, in conjunction with DC voltage and current readings, allowed the amplifier's performance to also be assessed in terms of both power and efficiency.

Referring to Figure 5.11, another difference between the set-up described in this section and that of the 4-tone test-bench shown in Figure 4.6 is that the output of the arbitrary waveform generator producing the low-frequency linearising signal is now DC coupled to the gate-bias feeding line at the input of the amplifier, whereas before it was applied through a DC-blocking capacitor. This change was necessary as the spectrum of the linearising signal is now continuous down to ω =0, and it is imperative that the effects of any frequency-selective components in the path of linearising signal are removed. Unfortunately, it was also found that this requirement meant that the amplifier used in the previous experiments (FLL351ME, see Section 3.1) was unsuitable, as direct connection of the HP33120A to the end of the quarter-wavelength gate bias-feed line caused oscillation to occur.

Due to time constraints, it was not possible to re-design the circuit, so it was decided to carry out the next series of experiments with a 4W LDMOS amplifier (MRF281S from Motorola) that had been designed at Nokia as part as a separate project. The measured small-signal S-Parameters of this amplifier have been included in Appendix B for reference. The circuit was ideal as it employed resistive gate-bias feeding using a 200Ω surface-mount resistor soldered directly to the input matching network, as shown in the photograph in Figure 5.12 below. This was a desirable feature as it was found that the HP33120A could be directly connected as shown, without causing oscillation.

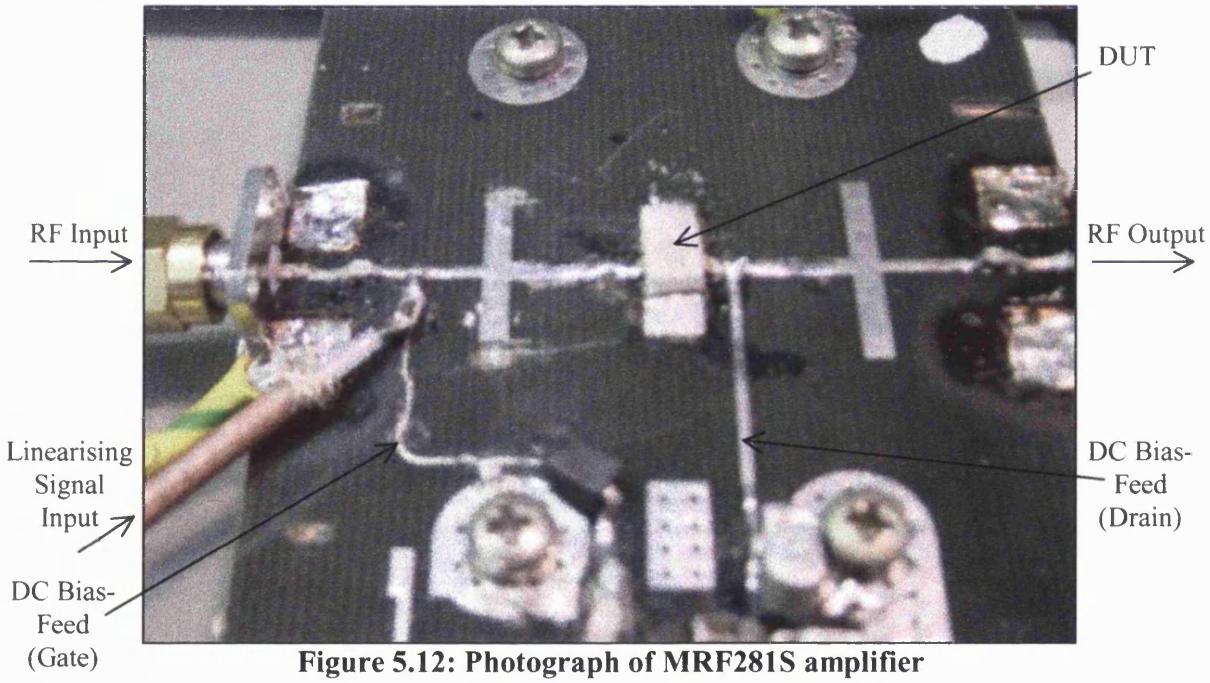


Figure 5.12: Photograph of MRF281S amplifier

As the output of the HP33120A can be 'floated' onto DC voltages of up to ±42V, it was possible to set the quiescent bias point independently of the linearising signal. Unless stated otherwise, this was chosen as I_{dq}=100mA ensuring Class AB operation.

It was also necessary to ensure that the linearising signal did not perturb the quiescent gate voltage, so the waveform was processed to remove any DC offset before being downloaded to the unit, as described overleaf.

5.3.1 Single Carrier Implementation and Testing

In a similar manner as described in Section 4.2.2, VisualBasic routines were written to create and manipulate waveforms, control the equipment and record measurement data. The following list describes the most important functions that were implemented, with the relevant code included in Appendix C for reference:

i) Random streams of I and Q pulses were created and stored (Appendix C, 1)

ii) The data streams were then convolved with an impulse response describing a raised-cosine baseband filter that had been designed using a digital filter designer tool available within *ADS*. Filters with cut-off frequencies of 200kHz, 1MHz and 2MHz were designed to allow testing with different bit-rates (Appendix C, 2).

iii) The peak to average ratio (PAR) of the generated signal was calculated.

iv) The two data streams were uploaded to the ESG4433B signal generator where they were stored, upconverted and output at RF.

v) The filtered I and Q waveforms were then squared individually and summed together to form the low-frequency linearising signal. This was downloaded to the AWG and injected to the PA via the gate bias network, combining with the RF carrier at the input of the device (Appendix C, 3).

Figure 5.13 overleaf shows example waveforms produced by the above steps.

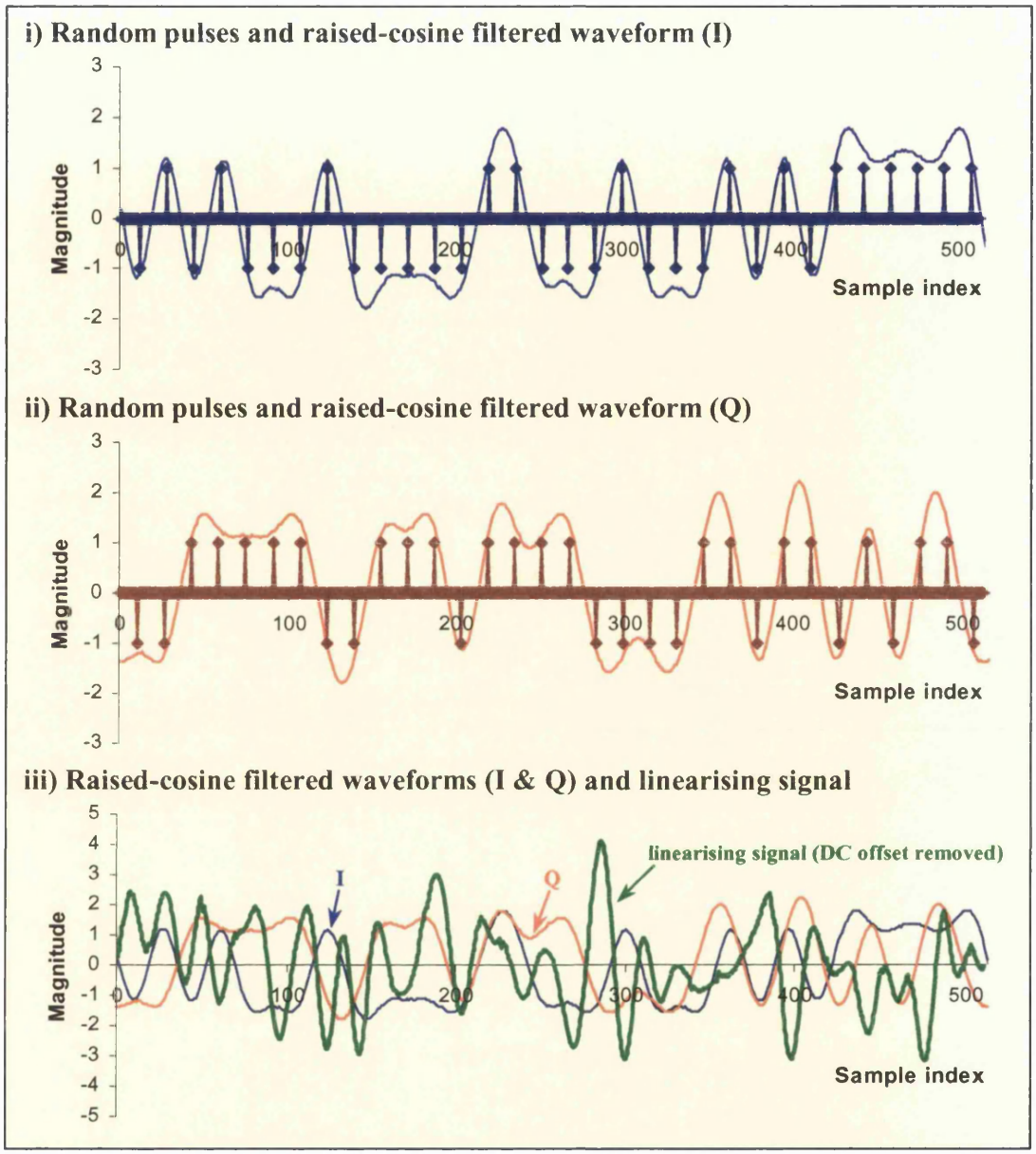


Figure 5.13: Example I, Q and linearising signal waveforms

In order to time-align the linearising waveform and the modulated carrier, it was again necessary to momentarily step the output frequency of the HP33120A up or down by a few tens of Hertz. After manually-optimising in this manner and adjusting the amplitude of the linearising signal, distortion improvement was observed. Figure 5.14 and Figure 5.15 show unlinearised and linearised output spectrums with a carrier at 1.85GHz, QPSK modulated by a periodic sequence of 256 symbols at a rate of 200kHz.

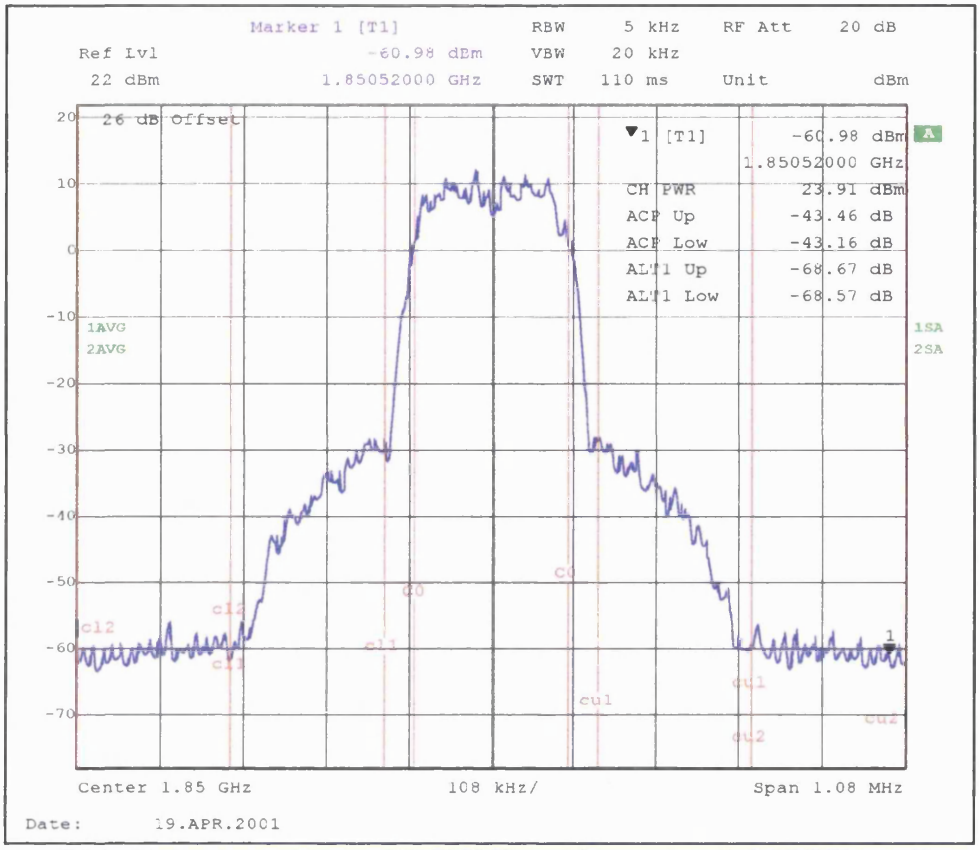


Figure 5.14: Output spectrum with 200kHz-wide carrier @ 1.85GHz, unlinearised

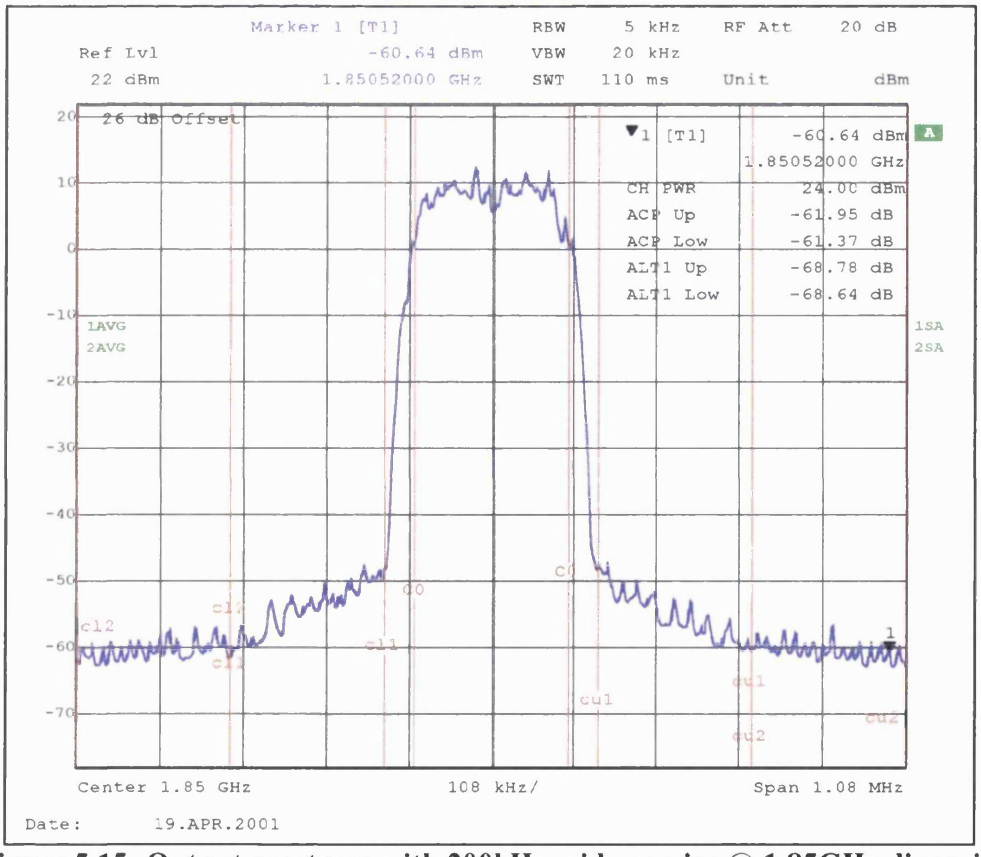


Figure 5.15: Output spectrum with 200kHz-wide carrier @ 1.85GHz, linearised

The peak-to-average ratio (PAR) of the test-signal was 4.37dB, and the Power-Added Efficiency (PAE) of the amplifier at the chosen operating point was 9.2%. Figure 5.14 shows the raw performance of the amplifier, where it can be seen that the upper and lower ACPR were approximately –43dBc. The output spectrum of same amplifier with the linearising signal applied is shown in Figure 5.15, where the upper and lower ACPR are reduced to –61.95dBc and –61.37dBc, respectively, giving distortion improvements of 18.49dB and 18.21dB. No change in Alternate Channel Power Ratio (ALCPR) was observed, though the measurement is dominated by the noise-floor of the test equipment.

In order to evaluate the technique over a wider bandwidth, a 2MHz QPSK carrier consisting of 256 symbols was then generated, having a PAR of 4.76dB. Figure 5.16 and Figure 5.17 show the raw and linearised output spectrums using this test-signal, at a power-level that gave an amplifier PAE of 13.3%. Again, the in-band distortion was reduced significantly by the action of the linearising signal, with the upper and lower ACPR improving from –40.59dBc and –39.96dBc to –53dBc and –53.11dBc, respectively. These reductions in adjacent-channel power (12.41dB and 13.15dB) were accompanied by small reductions in upper and lower ALCPR of approximately 0.3dB.

These findings show conclusively that the technique can be successfully applied to modulated carrier-signals, and good agreement with the two-tone and multi-tone investigations and simulation results in Section 3.3.1 was observed. Although this correlation is thought to be accidental as the nonlinear coefficients in the simulations were extracted from a different amplifier, it is possible that memory- and higher-order nonlinear effects will tend to produce similar levels of distortion improvement with other devices.

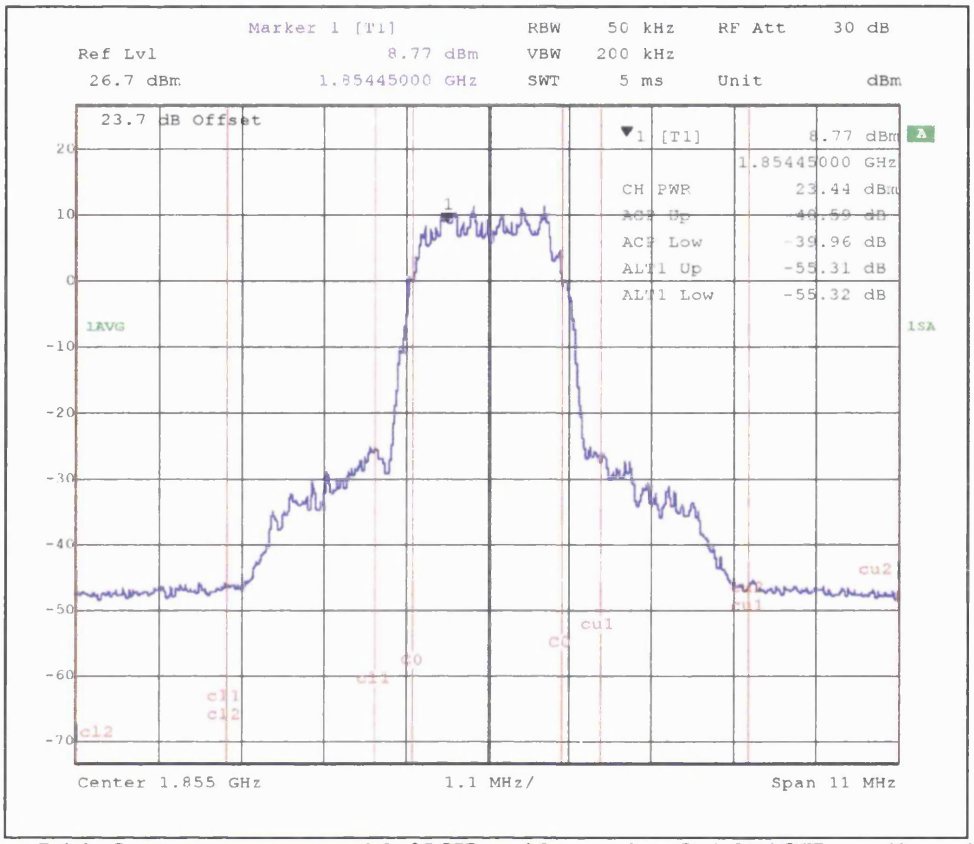


Figure 5.16: Output spectrum with 2MHz-wide carrier @ 1.854GHz, unlinearised

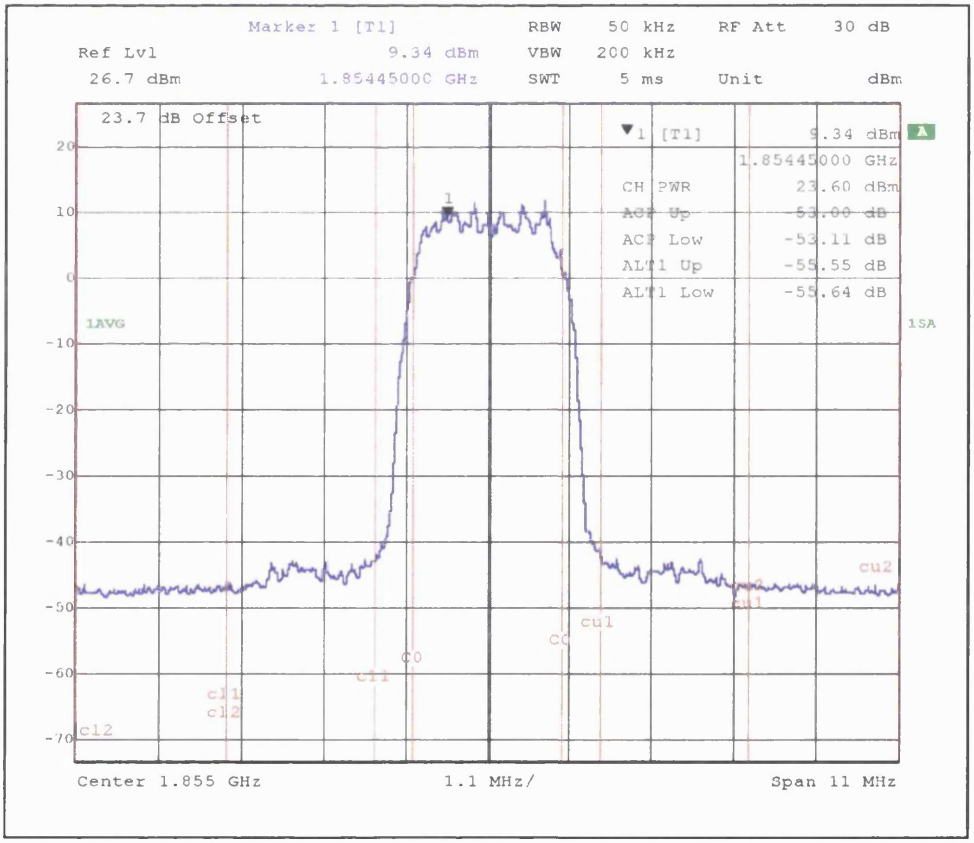


Figure 5.17: Output spectrum with 2MHz-wide carrier @ 1.854GHz, linearised

5.3.2 Measured Power Dependence

With this arrangement, it was also possible to measure the variation in distortion improvement against power level, as described and simulated in Section 5.2.3. In order to investigate this, the power of the 200kHz input signal shown in the previous section was swept over a range of 20dB, with the amplitude of the injected linearising signal re-optimised at each step. Figure 5.18 shows the variation in fundamental output power, efficiency and upper and lower ACPR (with and without linearisation applied).

