INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

**Ka-BAND VECTOR MODULATOR FOR LMCS
ADAPTIVE FEEDFORWARD LINEARIZER**

by

Douglas S. McPherson, B.Sc.E

A thesis submitted to the
School of Graduate Studies and Research
In partial fulfillment of the requirements for the degree of

Master of Engineering
in Electronics

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Electronics
Faculty of Engineering
Carleton University

January, 1998

©1998, Douglas S. McPherson, Ottawa, Canada

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitions et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-27020-3

Canada

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis

**“Ka-BAND VECTOR MODULATOR FOR LMCS
ADAPTIVE FEEDFORWARD LINEARIZER”**

submitted by

Douglas S. McPherson, B.Sc.E.

in partial fulfillment of the requirements of the degree of
Master of Engineering

Dr. J.S. Wight
Thesis Co-supervisor

Dr. M.G. Stubbs
Thesis Co-supervisor

Chair,
Department of Electronics

Carleton University

January 1998

Abstract

This work describes the development of a prototype Ka-band vector modulator for use in an LMCS adaptive feedforward linearizer. The device divides a 28 GHz RF signal into four orthogonal equi-amplitude components, amplifies each one by a prescribed amount, and adds the resultants in phase. By varying the gain in the four orthogonal channels, the transmission response can assume any phase and a continuous range of amplitudes. The vector modulator is a hybrid microwave circuit consisting of an alumina input phasesplitter, four MMIC variable gain amplifiers, and an alumina output power combiner. The design techniques, fabrication, and test results for the individual components and the final circuit are presented. Implemented, the vector modulator exhibits between -20 and -35 dB gain, less than -15 dB input return loss, and less than -12 dB output return loss between 27.35 GHz and 29.35 GHz.

The vector modulator is the primary RF component of a Ka-band adaptive feedforward linearizer. When operated correctly, this device will suppress distortion resulting from non-linear power amplification. The principles and requirements for the feedforward are reviewed. Based on this theory and on the measured vector modulator performance, the feasibility of the feedforward is assessed. The final calculations indicate that it is theoretically possible to reduce a power amplifier's level of intermodulation to signal power from -20 dB to -40 dB between 27.35 GHz and 27.85 GHz. The implementation of the entire wideband feedforward is also discussed.

Acknowledgements

I would like to take this opportunity to thank all the people who helped me complete this work. I am most grateful to Dr. J.S. Wight and Dr. M.G. Stubbs for making MMIC research available to me and for their valuable support and guidance. The field holds a great deal of appeal to me and I look forward to further work at the PhD level.

The research staff at the Communications Research Centre deserve a great deal of gratitude as well. I would like to thank C. Glaser for all her tireless efforts in the lab on my behalf, J. Bradley for his endless supply of practical advice, T. Laneve for his sober insight, H. Do-ky for answering my countless questions, and K. Verver for offering me many new ideas. I should add that John and Steve are to be commended for the tremendous amount of time and care spent assembling the vector modulator.

My thanks also go out to Dr. J. Cavers, S. Grant, and A. Smith at Simon Fraser University who taught me so much about feedforward linearization. I could not have fathomed it without their help.

Finally, I would like to acknowledge the unwavering support of Mum, Dad Jenny, and Gran. Mum and Dad's encouragement and counsel has been absolutely invaluable to me over these past two years. This work is dedicated to them.

Table of Contents

Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
List of Abbreviations and Acronyms	xii
List of Symbols	xiv
CHAPTER 1: Introduction	1
1.1 Broadband wireless communications	1
1.2 Local multipoint communications service (LMCS)	3
1.3 Linear power amplification	4
1.4 Thesis objectives	5
1.5 Thesis outline	6
1.6 Thesis contribution	6
CHAPTER 2: Background	8
2.1 Introduction	8
2.2 Microwave power amplifiers	9
2.2.1 Classification of microwave power amplifiers	9
2.2.2 Power amplifier efficiency	11
2.2.3 Amplifier gain compression	19
2.3 Characterization of amplifier distortion	22
2.3.1 AM/AM conversion	23
2.3.2 AM/PM conversion	30
2.3.3 Linear distortion	33
2.3.4 Non-linear distortion with memory	35
2.4 Linearization strategies	36
2.4.1 Cartesian feedback	37
2.4.2 Predistortion	40
2.4.3 Feedforward	44
2.5 Principles of adaptive feedforward	47
2.5.1 Adaptation of the signal cancellation coefficient	50
2.5.2 Adaptation of the error cancellation coefficient	52
2.5.3 Coefficient accuracy requirements	55
2.5.4 Delay mismatch and bandwidth	57
2.6 Broadband feedforward implementation considerations	58
2.7 Summary	63

CHAPTER 3: Vector modulator architectures **65**

3.1	Introduction	65
3.2	Phase split, amplitude adjust, and combine	65
3.3	Phase shift and amplitude adjust	67
3.4	Amplitude control	68
3.5	Review of active device technologies	70
3.6	Configuration selection	71
3.7	Summary	72

CHAPTER 4: Variable gain amplifier design **73**

4.1	Introduction	73
4.2	Fundamental amplifier design techniques	73
4.3	GaAs HBT description and modelling	81
4.3.1	Operation and description of the GaAs HBT	81
4.3.2	Large-signal HBT device model	87
4.4	Special design considerations	91
4.4.1	Bias network design	92
4.4.2	RF feedback design	95
4.4.3	Electromagnetic modelling of circuit elements	101
4.5	Matching networks	108
4.5.1	Input matching network	109
4.5.2	Output matching network	111
4.5.3	Interstage matching network	113
4.6	Simulated and measured results for the variable gain amplifier	115
4.6.1	Large-signal simulation	116
4.6.2	Amplifier stability	119
4.6.3	MMIC layout	121
4.6.4	Measured results	124
4.7	Summary	129

CHAPTER 5: Passive component design **130**

5.1	Introduction	130
5.2	Phasesplitter design	131
5.2.1	Description	131
5.2.2	Simulation and measurement	133
5.3	Wilkinson power combiner design	135
5.3.1	Description	135
5.3.2	Simulation and measurement	136
5.4	Passive element photomasks	138
5.5	Summary	142

CHAPTER 6: Vector modulator assembly and test	143
6.1 Introduction	143
6.2 Jig design	143
6.3 Measurements	145
6.4 Comments	150
6.5 Summary	155
CHAPTER 7: Conclusion	156
7.1 Summary	156
7.2 Future work	159
7.3 Final comments	160
REFERENCES	162
APPENDIX I: Fourier Expansions	165

List of Tables

Table	Description	Page
2.1	Parameters for Class A and B amplification	17
2.2	Power added efficiencies for equivalent class A, B and AB amplifiers	18
4.1	Conjugate matching coefficient and maximum gain for VGA gain blocks.	100
6.1	Bias states for the vector modulator.	147

List of Figures

Figure	Description	Page
1.1	High power wireless combiner.	2
1.2	Linear amplifier combiner.	2
2.1	Current waveforms for (a) class A (b) class B (c) class AB (d) class C PA.	10
2.2	Simplified power amplifier circuit for class A, B and AB.	11
2.3	Transfer characteristics of a class AB amplifier.	12
2.4	Output current waveform of a class AB amplifier.	14
2.5	Drain current for an overdriven class A amplifier.	20
2.6	Output power characteristic of a class A amplifier.	22
2.7	Non-linear transfer function showing gain compression.	25
2.8	Input and output spectrum for a third-order linearity.	28
2.9	Diagram showing third order intercept point.	28
2.10	Variation in output phase due to amplitude modulation.	33
2.11	Cartesian feedback amplifier.	38
2.12	Distortion cancellation model for a feedback amplifier.	39
2.13	Schematic of an adaptive predistortion linearizer.	41
2.14	Schematic of a feedforward linearizer.	45
2.15	Model of a feedforward linearizer.	48
2.16	Adaptation circuit for signal cancellation.	52
2.17	Adaptation circuit for distortion cancellation.	55
2.18	Signal cancellation loop with adaptive delay.	60
2.19	Feedforward linearizer with adaptive delay and partial gradient. ..	62
3.1	Schematic of a phase split, amplitude adjust and combine vector modulator.	65
3.2	Block diagram of a bi-phase amplitude controller.	67
3.3	Schematic of a phase shift and amplitude adjust vector modulator.	67
3.4	Analog phase shifter topology.	68
3.5	Variable attenuators, (a) T-network and (b) p-network.	69

3.6	Variable attenuator, reflection type.	69
4.1	A two-stage transistor amplifier with matching networks.	74
4.2	An N-port network characterized by S parameters.	74
4.3	A 2-port network excited by a generator at the input.	75
4.4	2 port network showing source and load reflection coefficients. . .	77
4.5	Load stability circle plotted on the Smith chart.	80
4.6	Steady-state velocity-field characteristics for electrons in GaAs and Si.	83
4.7	NPN transistor with associated energy-band diagram.	84
4.8	Nortel GaInP-GaAs HBT cross-section.	87
4.9	Standard BJT large-signal model (enclosing GP model).	88
4.10	Modified GP model for GaAs HBT.	89
4.11	DC I-V curves (a) v_{be} voltage source (b) current source at the base.	90
4.12	HBT Maximum available gain plotted versus V_{be} and V_{ce}	93
4.13	DC feedback bias loop.	94
4.14	DC-IV curves with DC feedback bias points superimposed.	95
4.15	The VGA gain block with RF feedback.	96
4.16	G_{max} versus R_{fb} and L_{fb} for the first gain block.	98
4.17	Γ_{sm} and Γ_{lm} versus R_{fb} and L_{fb}	98
4.18	RF feedback elements for (a) first and (b) second gain block. . . .	100
4.19	Nortel HBT interconnect layers and lumped elements.	102
4.20	1.2 nH choke inductor.	104
4.21	Comparison of spiral inductor models.	105
4.22	Lumped element model for a rectangular spiral inductor.	105
4.23	50 fF me1-me2 capacitor.	106
4.24	Comparison of me1-me2 capacitor models.	106
4.25	Complete feedback network for Em simulation.	108
4.26	Variable gain amplifier schematic showing matching networks. . .	109
4.27	Input matching network.	109
4.28	Input matching on the Smith chart.	110

4.29	Output matching network.	111
4.30	Output matching on the Smith chart.	112
4.31	Interstage matching network.	113
4.32	Interstage matching on the Smith chart.	114
4.33	Circuit simulator schematic of the variable gain amplifier.	115
4.34	Simulated VGA input reflection coefficient.	118
4.35	Simulated VGA transmission coefficient.	118
4.36	Simulated VGA output reflection coefficient.	119
4.37	Source and load stability circles at 17 GHz and $V_c = 2.7$ V.	120
4.38	MMIC layout of the variable gain amplifier.	122
4.39	Test set-up for on-wafer VGA characterization.	125
4.40	Measured VGA input reflection coefficient.	126
4.41	Measured VGA transmission coefficient.	126
4.42	Measured VGA output reflection coefficient.	127
4.43	VGA spectrum at the output port for $V_c = 2.9$ V.	128
5.1	Input phasesplitter schematic and layout.	132
5.2	Phasesplitter transmission response (a) simulated, (b) measured.	133
5.3	Simulated and measured phasesplitter matching results.	134
5.4	Output combiner schematic and layout.	135
5.5	Simulated and measured combiner transmission results.	137
5.6	Simulated and measured combiner matching results.	138
5.7	2" x 2" photomask for fabrication of passive elements.	139
5.8	Phasesplitter/Combiner with TRL calibration standards.	140
6.1	Vector modulator in test jig.	144
6.2	Schematic of vector modulator.	145
6.3	Vector modulator test setup.	146
6.4	Vector modulator transmission response at 28 GHz.	148
6.5	Vector modulator transmission response.	148
6.6	Vector modulator input return loss.	149
6.7	Vector modulator output return loss.	150
6.8	Roll-off insensitive feedforward linearizer with delay comp.	154

List of Abbreviations and Acronyms

AC	Alternating current
ACI	Adjacent channel interference
AlGaAs	Aluminium Gallium Arsenide
AM	Amplitude modulation
AM/AM	Amplitude modulation to amplitude modulation
AM/PM	Amplitude modulation to phase modulation
ATM	Asynchronous transfer mode
BJT	Bipolar junction transistor
BP	Bandpass
CITR	Canadian Institute for Telecommunications Research
CRC	Communications Research Centre
DAIE	Directorate of Antennas and Integrated Electronics
dBm	milliwatt expressed in decibel where one milliwatt equals zero decibel
DC	Direct current
DSP	Digital signal processor
EHF	Extremely High Frequency
Em	Refers to Sonnet™ Em version 4.0
FET	Field effect transistor
fF	Femtofarad
FM	Frequency modulation
GaAs	Gallium arsenide
GaInP	Gallium Indium Phosphide
GHz	Gigahertz
GP	Gummel Poon
HBT	Heterojunction bipolar transistor
HP	Hewlett Packard
IF	Intermediate frequency
IM	Intermodulation
IMD	Intermodulation distortion
IMSR	Intermodulation-signal ratio
InP	Indium Phosphide