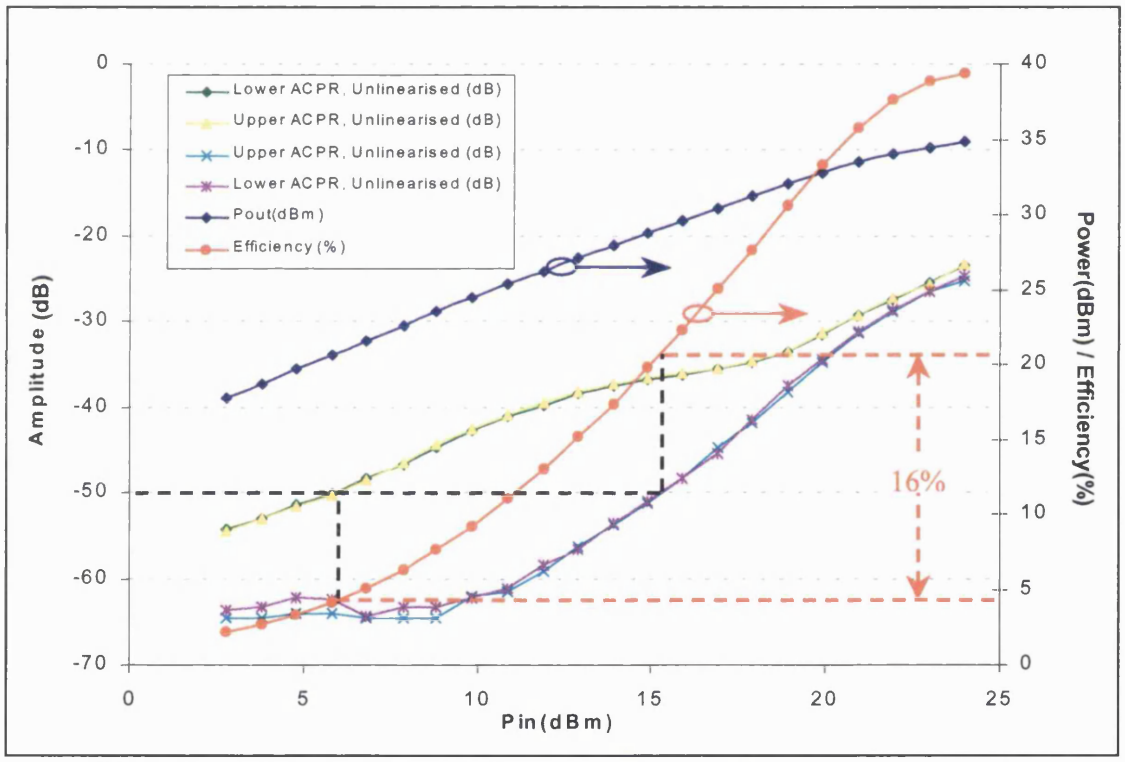


Figure 5.18: Measured amplifier performance with varying input power

Referring to the above figure, it can be seen that the reduction in ACPR was obtained across a wide dynamic range, with some improvement occurring at every power level. At least 10dB of ACPR improvement was obtained at all power levels below 30dBm, with a maximum reduction of 20dB occurring at an output power level of approximately 25dBm.

The real measure of the benefit that this improved linearity gives to a system designer is provided by the amplifier efficiency, as it is this that will dominate both DC supply current and thermal considerations in a practical transmitter. Referring again to Figure 5.18, a minimum linearity requirement of –50dBc has been highlighted as an example. Without linearisation, the amplifier is capable of meeting this specification if it is backed-off to an output power of 20dBm, giving a PAE of 4.2%. With the addition of the injection signal, however, the same relative level of distortion can be obtained at an output power of 29dBm, giving an increased PAE of 20.6%.

Figure 5.19 below shows additional data collected during the same series of measurements, with upper and lower ACPR improvement, upper and lower ALCPR improvement, gain compression and optimum injection-signal amplitude plotted against input power level. At low signal levels, the noise floor of the spectrum analyser dominates the measurement of ACPR and gives rise to the peak in distortion improvement shown.

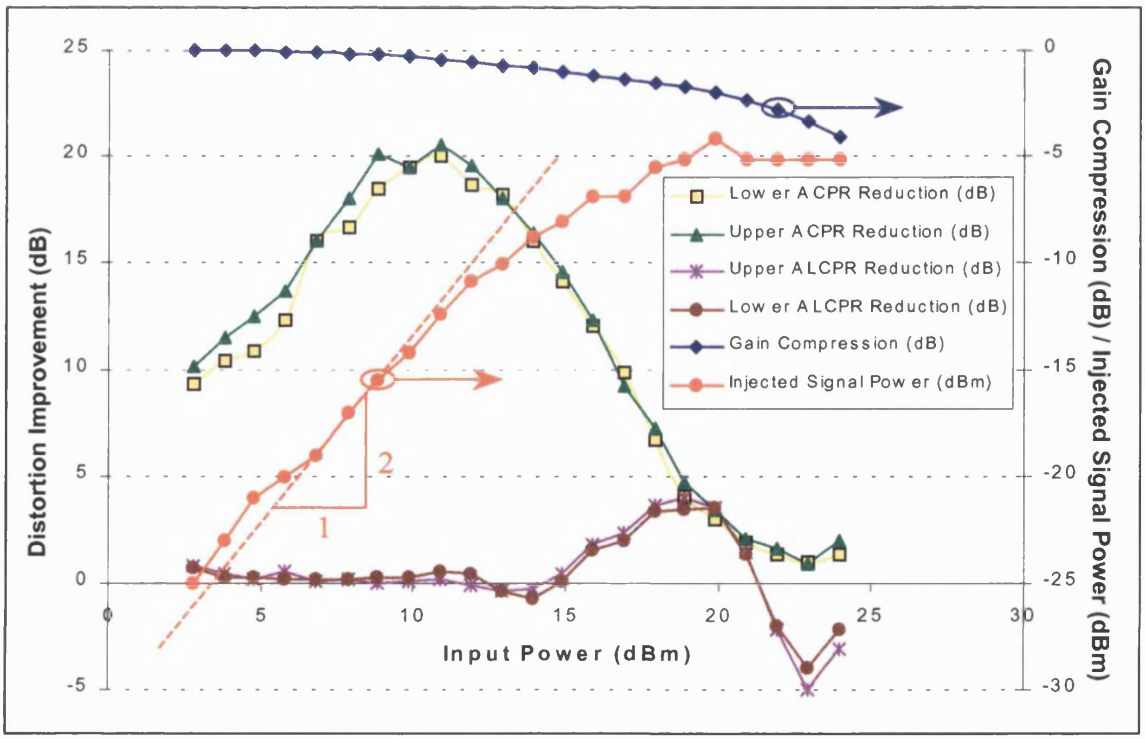


Figure 5.19: Upper and lower ACPR improvement, gain compression and optimum injection amplitude against input power level

With 1dB of gain-compression, roughly corresponding to the maximum power level at which such a PA would be used, the reduction in ACPR was found to be approximately 15dB. As can be seen in Figure 5.18, the improvement rapidly falls off at higher levels of compression, though this is expected as linearisation cannot correct for large-signal nonlinear effects caused by current and/or voltage-swing saturation.

The plots of ALCPR improvement on Figure 5.18 are also dominated by the noise floor of the spectrum analyser, only becoming valid beyond an input power of approximately 12dBm. Beyond this point, there is a small region over which a small degradation in ALCPR was observed, which is subsequently followed by a region of improvement that peaks at 3.5dB before dropping off again. For input power levels greater than 22dBm, the Alternate Channel Power Ratio is degraded by the addition of the linearising signal. The likely cause of this behaviour has already been suggested by the simulations that were discussed in Section 5.2.4.

The plot of optimum injected signal amplitude in Figure 5.18 is also interesting. Firstly, it shows that the level of the optimum injected signal is approximately 25dB below that of the RF input at any given power level, allowing the impact of linearisation on efficiency to be neglected. In addition to this, a 2:1 slope has been added to the plot to illustrate that, below compression, the relationship between carrier and injected-signal power predicted in Section 5.2.3 is correct. Beyond the onset of gain compression it can be seen that this linear relationship is no longer valid, but this was expected as the behaviour in this regime is governed by higher-order nonlinear effects that were not taken into account in the simulations presented earlier.

5.3.3 Measured Amplitude Tolerance

Unfortunately, the tolerance of the distortion reduction to phase deviations as simulated in Section 5.2.4 could not be measured, as this would require an idealised phase-shifter with a flat phase-response from DC to the bandwidth of the linearising signal. However, it was possible to evaluate the amplitude tolerance by optimising the performance at a static power level, then sweeping the power of the injected signal whilst recording the variation in upper and lower ACPR. The input signal used in these measurements was another 200kHz QPSK-modulated carrier consisting of 256 symbols centred at 1.85GHz, with the amplifier operating at an output power level of 25dBm with a PAE of 10.9%. Figure 5.20 below shows the result of this test, with ACPR plotted against injection signal amplitude offset from its optimised value of –13.5dBm.

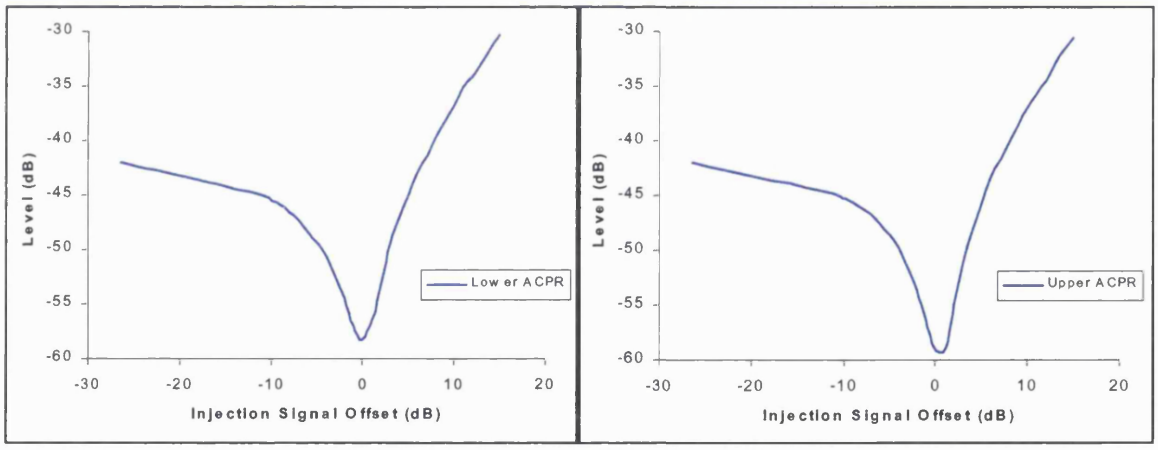


Figure 5.20: Measured ACPR reduction against linearising signal amplitude offset

Comparing Figure 5.20 with the simulated results shown in Figure 5.9, it can be seen that the overall shapes of the simulated and measured characteristics are similar, although the range over which the injected signal was swept was limited in practise by the minimum output power of the HP33120A Arbitrary Waveform Generator. The amount of ACPR reduction that occurred at the optimum injection power level was found to be greater in practice than in simulation, but this is to be expected as the nonlinear model and practical device were

unrelated. Despite this, both Figure 5.20 and Figure 5.9 show that a few dBs of distortion improvement still occurs in both cases at injection signal power levels over 20dB below the optimum. At power levels above the optimum point, both measured and simulated characteristics show rapid performance degradation, with ACPR reduction falling to zero in both cases beyond approximately +5dB. This suggests that the mechanisms dominating the distortion improvement are modelled adequately by only a third-order power series approximation.

5.3.4 EDGE Modulation and EVM

It was also possible to evaluate the application of the linearising signal to other modulation schemes, including EDGE (Enhanced Data Rates for Global Evolution). EDGE has a peak-to-average ratio of approximately 3.2dB and therefore suffers AM-AM and AM-PM distortion when amplified. The most commonly-employed measure of EDGE distortion is given by 'Error Vector Magnitude' (EVM) , a narrowband measurement that quantifies the deviation of the received, demodulated signal from its ideal path. As such, EVM provides a more accurate indication of amplifier distortion than ACPR as it quantifies the ease of signal recovery. Although it is possible to demodulate a transmitted EDGE carrier having an EVM of up to 7%, this allowance is usually distributed throughout a transmitter chain, resulting in a typical PA EVM requirement of approximately 2%.

A single-carrier EDGE test-signal was produced using Matlab code written separately within Nokia, and the VisualBasic software was adapted to allow for the upload of this waveform to the ESG4433B. The i and q waveforms were then processed in the manner described in Section 5.3.1 to produce the linearising signal which was transferred to the HP33120A arbitrary waveform generator.

The modulating waveform was chosen to be 256 symbols long with an oversampling ratio of 16, giving 4096 sample points for both i and q, as well as the linearising signal. Again, it was critical that the waveform cycled continuously without abrupt discontinuities, so this was ensured by repeating the same 256 symbols three times, convolving these longer pulse-trains with an EDGE transmit-filter impulse-response, and retaining only the middle set of waveform points. The resulting signal vector and its measured spectrum after being transferred to the ESG334B are shown below in Figure 5.21.

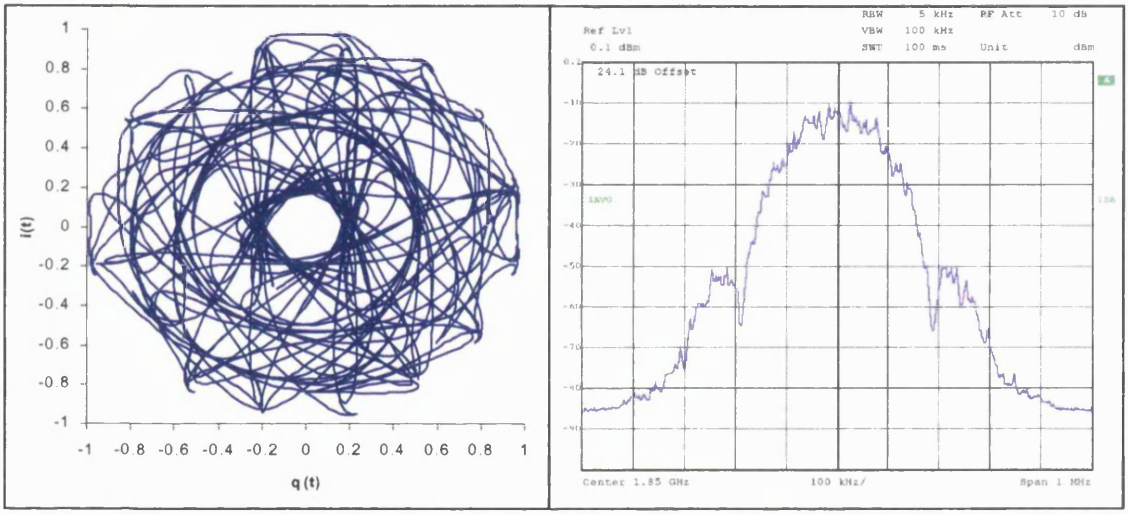


Figure 5.21: Edge signal vector and measured spectrum

With careful time- and amplitude alignment of the injected signal, it was again possible to observe distortion improvement, now quantified as a reduction in EVM. By sweeping the input power and manually re-optimising the amplitude of the injection signal, the results shown in Figure 5.22 were obtained. Referring to the plots, it can be seen that EVM is reduced across the entire range of the measurement sweep, typically by 2%, though this improvement was found to increase by up to 8% at the highest power levels. It should be noted, however, that the accuracy of EVM measurements becomes doubtful above approximately 15%.

Again, it is the increase in amplifier Power-Added Efficiency that is of most interest. A 2% EVM specification has been highlighted on Figure 5.22, giving an efficiency for the unlinearised PA of 12.2%. With an optimised injection signal, however, the amplifier output power can be increased by 6dB before the EVM requirement is exceeded, at which point the PAE was found to be 29%, giving an increase in efficiency of 16.8%.

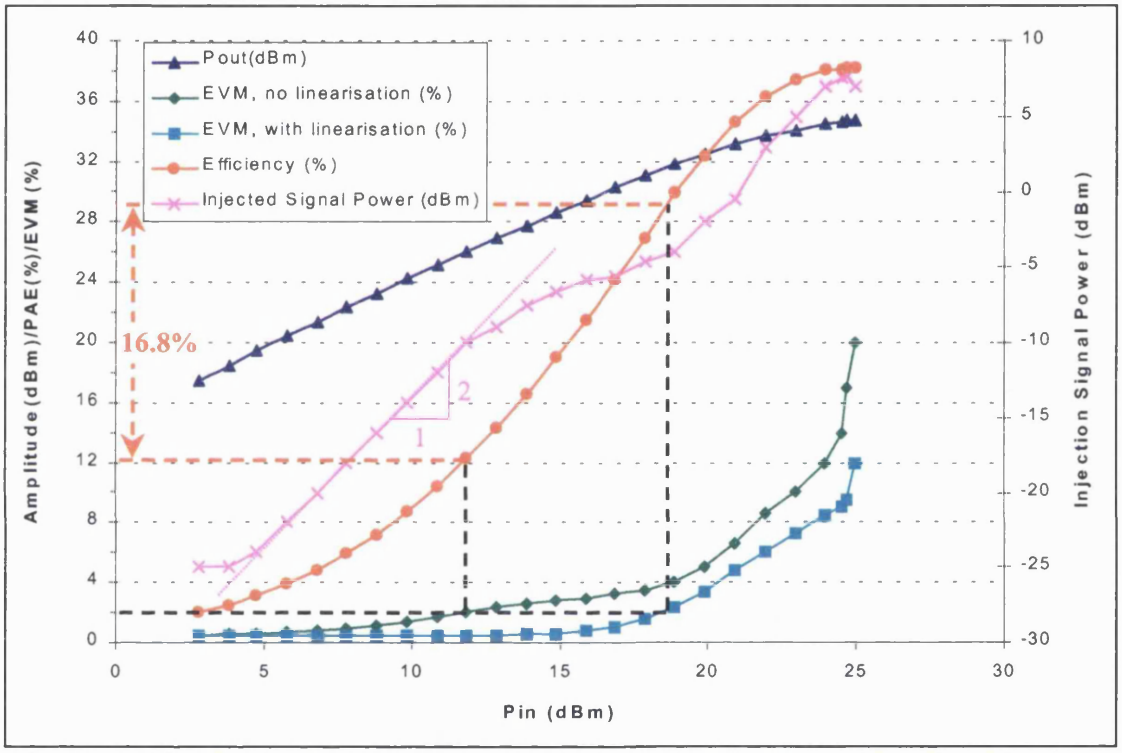


Figure 5.22: Measured EVM against input power, with and without linearisation

The amplitude of the optimised injection signal is also shown plotted in Figure 5.22, along with a line having a 2:1 gradient that shows that, as in Section 5.3.2, this relationship holds below compression. As in the earlier investigations with a QPSK input signal (see Figure 5.19), the gradient of the plot of injected signal power decreases as the amplifier begins to compress. The break point from the linear 2:1 relationship shown occurs at approximately 0.5dB of gain compression, with the gradient falling off briefly before increasing again to approximate a 2:1 relationship at higher power levels.

These findings show conclusively that second-order bias-injection reduces distortion occurring *within* the bandwidth of the carrier, correcting the trajectory of the signal vector. This enables accurate data recovery at higher levels of output power, and hence can increase amplifier efficiency.

5.4 Analysis of Multi-Carrier Linearisation with Digital Modulation

Again, it is convenient to begin this analysis in the frequency domain, with an input signal consisting of two carriers with amplitudes a and b and baseband spectrums $B_1(j\omega)$ and $B_2(j\omega)$, centred at carrier frequencies ω_1 and ω_2, respectively, where $\omega_2>\omega_1$ and $(\omega_2-\omega_1)<<\omega_1, \omega_2$:

$$x(t) \Rightarrow X(j\omega)=aB_1(j\omega)\otimes 1/2[\delta(\omega+\omega_1)+\delta(\omega-\omega_1)]+bB_2(j\omega)\otimes 1/2[\delta(\omega+\omega_2)+\delta(\omega-\omega_2)] \quad (5.10)$$

The products produced by the third-order nonlinearity can be found by cubing the above input signal and expanding, as follows:

$$\begin{aligned} g_3x^3(t) \Rightarrow\ & g_3X(j\omega)\otimes X(j\omega)\otimes X(j\omega) \\ & = 1/8g_3a^3B_1(j\omega)\otimes B_1(j\omega)\otimes B_1(j\omega)\otimes[\delta(\omega\pm 3\omega_1)+3\delta(\omega\pm\omega_1)] \\ & +1/8g_3b^3B_2(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm 3\omega_2)+3\delta(\omega\pm\omega_2)] \\ & +3/8g_3a^2bB_1(j\omega)\otimes B_1(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm 2\omega_1\pm\omega_2)+2\delta(\omega\pm\omega_1)] \\ & +3/8g_3ab^2B_1(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm 2\omega_2\pm\omega_1)+2\delta(\omega\pm\omega_1)] \end{aligned} \quad (5.11)$$

For clarity, the six in-band distortion components are now written separately:

$$\begin{aligned} & 3/8\,g_3a^3B_1(j\omega)\otimes B_1(j\omega)\otimes B_1(j\omega)\otimes[\delta(\omega\pm\omega_1)] && 1,1,1 @ \omega_1 \\ & 3/8\,g_3b^3B_2(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm\omega_2)] && 2,2,2 @ \omega_2 \\ & 3/4\,g_3ab^2B_1(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm\omega_1)] && 1,2,2 @ \omega_1 \\ & 3/4\,g_3a^2bB_1(j\omega)\otimes B_1(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm\omega_2)] && 1,1,2 @ \omega_2 \\ & 3/8\,g_3a^2bB_1(j\omega)\otimes B_1(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm(2\omega_1-\omega_2))] && 1,1,2 @ 2\omega_1\text{-}\omega_2 \\ & 3/8\,g_3ab^2B_1(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm(2\omega_2-\omega_1))] && 1,2,2 @ 2\omega_2\text{-}\omega_1 \end{aligned} \quad (5.12)$$

The notation after each component in Equation 5.12 shows its composition in terms of the two baseband spectra and the centre frequency; for example, 1,2,2@ω_1 denotes the convolution of $B_1(j\omega)$, $B_2(j\omega)$ and $B_2(j\omega)$, centred at ω_1. Referring to the six distortion products individually, the first two represent the spectral regrowth around the individual carriers as discussed earlier in the chapter. The second two, 1,2,2@ω_1 and 1,1,2@ω_2, are interesting in that they also appear on and around the carriers, but are composed of different combinations of the baseband spectrums. These terms therefore represent a new component of spectral regrowth arising from the interaction between the two carriers. If it is assumed that $B_1(j\omega)$ and $B_2(j\omega)$ are uncorrelated, the two sets of regrowth around each carrier will not sum simply. However, it is likely that the combined ACPR around the two carriers will be higher than that around each of the carriers if they were amplified individually. The final two terms of Equation 5.12 show the familiar intermodulation products, with $B_1(j\omega)\otimes B_1(j\omega)\otimes B_2(j\omega)$ appearing at $2\omega_1$-ω_2 and $B_1(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)$ at $2\omega_2$-ω_1.

All six of these in-band components must therefore be produced by the addition of the low-frequency second-order injection signal, which in this case may be written as:

$$I(j\omega)=\alpha e^{-j\theta}B_1(j\omega)\otimes B_1(j\omega)+\beta e^{-j\theta}B_2(j\omega)\otimes B_2(j\omega)+\gamma e^{-j\theta}B_1(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm(\omega_2-\omega_1))] \quad (5.13)$$

If this is added to the original input and amplified, the second-order amplifier nonlinearity $g_2x^2(t)$ produces new products, with the following in-band components being of relevance:

$$\begin{aligned}
1/2\,g_2a\alpha e^{-j\theta}B_1(j\omega)\otimes B_1(j\omega)\otimes B_1(j\omega)\otimes[\delta(\omega\pm\omega_1)] &\quad 1,1,1 @ \omega_1\\
1/2\,g_2b\beta e^{-j\theta}B_2(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm\omega_2)] &\quad 2,2,2 @ \omega_2\\
(1/2a\beta+1/4b\gamma)g_2e^{-j\theta}B_1(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm\omega_1)] &\quad 1,2,2 @ \omega_1\\
(1/2b\alpha+1/4a\gamma)g_2e^{-j\theta}B_1(j\omega)\otimes B_1(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm\omega_2)] &\quad 1,1,2 @ \omega_2\\
1/4\,g_2a\gamma e^{-j\theta}B_1(j\omega)\otimes B_1(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm(2\omega_1-\omega_2))] &\quad 1,1,2 @ 2\omega_1\text{-}\omega_2\\
1/4\,g_2b\gamma e^{-j\theta}B_1(j\omega)\otimes B_2(j\omega)\otimes B_2(j\omega)\otimes[\delta(\omega\pm(2\omega_2-\omega_1))] &\quad 1,2,2 @ 2\omega_2\text{-}\omega_1 \quad (5.14)
\end{aligned}$$

In order for the technique to be applied successfully, it must be possible to match the amplitudes of the new products shown in Equation 5.14 with those of the third-order distortion shown in 5.12, with fixed values for α, β and γ.

$$1,1,1@\omega_1:\ 3/8\,g_3a^3 = 1/2\,g_2a\alpha \Rightarrow \alpha = \frac{3g_3a^2}{4g_2} \tag{5.15}$$

$$2,2,2@\omega_2:\ 3/8\,g_3b^3 = 1/2\,g_2b\beta \Rightarrow \beta = \frac{3g_3b^2}{4g_2} \tag{5.16}$$

$$1,1,2@2\omega_1\text{-}\omega_2:\ 3/8\,g_3a^2b = 1/4\,g_2a\gamma \Rightarrow \gamma = \frac{3g_3ab}{2g_2} \tag{5.17}$$

$$1,2,2@2\omega_2\text{-}\omega_1:\ 3/8\,g_3ab^2 = 1/4\,g_2b\gamma \Rightarrow \gamma = \frac{3g_3ab}{2g_2} \tag{5.18}$$

The above expressions give the values of α, β and γ, showing them to be fixed for a particular pair of nonlinear coefficients g_2 and g_3. These may now be used to equate the amplitudes of the remaining components appearing at ω_1 and ω_2 to complete the proof:

$$1,2,2@\omega_1:\ (1/2a\beta + 1/4b\gamma)g_2 \equiv 3/4\,g_3ab^2 \Rightarrow \left(\frac{3g_3ab^2}{8} + \frac{3g_3ab^2}{8}\right) \equiv 3/4\,g_3ab^2 \tag{5.19}$$

$$1,1,2@\omega_2:\ (1/2b\alpha + 1/4a\gamma)g_2 \equiv 3/4\,g_3a^2b \Rightarrow \left(\frac{3g_3a^2b}{8} + \frac{3g_3a^2b}{8}\right) \equiv 3/4\,g_3ab^2 \tag{5.20}$$

Equations 5.15-5.20 show that the technique may also be applied to multiple modulated-carrier input signals, irrespective of the modulation scheme employed. Although this analysis only proves the case for a two-carrier input signal, it may be assumed from the multicarrier analysis of Chapter 4 that this would produce a similar result with any number of input signals.

The application of this technique to a 'real' multi-carrier signal is, as before, best examined in the time-domain. We will begin by considering an input signal consisting of two digitally-modulated carriers, which may be written as follows:

$$v_{in}(t) = i_1(t)\cos(\omega_1 t) - q_1(t)\sin(\omega_1 t) + i_2(t)\cos(\omega_2 t) - q_2(t)\sin(\omega_2 t) \tag{5.21}$$

Where $i_n(t)$ and $q_n(t)$ represent the in-phase and quadrature components, respectively, of the carrier denoted by the subscript n. By squaring this input signal, the required low-frequency second-order components may be found:

$$\begin{aligned} v_{in}^2(t) = {} & \frac{i_1^2(t)}{2}(1+\cos(2\omega_1 t)) + \frac{q_1^2(t)}{2}(1-\sin(2\omega_1 t)) + \frac{i_2^2(t)}{2}(1+\cos(2\omega_2 t)) + \frac{q_2^2(t)}{2}(1-\sin(2\omega_2 t)) \\ & - 2i_1(t)q_1(t)\cos(\omega_1 t)\sin(\omega_1 t) + 2i_1(t)i_2(t)\cos(\omega_1 t)\cos(\omega_2 t) - 2i_1(t)q_2(t)\cos(\omega_1 t)\sin(\omega_2 t) \\ & - 2q_1(t)i_2(t)\sin(\omega_1 t)\cos(\omega_2 t) + 2q_1(t)q_2(t)\sin(\omega_1 t)\sin(\omega_2 t) - 2i_2(t)q_2(t)\cos(\omega_2 t)\sin(\omega_2 t) \end{aligned} \tag{5.22}$$

Simplifying and retaining only the low-frequency components gives the required linearising signal as follows:

$$v_{lin}(t) = \underbrace{\frac{i_1^2(t)}{2} + \frac{q_1^2(t)}{2} + \frac{i_2^2(t)}{2} + \frac{q_2^2(t)}{2}}_{\text{Baseband components}} + \underbrace{i_1(t)i_2(t)\cos(\omega_2 - \omega_1) + q_1(t)q_2(t)\cos(\omega_2 - \omega_1) + q_1(t)i_2(t)\sin(\omega_2 - \omega_1) - i_1(t)q_2(t)\sin(\omega_2 - \omega_1)}_{\text{IF components}} \tag{5.23}$$

Equation 5.23 is shown schematically overleaf, where it can be seen that the linearising signal now consists of two separate components, one referred to as 'baseband' (centred at $\omega = 0$), the other denoted 'IF' (centred at the difference frequency, $\omega_2 - \omega_1$).

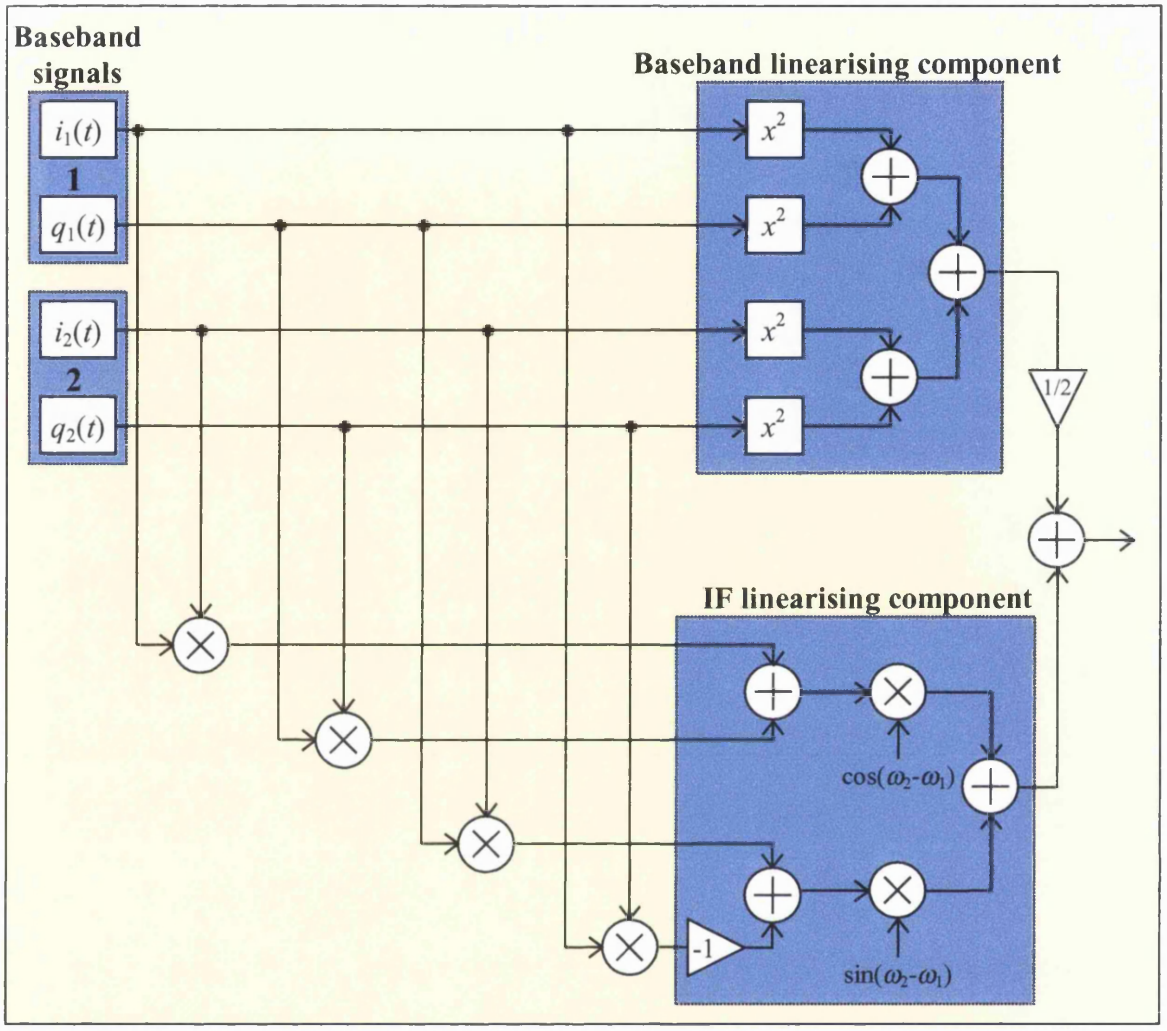


Figure 5.23: Schematic of linearising signal generation for two modulated carriers

5.5 Simulation of Multi-Carrier Linearisation with Digital Modulation

5.5.1 Idealised DSP/RF Cosimulation

The theory concerning the linearisation with multiple modulated-carriers shown in the previous section was used to develop a two-carrier simulation, again using the DSP/RF Co-Simulation capabilities of *ADS*. Four bit-streams (two for each carrier) were generated and then each was split into two paths. The first path carried the filtered bit-streams straight to the RF subcircuit for quadrature upconversion, with one carrier at 0.5MHz and the other at 0.6MHz with a bit-rate of 25kHz (the frequencies were kept low to reduce simulation times).

The second path carried the bit-streams to an *ADS* implementation of the network shown in Figure 5.23, using ideal multipliers and adders to generate the composite linearising signal which was then passed to the RF subcircuit. Finally, the upconverted carriers and linearising components were summed before being applied to the nonlinear amplifier model.

The required level for the IF linearising component was first determined by optimising for minimum intermodulation distortion. The left-hand trace of Figure 5.24 shows the output spectrum without linearisation, while the right-hand shows the output with only the IF linearising component added to the input signal. The reduction in the distortion is clear, with the components appearing at the IMD3 frequencies being reduced by over 25dB. The spectral regrowth around both carriers was also reduced, though only by 2-3dB.

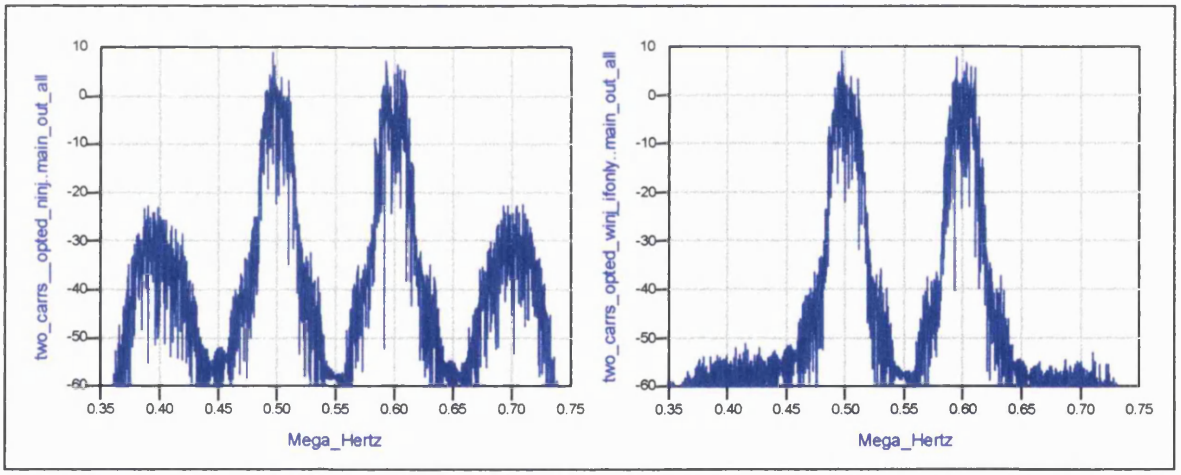


Figure 5.24: Output spectrum plots showing IMD3 reduction due to linearising signal

The baseband component was then added to the linearising signal, and its level optimised to give maximum reduction in the upper and lower ACPR around both carriers, producing the results shown in Figure 5.25 overleaf. After optimisation, it was found that the amplitude of the baseband component was typically 10dB less than that of the IF portion. This does not agree with the analysis of the preceding section, which suggests that the relative amplitude of the baseband component should be half that of the IF portion (see Equation 5.23). The

reasons for the discrepancy are unclear, though this will be further investigated in the experimental investigation in the next section.

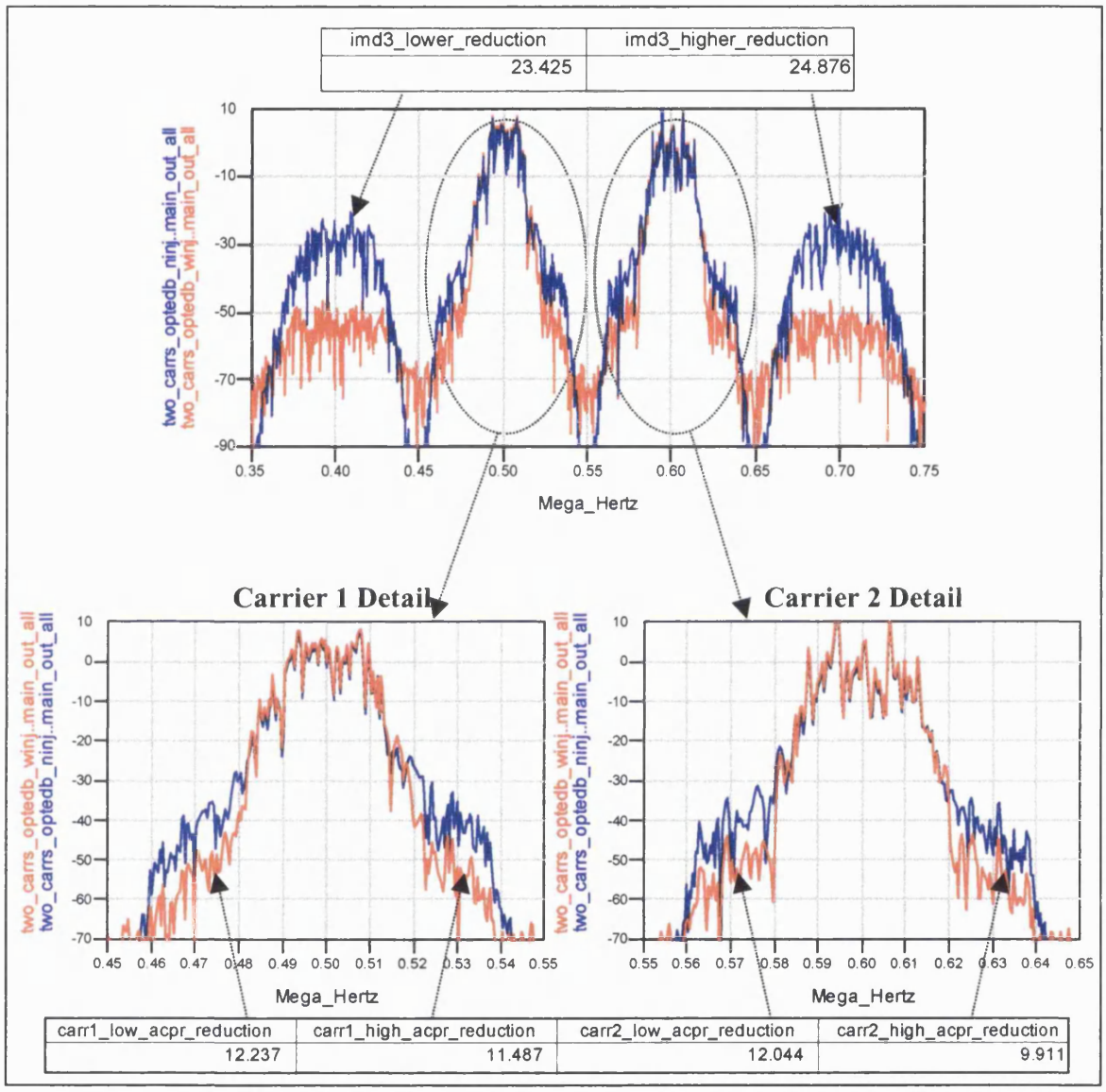


Figure 5.25: Simulated output spectrums showing distortion reduction due to linearising signal

These results show that the technique can also be applied to multiple modulated carriers, and that it can reduce both intermodulation and adjacent channel power simultaneously. This is another important finding, and verifies the theory presented in the preceding section. The level of reduction in the intermodulation products was found to be in excess of 20dB, while the ACPR was reduced by 10-12dB.