IV	Current Voltage
LEC	Low expansion coefficient
LMCS	Local multipoint communication system
LO	Local oscillator
MESFET	Metal semiconductor field effect transistor
MHZ	Megahertz
MIM	Metal-insulation-metal
MMIC	Monolithic microwave integrated circuit
MODEM	Modulation/demodulation
MRIND	Microstrip rectangular inductor
MTFC	Microstrip thin film capacitor
nH	Nanohenry
Nortel	Northern Telecom Ltd.
PA	Power amplifier
PAE	Power added efficiency
PDF	Probability density function
pF	Picofarrad
PHEMT	Pseudomorphic high electron mobility transistor
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase shift keying
RF	Radio frequency
SATCOM	Satellite communications
Si	Silicon
SiON	Silicon Oxinitride
TEM	Transverse electromagnetic mode
TFC	thin film capacitor
Ti	Titanium
TRL	Thru-reflect-line
VGA	Variable gain amplifier
VSWR	Voltage standing wave ratio
W	Tungsten
WLAN	Wireless local area network

List of Symbols

A	Sinusoidal wave amplitude (V) <i>or</i> gain
A_1, A_2	Gain from two amplifiers in Cartesian feedback analysis
A_{IP3}	Voltage amplitude at third-order intercept point (V)
A_m	Stopband attenuation of a rectangular band rejection filter
a_0	Average or DC value of an output AC current waveform (mA)
a_1	Fundamental component of an output AC current waveform (mA)
a_n	Incident wave at the n th port
α	Complex gain of a vector modulator
α_0	Complex linear gain
α_{mn}	Coefficient of nonlinear function F_1 or F_2
B	Bandwidth (GHz)
β	DC Current Gain for bipolar transistor <i>or</i> feedback factor <i>or</i> complex gain
β_n	Coefficient of nonlinear function $G\{x_d(t)\}$
b_n	Wave reflected from the n th port <i>or</i> bias state of vector modulator
C_{bc}	Base-collector depletion capacitance (pF)
C_{bcx}	Extrinsic base-collector junction capacitance in modified GP model (pF)
C_{fb}	DC blocking capacitance for VGA (pF)
C_L, C_S	Centre of load, source stability circles
c_n	Coefficient in Taylor series expansion
$c_1(t), c_2(t)$	Uncorrelated random processes
C_{th}	Thermal capacitance in modified GP model (pF)
Δ	Derived quantity defined by equation (2.21)
ΔE_g	Barrier height at a heterojunction (eV)
D_α	Estimate of the covariance of v_e and v_m
D_β	Estimate of the covariance of v_0 and v_e
D_E, D_B	Diffusion coefficients in emitter and base respectively
ϵ_α	Adjustment error for α
ϵ_β	Adjustment error for β
f	Frequency (GHz)
f_{max}	Maximum frequency of oscillation (GHz)
f_T	Cutoff frequency or gain-bandwidth product (GHz)

G	Gain
G_{max}	Maximum gain
G_0	Linear gain
G_{1dB}	Gain at 1 dB compression point
$G(x_m)$	Amplitude dependent complex gain
γ	Complex gain of auxiliary amplifier <i>or</i> emitter efficiency
γ_n	Coefficient of nonlinear function $K\{x_d(t)\}$
Γ_{IN}	Reflection coefficient at a transducer input
Γ_L	Load reflection coefficient
Γ_{LM}	Load reflection coefficient for simultaneous conjugate match
Γ_{OUT}	Reflection coefficient at a transducer output
Γ_S	Source reflection coefficient
Γ_{SM}	Source reflection coefficient for simultaneous conjugate match
$H(j\omega)$	Frequency dependent transfer function
$h(t)$	Impulse response of $H(j\omega)$
$h_1(\tau_1)$	First order impulse response
$h_2(\tau_1, \tau_2)$	Second order impulse response
$\hat{I}$	Maximum current peak before clipping (mA)
I_b	Base current (mA)
I_C	Collector current (mA)
I_{ce}	Collector-emitter current (mA)
I_{cn}	Collector current (due to negative carriers) (mA)
I_{DS}	DC drain-source current (mA)
$i_{ds, fund}$	Fundamental AC drain-source current waveform (mA)
I_{DSQ}	Quiescent DC drain-source current (mA)
$i_{ds, peak}$	Peak AC drain-source current waveform (mA)
I_{DSS}	DC drain-source current when $V_{GS} = 0$ (mA)
I_{en}	Emitter current (due to negative carriers) (mA)
I_{ep}	Emitter current (due to positive carriers) (mA)
I_{MAX}	Maximum DC drain-source current (mA)
$IMSR_a$	Ratio of IM power to signal power at the amplifier output
$IMSR_0$	Ratio of IM power to signal power at the feedforward output

I_{rn}	Recombination current (due to negative carriers) (mA)
I_{rp}	Recombination current (due to positive carriers) (mA)
k	Boltzmann's constant (1.38×10^{-38} J/K)
K	Gain of vector modulator (Magnitude) <i>or</i> Stability factor
λ	Wavelength (cm or degrees)
L_{fb}	Nominal RF feedback inductor in VGA (nH)
L_{feed}	Parasitic inductance associated with HBT feeds in VGA (nH)
L_E	Diffusion length of holes in the emitter (cm)
n_{bo}	Electron carrier density in the base (cm^{-3})
P_a	Average power in amplifier output signal <i>or</i> generator power available (W)
P_{DC}	DC power consumption (W)
P_e	Average power in error signal (W)
p_{eo}	Minority hole carrier density in the emitter (cm^{-3})
P_{IN}	Input power (dBm)
$P_{IN, LIN}$	Linear input power (dBm)
P_{IP3}	Power at third-order intercept point (dBm)
P_L	Average power delivered to the load Z_L (W)
P_m	Average power in modulating signal (W)
P_m	Average power at feedforward output (W)
P_{OUT}	Output power (dBm)
$P_{RF, MAX}$	Maximum FET output power (dBm)
P_{ω_1}	Power at first frequency ω_1 (dBm)
$P_{2\omega_1 - \omega_2}$	Power at frequency $2\omega_1 - \omega_2$ (dBm)
q	Fundamental unit of charge (1.60×10^{-19} C)
Q	Decorrelation matrix
R_b	Base resistance in GP model <i>or</i> Base resistor in VGA (Ω)
R_{bi}, R_{ci}	Intrinsic base and collector resistances of GP model (Ω)
R_{bc}	Base-collector DC feedback resistor in VGA (Ω)
R_{be}, R_{ce}, R_{bc}	Leakage resistances of GP model (Ω)
R_C	Collector resistor in VGA (Ω)
R_{fb}	Nominal RF feedback resistor in VGA (Ω)
r_L, r_S	Radius of load, source stability circles

R_L	Load resistance (Ω)
R_{th}	Thermal resistance in modified GP model (Ω)
S_{ij}	Non-generalized S parameter
S_{ij}'	Generalized S parameter
σ_{em}	Covariance of v_e and v_m
σ_{oe}	Covariance of v_0 and v_e
t	Time
t_1, t_2	Characteristic delays associated with first and second harmonics
T	Signal period <i>or</i> generalized transfer function <i>or</i> temperature (K)
τ_1	Group delay of PA in feedforward analysis
τ_α	Delay mismatch in the signal cancellation loop
τ_β	Delay mismatch in the error cancellation loop
t_d	Group delay
τ_{ec}	Emitter to collector transit time (ps)
$\theta(A)$	Amplitude dependent phase (rad)
U	Unilateral gain (dB)
$v_a(t)$	Amplifier output response (V)
V_b	Base voltage in VGA (V)
V_{be}	Base-emitter voltage (V)
V_c	Collector voltage in VGA (V)
V_{ce}	Collector-emitter voltage (V)
v_d	Instantaneous distortion voltage (V)
v_{d0}	Instantaneous distortion voltage after one loop of feedback circuit
v'_{d0}	Distortion at amplifier output without Cartesian feedback
V_{DGB}	DC drain-gate voltage at breakdown point (V)
V_{DS}	DC drain-source voltage (V)
$v_{ds}(t)$	AC drain-source voltage (V)
$V_{DS, MAX}$	DC drain-source voltage when V_{GS} is at its peak negative value (V)
$v_{ds, peak}$	Maximum AC drain-source signal swing (V)
V_{DSQ}	Quiescent DC drain-source voltage (V)
$v_e(t)$	Instantaneous error signal in feedforward analysis (V)
V_{eb}	Non-linear source modelling emitter-base tunnelling in GP model (V)

V_{GS}	DC gate-source voltage (V)
$v_{gs}(t)$	AC gate-source voltage (V)
V_{GSQ}	Quiescent DC gate-source voltage (V)
v_{in}	Instantaneous input voltage (V)
$V_{in}(j\omega)$	Fourier transform of v_{in}
$v_{ik}(t)$	Orthogonal output from phasesplitter (V)
V_K	Knee voltage (V)
$v_m(t)$	Modulating signal for linearizer analysis (V)
V_n	Potential seen by base electrons (V)
V_n^+	Wave incident on port with common, real reference impedance (V)
V_n^-	Wave reflected from port with common, real reference impedance (V)
v_{nb}	Average velocity of the electrons in the base (cm s^{-1})
$v_0(t)$	Instantaneous feedforward output signal (V)
v_{out}	Instantaneous output voltage (V)
$V_{out}(j\omega)$	Fourier transform of v_{out} (V)
V_P	Gate pinch-off voltage (V) <i>or</i> potential seen by holes (V)
V_ϕ	Small forward bias applied to gate-source contact (V)
v_{pe}	Average velocity of the holes in the emitter (cm s^{-1})
V_{pe}	Temperature controlled voltage source in modified GP model (V)
V_S	Amplitude of source voltage waveform (V)
u	Derived quantity defined by equation (2.30)
W	Width of base (μm)
ω	Frequency (rad/s)
ω_1, ω_2	Fundamental frequencies 1 and 2 (rad/s)
ω_m	Modulation frequency (rad/s)
x_{in}	Instantaneous input power (W)
x_{out}	Instantaneous output power (W)
Z_{IN}	Impedance seen looking into a transducer input
Z_L	Load impedance (Ω)
Z_0	Reference impedance (Ω)
Z_{OUT}	Impedance seen looking into a transducer output
Z_S	Source impedance (Ω)

CHAPTER 1: Introduction

1.1 Broadband wireless communications

In recent years, the world has witnessed explosive growth in the use of wireless communications, and the demand appears unlikely to subside. The next generation of systems endeavour to provide enhanced communication services such as data transfer, electronic mail, high resolution digital video and even full multimedia communication. Such services will require access to much greater amounts of radio spectrum than is currently available at L and X bands which cover 1-2 GHz and 8-12 GHz respectively. The main alternative is the Ka-band which encompasses frequencies ranging from 26 GHz to 40 GHz.

There are three major projects under way in Canada to take advantage of the upsurge in broadband wireless communications: advanced satellite communications (Advanced SATCOM), local multipoint communications systems (LMCS), and wireless local area networks (WLAN). All three are currently exploring the means and technology required to provide high frequency broadband services to a variety of commercial and private subscribers.

The successful implementation of broadband wireless systems requires considerable technical innovation in the areas of EHF propagation modelling, anti-multipath modulation, spectrum management, network control, multiple access protocol, and radio frequency (RF) technology. Underlying all of these aspects is the need for dynamic or adaptive capability. That is, the system parameters must be able to change real-time in accordance with varying user demands. This necessitates the use of components and sub-systems which allow for virtually instantaneous frequency re-planning of selected parts (or even all) of the wireless network [1]. From an RF point of view, the method used to combine multiple signals has a serious effect on the adaptability of the system. Many multi-channel systems currently in use accomplish the task of power combining using sep-

arate modulation and transmission chains for each narrowband channel [2]. As shown in Figure 1.1, the outputs of these chains are combined at high power using mechanically tuned resonant filters. Although the approach is excellent at suppressing adjacent channel interference (ACI), it is impossible to change frequency assignments without manually retuning the combiner. Moreover, about half of the transmit power is lost in this type of combiner.

Figure 1.1 High power wireless combiner.

These problems can be mitigated by combining the signals at low power and amplifying the sum with a single linear amplifier as shown in Figure 1.2. The combiner can be realized in many different technologies such as broadband stripline or microstrip. The task of channel reallocation is achieved, not by manual re-tuning, but by individual electronically controlled receiver circuits that are connected to the combiner. With recent advances in high speed digital signal processing (DSP), the means of allocating channels and bandwidth has become even more sophisticated. A DSP chip can output

Figure 1.2 Linear amplifier combiner.

a combined broadband signal at an intermediate frequency (IF) which can then be upconverted to RF and amplified by a linear amplifier. Using this scheme, frequency planning can be controlled entirely by software.

Both of the low power combining techniques allow for rapid and automatic reprogramming of the wideband signal. Whether analog or digital, the system can adapt frequency assignments without interruption to the users and without mechanical tuning. As the following section will indicate, the success or failure of this approach can be, to a large extent, dependent on the linearity of the power amplifier (PA). Even a slightly non-linear characteristic in the PA can lead to severe problems such as adjacent channel interference and increased bit error rates.

1.2 Local multipoint communications service (LMCS)

The work contained in this thesis was initiated as part of a research project aimed at developing an LMCS system concept and was sponsored by the Canadian Institute for Telecommunications Research (CITR). In order to place the work in the correct context, a review of the project aims and their implications is included here.