5.6 Multiple Modulated Carrier Proof-of-Concept

5.6.1 Test-Bench and Implementation Issues

In order to verify the analysis and simulations presented in the previous sections, and to complete the practical development of the linearisation technique, the single-carrier test-bench shown in Figure 5.11 was modified by the addition of another signal generator and driver amplifier, along with a hybrid combiner and isolators as shown in the figure below.

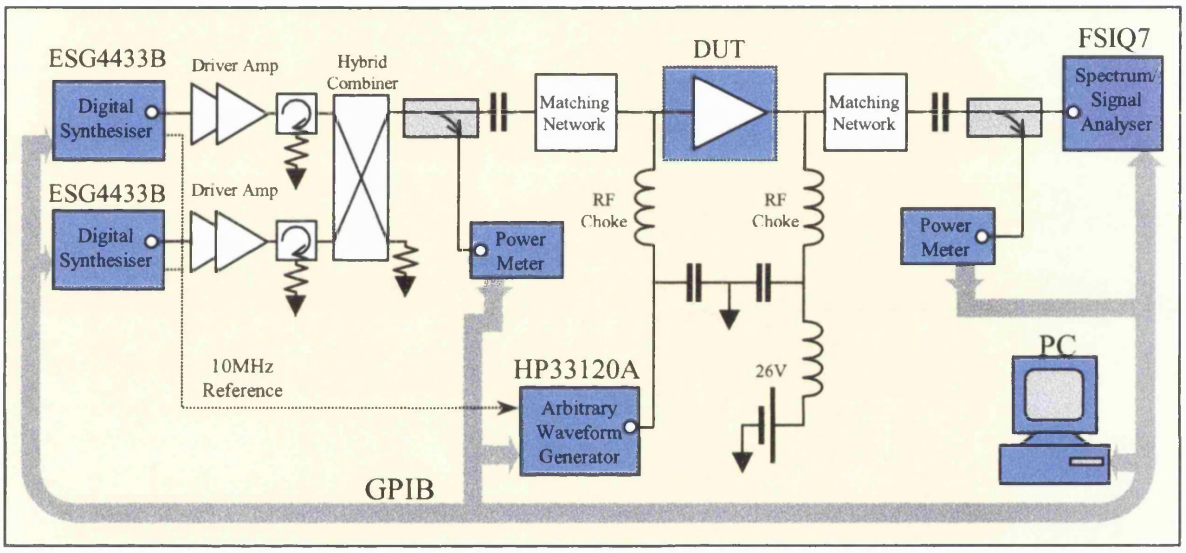


Figure 5.26: New two-carrier test bench

The VisualBasic code was modified to incorporate the generation of a two-carrier linearising signal as defined in Equation 5.2.3, consisting of both baseband and IF components (see Figure 5.23). This process was found to be considerably more complex than the single carrier case, as the linearising signal is now required to have a bandwidth that is greater than that of the individual carriers. The sampling rates used for the individual baseband waveforms were therefore insufficiently high to accurately reproduce the composite linearising waveform. To overcome this problem, software was written to perform oversampling of the four baseband waveforms during the generation of the linearising signal. The relevant code along with a more detailed description of the process can be found in Appendix C, 4.

Another issue arising from the use of a second signal generator was the synchronisation of the two modulating data streams. Although all the signal sources were connected to a common 10MHz oscillator, the absolute time reference of the arbitrary waveform generators within the two ESG3344Bs was not automatically synchronised. The linearising signal can only be formed if this time offset is known accurately, or preferably, if the offset is removed altogether by ensuring that the first and last points of the two sets of i and q data are correctly aligned. In a practical multicarrier transmitter, this would not be a problem as both signals would be generated in real-time and would thus be inherently synchronised.

Fortunately, the ESG4433B signal generators allow 'markers' to be placed at individual points on stored waveforms, and these can be used to send a trigger to a BNC connector on the rear panel of the unit. By connecting this output to the 'pattern trigger' input on the second signal generator and setting markers on the first few points of the first waveform, the two arbitrary waveform generators within the sources were automatically synchronised as required.

5.6.2 Linearisation with IF Injection Signal

Two separate 64-symbol 1MHz-wide carriers at 1.85GHz and 1.8525GHz (giving a spacing of 2.5MHz) were generated and downloaded to the ESG4433B signal generators as previously described. The carrier difference-frequency was then used to generate a linearising signal (as described in detail in Appendix C, 4), at first consisting of only the IF linearising components (see Section 5.5.1). This was downloaded to the HP33120A and applied to the DUT along with the carriers. After time-alignment of the signals, distortion improvement was again observed, with Figure 5.27 and Figure 5.28 overleaf showing typical results.

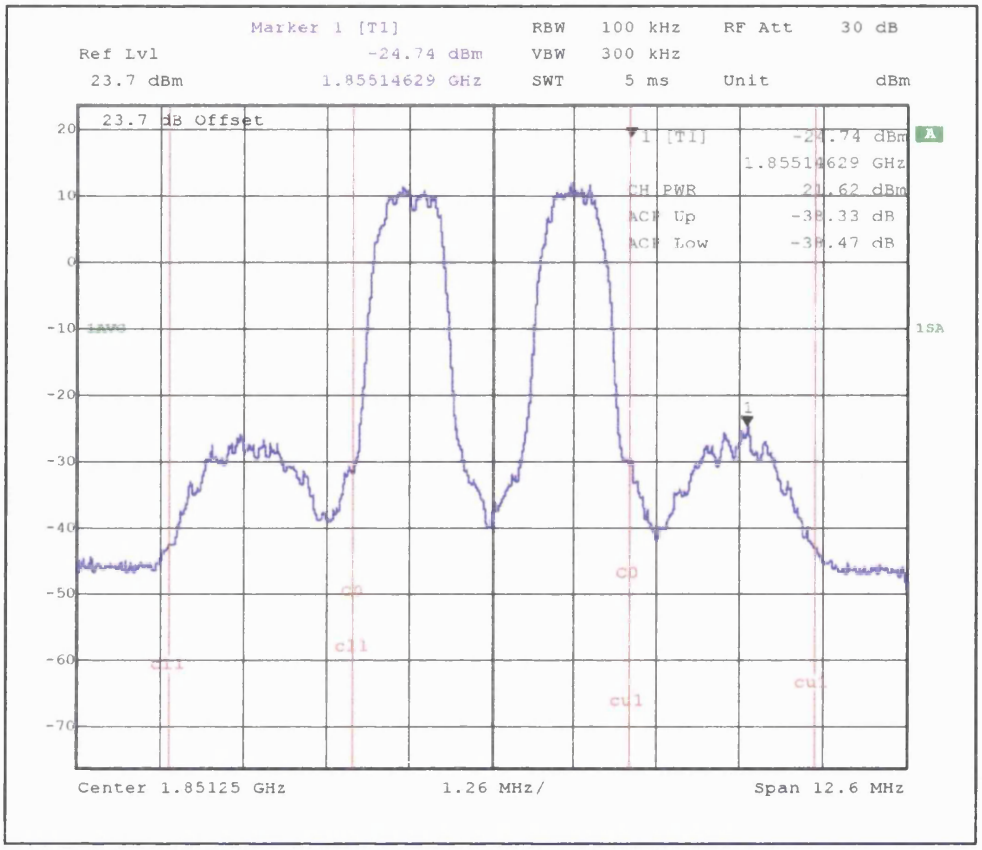


Figure 5.27: Output spectrum with two 1MHz-wide carriers, without linearisation

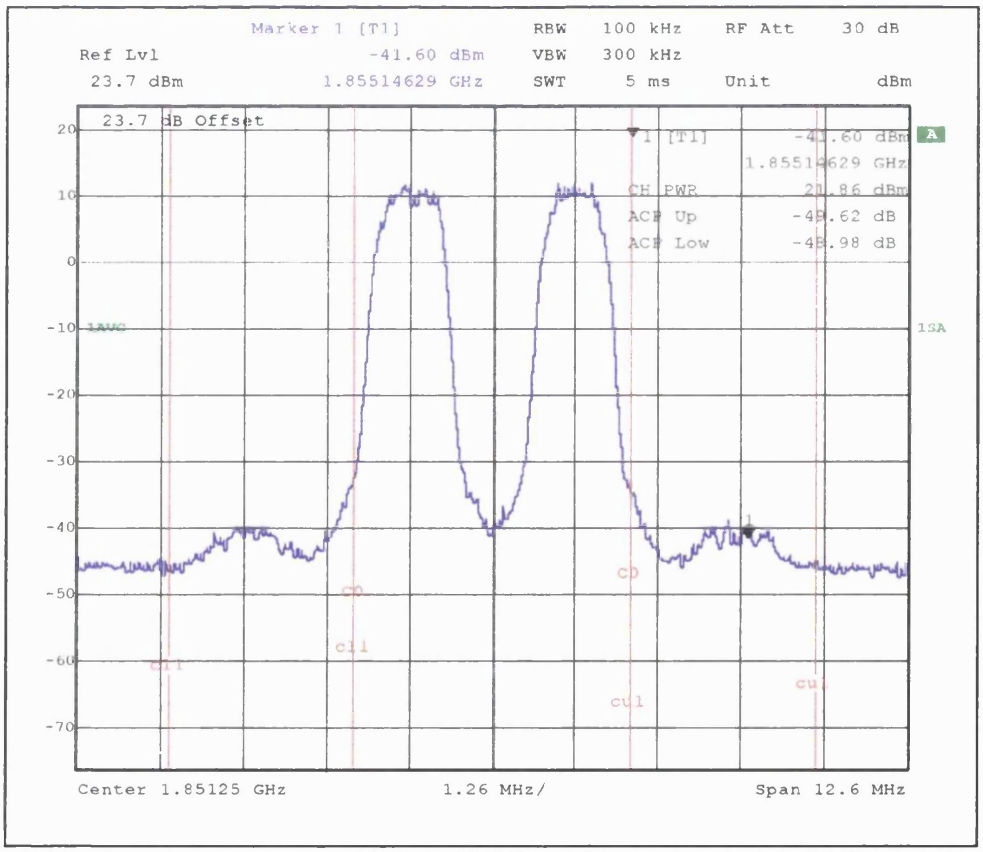


Figure 5.28: Output spectrum with two 1MHz-wide carriers, with IF linearising signal

Referring to the figures, it can be seen that, as in the results of the simulation shown in Figure 5.24, the addition of the IF linearising signal reduced the distortion appearing at the IMD3 frequencies. From the ACPR measurements shown on Figure 5.27 and Figure 5.28 it may be seen that the average reduction over both alternate-channel bandwidths was approximately 11dB, while the marker placed in the upper alternate channel shows that the peak level of distortion was reduced from –24.5dBm to –41.5dBm. Small improvements (~3-5dB) were also observed in the close-in ACPR around the carriers, which agrees well with the simulations of the same scenario.

5.6.3 Linearisation with IF and Baseband Composite Injection Signal

Adding the baseband components to the linearising signal did not prove to be as straightforward as suggested by the earlier two-carrier analysis and simulations. In order to optimise the distortion reduction, it was necessary to have independent control of the amplitudes of the baseband and IF linearising components. To enable this, a second HP33120A arbitrary waveform generator was added to the set-up shown in Figure 5.26. This allowed the amplitudes of the two signals to be independently tuned without having to generate and download a new composite waveform after each iteration.

Typical results after manual optimisation of the injection signal amplitudes are shown on the following pages; Figure 5.29 and Figure 5.30 show the ACPR around the individual carriers without linearisation; Figure 5.31 and Figure 5.32 show the same carriers after the introduction of optimally-tuned IF and baseband linearising signals; Figure 5.33 shows the output spectrum after linearisation over a wider frequency span.

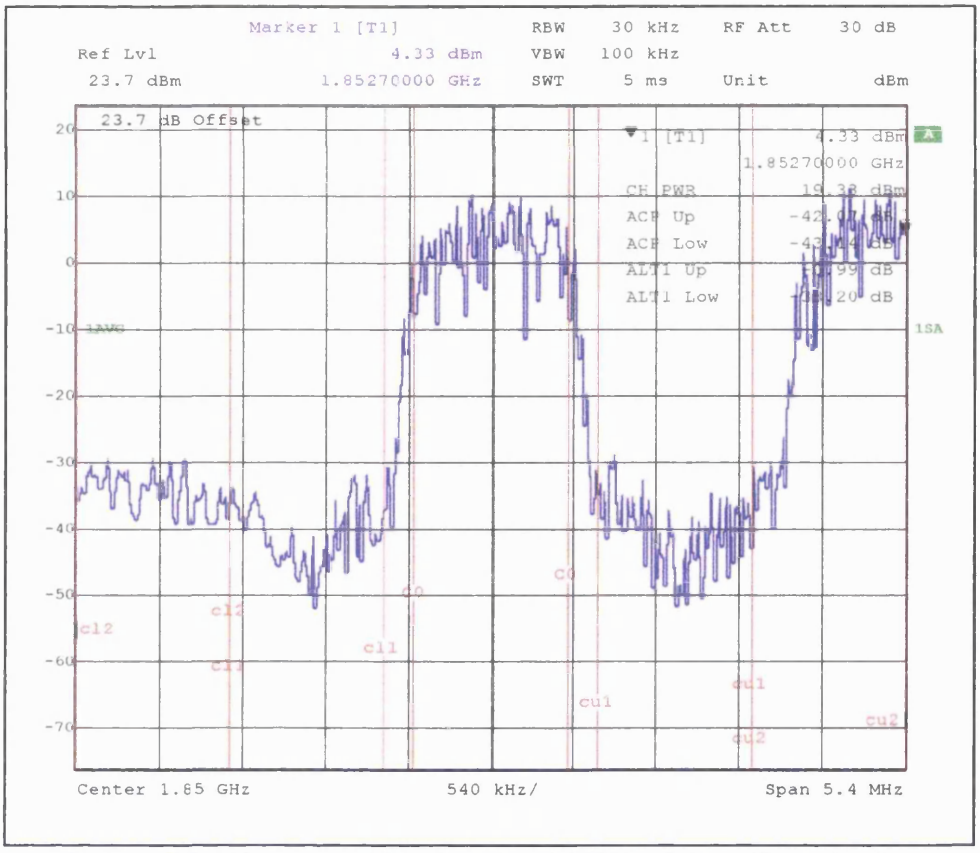


Figure 5.29: Lower carrier showing ACPR, without linearisation

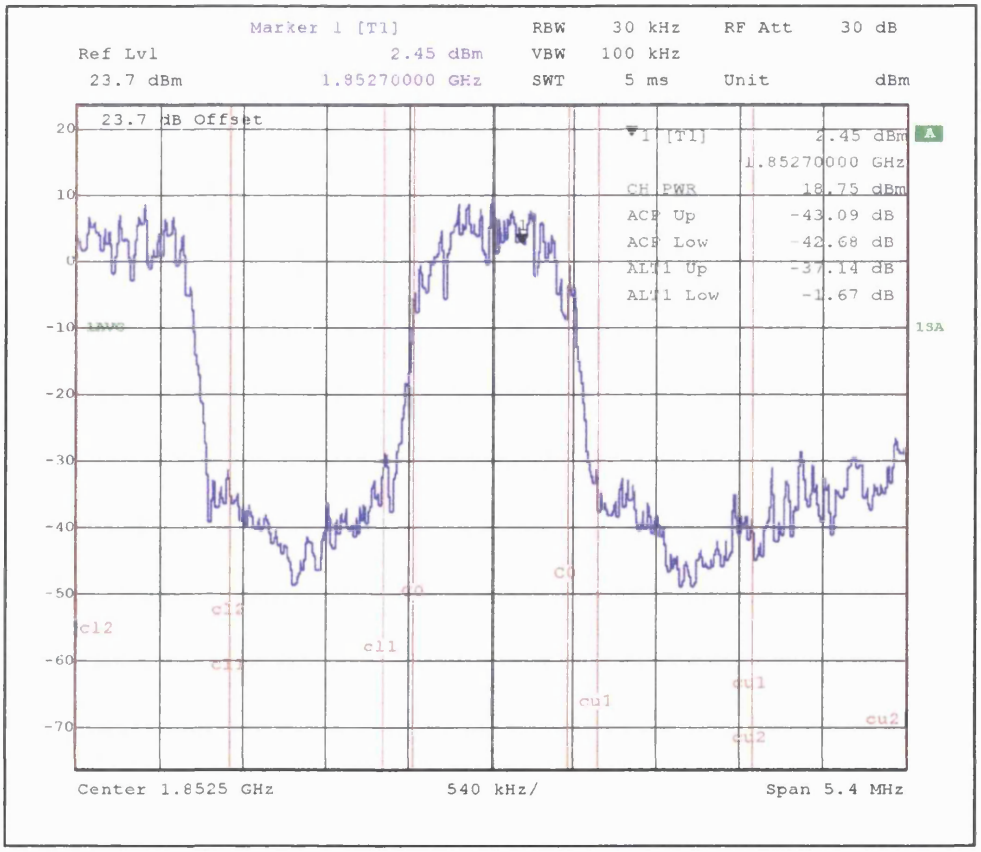


Figure 5.30: Upper carrier showing ACPR, without linearisation

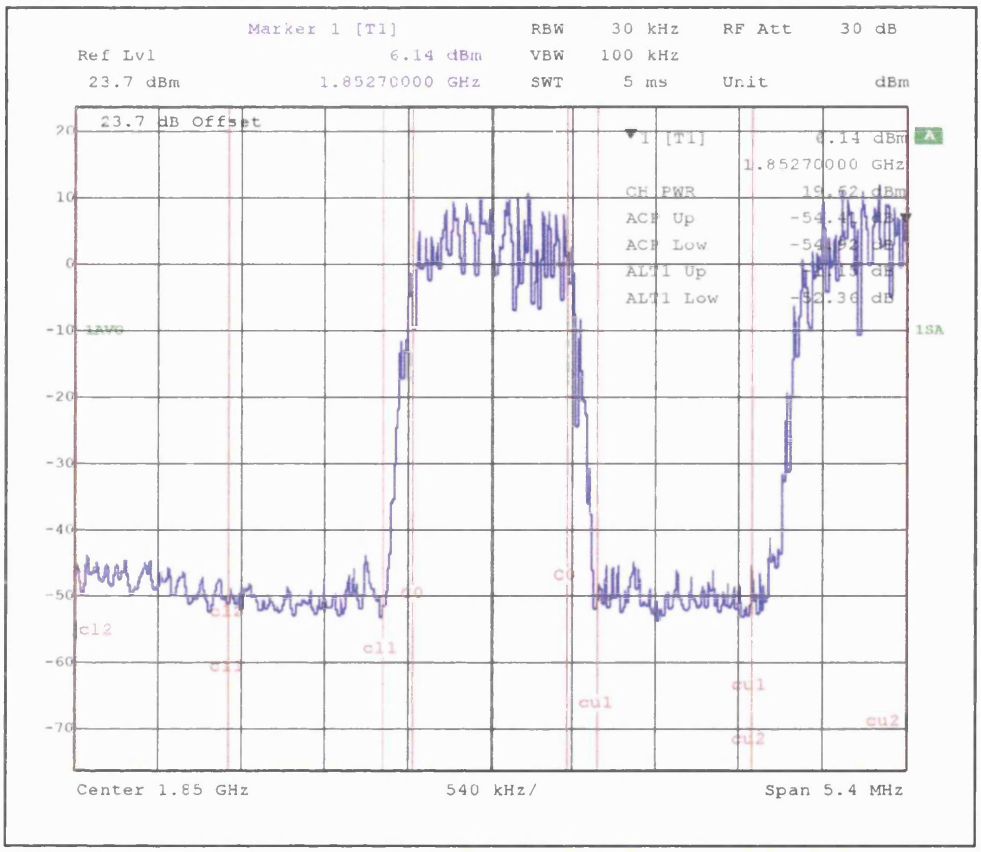


Figure 5.31: Lower carrier showing reduced ACPR, with linearisation

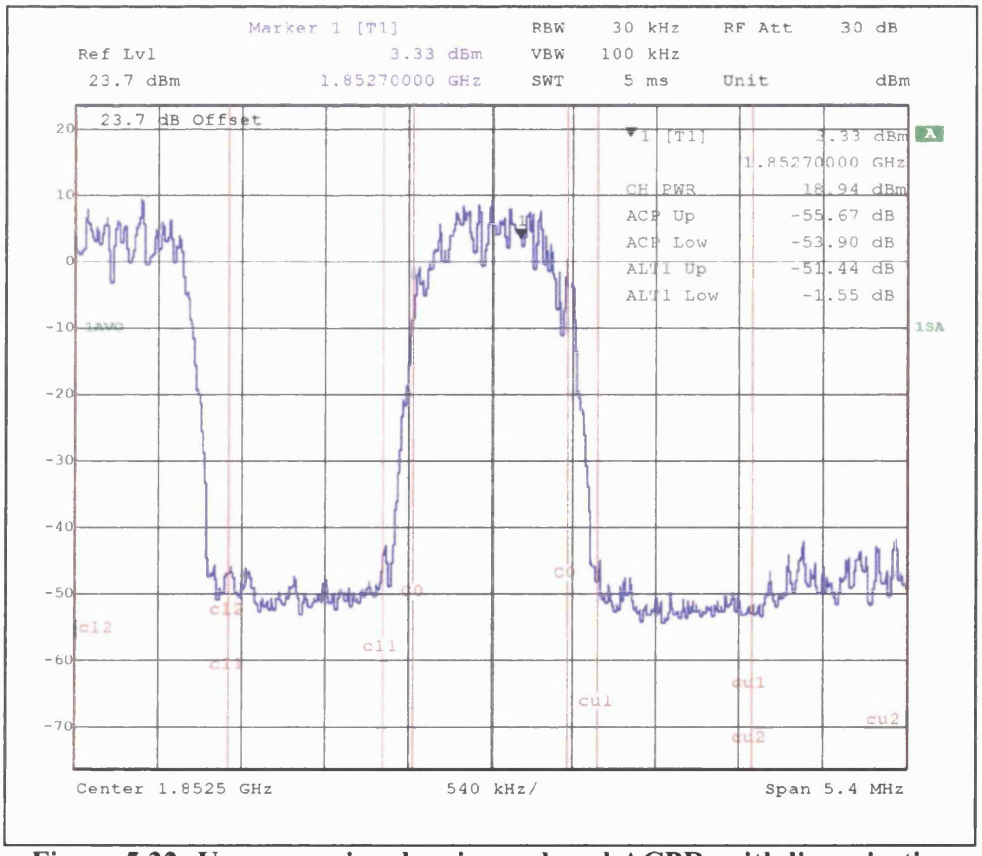


Figure 5.32: Upper carrier showing reduced ACPR, with linearisation

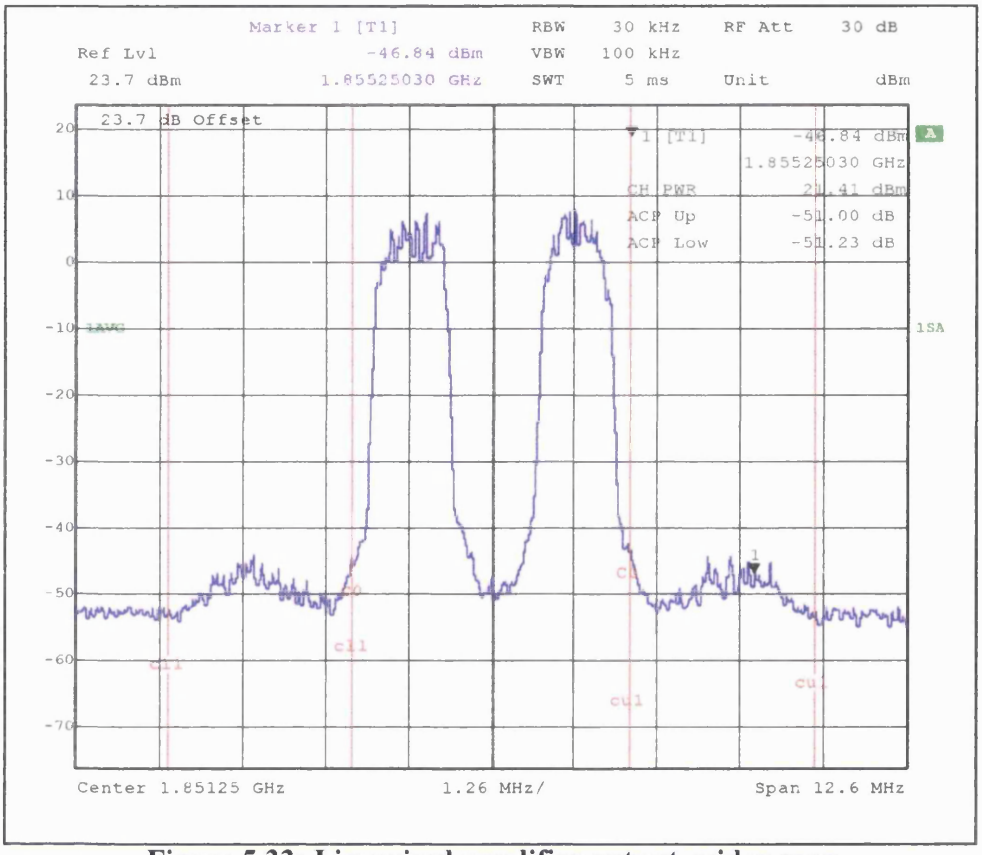


Figure 5.33: Linearised amplifier output, wider span

The results show that the addition of the baseband linearising component improved the ACPR around both carriers by 12-13dB, while the ALCPR on either side of the two carriers was reduced by approximately 14dB. It was found that the optimum amplitude for the IF component was almost exactly 10dB greater than that of the baseband component, which agrees well with the simulation of the same scenario (see Figure 5.25, Page 142) but not the analysis of Section 5.4, which gave a relative amplitude of 0.5 or 3dB. This discrepancy is a trivial problem, however, as in a digital implementation the required offset can easily be set and adjusted as required.

The power-added efficiency of the amplifier at this power level was 8.1%, somewhat smaller than would be acceptable for most practical applications. The figures on the following pages show results with the same two carriers at a higher power level, where the PAE was 17%.

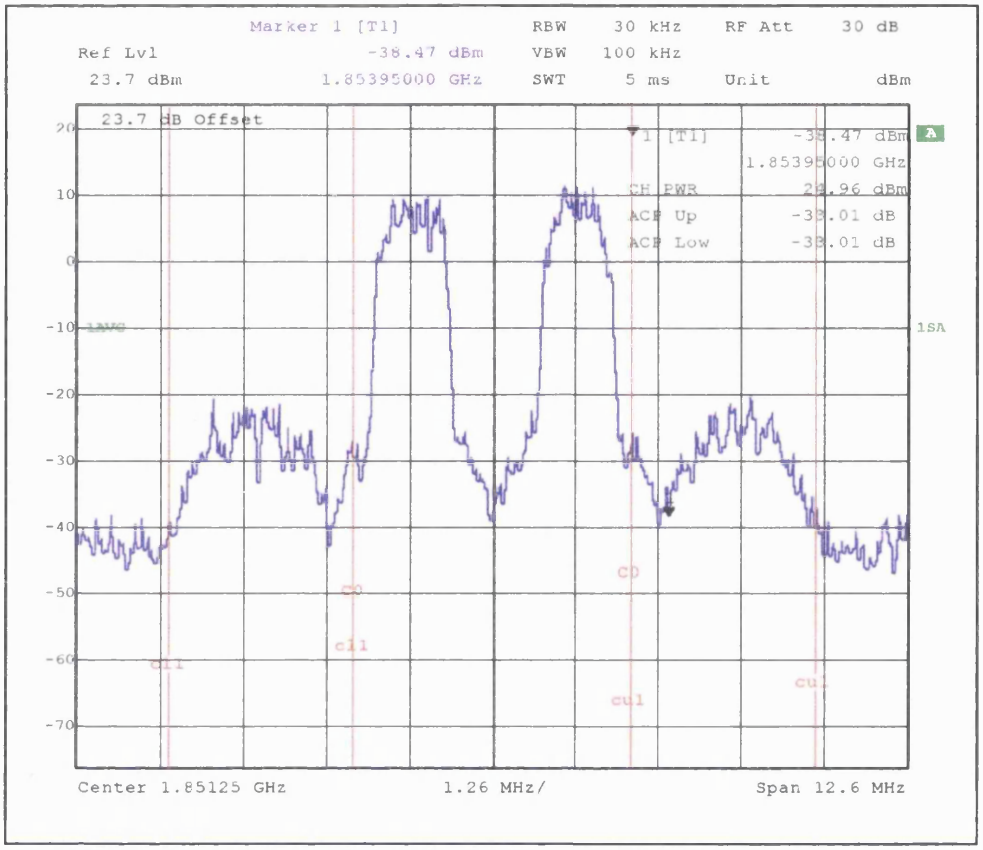


Figure 5.34: Unlinearised amplifier output, higher input power

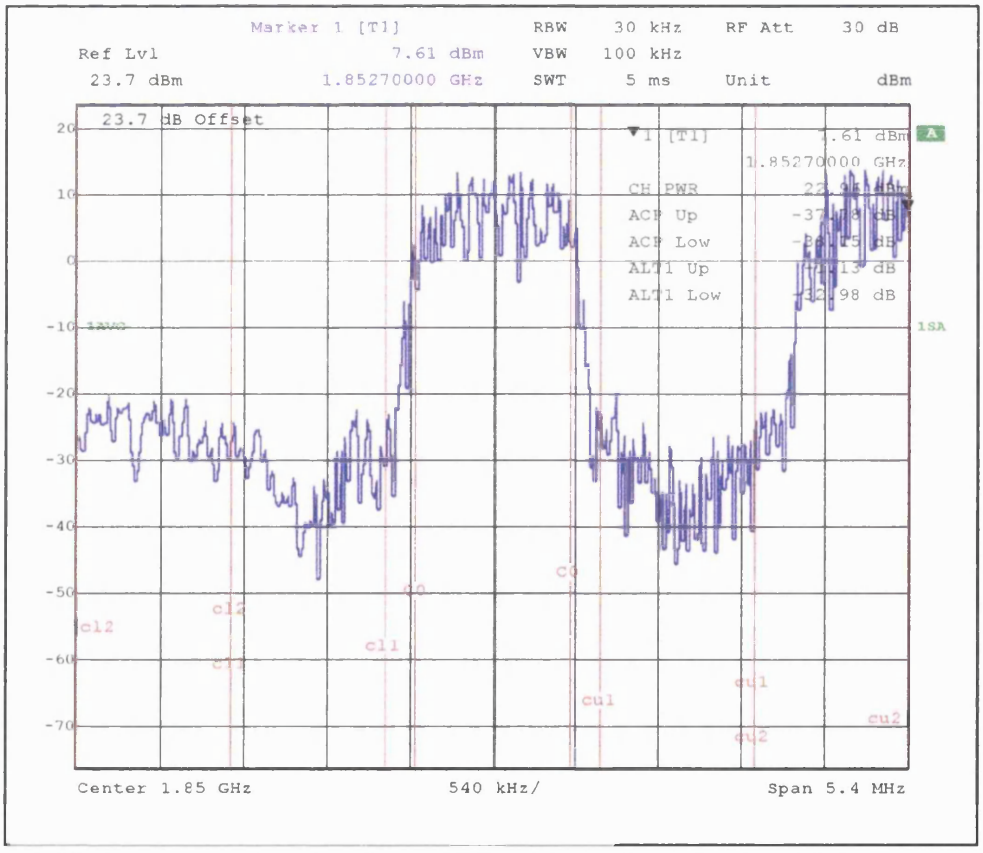


Figure 5.35: Lower carrier showing ACPR without linearisation, higher input power

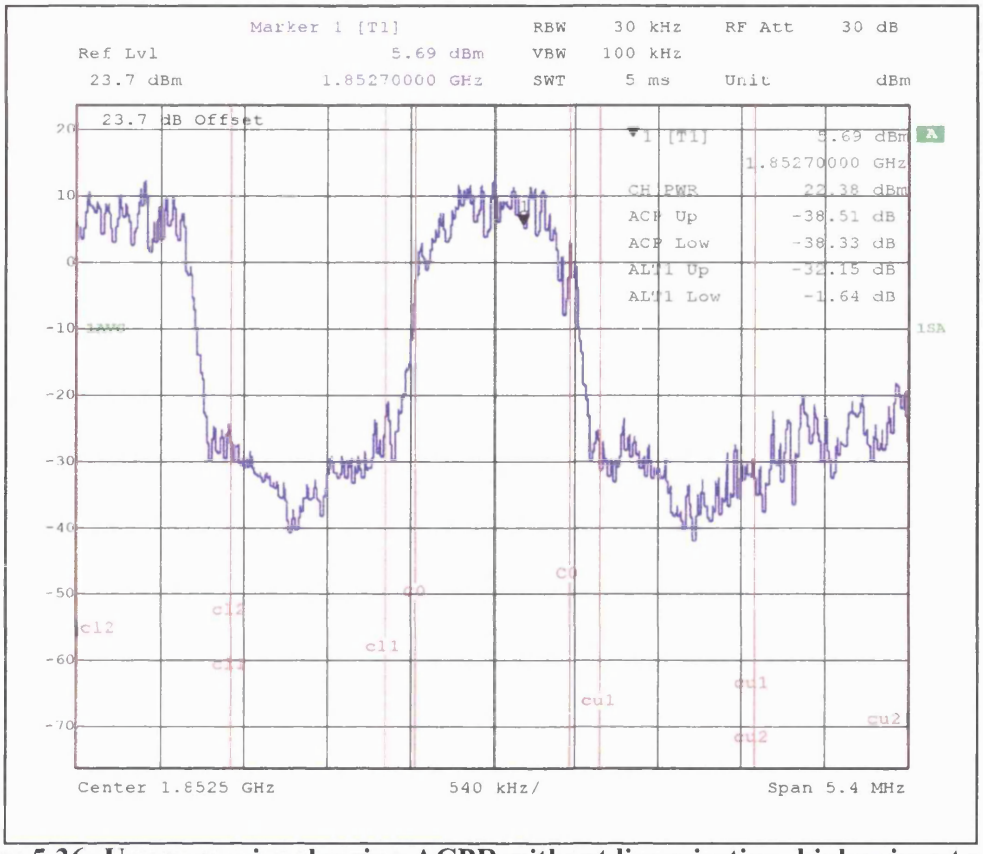


Figure 5.36: Upper carrier showing ACPR without linearisation, higher input power

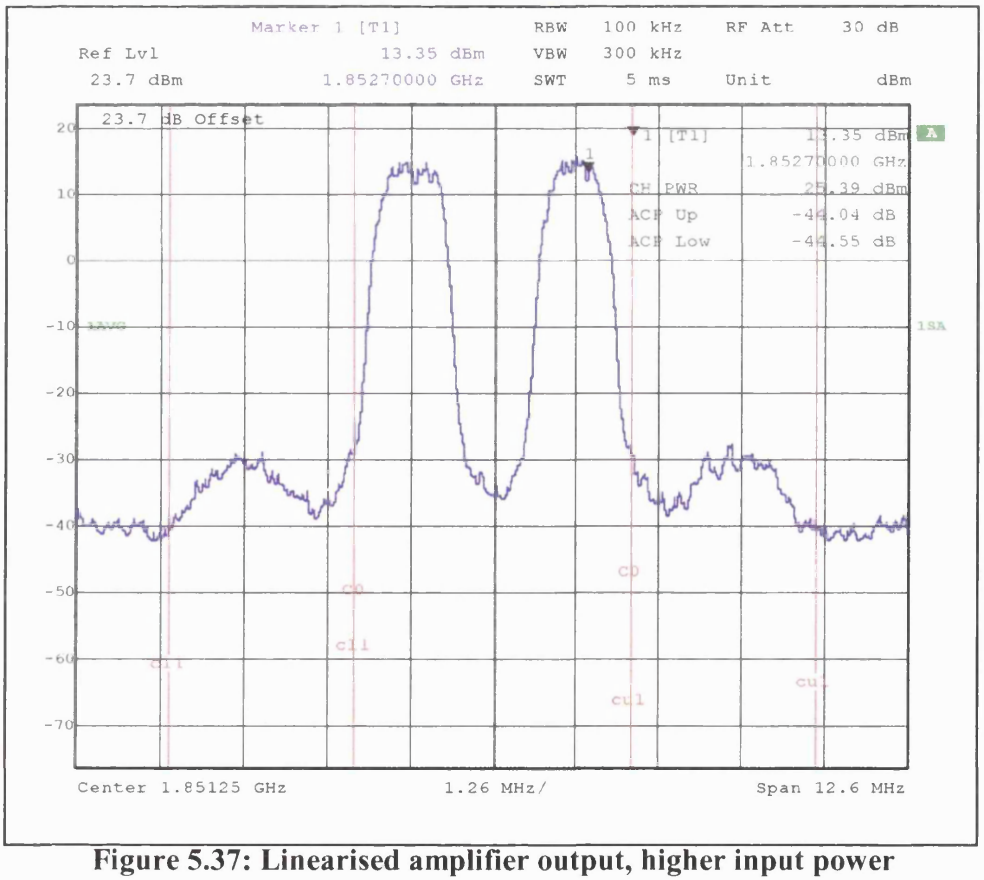


Figure 5.37: Linearised amplifier output, higher input power

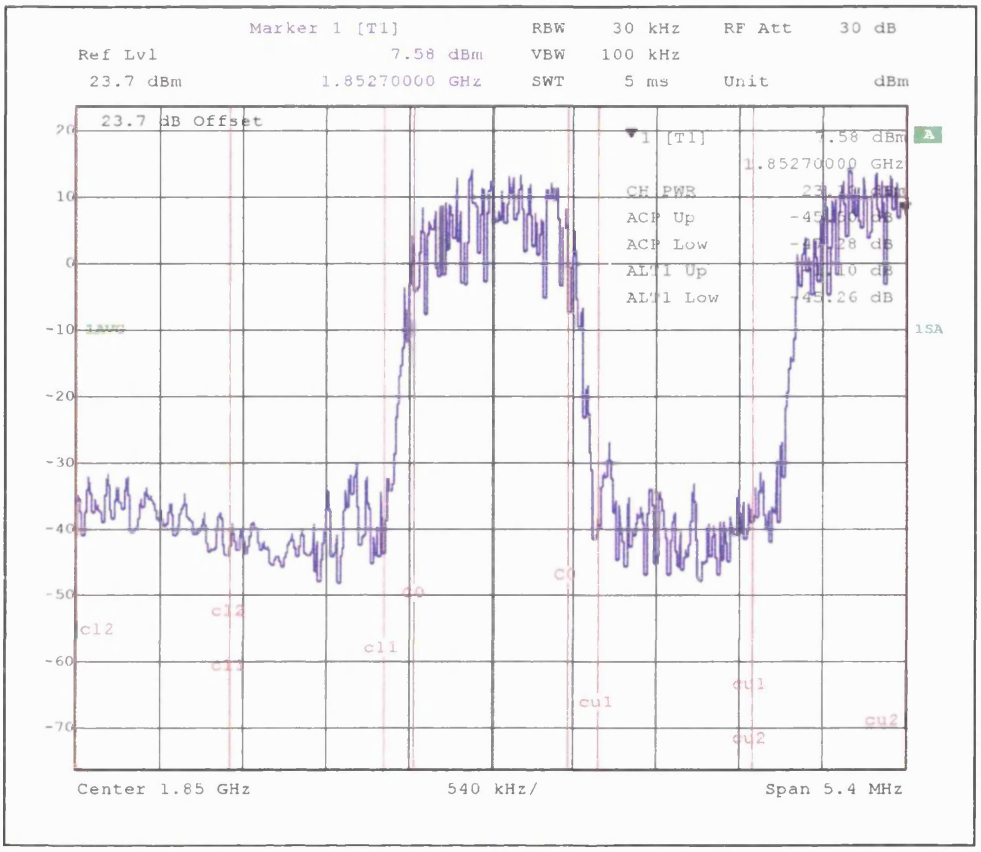


Figure 5.38: Lower carrier with linearisation, higher input power

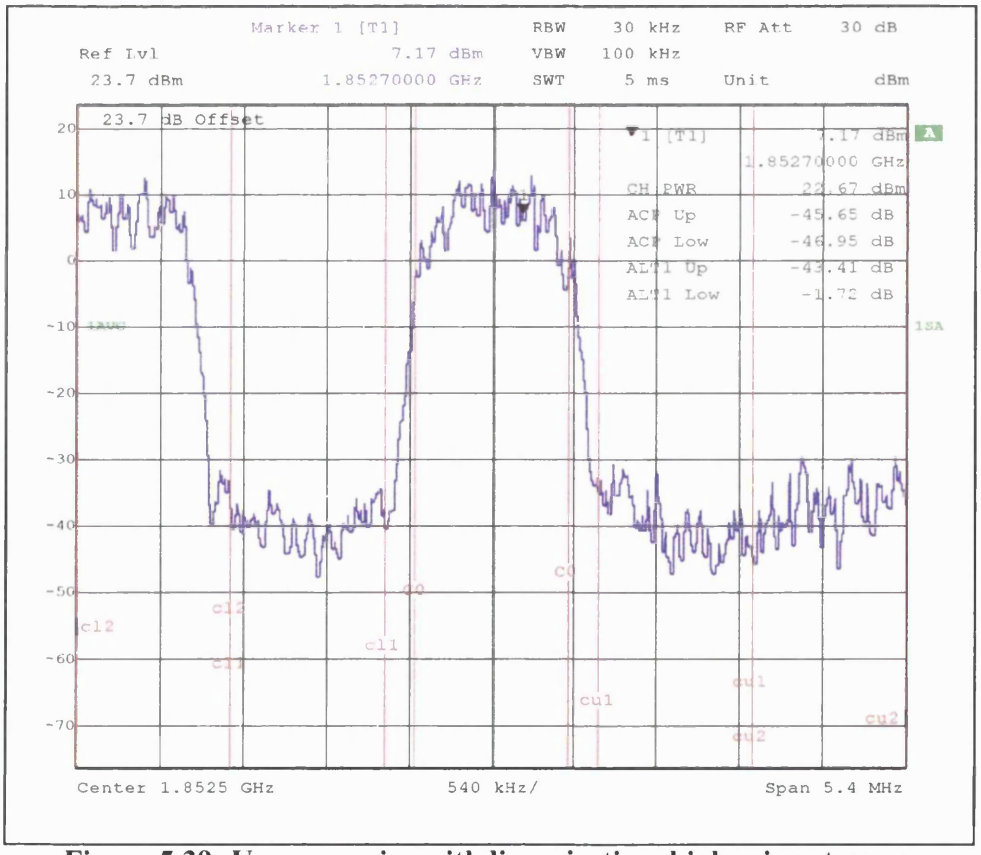


Figure 5.39: Upper carrier with linearisation, higher input power

The results show that, at the higher power level, the close-in ACPR was reduced by 7-8dB and the upper and lower alternate-channel powers were both reduced by approximately 11dB. The level of all the distortion components relative to the carriers with the addition of the linearising signal was found to be approximately –45dBc. Although the distortion improvement is not as great as that achieved at the lower power level, the amplifier is now being pushed into compression so this is expected.