LMCS is a broadband wireless telecommunications common carrier service in the 28 GHz range that will be capable of carrying basic and advanced communication services such as “wireless” cable TV, Internet access, video teleconferencing and various other multimedia products. At this point in time, Industry Canada has licensed 1 GHz of bandwidth in sixty-six major urban markets and 127 smaller communities [3]. This can be broken down into two 500 MHz blocks ranging from 27.35 GHz to 28.35 GHz. It is anticipated that another two blocks will be added bringing the maximum frequency up to 29.35 GHz. The system is based on a cellular arrangement of nodes designed to transmit and receive signals to and from homes and businesses within a 4-5 km radius. Several such nodes or hubs will be required in each city to service the total geographic area. Each one will be connected to an ATM backbone via fibre optic cable or a chain of

wireless repeater stations. Subscribers will send and receive signals to and from the hubs using a small unobtrusive antenna connected to a compact transceiver unit.

The CITR LMCS project is intended to be pre-competitive by developing an advanced system offering adequate and cost-effective coverage to subscribers over large areas and under varied geographical conditions [4]. In keeping with these objectives, the modulation/demodulation (MODEM) techniques employed must be simultaneously efficient in their use of the available bandwidth and resistant to channel fading. At this point in time, linear modulations with adaptive equalization are considered likely candidates. Examples such as quadrature phase shift keying (QPSK) and quadrature amplitude modulation (QAM) have demonstrated superior spectral efficiency when compared to nonlinear frequency modulation (FM) methods [4]. With appropriate channel characterization and judicious choice of the adaptive equalization parameters, the MODEMs should be resistant to the expected fading effects.

Unfortunately, linearly modulated RF signals have a fluctuating envelope that is susceptible to distortion and spectral spreading. The only way to preserve the envelope shape is if the signal is amplified by a PA wherein P_{OUT} varies linearly with P_{IN} . Without such a PA, spectral spreading will occur and this can lead to adjacent channel interference and increased bit-error rates [5][6]. Thus, if the flexibility of the low power combiner is to be realized along with spectral efficiency and multipath resistance, linear amplification is absolutely essential.

1.3 Linear power amplification

The degree of linearity available in the transmitter PA is a variable which can be controlled. However, to do so requires making sacrifices such as reduced efficiency and increased cost.

The simplest solution to the linearity problem is to employ a conventional class A amplifier and to back-off the input power. Unfortunately, the required amount of

power back-off can be considerable (25 to 30 dB), and the efficiency very low (1 to 2%) [7]. Moreover, to achieve a high level of average transmit power will require the use of a large and expensive power amplifier. Another alternative is to linearize a nonlinear class AB amplifier using either a negative feedback, predistortion, or feedforward technique. The choice of approach depends on the bandwidth, minimum acceptable efficiency, degree of required linearity, and cost. For wideband multicarrier applications such as LMCS, it turns out that feedforward is the only scheme inherently capable of linearizing wideband multicarrier signals. The comparison between schemes is carried out in further detail in section 2.4.

One of the principle costs associated with any linearization scheme is the need for additional RF circuitry. In the case of the adaptive feedforward proposed in this thesis, the amount of additional components is significant and it is preferable that all extra components be as cheap as possible and highly reliable. Assuming high volume production, monolithic microwave integrated circuits (MMICs) are the best alternative and are fully capable of operating in the bands of interest for high frequency broadband applications. Among the choices of technologies and devices available, the Gallium Arsenide (GaAs) heterojunction bipolar transistor (HBT) and pseudomorphic high electron mobility transistor (PHEMT) offer the best high power and low noise performance at the Ka-band. These devices have been utilized successfully in mixers, power amplifiers, oscillators, switches, and even for entire radio front-ends.

1.4 Thesis objectives

The purpose of this thesis is two-fold. The first objective is to present and describe a concept for an adaptive feedforward linearizer suitable for the LMCS application as described in section 1.2. The second objective is to investigate the possibility of using MMICs to realize one of the key RF components of the linearizer: a vector modulator. The challenges faced in the design and implementation of the vector modulator will

serve as a point of reference with which to assess the overall capability and feasibility of PA linearization at 28 GHz.

1.5 Thesis outline

The two thrusts in section 1.4 have been accommodated by dividing this thesis into two main segments. The first segment, comprising only chapter 2, reviews the characteristics of the non-linear power amplifier, characterizes the various forms of distortion, introduces the concept of PA linearization, establishes feedforward as the most practical of the possible strategies, and concludes with a section on the principles of adaptive feedforward. Also discussed in chapter 2 are the means by which this adaptivity can be applied to broadband multicarrier signals. The second segment traces the design of the key RF component of the feedforward, a Ka-band vector modulator. The task is divided among four chapters. Chapter 3 presents two possible configurations for the vector modulator and selects one. The rationale and implications of this choice are also outlined. Chapters 4 and 5 describe the components of the vector modulator in detail by means of background theory, step-by-step design procedures, circuit layouts, and electromagnetic simulation. Chapter 6 draws upon the development work in the previous chapters to present the final vector modulator design. Its operating bandwidth and dynamic range are assessed by means of electromagnetic simulation of all components. The seventh and final chapter summarizes the work on the vector modulator, suggests future projects, and makes some general comments on the challenges of high frequency wideband communications. Completing the design and implementation of the entire feedforward is by no means the intent of the work documented in this thesis. There remains a great deal of work to be done before a full demonstration could be assembled.

1.6 Thesis contribution

This thesis represents the first consideration of a wideband adaptive feedforward at 28 GHz. The cost and performance of the RF components is vital to the success

or failure of the eventual linearizer. The main question which this thesis endeavours to answer is whether or not it is possible to design and implement an RF chip-set capable of satisfying the requirements of an adaptive feedforward. The design of the vector modulator provides an excellent example of the problems which must be surmounted.

CHAPTER 2: Background

2.1 Introduction

In multicarrier wireless communications such as LMCS, linear power amplification is of the utmost importance. The ability to perform dynamic channel allocation, combat multipath delay spread, and conserve spectrum is contingent on the use of linear modulation formats and broadband linear power amplification. Unfortunately, all RF amplifiers contain some amount of power-dependent or nonlinear behaviour. If the envelope of the signal to be amplified has large amplitude fluctuations, it can induce significant variations in the PA transmission characteristics. With different portions of the signal now amplified by different amounts, the output will be a distorted replica of the input.

Of all the possible consequences of nonlinear amplification, the generation of intermodulation distortion (IMD) is perhaps the most serious. Even small amounts of IMD can cause unacceptable levels of interference with adjacent channels. Furthermore, since the majority of the intermodulation power appears at frequencies near the carriers, it is almost impossible to remove it once it is generated. It can only be reduced by backing-off the input power or by linearizing the amplifier.

The intent of this chapter is to identify the fundamental causes of non-linearity in power amplification and to present a wideband multicarrier linearization scheme capable of dealing with the resulting distortion. In order to support the transition from problem to proposed solution, sections on distortion classification and alternative linearization techniques are also included. The chapter concludes by presenting a schematic for a wideband feedforward which includes adaptive controllers and variable delay compensation.

2.2 Microwave power amplifiers

2.2.1 Classification of microwave power amplifiers

Wireless transmitting systems require that significant amounts of signal power be supplied to the transmitting antenna. This sort of power is only possible when transistors are biased at high current level and driven by large input signals. The amplifiers used for this application are described as large-signal or power amplifiers (PA). Unlike small signal or low noise amplifiers which have low power input signals, distortion can become appreciable. This occurs because the input signals can drive the amplifier into its cutoff or saturation regions causing the output to be a “clipped” version of the input. PAs also require sizable DC currents. These are a considerable drain on the system and it is desirable to ensure that the maximum amount of useable output power is obtained for the least amount of input power (i.e. the amplifier is efficient).

In practise, it is not possible to achieve conditions of low distortion and high efficiency simultaneously as one comes at the cost of the other. This trade-off is presented in greater detail in section 2.2.2. In order to distinguish power amplifiers, they are labelled as class A, B, AB, or C according to the output current waveform that results when an input signal is applied. These waveforms are illustrated in Figure 2.1 for a field effect transistor (FET) [8]. The class A power amplifier, Figure 2.1(a), is biased such that the DC current I_{DS} is greater than or equal to the amplitude of the output signal current. As a result, the transistor is able to conduct for the entire cycle of the input signal and has a conduction angle of 360° . The class B amplifier (Figure 2.1(b)), is biased at zero DC current and conducts for only half a cycle or 180° . The class AB amplifier (Figure 2.1(c)) represents a compromise between class A and class B with a conduction angle somewhere between 360° and 180° . Finally, class C amplifiers (Figure 2.1(d)) have conduction angles of less than 180° .

By simple inspection of the waveforms in Figure 2.1, it is clear that class A generates the least amount of distortion because its output is a scaled version of the input.

Figure 2.1 Current waveforms for (a) class A (b) class B (c) class AB (d) class C PA.

By comparison, class C amplifiers generate the most distortion because of the high degree of clipping. However, since this amplifier class has such a low conduction angle, the power supply outputs a zero current for most of the input signal swing. On average, this current will be substantially less than the DC current required for class A operation. Class C should, accordingly, be more efficient. As a final note, class C amplifiers will not be included in any subsequent analysis because the level of distortion they introduce is too high. Linearization of class C amplifiers is only possible for single-carrier narrowband signals.

2.2.2 Power amplifier efficiency

Numerous analyses of PA efficiency are available in the literature and in microwave amplifier textbooks. The following is based on that developed by J.L.B. Walker [9] with references to the thesis of Somsack Sychaleun [10]. Figure 2.2 shows a generic amplifier configuration which can be applied, without change, to either class A, B, or AB operation. It is comprised of a unipolar transistor, a resonant RLC circuit, two RF inductor chokes and two DC blocking capacitors. At Ka-band, the transistor will most likely be a GaAs PHEMT device. As far as this analysis is concerned, it is treated exactly the same as a FET. The device has infinite input and output impedance with DC-IV characteristics given by Figure 2.3. The implication of this approximation is that the microwave performance is being predicted using the transistor DC characteristics. Although this is a crude assumption, it is sufficient to establish that the efficiency is dependent on the mode of operation. The RLC network connected to the output is used to filter out higher order harmonics generated in class B and AB operation. The DC blocks are short circuits at the fundamental frequency and serve to direct DC to the device terminals. The RF chokes isolate the RF signals from the power supplies while passing the DC bias unimpeded.

Figure 2.2 Simplified power amplifier circuit for class A, B and AB.

The description of the class AB amplifier is sufficiently general in nature that it applies equally to class A and B amplifiers as well. In fact, it turns out that the two latter types are merely extreme cases of the general class AB amplifier. From Figure 2.3. The transistor is biased at xI_{MAX} , $V_{DS} = V_{DSQ}$, and $V_{GS} = V_{GSQ}$. At this point, the DC drain current is a fraction x of its maximum. The value of x determines the class of operation and ranges from 0.5 for class A to 0 for class B. Class AB amplifiers can assume any value of x between these two extremes. The peak current I_{MAX} is slightly higher than the GaAs FET data sheet specification of I_{DSS} for $V_{GS} = 0$ and is obtained by applying a small forward bias V_ϕ to the gate-source Schottky contact. This will increase the size of the channel opening and allow a current greater than I_{DSS} to flow. Depending on the device, this may or may not be viable. Nonetheless, the following analysis provides for the possibility anyway.

To obtain the maximum power output, the instantaneous value of $v_{ds}(t)$ must swing from the knee voltage V_K to a point just below the onset of drain-source ava-

Figure 2.3 Transfer characteristics of a class AB amplifier.

lanche breakdown. This voltage is normally specified in terms of a maximum drain to gate voltage V_{DGB} with the gate biased at pinch-off, $V_{GS} = -|V_P|$. The drain-source voltage reaches its maximum value $V_{DS,MAX}$ when the gate-source voltage is at its peak negative value of $V_\phi - 2(V_\phi - V_{GSQ})$. This latter limit is deduced from Figure 2.3 based on the fact that for maximum signal swing, $v_{gs}(t)$ should cover all values between the quiescent point V_{GSQ} and V_ϕ in one half cycle. It should be noted that drain-source breakdown does not present an absolute limit as it can be exceeded without causing any short term harm to the device [11]. However, over a long period of time, the continual stress on the device may cause a premature failure.

The gate-source quiescent point can be derived from the I_{DS} versus V_{GS} curve of Figure 2.3 using similar triangles.

$$V_{GSQ} = -(1-x)|V_P| + xV_\phi \quad (2.1)$$

As mentioned previously, the maximum drain-source voltage occurs when V_{GS} is at its peak negative value and the device is bordering on avalanche breakdown. Observing that $V_{DS} = V_{GS} + V_{DGB}$, the maximum possible output voltage is

$$\begin{aligned} V_{DS,MAX} &= [V_\phi - 2(V_\phi - V_{GSQ})] + [V_{DGB}] \\ &= V_{DGB} - 2(1-x)|V_P| - (1-2x)V_\phi \end{aligned} \quad (2.2)$$

Using this result, it is also possible to express the maximum output voltage swing $v_{ds,peak}$

$$\begin{aligned} v_{ds,peak} &= (V_{DS,MAX} - V_K)/2 \\ &= [V_{DGB} - 2(1-x)|V_P| - V_K - (1-2x)V_\phi]/2 \end{aligned} \quad (2.3)$$

and the drain-to-source quiescent voltage

$$\begin{aligned} V_{DSQ} &= (V_{DS,MAX} + V_K)/2 \\ &= [V_{DGB} - 2(1-x)|V_P| + V_K - (1-2x)V_\phi]/2 \end{aligned} \quad (2.4)$$

Linear RF output power is obtained when both the output voltage and current are pure single-frequency sinusoids and where an n-dB increase in input power results in an n-dB increase in output power. The linear RF power delivered to the load is maxi-

Figure 2.4 Output current waveform of a class AB amplifier.

mized when both the voltage and the current at the drain assume their peak values simultaneously.

$$P_{RF, MAX} = \frac{v_{ds, peak} \cdot i_{ds, fund}}{2} \quad (2.5)$$

where $i_{ds, fund}$ is the peak drain current at the fundamental frequency of the output current waveform, $i_{ds}(t)$. Unless the amplifier is operated as class A, $i_{ds, fund}$ will always be less than $i_{ds, peak}$.