5.6.4 Linearisation with Offset Carrier Amplitudes

In a practical transmitter the power levels of the individual carriers will rarely be equal, so it is necessary to investigate the application of the linearisation technique with input signals of this type. To enable this, the amplitude of the i and q waveforms used in the generation of the linearising signal were scaled to correspond with the amplitude of the carriers. The results in Figure 5.40 overleaf show unlinearised and linearised output spectrums with the power of the upper carrier chosen to be 10dB below that of the lower. To ensure that the linearising signal was generated correctly in this case, the VisualBasic routine that formed the composite signal (see Appendix C, 4) was modified to scale the amplitudes of the $i_2(t)$ and $q_2(t)$ baseband waveforms by a factor of 0.1.

The blue trace on Figure 5.40 shows that, as expected, the difference in carrier amplitudes produces distortion components whose levels are offset accordingly. The red trace on the same figure shows the linearised amplifier output, clearly demonstrating that both the close-in ACPR and ALCPR either side of the carriers are reduced. This confirms that the relationship between the amplitudes of the input signals and those of the components used in the formation of the linearising signal predicted by the analysis in Section 5.4 is correct.

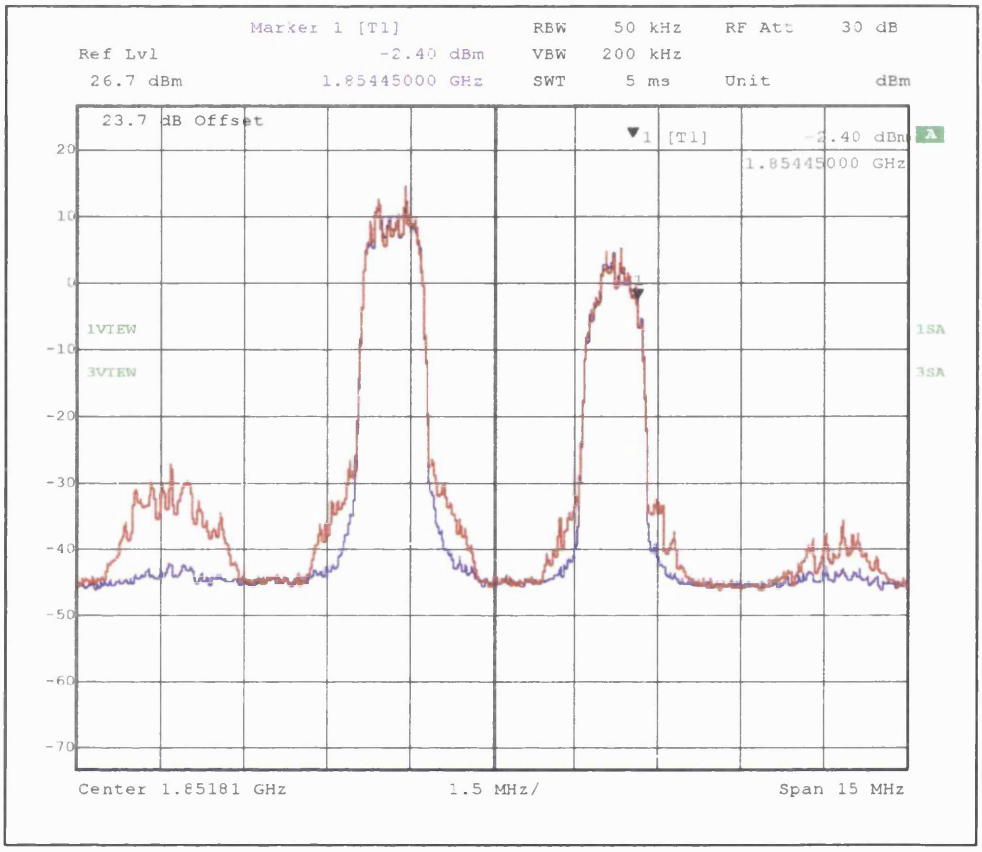


Figure 5.40: Unlinearised and linearised output spectrum for two carriers, with upper carrier power reduced by 10dB

5.7 Summary

The analysis, simulation results and measurements presented in this chapter have shown that the technique may be applied to both single- and multi-carrier modulated input-signals, producing a levels of distortion improvement in accordance with those expected. The analysis in Sections 5.1 and 5.4 gave the required composition of the linearising signal for both single and two-carrier inputs, illustrating that the generation becomes more considerably more complex as the number of carriers increases. However, the mathematical operations involved are all straightforward, consisting solely of multiplications and additions. Therefore, the processing overhead to produce multicarrier linearising signals should not be significantly greater than that required in single-carrier applications.

6. Summary, Future Work and Conclusions

6.1 Summary

In Chapters 1 and 2, the issue of linearity was discussed, along with theory to illustrate the origins of amplifier distortion. This highlighted the need for cost-effective linearisation techniques, and existing methods were described and appraised in terms of efficacy, cost, and viability. It was shown that the effectiveness of a linearisation scheme is, in general, proportional to its complexity, that all techniques have advantages and disadvantages, and that the optimum solution is strongly dependent upon the application.

In Chapters 3 and 4, the preliminary practical verification of the technique was described. This began by applying narrowband feedback with two-tone input signals, and progressed to become an 'external generation and injection' technique that was successfully demonstrated with multi-tone inputs. The practical measurements showed good agreement with simulated predictions, and verified the analysis presented.

Chapter 5 contained analysis, simulations and measurements to demonstrate the application of the linearisation technique to modulated input signals. The findings of this work showed conclusively that a low-frequency second-order injection signal, if generated correctly, can reduce in-band distortion appreciably. With QPSK input signals, achievable ACPR improvement was found to be typically 10-20dB, giving effective efficiency increases of approximately 15%. With single-carrier EDGE signals, the distortion improvement was observed as a reduction in EVM, and similar gains in efficiency were observed. The application of the technique to two-carrier QPSK inputs was also analysed, simulated and demonstrated successfully, employing both equal- and offset-power carriers.

In practical applications, these increases in linearity and/or efficiency would be translated into reductions in power consumption, heat-dissipation, amplifier size and hence overall cost (if it is assumed that the additional cost and power-consumption of the linearising circuitry is negligible).

6.2 Practical Implementation Issues

6.2.1 Single-Carrier

An example of how this technique might be applied in a practical single-carrier transmitter is shown overleaf in Figure 6.1. If the power control is implemented digitally as shown, it would be straightforward to store a look-up table of linearising signal amplitude coefficients (describing a characteristic similar to those shown in Figure 5.19 and Figure 5.22) to be referenced as the power of the main carrier varies.

As the linearising circuitry is primarily digital, its characteristics are repeatable and will not drift with time and temperature, giving it an advantage over analogue linearisers. Although it is likely that the nonlinear performance of the PA *will* be time-varying, it is unclear at this time whether adaptation of the linearising signal characteristics (i.e. amplitude and time offset) will be required. To obtain distortion improvement with Second-Order Bias Injection, two conditions must be satisfied:

i) The amplitude of the linearising signal must be such that the new signal components are generated with the same amplitude as the existing in-band third-order distortion. This was investigated both in simulation (Section 5.2.4) and in practise (Section 5.3.3), where it was found that the tolerance to amplitude deviations was considerably larger than with other linearisation techniques. From Equation 2.8, a Feedforward cancellation loop with an amplitude imbalance of 4.65dB gives a theoretical distortion suppression

of only 3dB. As shown in Figure 5.20, SOBI produces 3dB of distortion reduction at offsets of up to 20dB below or 7dB above the optimum amplitude.

ii) The time-alignment of the linearising and carrier waveforms. If the two signals are generated in the same IC, they will remain in synchronisation until they pass into the analogue domain, so any time-offset at the PA input will be purely dependent upon the delay-variation between the two paths indicated on Figure 6.1. In a commercially-available transmitter, carrier upconversion requires several stages of filtering and pre-amplification, and the time delay through these components will almost certainly vary during normal operation. Until a more advanced prototype is developed it is not possible to evaluate how detrimental to the performance of the technique this variation will be. However, investigations into the tolerance of the distortion reduction to injected-signal phase deviation (see Figure 5.7, page 117) suggest that Second-Order Bias Injection is more robust than other comparable linearisation schemes.

If the time-delay and amplitude variations in a practical system are such that closed-loop control is required to maintain performance, one possible solution is shown in Figure 6.1 as a feedback loop (dashed). By downconverting and recovering as shown, the input and output signal vectors can be compared, allowing the performance of the lineariser to be accurately monitored (variants of this type of control loop are often used to optimise and adapt the performance of Digital Predistorters). This control information can then be used to adapt the digital time delay and amplitude offset to maintain performance over changing operating conditions.

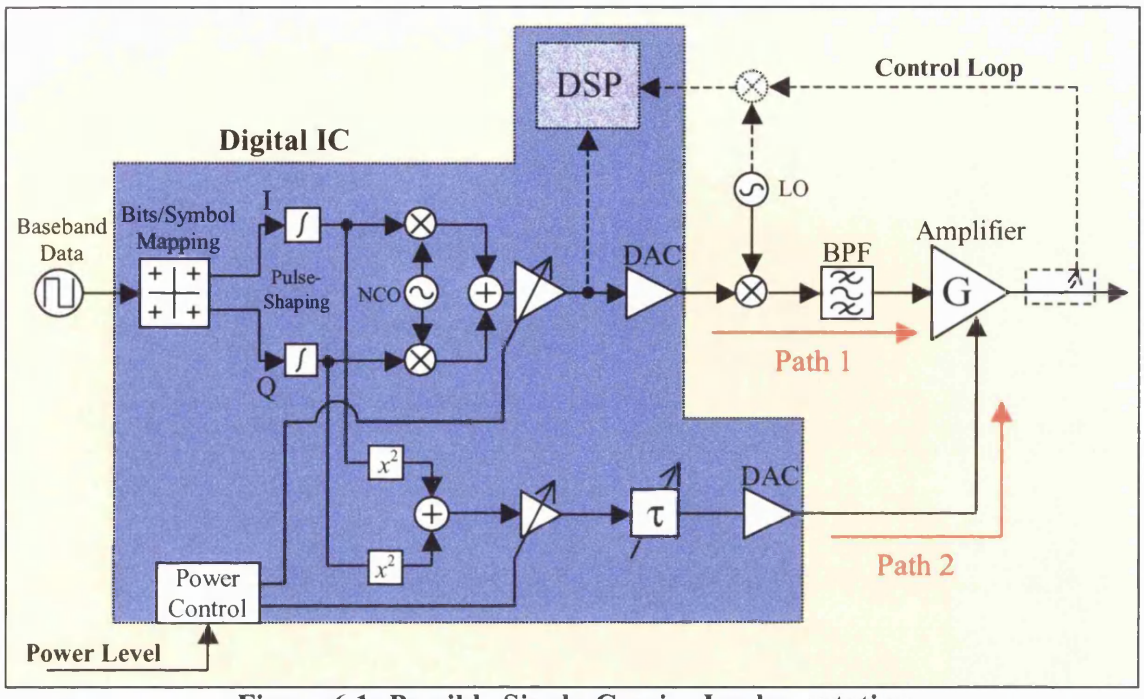


Figure 6.1: Possible Single-Carrier Implementation

6.2.2 Multicarrier

The use the new linearisation technique in a multicarrier transmitter will be more challenging than in a single carrier application, requiring a considerable increase in the complexity of the digital circuitry that forms the linearising signal. Although the bias de-coupling at the input of the amplifier must be designed with the increased bandwidth of the injection signal taken into account, this is the only analogue modification required. An example of how the technique might be employed in an integrated multicarrier transmitter is shown below in Figure 6.2.

Referring to the figure, the carriers are formed, upconverted to an IF and combined in the digital domain before being passed to a wideband DAC. In order to form the linearising signal, it is necessary to have access to all the individual baseband waveforms as well as information concerning carrier spacings, carrier power levels and possibly temperature, as indicated. If the lineariser were integrated into the digital portion of the transmitter as shown

in Figure 6.2, the relevant data would be readily available. Although the required information would still be available if the carriers were formed in separate ICs before being combined in the analogue domain, it would be harder to obtain and the synchronisation of the various modulators and upconverters would be difficult. Both of these scenarios (digital- or analogue-combined carriers) require a fully-integrated system architecture, and as such they are only feasible if the technique is applied as part of a linearised transmitter and not as a modular linear amplifier with RF inputs and outputs.

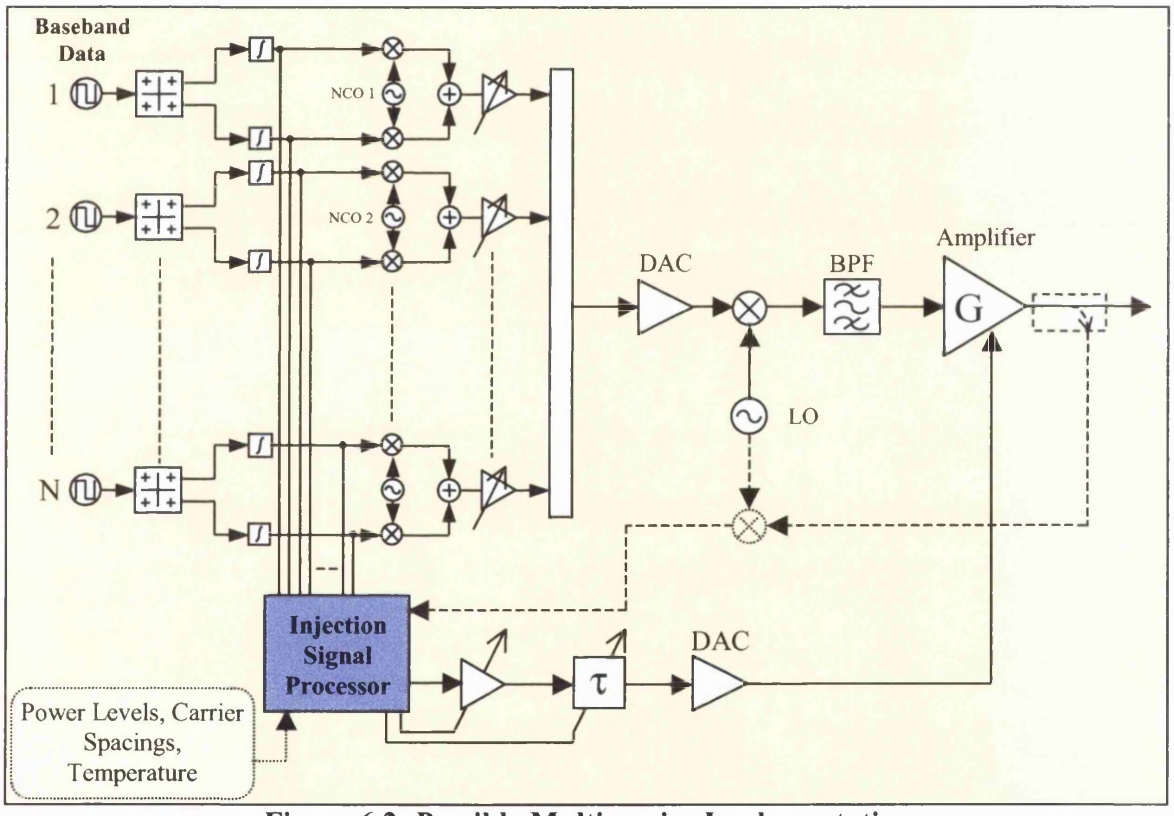


Figure 6.2: Possible Multicarrier Implementation

6.2.3 Integration with Other Techniques

It is also possible to use the technique in conjunction with linearisation methods such as Feedforward, as illustrated in Figure 6.3 overleaf. Currently, Feedforward is the only

practical solution for multicarrier GSM/EDGE transmitters, but it has inherent limitations as outlined earlier in Section 2.2.2. Typically, a well-tuned Feedforward amplifier can provide between 25dB and 30dB of intermodulation distortion suppression [40], so in order to the achieve the GSM/EDGE linearity specification of –75dBc, the Main Amplifier shown in Figure 6.3 must have a raw unlinearised performance of no worse than –45dBc. This requires a device biased for Class A operation, and the high peak-to-average ratios of multicarrier signals coupled with the high linearity requirement typically force average output powers to be backed-off by up to 10dB, giving main amplifier efficiencies of only 10-15%. This dominates the performance of Feedforward amplifiers, and additional component losses along with the need for an auxiliary Error Amplifier can reduce overall efficiencies to between 5% and 10% [40].

One of the methods employed to improve Feedforward efficiency is to use an analogue predistorter in front of the Main Amplifier, improving its raw linearity by 10dB or so. This either allows for a higher average output power, or enables the use of more a efficient quiescent operating point; either of these alternatives will result in increased efficiency for the Main Amplifier and also, correspondingly, for the overall system.

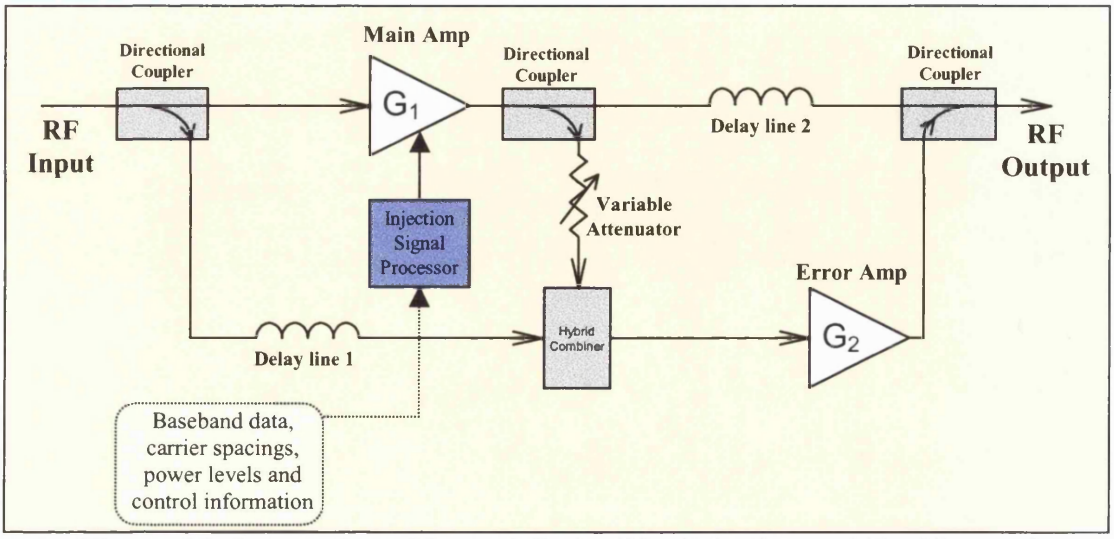


Figure 6.3: Integration of Second-Order Bias Injection and Feedforward

In a similar manner, Second-Order Bias-Injection can also be utilised within a Feedforward loop to increase the linearity of the Main Amplifier and improve the overall efficiency. As in the multicarrier application outlined in the previous section, this will only be possible if the MCPA is fully integrated into a linear transmitter, as the generation of the second-order linearising signal requires access to all the baseband waveforms as well as carrier spacing and power-level information.

6.3 Comparison with Other Linearisation Techniques

The newly-developed linearisation technique will now be evaluated against the three alternatives that are most commonly implemented in current systems – Analogue Predistortion, Digital Predistortion and Feedforward.

6.3.1 Analogue Predistortion

The new technique gives linearity improvements of 10-20dB, similar to those produced by a well-tuned analogue lineariser. However, the performance of analogue linearisers is defined by the nonlinear behaviour of the components within them (often smaller amplifiers biased to give gain-expansive transfer functions), and as such the resulting predistortion characteristic does not lend itself to accurate and straightforward control. It is therefore very challenging to ensure that the lineariser and amplifier remain aligned over the necessary dynamic range and also with time and temperature. In comparison, the nonlinear characteristics of the digital circuit in a SOBI lineariser are trivial to adjust, and will not vary over time (unless required to do so). Tuning in a production environment would also be simple in comparison, and as no physical adjustment of analogue circuitry is required, the process would be suited to automation.

The tolerance of an analogue-predistorted amplifier to amplitude and phase deviations within the circuit is the same as that in a single Feedforward loop (see Figure 2.7), as both techniques rely upon aligning a distorted amplifier output with an error signal. Of the two, a predistorter is less prone to phase and amplitude imbalance problems as the RF error signal and carrier are applied to the amplifier along the same path. However, any offsets in the amplitude or phase of the error signal produced by an incorrectly-tuned analogue predistorter are translated directly into imperfect cancellation, and performance degradation. In contrast, as discussed in Section 5.2.4, it has shown that the tolerance to errors in the phase and amplitude of the signal produced by a SOBI lineariser is approximately an order of magnitude wider than that shown in Figure 2.7.

The main advantage of analogue predistorters is that they do not need to be integrated into a linear transmitter, as they can be deployed within a linearised PA module with RF interfaces. Although the performance improvements are relatively modest, the low cost, flexibility and ease of implementation offered by a modular 'drop-in' PA mean that they are an attractive solution for many applications.

6.3.2 Digital Predistortion

Digitally-Adaptive Digital Predistortion (DAPD) is currently one of the fastest-growing areas of linearisation development, with several companies now offering complete solutions for both single- and multi-carrier transmitters. These systems are versatile and self-adapting, requiring little or no human intervention, and are currently reported to yield closed-loop distortion improvements in excess of 20dB. Unlike Analogue Predistortion and Second-Order Bias Injection, DAPD can take amplifier memory effects into account, so it is likely that the performance of currently-available systems will be surpassed as both technology and techniques evolve further.

Digital Predistortion is by no means simple to implement, and there are fundamental bottlenecks in critical circuit components arising from to the fact that both the error signal and carrier(s) are formed and combined in the digital domain. This composite signal has a bandwidth that is typically five times wider than that of the original combined carriers, as it contains both third- and fifth-order nonlinear components. Current state-of-the-art DAC performance limits the bandwidth of multi-carrier signals to approximately 15MHz, and a commercially-available system with three WCDMA carriers has recently been reported [41]. After the DAC, the complete wideband signal must be upconverted, filtered and pre-amplified before it reaches the final amplifier stage. This also limits the achievable operating bandwidth, and makes the design of the analogue circuitry much more challenging.

The main advantage of the new linearisation technique over DAPD is that the bandwidth requirements for both the transmit DAC and upconverter are unaffected, as the low-frequency and RF signals are only combined at the input of the final-stage amplifier. Although the generation of the linearising signal does require a DAC bandwidth that is twice that of the composite transmit bandwidth, this is less the three- or five-times bandwidth requirement of DAPD. The design of the input bias-circuitry does require care, but this is a trivial modification in comparison to the cost of increasing the bandwidth of an entire upconversion chain by a factor of five.

A digital predistorter produces impressive results, but it does so with a complex array of hardware and software that is expensive to develop and difficult to implement. In comparison, a SOBI lineariser produces more modest improvements in linearity, but it does so with very straightforward mathematical operations and only small changes to analogue circuitry. The disadvantage of both techniques with respect to Analogue Predistortion and

Feedforward is that they can only be implemented as part of a complete transmitter, and not as a modular linearised PA.

6.3.3 Feedforward

Feedforward is one of the oldest and most well-established linearisation techniques. Until recent advances in Digital Predistortion, it was the only viable solution for most commercial multi-carrier applications, and is still the only method capable of achieving the linearity specifications required in multi-carrier GSM/EDGE transmitters. However, as already discussed, Feedforward amplifiers are very inefficient, large and are horrendous to manufacture in large volumes, so their use is only justified in applications where large operating bandwidths (>20MHz) and ultra-linear performance (60-75dBc) is required. In WCDMA transmitters, where linearity requirements are not as stringent as those in a comparable GSM system, the use of a Feedforward amplifier may not be the optimal solution, and there is much scope to examine the trade-off between performance, cost, ease of manufacture, running costs, size and complexity. In applications where only a moderate improvement in linearity is necessary, and high efficiency or minimum size is required, Second-Order Bias Injection may, overall, prove to be the more cost-efficient solution.

6.4 Future Work

6.4.1 IC Prototyping

In order to properly evaluate the new linearisation technique, it must now be implemented in a form that will allow testing under realistic operating conditions. In order to do this, a digital IC and upconversion chain with functionality similar to that shown in Figure 6.1 will be developed as part of further work to be undertaken at Nokia Networks, Camberley. By generating both the baseband and SOBI waveforms in the same IC, it will be possible to quantitatively assess whether the necessary time- and amplitude-alignment of the RF and second-order signals can be maintained over time and temperature without closed-loop control. It will also be possible to investigate whether the table of injection-signal amplitude coefficients versus input power can be fixed, or if it needs to be updated periodically under closed-loop control to maintain performance as amplifier characteristics vary.

If possible, the lineariser IC will be designed with enough flexibility to also allow multi-carrier applications to be investigated, with additional NCOs and digital quadrature-modulators included to enable the formation of more complex linearising waveforms.

6.4.2 Integration with Other Techniques

The integration of the new technique with other linearisation schemes such as Feedforward should also now be explored. This will allow the expected linearity and efficiency improvements to be accurately quantified, and will also enable any unseen integration issues to be investigated.

6.4.3 Investigation of Performance Limitations

Over wide bandwidths, the performance of a SOBI lineariser will be limited by three factors:

1. The bandwidth of the DAC used to generate the low-frequency signal

2. The bandwidth of the bias-feeding network at the amplifier input
3. Amplifier memory effects

The first of these issues is not likely to be the primary concern, as DACs having bandwidths in excess of 50MHz are already available and performance is steadily improving. The bias-feeding circuitry and amplifier memory effects will, however, place fundamental limitations on the bandwidth and performance of the technique. The input bias network is required to have a low-pass characteristic with a cut-off frequency greater than the bandwidth of the linearising signal, but it must also be ensured that the RF matching conditions are not unduly affected as a result. Although this issue will be strongly application-dependent, it can never be ignored and should be considered as an integral part of the amplifier design from an early stage. As discussed in Section 3.3.1 the impact of memory effects can be minimised with bias-circuit optimisation at the amplifier output, and this is usually designed to provide a near short-circuit at the envelope frequencies. Again, the severity of the problem will depend on factors such as the modulation bandwidth, device technology and supply-current requirements, but the issue cannot be ignored if the performance is to be optimised.

6.5 Conclusions

The original aim of the study as outlined in the introduction has now been realised – namely, a new linearisation technique has been developed. It has been demonstrated that Second-Order Bias Injection (SOBI) is capable of significant improvements in linearity with a relatively moderate increase in circuit complexity, and an International Patent Application was filed in November 2000. Although the linearity improvements recorded in this investigation are not as great as those achieved by more well-established techniques such as Feedforward and Digital Predistortion, useful performance improvements are attainable with considerably less complex circuitry. Due to the anticipated implementation simplicity, low

cost, size and power consumption, it is believed that the new technique has great potential for use in both handsets and basestations in next-generation mobile networks.

Appendix A

MDS FLL351ME Nonlinear Model Implementation

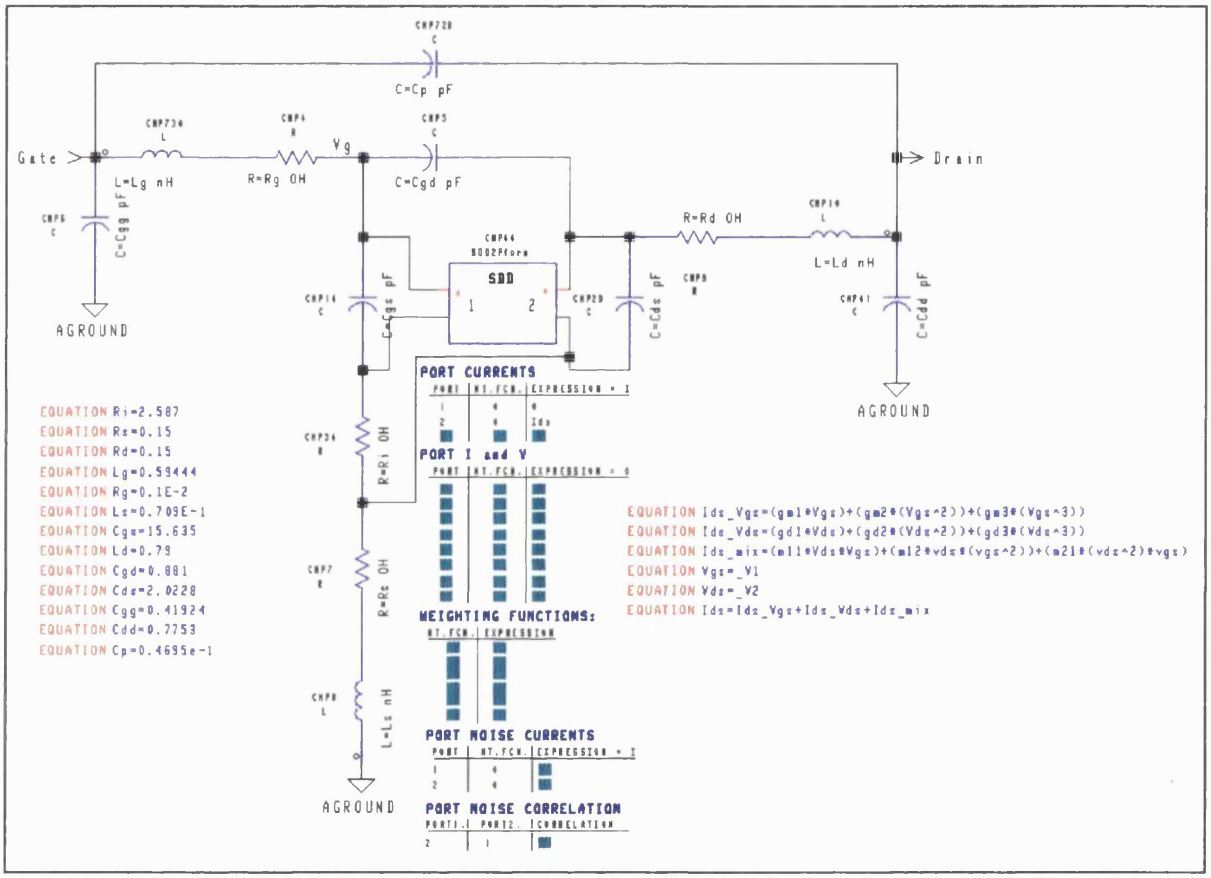


Figure 7.1: *MDS* implementation of nonlinear FET model showing extracted 2-D coefficients

Appendix B

Measured MRF281 Test Amplifier S-Parameters

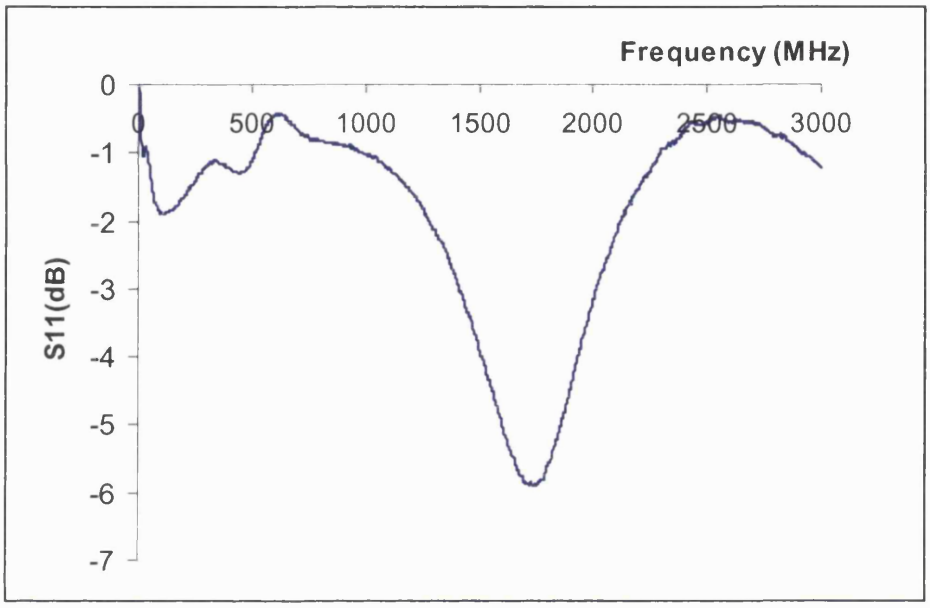


Figure 7.2: Measured MRF281 S_{11}

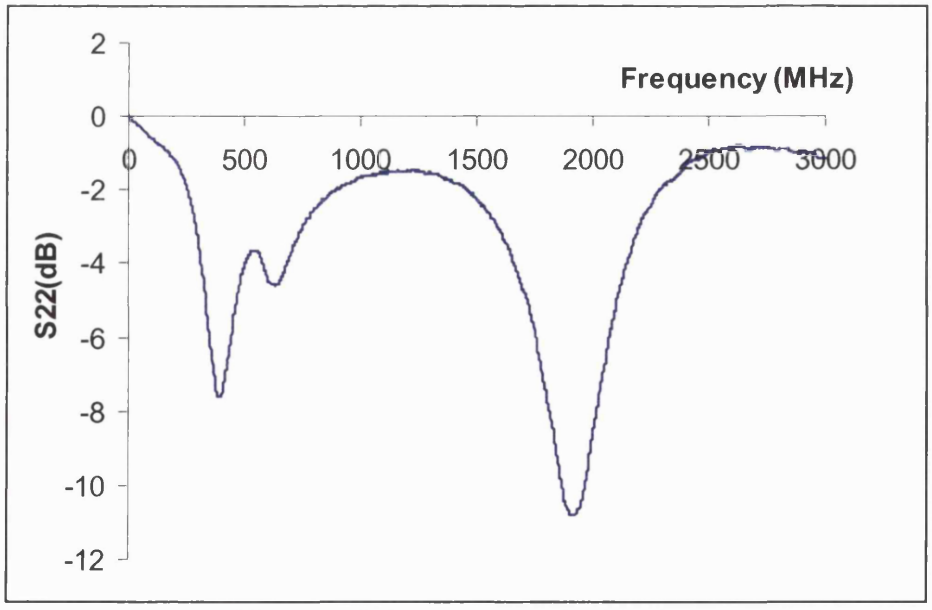


Figure 7.3: Measured MRF281 S_{22}

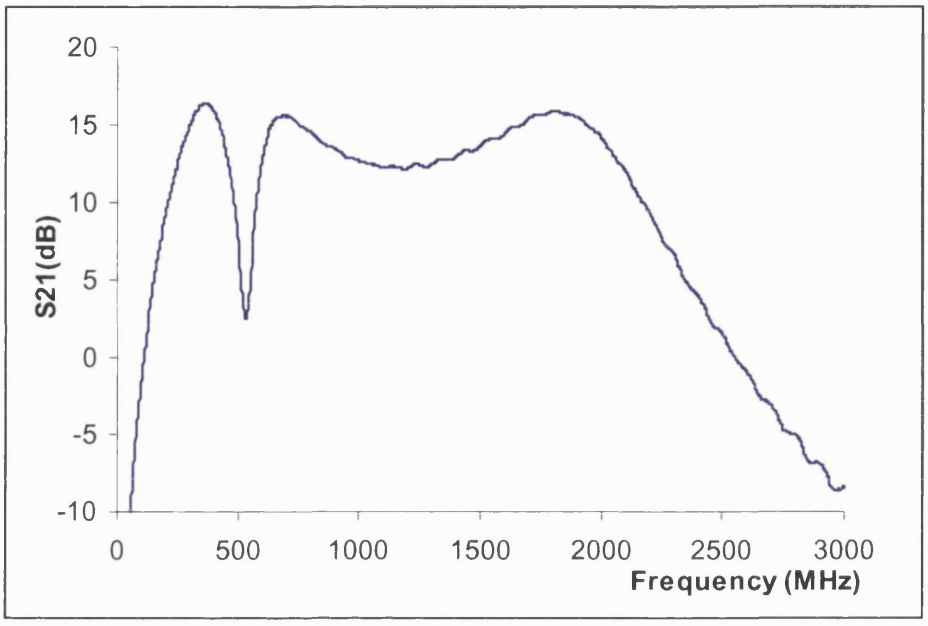


Figure 7.4: Measured MRF281 S_{21}

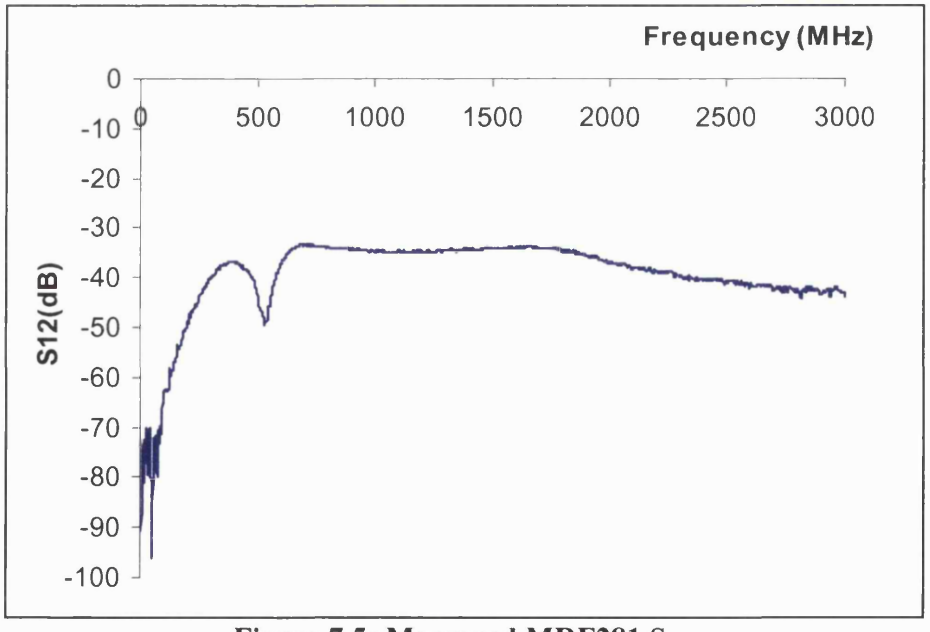


Figure 7.5: Measured MRF281 S_{12}

Appendix C

1. Pseudo-Random Bit-Stream Generator

The following subroutine generates two pseudorandom bit-streams of length `num_bits%` oversampled by a ratio of `oversample%` and stores them in two arrays, `i()` and `q()`, both of length `num_points%` (=`num_bits%`x`oversample%`). Finally, the values are copied three times to force the middle set of values to be periodic after subsequent filtering.

```
Public Sub bit_init(num_bits%, oversample%)
    Static n%       'initialise counter (static between calls to numgen)
    n% = 0
    num_points% = num_bits% * oversample%    'set number of sample points
    'limit num_points to <= 4096 (AWG memory depth)
    If num_points% > 4096 Then
        txtnumbits.Text = 4096 / oversample%
        num_points% = val(txtnumbits.Text) * oversample%
    End If
    'initialise arrays for i and q
    ReDim i(num_points%)
    ReDim q(num_points%)
    'generate 2x random pulse-streams and store in i() and q()
    Do While n% < num_points%
        i(n%) = numgen            'call random number generator, returns 1 or 0
        q(n%) = numgen            'call random number generator, returns 1 or 0
        n% = n% + oversample%     'leave oversample% - 1 zeros
    Loop
    'now copy pulse-stream 3 times ie [a b c] -> [a b c a b c a b c]
    a% = num_points%
    num_points% = num_points% * 3
    ReDim Preserve i(num_points%)
    ReDim Preserve q(num_points%)
    For n% = a% To num_points%
        i(n%) = i(n% - a%)
        q(n%) = q(n% - a%)
    Next n%
End Sub
```

2. Pulse-Shaping

The following code loads a raised-cosine filter response from a text file (`filestr$`) into array `filterweights()` and convolves it with the arrays `i()` and `q()` that were populated by the subroutine `bit_init`, shown above. The first two lines of the filter response contain the sampling frequency and reconstruction filter frequency that are subsequently used to set up the arbitrary waveform generators within the ESG4433B. After the convolution, only the middle third of the waveform is retained as this ensures periodicity without discontinuities.

```
Public Sub dofilter()
    'digital filtering by convolution
    filestr$ = "c:\will\digfilters\filter1bscaled.txt"
    'load filter weights into end of filterweights array i.e. [0 0 ... 3 3 7]
    Dim filterweights() As Double        'declare filterweights array
    weight_index% = num_points%          'num_points = number of i (or q) samples
    ReDim filterweights(weight_index%)   'set array size to num_points%
    Open filestr$ For Input As #1        'open filter coefficients file
    Line Input #1, txtstr$               ' read sampling frequency line
    esg_fs$ = txtstr$ & " MHZ"
    esg_arbfreq! = val(txtstr$)
    Line Input #1, txtstr$               'read reconstruction filter line
    esg_filt$ = txtstr$ & " MHZ"
    Line Input #1, txtstr$               'read first filter coefficient
    Do While Not EOF(1)                  'file read loop
        filterweights(weight_index%) = val(txtstr$)   'store value
        weight_index% = weight_index% + 1             'increment array index
        ReDim Preserve filterweights(weight_index%)   're-size array
        Line Input #1, txtstr$                        'read next line
    Loop                                              'loop until end of file
    total_points% = weight_index% - 1   'set overall convolution length (a+b-1)
    ReDim Preserve filterweights(total_points%) 'redefine arrays to be same size
    ReDim Preserve i(total_points%)             'resize i[] i.e. [1 3 ... 0 0 0]
    ReDim Preserve q(total_points%)             'resize q[] i.e. [1 5 ... 0 0 0]
    ReDim i_out(total_points%)                  'initialise arrays for results
    ReDim q_out(total_points%)
    '*** perform convolution ***
    For r% = 1 To total_points%         'outer convolution loop
        i_curr# = 0
        For p% = 1 To total_points%
            i_temp# = filterweights(p%) * i(p%)
            i_curr# = i_curr# + i_temp#
        Next p%
        q_curr# = 0
        For p% = 1 To total_points%
            q_temp# = filterweights(p%) * q(p%)
            q_curr# = q_curr# + q_temp#
        Next p%
        'loop to shift filter weights right 1 place
       For t% = total_points To 2 Step -1
            i(t%) = i(t% - 1)
            i(t% - 1) = 0
            q(t%) = q(t% - 1)
            q(t% - 1) = 0
        Next t%
        i_out(r%) = i_curr#
        q_out(r%) = q_curr#
    Next r%
    're-size arrays and write middle set of points to i[] and q[]
    total_points% = num_bits% * oversample% - 1
    ReDim i(total_points%)
    ReDim q(total_points%)
    Open "c:\will\data\convolve.txt" For Output Access Write As #2
    For n% = 0 To total_points%
        i(n%) = i_out(total_points% + n%)
        q(n%) = q_out(total_points% + n%)
        Print #2, "i " & i(n%) & " q " & q(n%) 'output data to file for checking
    Next n%
    Close #1
    Close #2
End Sub
```