The DC power dissipated by the transistor is given by straight multiplication of the quiescent voltage and current values

$$P_{DC} = V_{DSQ} I_{DSQ} \quad (2.6)$$

Where I_{DSQ} is the DC component of the output current waveform. The values of $i_{ds, peak}$ and I_{DSQ} can both be extracted from $i_{ds}(t)$ using Fourier series expansion. By judicious selection of the time scale, the output current waveform of Figure 2.3 can be represented with even symmetry as shown in Figure 2.4. It is periodic over the interval $(-T/2, T/2)$, equal to zero in the intervals $(-T/2, -t_0)$ and $(t_0, T/2)$, and defined by the following piece-wise equation

$$i_{ds}(t) = \begin{cases} 0 & -T/2 \leq t < -t_0 \\ xI_{MAX} + i_{ds, peak} \cos(\omega t) & -t_0 \leq t < t_0 \\ 0 & t_0 \leq t < T/2 \end{cases} \quad (2.7)$$

Here, the fundamental frequency (ω) is equal to $2\pi/T$ and $i_{ds, peak}$ can be expressed as $(I_{MAX} - xI_{MAX})$ or $(1 - x)I_{MAX}$. The zero crossings can be found by solving for t_0 when $i_{ds}(t) = 0$

$$xI_{MAX} + (1 - x)I_{MAX} \cos(\omega t_0) = 0 \quad (2.8)$$

Re-arranging yields the following solutions for t_0 and $-t_0$

$$t_0 = \pm \frac{1}{\omega} \cos^{-1} \left(-\frac{x}{1 - x} \right) \quad (2.9)$$

Since the waveform has even symmetry and the amplifier has a filter to remove higher order harmonics ($2\omega, 3\omega, \dots$), there are only two components to the Fourier series that need to be considered.

$$i_{ds}(t) = a_0 + a_1 \cos \omega t \quad (2.10)$$

The coefficient a_0 is the average or DC value of the output current, and hence, it must be equal to I_{DSQ} . The second coefficient a_1 represents the amplitude of the fundamental component of the output current or $i_{ds, fund}$. From the definition of the Fourier series, the coefficients for a periodic function with even symmetry are:

$$a_0 = \frac{2}{T} \int_0^{T/2} i_{ds}(t) dt \quad (2.11)$$

$$a_1 = \frac{4}{T} \int_0^{T/2} i_{ds}(t) \cos(\omega t) dt \quad (2.12)$$

Substituting equations (2.7) and (2.9) and evaluating the integrals yields the following pair of Fourier coefficients

$$a_0 = \frac{I_{MAX}}{\pi} \left[\sqrt{1 - 2x} + x \cos^{-1} \left(-\frac{x}{1 - x} \right) \right] \quad (2.13)$$

$$a_1 = \frac{I_{MAX}}{\pi} \left[\frac{x\sqrt{1-2x}}{1-x} + (1-x)\cos^{-1}\left(-\frac{x}{1-x}\right) \right] \quad (2.14)$$

The details of these expansions are shown in Appendix I. Having determined the amplitude of the DC term and the first harmonic of the output current waveform, equations (2.5) and (2.6) can now be updated as follows:

$$P_{RF, MAX} = \frac{v_{ds, peak} \cdot a_1}{2} \quad (2.15)$$

$$P_{DC} = V_{DSQ} \cdot a_0 \quad (2.16)$$

The power added efficiency of an amplifier is defined as the change in RF power from input to output divided by the DC power consumed in the process

$$PAE = \frac{P_{RF} - P_{IN}}{P_{DC}} \quad (2.17)$$

The inclusion of the input power in the above definition is intended to offset a bias in the definition towards low-gain devices. If the efficiency were defined as the ratio of output RF power to DC power only, amplifiers with low gain would appear artificially more efficient. The definition would not account for the fact that the majority of the observed output power comes from transferred input power. Equation (2.17) can be written in an alternative form

$$PAE = \left(1 - \frac{1}{G}\right) \frac{P_{RF}}{P_{DC}} \quad (2.18)$$

Where the *linear* power gain is $G = \frac{P_{RF}}{P_{IN}}$ (2.19)

Substituting equations (2.15) and (2.16) into (2.18), the efficiency for a class A, B or AB amplifier is expressed

$$PAE = \frac{1}{2} \cdot \left(1 - \frac{1}{G}\right) \cdot \frac{(1-\Delta)}{(1+\Delta)} \cdot \frac{a_1}{a_0} \quad (2.20)$$

where

$$\Delta = \frac{V_K}{V_{DGB} - 2(1-x)|V_P| - (1-2x)V_\phi} \quad (2.21)$$

To verify this result, consider the cases of class A and class B amplification. Substituting $x = 0.5$ and $x = 0$ into equations (2.13), (2.14), and (2.21) the equations specific to these respective amplifiers are obtained

Table 2.1 Parameters for Class A and B amplification

class	a_0	a_1	Δ
A, $x = 0.5$	$\frac{I_{MAX}}{2}$	$\frac{I_{MAX}}{2}$	$\alpha = \frac{V_K}{V_{DGB} - V_P }$
B, $x = 0$	$\frac{I_{MAX}}{\pi}$	$\frac{I_{MAX}}{2}$	$\beta = \frac{V_K}{V_{DGB} - 2 V_P - V_\phi}$

Using these expressions, the power added efficiency for a class A and a class B amplifiers are as follows:

$$PAE_A = \frac{1}{2} \cdot \left(1 - \frac{1}{G}\right) \cdot \frac{(1-\alpha)}{(1+\alpha)} \cdot 100\% \quad (2.22)$$

$$PAE_B = \frac{\pi}{4} \cdot \left(1 - \frac{1}{G}\right) \cdot \frac{(1-\beta)}{(1+\beta)} \cdot 100\% \quad (2.23)$$

These equations agree with those developed by J.B. Walker [9]. Using both the specific equations and the general class AB equation, a comparison of the power added efficiency of all classes is possible. Consider an amplifier with the following parameters: $V_K = 1$ V, $V_{DGB} = 20$ V, $V_\phi = 0.7$ V, and $V_P = 4$ V. Table 2.2 presents the power added efficiencies for a set of amplifiers with these common characteristics. The value of x is varied from 0.5 to 0.

Table 2.2 Power added efficiencies for equivalent class A, B and AB amplifiers

x	PAE ($G = \infty$)	PAE ($G = 5$)
0.5	44.1%	35.3%
0.4	53.7%	43.0%
0.3	59.0%	47.2%
0.2	62.3%	49.9%
0.1	64.4%	51.6%
0.0	65.8%	52.6%

The first set of values for power added efficiency are for the case of infinite gain. These represent the maximum *theoretical* efficiencies achievable using the transistor specified above. Class A is the least efficient because it draws the most current from the power supply to maintain the proper bias conditions. Curiously, it would appear that the efficiency improves markedly even for a small deviation from class A operation. There is a 10% improvement in the efficiency of a class AB amplifier with $x = 0.4$. As expected, the class B amplifier is most efficient and is 22% better than the equivalent class A. This is only logical, because the DC power supply spends half a cycle turned off.

The second set of values are intended to give a more realistic impression of power added efficiency. For an amplifier with $G = 5$, these results are nonetheless optimistic. The device is being driven at its operating limits, and in order to avoid long term reliability problems, it may be necessary to back-off the input power. The resulting decrease in P_{RF} will in turn degrade the PAE. Finally, the results overstate the efficiency because they do not include the reduction in P_{RF} due to circuit losses.

Assuming the transistor can be driven at its limits and there are no circuit losses, the power added efficiencies given by Table 2.2 can only be achieved by presenting an optimum load impedance to the output. The optimal load will ensure that P_{RF} is maximized to $P_{RF, MAX}$. Referring to equation (2.5), it is clear that it should be

$$R_L = v_{ds, peak}/a_1 \quad (2.24)$$

This load impedance is not necessarily the value required to achieve good output match. It is more than likely that the maximum efficiency will have to be sacrificed in order to achieve an adequate matching impedance.

As a final comment, it should also be pointed out that for a given gain and input power, a class B amplifier outputs half the RF power of a class A amplifier. Hence, the results presented in Table 2.2 are slightly misleading. For a class B amplifier to be of any practical use, it must have at least 10-dB gain in class A operation [9]. At Ka-band, where gain is scarce, class B operation is almost completely unviable. In practise, Class AB biased at $x = 0.25$ or greater represents the most efficient amplifier possible.

2.2.3 Amplifier gain compression

If $v_{ds}(t)$ swings above $V_{DS, MAX}$ and below V_K without causing any short term damage to the device, the RF output power can be increased beyond $P_{RF, MAX}$. However, the increased voltage cannot induce a commensurate rise in the drain current because of the limit existing at I_{MAX} . The result is the distorted drain-current shown in Figure 2.5. For simplicity, the following discussion will be limited to the case of a class A amplifier. A similar analysis can be developed for class AB and class B amplifiers, but it is

Figure 2.5 Drain current for an overdriven class A amplifier.

more complicated and yields the same general conclusions. The distorted current waveform of Figure 2.5 is expressed in piece-wise form as shown on the following page.

$$i_d(t) = \begin{cases} -I_{MAX}/2 & -T/2 \leq t < -t_2 \\ \hat{I} \cos(\omega t) & -t_2 \leq t < -t_1 \\ I_{MAX}/2 & -t_1 \leq t < t_1 \\ \hat{I} \cos(\omega t) & t_1 \leq t < t_2 \\ -I_{MAX}/2 & t_2 \leq t < T/2 \end{cases} \quad (2.25)$$

Where t_1 is determined from the point at which the sine wave of amplitude $\hat{I}$ reaches the cutoff point $I_{MAX}/2$.

$$\hat{I} \cos(\omega t_1) = \frac{I_{MAX}}{2} \quad (2.26)$$

Solving for t_1

$$t_1 = \frac{1}{\omega} \cos^{-1} \left(\frac{I_{MAX}}{2\hat{I}} \right) \quad (2.27)$$

For clarity, the argument of the inverse cosine will be replaced by the variable u . The remaining unknown t_2 is related to t_1 and must be $(T/2 - t_1)$. As before, the current waveform can be expanded as a standard Fourier series with even symmetry:

$$I_{ds}(t) = a_0 + a_1 \cos(\omega t) \quad (2.28)$$

Again, all higher order terms are neglected owing to the filter at the amplifier output. The bias term a_0 can also be ignored because the waveform is symmetrical about the time axis. Carrying out the Fourier expansion yields the following coefficient

$$a_1 = \hat{I} \left[1 - \frac{2}{\pi} \cos^{-1}(u) - \frac{1}{\pi} \sin \{ 2 \cos^{-1}(u) \} \right] \quad (2.29)$$

The details of this expansion are documented in Appendix I. The variable $\hat{I}$ represents the drain current that would flow if the I_{MAX} limit did not exist. Thus, it is linearly proportional to v_{gs} and proportional to the square root of the input power P_{IN} , or $\hat{I} = k\sqrt{P_{IN}}$ where k is a constant. If $P_{IN, LIN}$ is defined as the input power corresponding to v_{gs} at its peak value just before the onset of current cutoff, then $\hat{I} = I_{MAX}/2$ when $P_{IN} = P_{IN, LIN}$. Based on these observations, the variable u is expressed as follows

$$u = I_{MAX}/2\hat{I} = \sqrt{P_{IN, LIN}/P_{IN}} \quad (2.30)$$

Since the filter eliminates all but the fundamental frequency, the output power is defined

$$P_{OUT} = a_1^2 R_L / 2 \quad (2.31)$$

Where R_L is the load impedance for maximum power transfer. The amplifier's output response can be divided into two operating regions depending on the output power level. Below the current limit, there is no distortion of the current waveform. The amplitude of the fundamental is equal to $\hat{I}$ where $\hat{I} \leq I_{MAX}/2$. Furthermore,

$$\hat{I}^2 R_L / 2 = G P_{IN} \quad (2.32)$$

Where G is the linear gain. Substituting equation (2.32) into (2.31), the output power in the linear region, before the onset of cut-off, is

$$P_{OUT} = G P_{IN} \leq P_{RF, MAX} \quad P_{IN} \leq P_{IN, LIN} \quad (2.33)$$

Once $\hat{I} \geq I_{MAX}/2$, some of the output power at the fundamental is lost because it appears as higher order harmonics. Consequently, it no longer increases in direct proportion to the input power. Substituting equation (2.29) and (2.32) in equation (2.31) the output power is now given by

Figure 2.6 Output power characteristic of a class A amplifier.

$$P_{OUT} = GP_{IN} \left(1 - \frac{2}{\pi} \cos^{-1} \sqrt{P_{IN,LIN}/P_{IN}} - \frac{1}{\pi} \sin(2 \cos^{-1} \sqrt{P_{IN,LIN}/P_{IN}}) \right)^2 \quad (2.34)$$

$$P_{IN} \geq P_{IN,LIN}$$

Using equations (2.33) and (2.34), it is possible to describe the output power of the amplifier as a function of the input power in both linear and non-linear regions. A plot of this power transfer function is shown in Figure 2.6. Below $P_{IN,LIN}$ the output power is proportional to the input power with a constant of proportionality equal to G . As the input power is increased, distortion occurs and higher order harmonics appear. These higher order terms do not contribute to the measure of output power because they are filtered out by the resonant circuit. The observed departure from linearity is referred to as gain compression. This characteristic has grave consequences and forms the starting point of the next section.