3. Single-Carrier Linearising Signal Generation

This routine forms the low-frequency linearising signal described in Section 5.1 as $i(t)^2+q(t)^2$. After the formation of the signal, DC offsets are removed and the waveform normalised to an amplitude of ±1 as required by the HP33120A. A null waveform consisting of zeros is also created to provide a convenient means of switching off the linearising signal. The output frequency of the HP33120A is calculated and sent to the unit, before the linearising and null signals are downloaded and stored as `LIN` and `ZEROS`, respectively.

```
Public Sub cmdmakelinsig_Click()
    'form linearising signals from i and q waveforms
    'and download to HP33120A
    ReDim linsig(total_points%) As Double
    'form linsig as i()^2 + q()^2
    average1# = 0
    For n% = 0 To total_points%
        linsig(n%) = (i(n%) ^ 2) + (q(n%) ^ 2)
         average1# = average1# + linsig(n%)
    Next n%
    'find DC average
    average1# = average1# / (total_points% + 1)
    'subtract DC and find max and min values for normalisation
    For n% = 0 To total_points%
        linsig(n%) = linsig(n%) - average1#
        If linsig(n%) > maxval1# Then maxval1# = linsig(n%) ' find max value
        If linsig(n%) < minval1# Then minval1# = linsig(n%) ' find min value
     Next n%
    'calculate normalisation factor
    If Abs(maxval1#) >= Abs(minval1#) Then
        normfact1# = Abs(maxval1#)
    Else
        normfact1# = Abs(minval1#)
    End If
    'normalise waveform and put into string form for downloading
    'and create a null waveform consisting of zeros
    For n% = 0 To total_points% - 1
        linsigsend1$ = linsigsend1$ & (linsig(n%) / normfact1#) & ","
       zerosigsend$ = zerosigsend$ & "0" & ","
    Next n%
    'omit comma for last entry
    linsigsend1$ = linsigsend1$ & (linsig(total_points%) / normfact1#)
    zerosigsend$ = zerosigsend$ & "0"
    'calculate awg output freq and send frequency command to HP33120A
    awg_freq!=(val(txtsymbolrate.Text)*val(txtoversamp.Text))/(total_points%+1)
    fstr$ = "FREQ " & awg_freq! & "KHZ"
    Call Send(0, awg_addr%, fstr$, NLend)
    'download and store linsig as LIN
    downsend$ = "DATA VOLATILE, " & linsigsend1$
    Call Send(0, awg_addr%, downsend$, NLend)
    Call Send(0, awg_addr%, "DATA:COPY LIN", NLend)
    'download and store zeros
    downsend$ = "DATA VOLATILE, " & zerosigsend$
    Call Send(0, awg_addr%, downsend$, NLend)
    Call Send(0, awg_addr%, "DATA:COPY ZEROS", NLend)
    'select linsig
    Call Send(0, awg_addr%, "FUNC:USER LIN", NLend)
    Call Send(0, awg_addr%, "FUNC:SHAP USER", NLend)
End Sub
```

4. Two-Carrier Linearising Signal Generation

The following routine forms a two-carrier linearising signal, defined in Section 5.4 as $\frac{1}{2}\{i_1(t)^2+q_1(t)^2+i_2(t)^2+q_2(t)^2\}+i_1(t)i_2(t)\cos(\omega_2-\omega_1)+q_1(t)q_2(t)\cos(\omega_2-\omega_1)+q_1(t)i_2(t)\sin(\omega_2-\omega_1)-i_1(t)q_2(t)\sin(\omega_2-\omega_1)$. As in the simulations of Section 5.5.1, two waveforms are generated, the first consisting of only IF linearising components and the second consisting of both IF and baseband (see Figure 5.23). The bandwidth of the linearising signal is greater than the bandwidth of the carriers themselves as it now includes IF components centred at the carrier spacing frequency. As such, the sample rates used in the formation of the individual modulating waveforms are insufficient for the linearising signal. To avoid aliasing, the four *i* and *q* waveforms are oversampled in this subroutine before being used to form the composite signal. To check that sufficienct oversampling has been applied, the number of waveform points per cycle of the difference frequency component is calculated as `points_per_cycle!` and displayed on the user interface. The oversampling is accomplished by zero-padding the waveforms before they are low-pass digital filtered with `dofilter2()` to complete the process.

The variable `points_per_cycle!` is then used to calculate the phase increment that is required in the formation of the cosine and sine waveforms corresponding to the difference frequency (carrier-spacing). A simple FOR loop creates the two sinusoids and stores them in separate arrays, from which they are recalled during the formation of the linearising signals.

DC offsets are then removed and the waveforms normalised before being downloaded to the HP33120A. The downloading of the linearising waveforms is performed differently than in the single-carrier case, with values sent as 2-byte words as this reduces the time taken to transfer the data. A null waveform consisting of zeros is again created to provide a convenient means of switching off the linearising signal. The output frequency of the

HP33120A is calculated and sent to the unit, before the two linearising waveforms and the null signal are downloaded and stored as LIN, LINB and ZEROS, respectively.

```
Public Sub cmdMakelinsig2_Click()
    pi# = 3.14159265358979
    'oversample i, q, i2 and q2 and form linearising signals from the waveforms
    ReDim i_linsig(total_points%) As Double  'initialise 4 arrays for
    ReDim q_linsig(total_points%) As Double  'the oversampled I and Q
    ReDim i2_linsig(total_points%) As Double 'waveforms
    ReDim q2_linsig(total_points%) As Double 'and copy the original
    For n% = 0 To total_points%              'i and q into them
        i_linsig(n%) = i(n%)
        q_linsig(n%) = q(n%)
        i2_linsig(n%) = i2(n%)
        q2_linsig(n%) = q2(n%)
    Next n%
    'find oversampling ratio from user interface
    oversampling% = val(txtLinsigoversample.Text)
    If oversampling% = 2 Then num_loops% = 1
    If oversampling% = 4 Then num_loops% = 2
    If oversampling% = 8 Then num_loops% = 3
    If oversampling% = 16 Then num_loops% = 4
    If oversampling% = 32 Then num_loops% = 5
    If oversampling% = 64 Then num_loops% = 6
    current_size% = UBound(i_linsig) + 1          'find number of waveform points
    new_size% = (oversampling% * current_size%)   'calculate new size
    ReDim i_temp(new_size% - 1) As Double  'create four temp signal arrays
    ReDim q_temp(new_size% - 1) As Double  'of size new_size%
    ReDim i2_temp(new_size% - 1) As Double
    ReDim q2_temp(new_size% - 1) As Double
    'zero padding loop
    For m% = 0 To current_size% - 1
         i_temp(m% + num_loops%) = i_linsig(m%)  'copy i value
         q_temp(m% + num_loops%) = q_linsig(m%)  'copy q value
         i2_temp(m% + num_loops%) = i2_linsig(m%)'copy i2 value
         q2_temp(m% + num_loops%) = q2_linsig(m%)'copy q2 value
         'pad waveforms with num_loops% zeros
         for p%=1 to num_loops%
             i_temp(m% + num_loops%) = 0
             q_temp(m% + num_loops%) = 0
             i_temp2(m% + num_loops%) = 0
             q_temp2(m% + num_loops%) = 0
         next p%
    Next m%
    'resize 4 arrays for waveform storage
    ReDim i_linsig(new_size% - 1) As Double
    ReDim q_linsig(new_size% - 1) As Double
    ReDim i2_linsig(new_size% - 1) As Double
    ReDim q2_linsig(new_size% - 1) As Double
    'copy values to i_linsig, q_linsig, i2_linsig and q2_linsig
    For o% = 0 To new_size% - 1
        i_linsig(o%) = i_temp(o%)
        q_linsig(o%) = q_temp(o%)
        i2_linsig(o%) = i2_temp(o%)
        q2_linsig(o%) = q2_temp(o%)
    Next o%
    'Now low-pass filter the 4 waveforms to complete the oversampling
    'leaving i_linsig(o%), q_linsig(o%), i2_linsig(o%) and q2_linsig(o%)
    Call dofilter2(oversampling%)
    'calculate and send output freq to HP33120A
    symbol_rate% = val(txtsymbolrate.Text)
    oversample% = val(txtoversample.Text)
    awg_reprate# = (symbol_rate% * oversample% * Linsigovsam.Text) / new_size%
    fstr$ = "FREQ " & awg_reprate# & "KHZ"
    Call Send(0, awg_addr%, fstr$, NLend)
```

```
'calculate and check linsig properties
diff_freq! = val(txtCarrierspacing.Text) * 1000 'desired diff freq (in kHz)
no_of_cycles! = diff_freq! / awg_reprate#  'find number of cycles for AWG
points_per_cycle! = new_size% / no_of_cycles! 'find number of points/cycle
'***********************************
'*** Build linearising waveforms ***
'***********************************
ReDim linsig1(new_size% - 1) As Double 'linsig1 contains only IF components
ReDim linsig2(new_size% - 1) As Double 'linsig2 contains BB + IF components
'form sin and cos waveforms corresponding to the chosen carrier spacing
'of length new_size%
phase_increment# = 2 * pi# / points_per_cycle!
ReDim my_sin(new_size% - 1) As Double
ReDim my_cos(new_size% - 1) As Double
For n% = 0 To new_size% - 1
    my_sin(n%) = Sin(n% * phase_increment#)
    my_cos(n%) = Cos(n% * phase_increment#)
Next n%
average1# = 0
average2# = 0
'form linearising signal in 3 stages, a#, b# and c#
For n% = 0 To new_size% - 1
   a#=my_cos(n%)*((i_linsig(n%)*i2_linsig(n%))+(q_linsig(n%)*q2_linsig(n%)))
   b#=my_sin(n%)*((i2_linsig(n%)*q_linsig(n%))-(i_linsig(n%)*q2_linsig(n%)))
   c#=0.5*((i_linsig(n%)^2)+(q_linsig(n%)^2)+(i2_linsig(n%)^2)+(q2_linsig(n%)^2))
   linsig1(n%) = a# + b#
   linsig2(n%) = a# + b# + c#
   average1# = average1# + linsig1(n%)
   average2# = average2# + linsig2(n%)
Next n%
'calculate DC average
average1# = average1# / new_size%
average2# = average2# / new_size%
'remove DC average and find signal limits for normalisation
For n% = 0 To new_size% - 1
    linsig1(n%) = linsig1(n%) - average1#
    linsig2(n%) = linsig2(n%) - average2#
    If linsig1(n%) > maxval1# Then maxval1# = linsig1(n%) ' find max value
    If linsig1(n%) < minval1# Then minval1# = linsig1(n%) ' find min value
    If linsig2(n%) > maxval2# Then maxval2# = linsig2(n%) ' find max value
    If linsig2(n%) < minval2# Then minval2# = linsig2(n%) ' find min value
Next n%
    'find normalisation factors
If Abs(maxval1#) >= Abs(minval1#) Then
    normfact1# = Abs(maxval1#)
Else
    normfact1# = Abs(minval1#)
End If
If Abs(maxval2#) >= Abs(minval2#) Then
    normfact2# = Abs(maxval2#)
Else
    normfact2# = Abs(minval2#)
End If
normfact1# = normfact1# / 2047
normfact2# = normfact2# / 2047
'************************************************
'***** download waveforms to HP33120A ***********
'************************************************
'find info needed for header
num_bytes% = 2 * (new_size%)
num_digits% = Len(CStr(num_bytes%))
int_num_bytes! = val(num_bytes%)
'form header and send linsig bytes (normalised)
strsend$ = "DATA:DAC VOLATILE, #" & num_digits% & int_num_bytes!
Call Send(0, awg_addr%, strsend$, NULLend) 'send data header
For n% = 0 To new_size% - 1
    msb2 = Chr(CDec("&H" & msb(Hex(CInt((linsig1(n%) / normfact1#))))))
    lsb2 = Chr(CDec("&H" & lsb(Hex(CInt((linsig1(n%) / normfact1#))))))
    Call Send(0, awg_addr%, msb2, NULLend) 'format values using functions
    Call Send(0, awg_addr%, lsb2, NULLend) 'msb and lsb
```

```
     Next n%
     Call Send(0, awg_addr%, "", NLend) 'Terminate transfer

    Call Send(0, awg_addr%, "DATA:COPY LIN", NLend) 'store linsig
    'send same header followed by linsig2 bytes
    Call Send(0, awg_addr%, strsend$, NULLend) 'send data header
    For n% = 0 To new_size% - 1
        msb2 = Chr(CDec("&H" & msb(Hex(CInt((linsig2(n%) / normfact2#))))))
        lsb2 = Chr(CDec("&H" & lsb(Hex(CInt((linsig2(n%) / normfact2#))))))
        Call Send(0, awg_addr%, msb2, NULLend) 'format values using functions
        Call Send(0, awg_addr%, lsb2, NULLend) 'msb and lsb
     Next n%
    Call Send(0, awg_addr%, "", NLend) 'Terminate transfer
    'store linsig2
    Call Send(0, awg_addr%, "DATA:COPY LINB", NLend)
    'form a zero signal of 100 points
    For n% = 0 To 100
        zerosigsend$ = zerosigsend$ & "0,"
    Next n%
    zerosigsend$ = zerosigsend$ & "0"
        'store zeros
    downsend$ = "DATA VOLATILE, " & zerosigsend$
    Call Send(0, awg_addr%, downsend$, NLend)
    Call Send(0, awg_addr%, "DATA:COPY ZEROS", NLend)
    'select linsig
    Call Send(0, awg_addr%, "FUNC:USER LIN", NLend)
    Call Send(0, awg_addr%, "FUNC:SHAP USER", NLend)
End Sub
```

REFERENCES

1. EMC World Cellular Database (http://www.gsmworld.com/membership/ass_sub_stats.html)
2. 3rd Generation Partnership Project (3GPP) Technical Specification Group (TSG) RAN WG4 UTRA (BS) FDD; Radio Transmission and Reception – TS 25.104 v3.1.0
3. Gola, P and J-C Nana, "GSM/EDGE: Device Characterization for RF Power Amplifiers", Microwave Engineering Europe, March 2000
4. Kennington, P. B., "High-Linearity RF Amplifier Design", Artech House, 2000
5. European Telecommunications Standards Institute, "Digital cellular telecommunications system (Phase 2+); Radio transmission and reception", GSM 05.05 version 8.4.0, 1999 (http://www.etsi.org)
6. Maas, S. A., "Nonlinear Microwave Circuits", Artech House, 1988
7. Cripps, S. C., "RF Power Amplifier Design", Training course at Nokia Networks, Camberley, 1997
8. Black, H. S., "Translating Systems", US Patent 1686792; issued Oct 29, 1928, and U.S. Patent 2,102,671, issued Dec. 1937.
9. Seidel, H., "A Microwave Feedforward Experiment", Bell Syst. Tch. Journal, 50, (9), 1971, pp. 2879-2916
10. Kenington, P. B., K. J. Parsons, A. Bateman and J. P. McGeehan, "High-Efficiency Power Amplifier Linearisation for Mobile Communications", Microwaves and RF Conf. Proc., Nexus Media, Swanley, UK, 1995, pp. 24-7
11. Talwar, A. K., "Reduction of Noise and Distortion in Amplifiers Using Adaptive Cancellation" IEEE Trans. MTT, Vol. 42, No. 6, June 1994, pp. 1086-8
12. Meyer, R. G., R. Eschenbach, and W. M. Edgerley, "A Wideband Feedforward Amplifier," IEEE Journal of Solid-State Circuits, Vol. 9, pp. 422-488, Dec. 1974
13. Eid, E. E., F. M. Ghannouchi, and F. Beauregard, "Optimal Feedforward Linearization System Design," Microwave J., pp. 78-86, Nov. 1995
14. Cripps, S. C., "RF Amplifiers for Wireless Communications", Artech House, 1999

15. Kennington, P. B., and D. W. Bennett, "Linear Distortion Correction Using a Feedforward System", IEEE Trans. on Vehicular Technology, Vol. 45, No. 1, February 1996, pp. 74-81

16. Yamouchi, K., et al., "A Novel Series Diode Linearizer for Mobile Radio Power Amplifiers," IEEE Intl. Microwave Symp., MTT-S, 1996, pp. 831-834

17. Cavers, J., "A Linearizing Predistorter with Fast Adaption," Proc. 40th IEEE Vehicular Tech. Conf., 1990, pp. 41-47

18. Mansell, A., and A. Bateman, "Adaptive Digital Predistortion Linearisation," Proc. Microwaves & RF Conf., London, October 1996, pp. 270-275

19. Ballesteros, E., F. Perez, and J. Perez, "Analysis and Design of Microwave Linearised Amplifiers Using Active Feedback", IEEE Trans. MTT, Vol. 36, No. 3, March 1988, pp. 499-504

20. Ballesteros, E., F. Perez, and J. Perez, "Linearisation of Microwave Power Amplifiers Using Active Feedback Networks", Electronics Letters, 3 January 1985, Vol. 21, pp. 9-10

21. Arthanayake, T., and H. B. Wood, "Linear Amplification Using Envelope Feedback," Electronics Letters, Vol. 7, No. 7, pp. 145-146, 1971

22. Petrovic, V., and C. N. Smith, "Reduction of Intermodulation Distortion by Means of Modulation Feedback," IEE Conf. On Radio Spectrum Conservation Techniques, Sept. 1983, pp. 44-49

23. Bateman, A., and D. M. Haines, "Direct Conversion Transceiver Design for Compact Low Cost Portable Mobile Radio Terminals," 39th IEEE Vehicular Techn. Conf., San Francisco, 1989, pp. 57-62

24. Johansson, M., and T. Mattson, "Linearised High Efficiency Power Amplifier for PCN," Electronics Letters, Vol. 27, No. 9, 1991, pp. 762-764

25. Cox, D. C., "Linear Amplification with Nonlinear Components," IEEE Trans. Comm., Vol. 22, pp. 1942-1945, December 1974

26. Cox, D. C., and R. P. Leek, "Component Signal Separation and Recombination for Linear Amplification with Nonlinear Components," IEEE Trans. Comm., Vol. 23, pp. 1281-1287, Nov. 1975

27. Casadevall, "The LINC Transmitter," RF Design, pp. 41-48, February 1990

28. Moazzam, M. R., and C. S. Aitchison, "A low third-order intermodulation amplifier with harmonic feedback circuitry", IEEE MTT-S Int. Micr. Symp. Dig., San Francisco, CA, June 1996, pp. 827-830

29. Jing, D. et al, "New Linearization Method Using Interstage Second-Harmonic Enhancement", IEEE Microwave Guided Wave Letters, Vol. 8, No. 11, Nov. 1998, pp. 402-4

30. Hu, Y., J. C. Mollier and J. Obregon, "A New Method of Third-Order Intermodulation Reduction in Nonlinear Microwave Systems", IEEE Trans. MTT, Vol. MTT-34, No. 2, Feb. 1986, pp. 245-50

31. Jenkins, W. J., "Linearised Power Amplifiers", MSc Thesis, UCL, October 1997

32. Pedro, J. C. C. and Perez, J., "Accurate Simulation of GaAs MESFET Intermodulation Distortion Using a New Drain-Source Current Model", IEEE Trans. MTT, Vol. 42, Vo. 1, January 1994, pp. 25-31

33. Qu, G., Parker, A. E., "New Model Extraction for Predicting Distortion in HEMT and MESFET Circuits", IEEE-Microwave and Guided Wave Letters, Vol. 9, No. 9, Sept 1999

34. Higgins, J. A. and Kuvas, R. L., "Analysis and Improvement of Intermodulation Distortion in GaAs Power FETs", IEEE Trans. MTT, Vol. MTT-28, No. 1, January 1980, pp. 9-17

35. Youngoo, Y. and Bumman, K., "A new linear amplifier using low-frequency second-order intermodulation component feedforwarding", IEEE Microwave and Guided Wave Letters, Vol. 9, Issue 10 , Oct. 1999, pp. 419 -421

36. Vuolevi, J., et al, "Measurement technique for characterizing memory effects in RF power amplifiers", IEEE Radio and Wireless Conference, 2000, pp. 195 -198

37. Sevic, J. F., et al, "A Novel Envelope-Termination Load-Pull Method for ACPR Optimization of RF/Microwave Power Amplifiers", IEEE MTT-S Int. Micr. Symp. Digest, 1998, Vol. 2, pp. 723 -726

38. Jenkins, W. J. and Khanifar, A., "A Multicarrier Amplifier with Low Third-Order Intermodulation Distortion", IEEE MTT-S Int. Micr. Symp. Digest, 2000, Vol. 3, pp. 1515-1518

39. Modeste, W., et al, "Analysis and Practical Performance of a Difference Frequency Technique for Improving Multicarrier IMD Performance of RF Amplifiers", IEEE MTT-S on Technologies for Wireless Applications Digest, 1999, pp. 53-56

40. Pothecary, N., "Feedforward Linear Power Amplifiers", Artech House, 2000

41. PMC-Sierra, "Paladin-15", http://www.pmc-sierra.com