2.3 Characterization of amplifier distortion

The analysis of the class A, AB, and B amplifiers of section 2.2 is based entirely on the DC characteristics of the transistor. Further simplifications such as infinite input and output impedance are also assumed. In practise, this is not the case. Real power devices exhibit finite terminal impedances and gain, both of which vary with input signal

level. As a result, even the most “linear” class A power amplifier will generate harmonics, intermodulation products, and other unwanted modulations. Unlike the idealized model of section 2.2.3, these effects do not disappear below some specific drive power level. To completely remove them, the power must be backed-off considerably. Depending on the amplifier in question, this input back-off could be as much as 25 to 30 dB.

To make matters worse, all power amplifiers exhibit frequency and time dependence in their behaviour. Such things as gain roll-off, variable group delay response, and memory effects all serve to distort or disperse the transmitted signal. The only sure way to limit their occurrence is by reducing the signal bandwidth. A narrowband signal does not see variations in gain and should have a reciprocal bandwidth significantly higher than the memory time of the amplifier.

Later on in this chapter, all of the distortion described above will be epitomized by a single variable. It is important that the exact nature of the term be recognized. Hence, the intent of this section is to divide it into classes and to explain each one by means of simple examples, and/or models. The following four subsections review AM/AM conversion, AM/PM conversion, linear distortion, and memory effects.

2.3.1 AM/AM conversion

By definition, amplitude-amplitude (AM/AM) conversion describes how a device transfers the amplitude modulation of a carrier from the input to the output. This relationship is illustrated by the following base-band equation:

$$v_{out} = v_{in} |G(v_{in})| \quad (2.35)$$

Where G is the complex level-dependent gain of the device. The exact value of G is specified by the input amplitude or the instantaneous input power x_{in} . AM/AM characteristics are assessed by observing the device’s amplitude response to single, double, and sometimes triple-tone test signals of varying amplitude. Based on these measurements, it is possible to fit a power series expression which can be used to predict how the device will

respond to other, more complex signals. This section describes the single and double tone tests which are used to investigate and characterize harmonic and intermodulation distortion. Both are manifestations of non-ideal AM/AM conversion.

Harmonic distortion

If a single tone is injected into the amplifier, the output should be a tone at the same frequency, but with a greater amplitude. As the amplitude of the input power is increased, some of the output power appears at the harmonics of the input frequency ($2\omega, 3\omega, \dots$). This siphoning of power away from the fundamental leads to gain compression as shown in Figure 2.6.

For devices with weak non-linear behaviour, their output voltage can be modelled as a power series of the input voltage:

$$v_{out} = c_1 v_{in} + c_2 v_{in}^2 + c_3 v_{in}^3 + \dots \quad (2.36)$$

Since this section deals with amplitude non-linearities only, the c_i 's cause no phase shift and are real. If the non-linearity is particularly slight, it can be represented adequately with only the first three terms of (2.36). Substituting the input signal $v_{in} = A \cos \omega t$ the output becomes

$$\begin{aligned} v_{out} &= c_1 A \cos \omega t + c_2 A^2 \cos^2 \omega t + c_3 A^3 \cos^3 \omega t \\ &= c_1 A \cos \omega t + c_2 A^2 \left(\frac{1}{2} + \frac{1}{2} \cos 2\omega t \right) + c_3 A^3 \left(\frac{3}{4} \cos \omega t + \frac{1}{4} \cos 3\omega t \right) \quad (2.37) \\ &= \frac{1}{2} c_2 + \left(c_1 A + \frac{3}{4} c_3 A^3 \right) \cos \omega t + \frac{1}{2} c_2 A^2 \cos 2\omega t + \frac{1}{4} c_3 A^3 \cos 3\omega t \end{aligned}$$

The output signal, as described by equation (2.37) is comprised of the desired signal at ω and spurious responses at DC, the second harmonic frequency 2ω , and the third harmonic frequency 3ω . At first glance, it would appear that the amplitude of the component at the fundamental frequency, $c_1 A \left[1 + \frac{3}{4} (c_3/c_1) A^2 \right]$, offers increased amplification compared to a linear two-port with transfer function $v_{out} = c_1 v_{in}$. However, in most cases $c_3 < 0$, and the device is termed compressive.

Figure 2.7 Non-linear transfer function showing gain compression.

The output power at the fundamental frequency for weakly non-linear devices is characterized in terms of the 1 dB compression point, the point at which the fundamental output power is 1 dB below its extrapolated linear behaviour (see Figure 2.7). Measurement of this point can be used to infer the parameters c_1 and c_3 of the non-linear device under test (DUT). From equation (2.37), the gain at the fundamental frequency ω is given as

$$G = 20 \log \frac{c_1 A + (3/4)c_3 A^3}{A} = 20 \log \left(c_1 + \frac{3}{4}c_3 A^2 \right) \quad (2.38)$$

By comparison, the linear gain, G_0 , is defined as

$$G_0 = 20 \log \frac{c_1 A}{A} = 20 \log c_1 \quad (2.39)$$

From the definition of the 1 dB compression point

$$G_{1dB} = G_0 - 1 \quad (2.40)$$

Substituting (2.39) and (2.38) yields

$$c_1 + \frac{3}{4}c_3 A^2 = 0.891 c_1 \quad (2.41)$$

Hence, the input signal amplitude at the 1 dB compression point is

$$A^2 = 0.145 \frac{c_1}{|c_3|} \quad c_3 < 0 \quad (2.42)$$

If the input and output impedances of the two-port are assumed equal to $R \Omega$, then the input and output powers P_{IN} and P_{OUT} can be stated in dBm as:

$$P_{IN} = 10 \log \left\{ \left(\frac{A}{\sqrt{2}} \right)^2 \frac{10^3}{R} \right\} \quad \text{dBm} \quad (2.43)$$

$$\begin{aligned} P_{OUT} &= 10 \log \left\{ \left(\frac{c_1 A + (3/4)c_3 A^3}{\sqrt{2}} \right)^2 \frac{10^3}{R} \right\} \quad \text{dBm} \\ &= G + P_{IN} \quad \text{dBm} \end{aligned} \quad (2.44)$$

Using the definition of G at the 1 dB compression point in (2.44) gives

$$P_{1dB} = G_{1dB} + P_{IN} = G_0 - 1 + P_{IN} \quad (2.45)$$

Substituting (2.39), (2.42) and (2.43) into (2.45)

$$\begin{aligned} P_{1dB} &= G_0 - 1 + 10 \log \left[\frac{0.145 c_1 10^3}{2 |c_3| R} \right] \\ &= 10 \log \frac{57.7 c_1^3}{|c_3| R} = 10 \log \left\{ \left(\frac{1}{17.3 |c_3|} \right) \frac{c_1^3}{R} \right\} \quad \text{dBm} \end{aligned} \quad (2.46)$$

And, for the case of $R = 50 \Omega$

$$P_{1dB} = 10 \log \frac{c_1^3}{|c_3|} + 0.62 \quad \text{dBm} \quad (2.47)$$

In section 2.2.3, the consequence of overdriving an amplifier was illustrated in Figure 2.5 and expressed by equation (2.34). As the input power was increased such that the drain current exceeded its maximum value, a clipped current waveform resulted. Once this occurred the output power and the input power were no longer linearly related and gain compression was observed. Physically, equation (2.34) is only an approximation of the device behaviour and it is also cumbersome to work with. For weak amplitude non-linearities, a truncated power series is easy to manipulate and will represent the P_{IN} vs P_{OUT}

characteristic reasonably accurately. By measuring the linear gain of the amplifier and extrapolating to find the 1 dB compression point, the parameters c_1 and c_3 of equation (2.36) can be deduced. Using a spectrum analyser, the amplitude of the second harmonic can be measured and related via equation (2.37) to c_2 . It is now possible to obtain an estimate of how the amplifier will affect other more complicated modulations.

Intermodulation distortion

This form of distortion is of greater concern than harmonic distortion because it lies very close to the passband of the amplifier. Intermodulation is best described as the intermixing of tones when simultaneously transferred by a non-linearity. Two signals, at frequencies ω_1 and ω_2 are combined and injected into the device. The output is then viewed on a spectrum analyser and contains a family of harmonic products and intermodulation products. Using the same transfer function as for harmonic distortion and a two-tone input signal $v_{in} = A(\cos \omega_1 t + \cos \omega_2 t)$ the output becomes

$$v_{out} = c_1 A(\cos \omega_1 t + \cos \omega_2 t) + c_2 A^2(\cos \omega_1 t + \cos \omega_2 t)^2 + c_3 A^3(\cos \omega_1 t + \cos \omega_2 t)^3 \quad (2.48)$$

Expanding,

$$\begin{aligned} v_{out} = & c_2 A^2 + c_2 A^2 \cos(\omega_1 - \omega_2)t + \left(c_1 A + \frac{9}{4}c_3 A^3\right) \cos \omega_1 t \\ & + \left(c_1 A + \frac{9}{4}c_3 A^3\right) \cos \omega_2 t + \frac{3}{4}c_3 A^3 \cos(2\omega_1 - \omega_2)t \\ & + \frac{3}{4}c_3 A^3 \cos(2\omega_2 - \omega_1)t + c_2 A^2 \cos(\omega_1 + \omega_2)t + \frac{1}{2}c_2 A^2 \cos 2\omega_1 t \\ & + \frac{1}{2}c_2 A^2 \cos 2\omega_2 t + \frac{3}{4}c_3 A^3 \cos(2\omega_1 + \omega_2)t + \frac{3}{4}c_3 A^3 \cos(2\omega_2 + \omega_1)t \\ & + \frac{1}{4}c_3 A^3 \cos 3\omega_1 t + \frac{1}{4}c_3 A^3 \cos 3\omega_2 t \end{aligned} \quad (2.49)$$

This signal consists of components at the fundamental frequencies ω_1 and ω_2 , a DC component, second-order intermodulation products $\omega_1 \pm \omega_2$, second and third order harmon-

Figure 2.9 Diagram showing third order intercept point.

ics $2\omega_1$, $2\omega_1$ and $3\omega_1$, $3\omega_2$, and third order intermodulation products $2\omega_1 \pm \omega_2$ and $2\omega_2 \pm \omega_1$. Of the spurious responses, only $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ pose a problem as they fall within the passband; that is, they are very close in frequency to the fundamental components. The input and output spectrum corresponding to v_{in} and v_{out} is shown in Figure 2.8. Before proceeding, it should be noted that the IM products arise because of the odd-ordered term, $n = 3$. The even-ordered term $n = 2$, on the other hand, only produces out of band responses. As a result, it will be dropped from any subsequent power series expansions. A common measure of the level of intermodulation distortion is the intercept point. By definition, it is the output power level P_{IP3} for which the output power of the $2\omega_1 - \omega_2$ component would intercept the linear extrapolation of the output power at the fundamental frequency ω_1 . This relationship is shown in Figure 2.9. As in the single-tone measurement at low input power levels, the fundamental output power increases line-

Figure 2.8 Input and output spectrum for a third-order linearity.

arly with input power. The third order intermodulation power, however, increases to the cube of the input power and on a log-log scale this corresponds to a straight line with a slope of three.

From equation (2.49), the amplitude of the output at the frequency ω_1 is $c_1 A + \frac{9}{4} c_3 A^3$. where $c_3 < 0$ for compressive transducers. For power levels below the 1 dB compression point, the response at ω_1 appears very close to the linear approximation $P_{OUT, LIN}$. This is because $c_1 A \gg \frac{9}{4} c_3 A^3$ for weakly non-linear devices. Such is not the case when the device enters into compression, hence the power is denoted by P_{ω_1} . From equation (2.49), the following expressions are derived

$$P_{OUT, LIN} = 10 \log \left\{ \left(\frac{c_1 A}{\sqrt{2}} \right)^2 \frac{10^3}{R} \right\} \quad \text{dBm} \quad (2.50)$$

$$P_{\omega_1} = 10 \log \left\{ \left(\frac{c_1 A + \frac{9}{4} c_3 A^3}{\sqrt{2}} \right)^2 \frac{10^3}{R} \right\} \quad \text{dBm} \quad (2.51)$$

$$P_{2\omega_1 - \omega_2} = 10 \log \left\{ \left(\frac{\frac{3}{4} c_3 A^3}{\sqrt{2}} \right)^2 \frac{10^3}{R} \right\} \quad \text{dBm} \quad (2.52)$$

At the intercept point, P_{IP3} the third-order intermodulation power is equal to the extrapolated linear power, $P_{OUT, LIN} = P_{2\omega_1 - \omega_2}$. By comparing equations (2.50) and (2.52), the input signal amplitude at the intercept point is

$$A_{IP3}^2 = \frac{4 c_1}{3 |c_3|} \quad (2.53)$$

Substituting in (2.50),

$$P_{IP3} = 10 \log \left\{ \left(\frac{2 c_1^3}{3 |c_3|} \right) \frac{10^3}{R} \right\} \quad \text{dBm} \quad (2.54)$$

For the case of $R = 50 \Omega$

$$P_{IP3} = 10 \log \frac{c_1^3}{|c_3|} + 11.3 \quad \text{dBm} \quad (2.55)$$

As with the 1 dB compression point, the third-order intercept is a function of the non-linearity only which makes it a useful measurement. Now, comparing equations (2.50), (2.52) and (2.54), the output power at the intermodulation frequency is given as

$$P_{2\omega_1 - \omega_2} = 3P_{OUT, LIN} - 2P_{IP3} \quad \text{dBm} \quad (2.56)$$

Assuming that the input power is low, $P_{OUT, LIN} \approx P_{\omega_1}$

$$P_{2\omega_1 - \omega_2} \approx 3P_{\omega_1} - 2P_{IP3} \quad \text{dBm} \quad (2.57)$$

It is now possible to predict the intermodulation power for a given input signal level knowing only the intercept power level. This is a handy relation for making quick predictions regarding the intermodulation distortion level.

As a final remark, equation (2.57) indicates that IMD exists even for low input powers. This has serious implications so far as linear power amplification is concerned because even at medium input power levels, IMD tends to dominant all other forms of in-band distortion. It is also difficult to suppress. Towards the end of this chapter, it will become apparent that the primary aim of applying linearization to an amplifier is to remove intermodulation products.

2.3.2 AM/PM conversion

The previous section examined the amplitude behaviour of a non-linear amplifier in response to a signal with a time-varying envelope. The next obvious extension of this theory is to develop a means of characterizing a non-linearity where the output phase of the signal is dependent on the input amplitude. As in the previous section, this relationship is illustrated by a simple base-band equation:

$$v_{out} = A(v_{in}) \cos(\omega t + \theta(v_{in})) \quad (2.58)$$

Where θ is the level dependent phase. This form of non-linearity has serious consequences for spectrally efficient QPSK or QAM signals. For these modulation formats,

control of the signal phase is of the utmost importance. Slight deviations can cause the receiver to make incorrect decisions which in turn increase the bit-error-rate. The following derivation is a simplified version of that developed by G.L. Heiter [13].

As before, assume that the transfer function of a device with slightly non-linear behaviour can be expressed in terms of a power series with real coefficients. However, unlike in equation (2.36), the implicit assumption that the time delays of the respective orders are equal will not be made. Hence,

$$v_{out} = c_1 v_{in}(t - t_1) + c_2 v_{in}^2(t - t_2) + c_3 v_{in}^3(t - t_3) \quad (2.59)$$

As with the single-tone test, the device is excited with $v_{in} = A \cos \omega t$. From the previous work on AM/AM conversion, it is known that only odd-ordered terms result in intermodulation distortion. Thus, the second order term can be ignored

$$v_{out} = c_1 v_{in}(t - t_1) + c_3 v_{in}^3(t - t_3) \quad (2.60)$$

Substituting the input signal yields

$$v_{out} = c_1 A \cos \omega(t - t_1) + c_3 A^3 \cos^3 \omega(t - t_3) \quad (2.61)$$

If all measurements are referenced to the time-delayed input signal at the fundamental frequency, then $t_1 = 0$ without loss of generality. Expanding (2.61)

$$\begin{aligned} v_{out} = c_1 A \cos \omega t + \frac{3c_3 A^3}{4} (\cos \omega t \cos \omega t_3 + \sin \omega t \sin \omega t_3) \\ + \frac{c_3 A^3}{4} (\cos 3\omega t \cos 3\omega t_3 + \sin 3\omega t \sin 3\omega t_3) \end{aligned} \quad (2.62)$$

Separating the fundamental output and expressing it in terms of its inphase and quadrature components yields

$$v_{out} = A_\omega \cos \omega t + B_\omega \sin \omega t \quad (2.63)$$

where

$$A_{\omega} = c_1 A + \frac{3}{4} c_3 A^3 \cos \omega t_3 \quad (2.64)$$

$$B_{\omega} = \frac{3}{4} c_3 A^3 \sin \omega t_3$$

For clarity, (2.63) is arranged into amplitude and phase format

$$v_{out} = \hat{v}_{out} \cos(\omega_0 t + \theta) \quad (2.65)$$

where

$$\hat{v}_{out} = \sqrt{A^2 + B^2} \approx A \quad (2.66)$$

$$\theta = -\text{atan} \frac{B}{A} \approx -\frac{B}{A} \quad (2.67)$$

The approximations made in equations (2.66) and (2.67) are valid assuming that t_3 is small. When this is true, $A_{\omega} \gg B_{\omega}$. If the device is being operated at lower power, then the linear gain term at the fundamental frequency is $A_{\omega} \approx c_1 A$. Substituting this approximate value for A_{ω} and the value for B_{ω} as given by equation (2.64) into (2.67) gives the following equation for the phase of the output signal at the fundamental

$$\theta(A) = -\frac{\frac{3}{4} c_3 A^3 \sin \omega t_3}{c_1 A} = -\frac{3 c_3}{4 c_1} A^2 \sin \omega t_3 \quad (2.68)$$

Thus, the phase shift varies as the square of the input amplitude: $\theta(A) = k A^2$. Consider the case of a signal with time varying amplitude given by $v_{in}(t) = A(1 + M \cos \omega_m t)$. In this case,

$$\theta(A) = k A^2 (1 + 2M \cos \omega_m t + M^2 \cos^2 \omega_m t) \quad \text{radians} \quad (2.69)$$

Assuming $M \ll 1$

$$\theta(A) \approx (1 + 2M \cos \omega_m t) = \theta_0 + \theta_p \cos \omega t \quad (2.70)$$

The implication of equation (2.70) is that when order-dependent time delays exist in the device, the phase shift of the output becomes a function of the instantaneous power of the input signal. In the simplistic example shown here, the phase will rip-

Figure 2.10 Variation in output phase due to amplitude modulation.

ple around a mean value as shown in Figure 2.10. This effect is called AM-to-PM conversion.

2.3.3 Linear distortion

Power amplifiers, like all microwave devices, exhibit frequency dependent behaviour. When transmitting a signal with non-zero bandwidth, they will alter the amplitude of each frequency component by a different amount. Likewise, the phase shift will not necessarily be proportional to frequency. The resulting output signal is said to be linearly distorted because the exact shape of the distorted signal is independent of the input power level. If the amplifier has a frequency dependent transfer function, $H(j\omega)$, then the output signal $v_{out}(t)$ is

$$v_{out}(t) = \mathcal{F}^{-1}\{H(j\omega)V_{in}(j\omega)\} \quad (2.71)$$

Where $V_{in}(j\omega)$ is the Fourier transform of the input $v_{in}(t)$. To highlight this form of distortion, consider the following level-independent transfer function:

$$H(j\omega) = K(1 + a\cos\tau\omega)e^{-j\omega t_0} \quad (2.72)$$

In this case, the device adds a small frequency dependent ripple and a linear phase shift to the input signal. Applying Euler's relation allows (2.72) to be expressed

$$H(j\omega) = K \left[e^{-j\omega t_0} + \frac{a}{2} (e^{j\omega(\tau - t_0)} + e^{j\omega(\tau + t_0)}) \right] \quad (2.73)$$

taking the inverse Fourier transform

$$h(t) = K \left[\delta(t_0 - t) + \frac{a}{2} (\delta(t_0 - \tau - t) + \delta(t_0 + \tau - t)) \right] \quad (2.74)$$

and convolving with the input signal $v_{in}(t)$ gives

$$y(t) = h(t) \otimes v_{in}(t) = K v_{in}(t - t_0) + \frac{Ka}{2} [v_{in}(t - t_0 + \tau) + v_{in}(t - t_0 - \tau)] \quad (2.75)$$

Thus, the ripple in the amplitude response produces a pair of scaled and delayed versions of the input signal $v_{in}(t)$. The phase shift of these “echoes” is frequency dependent. Hence, when a nonzero bandwidth signal is passed by a transducer with variable amplitude response, echoes will appear everywhere in the band.

In the above case, the phase shift is a linear function of the input frequency. Such devices are usually characterized by their group delay, which is the negative of the slope of the derivative of frequency with phase.

$$t_d = -\frac{d\phi}{d\omega} \quad (2.76)$$

For devices with a linear phase response, the group delay is constant and there is no phase distortion.

If the phase shift is a nonlinear function of the input frequency, then different frequency components will experience different group delays. Consider the case of an amplitude modulated signal which is transferred through a device in which the group delay varies with excitation frequency and the gain is flat.

$$\begin{aligned} v_{in}(t) &= A(1 + M \cos \omega_m t) \cos \omega t \\ &= A \cos \omega t + \frac{MA}{2} [\cos(\omega + \omega_m)t + \cos(\omega - \omega_m)t] \end{aligned} \quad (2.77)$$

where M and ω_m are the amplitude and frequency of the envelope. If the group delay is not a constant, then the delays for the frequency components at ω , $\omega + \omega_m$, and $\omega - \omega_m$,

will be different. When the signal emerges from an amplifier with constant gain K and variable group delay, it appears as

$$v_{out}(t) = KA \cos \omega(t - t_d) + \frac{KMA}{2} [\cos(\omega + \omega_m)(t - t_d - \delta t_d) + \cos(\omega - \omega_m)(t - t_d + \delta t_d)] \quad (2.78)$$

where it is assumed that the group delay varies linearly with frequency. As a result, the delay is slightly greater at the upper end of the band $(+\delta t_d)$ and slightly less at the lower end $(-\delta t_d)$. With the two sideband components at $\omega \pm \omega_m$ delayed by different amounts it is no longer possible to manipulate equation (2.78) into a scaled form of equation (2.77). Thus, the non-constant group delay has yielded a distorted version of the input.

2.3.4 Non-linear distortion with memory

In the previous sections, the implicit assumption has been that the transfer characteristics of the two-port vary only with the instantaneous amplitude and frequency of the input signal. However, most practical circuits contain some amount of memory; that is, the response of the network at time t may be dependent on past inputs. Characterizing such a network requires a sophisticated mathematical model. The preferred approach is to use a Volterra functional series, which is likened to a power series wherein the amplitude and delay coefficients vary as a function of time as well as frequency.

For the general case of $v_{out}(t) = T[v_{in}(t)]$, there exists a series which can represent all memory, frequency and power effects represented within T . Basically, the Volterra functional series is a generalization of the convolution procedure used in a linear system.

$$v_{out}(t) = h(t) \otimes v_{in}(t) = \int_{-\infty}^{\infty} h(\tau) v_{in}(t - \tau) d\tau \quad (2.79)$$

The transfer function with the impulse response $h(t)$ is adequate to describe a simple system consisting of ideal lumped capacitors, inductors and resistors. In terms of a

Volterra functional series, equation (2.79) is merely the first order term. Like a power series expansion, additional terms can be added add infinitum. For the case of three terms, $n = 1 \dots 3$

$$\begin{aligned}
 v_{out} = & \int_{-\infty}^{\infty} h_1(\tau_1) v_{in}(t - \tau_1) d\tau_1 \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2) v_{in}(t - \tau_1) v_{in}(t - \tau_2) d\tau_1 d\tau_2 \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) v_{in}(t - \tau_1) v_{in}(t - \tau_2) v_{in}(t - \tau_3) d\tau_1 d\tau_2 d\tau_3
 \end{aligned} \tag{2.80}$$

In this example, $v_{out}(t)$ is the response of a third order non-linear network characterized by the n -th impulse responses. The linear character is described by the first order impulse response, $h_1(\tau_1)$, the quadratic character is expressed by the second order impulse response, $h_2(\tau_1, \tau_2)$, and the cubic character is represented by the third order impulse response, $h_3(\tau_1, \tau_2, \tau_3)$. For more complex systems, additional terms can be included.

Although the Volterra functional series can theoretically describe any and all sources of distortion, including the ones reviewed in the previous sections, its sheer complexity imposes limits to its usefulness. Most calculations become completely intractable if more than three terms are included. As far as the work contained in this thesis is concerned, only a passing understanding of the Volterra functional series is necessary. Since most linearization techniques are bandwidth limited, memory effects are of negligible effect compared to other sources of distortion. However, if a more comprehensive grasp of this analytical approach is desired, the reader is referred to the book by Tri. T. Ha [12].

2.4 Linearization strategies

For the remainder of this thesis, it will be assumed that all RF amplifiers can be characterized entirely by measurement of their AM/AM and AM/PM conversion

characteristics. This implies that the amplifiers are nonlinear, memoryless and can be completely described by a frequency-independent complex gain

$$G(x) = g(x)e^{j\phi(x)} \quad (2.81)$$

Here, the magnitude and phase of $G(x)$ are both functions of the instantaneous input power x . It should be pointed out that the frequency dependence is omitted for analytical convenience rather than for physical reasons. The approach taken in this thesis and in much of the literature is to first consider how the linearization strategy behaves when a single tone is injected into the system. Once the response is understood, due consideration is given to the effects of non-zero bandwidth signals. In this section, bandwidth considerations will be addressed by means of qualitative remarks. Section 2.5, which is more quantitative, will handle frequency dependence by establishing accuracy requirements for the adaptation parameters of the linearizer. For more details, the reader is referred to sections 2.5.3 and 2.5.4.

2.4.1 Cartesian feedback

Negative feedback has long been used to linearize nonlinear systems. It was first proposed by H.S. Black in 1927 [14] a full four years after he conceived of feed-forward linearization. The main drive behind both inventions was to realize long-distance repeater systems for modulated carrier telephony. The principle behind feedback is to cancel the distortion of the output signal by using the same nonlinear amplifier twice [15].

An effective way of implementing the principle is to use a Cartesian feed-back loop. As illustrated in Figure 2.11, this technique compares the baseband Cartesian inputs with the baseband Cartesian signals demodulated from the output and presents the PA with a predistorted input signal. The loop control characteristics are established by the gain and the compensation filters inserted in the I and Q channels. The level of intermodulation distortion reduction is essentially governed by the loop gain. However, the gain must be carefully selected in order to provide sufficient phase margin. The modulator and the demodulator are synchronized by carefully adjusting the phase of a common oscillator. This should compensate for any differences in the forward and feedback paths. In an additive, the phase adjustment can be made adaptive to compensate for temperature, age, or power-induced changes in the amplifier characteristics.

Figure 2.12 shows a model of a Cartesian feedback system. In fact, it could represent any simple negative feedback loop. For the case of a slight non-linearity, the distortion introduced by the amplifier can be represented by a generator v_d which is inserted

Figure 2.11 Cartesian feedback amplifier.

between two linear amplifiers with gains of A_1 and A_2 . The modulation/demodulation

Figure 2.12 Distortion cancellation model for a feedback amplifier.

processes are summarized by a simple feedback factor β . By indicating subtractive combination at the input, the model assumes that the phase adjuster of Figure 2.11 is set such that the returned signal is exactly 180° out of phase with the input.

Considering distortion components only, the portion of the output v_{d0} return to the point of injection (i.e. between A_1 and A_2) is evidently $-A_1\beta v_{d0}$. The input to the A_2 amplifier is $v_d - A_1\beta v_{d0}$, whence

$$v_{d0} = A_2(v_d - A_1\beta v_{d0}) \quad (2.82)$$

Re-arranging
$$v_{d0} = \frac{A_2}{1 + A_1 A_2 \beta} v_d \quad (2.83)$$

By comparison, when there is no feedback $v'_{d0} = A_2 v_d$. Taking the ratio of the output distortion with and without the feedback network gives a familiar result

$$\frac{v_{d0}}{v'_{d0}} = \frac{1}{1 + A_1 A_2 \beta} = \frac{1}{1 + A\beta} \quad (2.84)$$

Hence, the distortion level at the output of the feedback amplifier is $1 + A\beta$ lower than that of the equivalent amplifier without feedback.

Although simple in principle, Cartesian feedback linearization has some significant drawbacks. Firstly, unless β can somehow be reduced for the signal components, the use of feedback will reduce the overall gain of the amplifier by an amount

equivalent to the reduction in the distortion level. A second problem with the Cartesian feedback amplifier is its susceptibility to instability [16]. If the magnitude of the loop transfer function $|A\beta|$ becomes greater than unity when its phase is 180° , a signal transmitted around the loop will increase in magnitude until it reaches saturation. Henceforth, the amplifier becomes an oscillator. Thirdly, it has been assumed to this point that the returned signal is transmitted instantaneously around the loop to permit simultaneous comparison with the input signal. In practical terms, this implies that the feedback amplifier must be able to respond much faster than the envelope of the input signal. It is recommended that the amplifier and feedback circuit have a bandwidth as much as ten times that of the input signal for effective operation [15]. Finally, Cartesian feedback is generally limited to single channel linearization. Reports of multi-carrier feedback exist [17], but these lack the capacity for dynamic channel allocation. Hence, feedback linearization is bandwidth limited and not suitable for broadband multi-channel applications.

2.4.2 Predistortion

The principle of predistortion is to insert a nonlinear circuit between the input signal and the power amplifier. The nonlinear circuit contains devices which generate distortion that is in anti-phase to that produced by the power amplifier. The corrupted input signal will, in effect, be corrected by the amplifier's inverse nonlinear transfer characteristic. Although capable of removing all unwanted AM/AM and AM/PM effects, most authors measure the effectiveness of the predistortion linearizer by its capacity to lower

intermodulation levels. Hence, it is common to refer to general distortion suppression as merely intermodulation suppression.

Exactly how the nonlinear predistortion circuit behaves is the subject of a large number of papers. The predistortion linearizer investigated by S.P. Stapleton et al. provides a good example of the technique [18]; it is reproduced in the schematic of Figure 2.13. The input signal $v_m(t)$ is predistorted by multiplying its inphase and quadrature components by two nonlinear functions F_1 and F_2 .

$$\begin{aligned} F_1\{x_m(t)\} &= \alpha_{11} + \alpha_{13}x_m(t) + \alpha_{15}x_m^2(t) \\ F_2\{x_m(t)\} &= \alpha_{21} + \alpha_{23}x_m(t) + \alpha_{25}x_m^2(t) \end{aligned} \quad (2.85)$$

Where the independent variable is the input power, i.e. $x_m(t) = |v_m(t)|^2$. The α coefficients are complex quantities which can be varied in order to obtain the maximum intermodulation suppression. The transfer function for the predistorter circuit can be expressed as shown in (2.86).

$$\begin{aligned} F\{x_m(t)\} &= F_1\{x_m(t)\} + jF_2\{x_m(t)\} \\ &= \alpha_1 + \alpha_3x_m(t) + \alpha_5x_m^2(t) \end{aligned} \quad (2.86)$$

This yields a distortion signal $v_d(t)$ which appears as follows

Figure 2.13 Schematic of an adaptive predistortion linearizer.

$$v_d(t) = v_m(t) \cdot F\{x_m(t)\} = \alpha_1 v_m(t) + \alpha_3 v_m(t)|v_m(t)|^2 + \alpha_5 v_m(t)|v_m(t)|^4 \quad (2.87)$$

The predistorter output contains a fundamental term and terms of third and fifth order. Recall from sections 2.3.1 and 2.3.2 that it was the presence of these odd-ordered terms that lead to intermodulation distortion. If the α coefficients are selected wisely, then the third and fifth order components of (2.87) can be used to offset the third and fifth order responses of the amplifier. For most weakly non-linear devices, fifth order IMD suppression is sufficient to achieve satisfactory linearity. Using the distorted input signal, the amplifier's output response becomes

$$v_a(t) = v_d(t)G\{x_d(t)\} \quad (2.88)$$

As per the same arguments of pervious sections, the complex gain of the amplifier, $G\{x_d(t)\}$, can be expressed as a power series in $x_d(t)$. To include the AM/PM conversion characteristics, the coefficients of this series must now assume complex values.

$$\begin{aligned} G\{x_d(t)\} &= \beta_1 + \beta_3 x_d(t) + \beta_5 x_d^2(t) \\ &= G_1\{x_d(t)\} + jG_2\{x_d(t)\} \end{aligned} \quad (2.89)$$

Cascading the predistorter and power amplifier, the output signal, $v_a(t)$, becomes

$$\begin{aligned} v_a(t) &= v_m(t)F\{x_m(t)\}G\{x_d(t)\} \\ &= v_m(t)F\{x_m(t)\}G\{x_m(t)|F[x_m(t)]|^2\} \end{aligned} \quad (2.90)$$

hence, $v_m(t)$ and $v_a(t)$ are related by the complex gain function $K\{x\}$

$$K\{x_m(t)\} = F\{x_m(t)\}G\{x_m(t)|F[x_m(t)]|^2\} \quad (2.91)$$

This complex gain can be represented by yet another truncated power series

$$\begin{aligned} K\{x_m(t)\} &= \gamma_1 + \gamma_3 x_m(t) + \gamma_5 x_m^2(t) \\ &= K_1\{x_m(t)\} + jK_2\{x_m(t)\} \end{aligned} \quad (2.92)$$

Where

$$\begin{aligned}
K_1\{x_m(t)\} &= \gamma_{11} + \gamma_{13}x_m(t) + \gamma_{15}x_m^2(t) \\
K_2\{x_m(t)\} &= \gamma_{21} + \gamma_{23}x_m(t) + \gamma_{25}x_m^2(t)
\end{aligned} \tag{2.93}$$

Expressed in terms of the γ coefficients, $v_a(t)$ is

$$v_a(t) = v_m(t) \cdot K\{x_m(t)\} = \gamma_1 v_m(t) + \gamma_3 v_m(t)|v_m(t)|^2 + \gamma_5 v_m(t)|v_m(t)|^4 \tag{2.94}$$

If the right-hand side of equation (2.91) is expanded and related to (2.94), it can be shown that the coefficients α , β , and γ are related.

$$\begin{aligned}
\gamma_1 &= \alpha_1 \beta_1 \\
\gamma_3 &= \alpha_3 \beta_1 + \alpha_1 \beta_3 |\alpha_1|^2 \\
\gamma_5 &= \alpha_5 \beta_1 + \alpha_3 \beta_3 |\alpha_1|^2 + \alpha_1 \beta_5 |\alpha_1|^4 + 2\alpha_1 \beta_3 \text{Re}\{\alpha_1 \alpha_3^*\}
\end{aligned} \tag{2.95}$$

By careful selection of α_1 , α_3 and α_5 , the third- and fifth-order distortion terms, γ_3 and γ_5 , can be nulled. However, as the amplifier ages or is subjected to temperature variations, its complex gain can be expected to change. Anticipating these deviations, a predistortion linearizer will include an adaptive controller to monitor the intermodulation power and make corrections as necessary. For the purposes of this discussion, the adaptive control circuit is represented in Figure 2.13 by a micro-controller box. A downconverter, bandpass filter, and power detector are also shown to indicate how the intermodulation products are extracted from the output signal and fed to the controller.

Predistortion linearization is an effective way to reduce distortion in a power amplifier and requires only a few additional RF components. Its main advantage over feedback linearization is its inherent stability. However, there are drawbacks too. One of the most common, which is cited in numerous papers [19], is the need for very complicated adaptation algorithms to adjust the coefficients. As a result, not only is a fast processor needed to perform the adaptation, but convergence times tend to be quite long. For this reason, predistortion linearization can only be applied to signals with two carriers or less.

One last drawback to predistortion linearization is the stipulation that the starting values of the coefficients of F_1 and F_2 be selected at or very near to the convergence point. This requires prior knowledge of the AM/AM and AM/PM characteristics of the PA. So far, not a problem. However, if the predistorter must be carefully initialized, it cannot be very resistant to transient effects. This is because the optimization space has many dimensions and many local minima. Sudden irregularities in the input signal could force the circuit into one of these minima causing non-optimal distortion cancellation.

2.4.3 Feedforward

As mentioned in section 2.4.1, H.S. Black conceived the idea of feedforward linearization about four years before he devised the negative feedback technique [14]. A feedforward amplifier contains a signal cancellation loop and a distortion cancellation loop as shown in Figure 2.14. The signal cancellation loop generates an error signal based on the PA distortion. The distortion cancellation loop takes this error signal, and by appropriate selection of attenuation and phase shift, adds it back to the amplifier output. If both loops are adjusted correctly, the amplifier distortion will be cancelled.

Before commencing the mathematical description of the feedforward, a few comments are in order regarding an alternative representation for the power amplifier. Prior to this point, the PA has been represented as memoryless non-linearity with complex gain equal to $G(x)$. For the purpose of this section and section 2.5, the analysis is greatly simplified if the PA output is described by a linearly amplified fundamental term plus an additive distortion term:

$$v_a(t) = \alpha_0 v_m(t - \tau_1) + v_d(t) \quad (2.96)$$

Figure 2.14 Schematic of a feedforward linearizer.

Here, α_0 and τ_1 are the complex linear gain and group delay of the PA at the centre frequency and $v_d(t)$ is additive distortion. Although this latter term can include any combination of the four types of distortion described in section 2.3, it is often specified as IM distortion only. The justification lies in the fact IM products tend to be the dominant form of distortion and as such are an excellent measure of the overall level of distortion suppression. This is the same reason as given in section 2.4.2.

From Figure 2.14, an input modulated signal, designated as $v_m(t)$ is split into two paths. The upper path contains the PA characterized by equation (2.96). The lower reference path contains a group delay element, such as a length of line, and a vector modulator. The purpose of the reference path is to present a delay, gain and phase adjusted copy of the input signal to the subtraction circuit. The combined effect is to yield a reference signal of the form $\alpha v_m(t - \tau_1)$ where α is the complex gain selected by the vector modulator and τ_1 is the delay introduced by the length of line.

Returning to the upper branch, the output of the PA is split into two components of unequal amplitude by a directional coupler. The component with the smaller amplitude is attenuated by a factor $1/K$ and added to an inverted version of the lower path reference signal. The result is an error signal $v_e(t)$

$$v_e(t) = \frac{\alpha_0}{K} v_m(t - \tau_1) - \alpha v_m(t - \tau_1) + \frac{v_d(t)}{K} \quad (2.97)$$

If α is chosen to exactly equal α_0/K then the error signal continuing on in the lower branch of the second loop will contain only distortion. Next, $v_e(t)$ is passed through a second vector modulator and an auxiliary amplifier. These elements are both linear with complex gains given as β and γ respectively. The error signal now becomes $\beta\gamma v_e(t - \tau_2)$ where τ_2 is the characteristic group delay associated with the auxiliary amplifier.

The component of the PA output in the upper branch of the second loop is delayed by an amount τ_2 so that it matches the delay of the auxiliary amplifier. The results of the upper and lower branches of the second loop are then subtracted to yield the final output signal:

$$v_o(t) = \alpha_0 v_m(t - \tau_{total}) + v_d(t - \tau_2) - \beta\gamma v_e(t - \tau_2) \quad (2.98)$$

If the first loop is properly synchronized, then $v_d(t)$ can be replaced by $v_e(t)$ in the preceding equation. If β is adjusted such that its product with γ is unity, then the latter terms in equation (2.98) will be cancelled leaving only $\alpha_0 v_m(t - \tau_{total})$ where $\tau_{total} = \tau_1 + \tau_2$. Hence, the distortion has been expunged from the amplifier output.

The preceding analysis appears to indicate that the feedforward network is inherently superior to all previous linearization strategies. Unlike predistortion or negative feedback, its performance is not linked to any particular signal or distortion type. It can theoretically accommodate any number of carriers and any class of distortion. However, to fully exploit this capability, some sacrifices must be made in the areas of efficiency and cost. For example, the auxiliary amplifier must meet very stringent linearity requirements or it will introduce incorrigible distortion of its own. This problem can only be averted by using an expensive high-power class A amplifier that is backed-off well below its 1 dB compression point. This is a wasteful use of an costly MMIC component.

Feedforward linearization also requires a large number of additional analog and control components. From Figure 2.14, it is clear that the linearizer contains couplers, combiners, splitters and vector modulators. In order to cancel all the distortion, every one of them must have a bandwidth that exceeds that of the main signal. Providing this is pos-

sible, there are still other considerations. For wideband multicarrier signals, the amount of circuitry required to perform the adaptation of the complex α and β coefficients can be prohibitive both in terms of cost and DC power consumption. In contrast to predistortion linearization, however, the adaptation algorithms for feedforward are relatively simple and convergence is less dependent on initial conditions. For more details on the implementation of the wideband feedforward linearizer, the reader is referred to section 2.6.

In summary, feedforward offers the only inherently broadband approach to PA linearization. The price for this capability is a circuit with a greatly enlarged number of RF components and high DC power consumption.

2.5 Principles of adaptive feedforward

A circuit is termed adaptive when it can automatically adjust its parameters in response to changing operating conditions. In the case of the feedforward linearizer, adaptation is used to match the complex gain of the lower path of the signal cancellation circuit to the PA and to match the linear gain of the AA to the upper path of the distortion cancellation circuit. This section provides an overview of the gradient adaptation strategy proposed by J.K. Cavers [19][20][21]. Put simply, gradient signals generated from the RF signals in the loops are used to indicate the direction of adjustment for the α and β coefficients. The technique is proven to be rapid and accurate and of medium complexity both from the analog and control points of view [22].

Gradient adaptation is by no means the only approach to control feedforward parameters. In pilot guided adaptation, two out-of-band pilot tones are inserted at the input to the feedforward and at the output of the PA respectively. The control circuits adjust the gain and phase of the lower branches until the first tone is cancelled at the input

Figure 2.15 Model of a feedforward linearizer.

to the β vector modulator and the second tone is cancelled at the output. If the pilot tones are within the bandwidth of the linearizer, the final output signal should now be free of distortion. For further details on this technique, the reader is referred to references [23] and [24]. A third strategy involves minimizing the power in certain frequency bands at key points in the circuit using a perturbation technique. For example, to extract an optimal error signal from the PA output, the controller will periodically adjust the α parameter. After each adjustment, the controller observes the resulting change in signal power at the input to the β vector modulator. If the power level has decreased, the vector modulator will be directed to assume the new operating condition. Otherwise, it will return to its original state and be perturbed in a different direction at a later time. More information on this technique can be obtained from reference [25].

For consistency, this section employs the same notation as that used in section 2.4.3. However, for the sake of simplifying the analysis to follow, the schematic of the feedforward of Figure 2.14 is replaced by Figure 2.15. Here, it is assumed that the upper and lower paths of both cancellation loops have been exactly delay matched. Thus, the group delays are ignored for the time being. They will be re-introduced in subsection 2.5.4. A second simplification to the analysis is to lump the linear gain of the auxiliary amplifier together with the gain of the β vector modulator. The composite term shall henceforth be referred to as β only. Finally, the fixed attenuator denoted by $1/K$ will be

omitted to avoid carrying unnecessary multiplicative constants. Strictly speaking, this attenuation is absorbed by the adaptation coefficients α and β .

In this section, all bandpass signals are represented by their complex envelopes, which are each referred to as $v(t)$ with a subscript. By convention, the instantaneous power of the complex envelope is half its amplitude squared: $x_m(t) = \frac{1}{2}|v_m(t)|^2$. The average power is then calculated by taking the expectation of $x_m(t)$: $P_m = E[x_m(t)]$. Using the simplified schematic and the aforementioned notation, the principal equations for the feedforward linearizer can be derived.

As in the previous section, the PA output can be represented by a linearly amplified component and a predominantly IM based distortion term.

$$v_a(t) = \alpha_0 v_m(t) + v_d(t) \quad (2.99)$$

where the linear gain is complex and represented by α_0 . In the lower branch, a copy of the input signal is amplitude and phase adjusted before being subtracted from the amplifier output. The resultant is the error signal

$$v_e(t) = v_a(t) - \alpha v_m(t) \quad (2.100)$$

If signal cancellation loop is properly synchronized, $\alpha = \alpha_0$ and $v_e(t)$ is replaced by the distortion term $v_d(t)$. The final output signal is formed by subtracting an amplitude and phase adjusted error signal from the amplifier output

$$v_o(t) = v_a(t) - \beta v_e(t) \quad (2.101)$$

With $\beta = 1$ and α also at its optimum value, the feedforward will completely cancel the distortion because $v_o(t) = v_a(t) - v_d(t) = \alpha_0 v_m(t)$. The aim at this point is to illustrate how the gradient based adaptation strategy will seek out the optimum values of α and β without prior knowledge of α_0 .

As always, the PA is assumed to be a memoryless nonlinearity and is completely represented by its AM/AM and AM/PM characteristics. Consequently, $v_a(t)$ of equation (2.101) can also be represented by

$$v_a(t) = v_m(t)G(x_m(t)) \quad (2.102)$$

Using equations (2.99) to (2.102), it is now possible to explain the principles of the gradient feedforward linearizer. For clarity, the time dependence will no longer be included explicitly.

2.5.1 Adaptation of the signal cancellation coefficient

The gradient for the adaptation of the signal cancellation coefficient is simply the first derivative of the average power of the error signal with the α coefficient. This power is given by

$$P_e = E[x_e] = E\left[\frac{1}{2}|v_e|^2\right] \quad (2.103)$$

Substitute for v_e using equation (2.100)

$$\begin{aligned} P_e &= \frac{1}{2}E[|v_m G(x_m) - \alpha v_m|^2] \\ &= \frac{1}{2}E[2x_m |G(x_m)|^2 - 4\text{Re}\{\alpha^* x_m G(x_m)\} + 2|\alpha|^2 x_m] \\ &= P_a - 2\text{Re}\{\alpha^* E[x_m G(x_m)]\} + |\alpha|^2 P_m \end{aligned} \quad (2.104)$$

and take its derivative with respect to the complex variable α

$$\frac{\partial P_e}{\partial \alpha} = 2\alpha P_m - 2E[x_m G(x_m)] \quad (2.105)$$

When this expression is set to zero, the power in the error term v_e is minimized and it follows that $\alpha = \alpha_0$. Thus,

$$\alpha_0 = \frac{1}{P_m}E[x_m G(x_m)] = \frac{1}{P_m} \int_{-\infty}^{\infty} x_m G(x_m) p_m(x_m) dx_m \quad (2.106)$$

where the expectation is taken with respect to the probability density function (PDF) of x_m . This is an interesting result because it indicates that the linear gain of the amplifier

depends on the modulation format of the input and on the input power backoff. It also confirms that the amplifier is sensitive to changes in the number, location and bandwidth of the individual channels in a multicarrier signal.

The question now arises, how is the gradient of the error signal calculated? Adaptation of α relies on the fact that the lower branch of the signal cancellation loop will form an estimate of the desired PA output component, $\alpha_0 v_m$. If the estimate is exactly equal, then v_e will be a faithful replica of v_d . It is well established in linear estimation theory that the covariance of the error v_e and the basis of the estimate v_m form a gradient signal which can be used for the adaptation of the variable α . Mathematically, this relationship is expressed as

$$\frac{\partial P_e}{\partial \alpha} \propto \sigma_{em}^2 = E[v_e v_m^*] \quad (2.107)$$

At this point, the above result is merely being asserted. However, if the right hand side is expanded using equations (2.100) and (2.102),

$$\begin{aligned} E[v_e v_m^*] &= E[(v_a - \alpha v_m) v_m^*] \\ &= E[v_a v_m^* - \alpha |v_m|^2] \\ &= E[v_m G(x_m) v_m^* - \alpha |v_m|^2] \\ &= 2E[x_m G(x_m)] - 2\alpha P_m \end{aligned} \quad (2.108)$$

it is clear by substitution of (2.105) that the left hand side of (2.107) results in

$$\sigma_{em}^2 = -\frac{\partial P_e}{\partial \alpha} \quad (2.109)$$

In practise, it is impossible to compute the exact gradient signal as covariance requires that the expectation of v_e and v_m^* be calculated over all time. For fast adaptation, an approximate gradient must be computed. The simplest estimate is simply the product of the error signal with the basis.

$$D_\alpha = v_e v_m^* \quad (2.110)$$

Figure 2.16 illustrates a circuit which will implement the above operation. It includes a complex mixer to form D_α , an integrator to cancel out some of its self noise and a gain element K_α to control the time constant of adaptation.

Figure 2.16 Adaptation circuit for signal cancellation.

2.5.2 Adaptation of the error cancellation coefficient

The approach to adapting the error cancellation loop is identical to that used for signal cancellation. The analysis proceeds by assuming that α is exactly adjusted to α_0 , or alternatively, that $v_e = v_d$. The aim of the second loop is to minimize the power in the output signal

$$P_o = E\left[\frac{1}{2}|v_o|^2\right] \quad (2.111)$$

By manipulating the equations (2.99)-(2.101), the following expression for v_0 can be obtained

$$v_o = (1 + \beta\epsilon_\alpha)\alpha_0 v_m - \epsilon_\beta v_d \quad (2.112)$$

Where ϵ_α and ϵ_β are useful constants which measure the percentage error in the adjustment of α and β respectively.

$$\epsilon_\alpha = (\alpha - \alpha_0)/\alpha_0 \quad (2.113)$$

and

$$\epsilon_\beta = (\beta - \beta_0)/\beta_0 = \beta - 1 \quad (2.114)$$

When the signal cancellation loop is exactly adjusted, ε_α becomes zero and part of the first term in equation (2.112) disappears. The remainder is substituted into equation (2.111) to obtain

$$\begin{aligned} P_o &= E\left[\frac{1}{2}|\alpha_0 v_m - \varepsilon_\beta v_d|^2\right] \\ &= |\alpha_0|^2 P_m + |\beta - 1|^2 P_d \end{aligned} \quad (2.115)$$

There are no cross-terms in this result because, by allowing α to equal α_0 , v_m and v_d are decorrelated and hence the expectation of their product is zero. As before, the minimum value of P_o occurs when the gradient of this quadratic power surface equals zero. This minimum is found by taking the derivative of P_o with respect to β and equating the result to zero

$$\frac{\partial P_o}{\partial \beta} = 2(\beta - 1)P_d = 0 \quad (2.116)$$

As expected, the minimum power occurs for $\beta = 1$.

Now, the lower branch of the error cancellation loop forms an estimate of the distortion, v_d . As this estimate approaches the actual result, the PA output will contain only the desired linearly amplified component, $\alpha_0 v_m$. Based on linear estimation theory, it is asserted that the covariance of the error, now v_o , and the basis of the estimate v_e form a gradient signal for the adaptation of the variable β .

$$\frac{\partial P_o}{\partial \alpha} \propto \sigma_{oe}^2 = E[v_o v_e^*] \quad (2.117)$$

For proof of this assertion, the right side of this equation can be expanded using equation (2.112) for v_o and equations (2.99) and (2.100) for v_e . The latter two equations can be combined to yield

$$v_e = -\varepsilon_\alpha \alpha_0 v_m + v_d \quad (2.118)$$

Substituting for the right hand side of (2.117),

$$\begin{aligned} E[v_o v_e^*] &= E[((1 + \beta \epsilon_\alpha) \alpha_0 v_m - \epsilon_\beta v_d)(-\epsilon_\alpha \alpha_0 v_m + v_d)^*] \\ &= -2(1 + \beta \epsilon_\alpha) \epsilon_\alpha^* |\alpha_0|^2 P_m - 2(\beta - 1) P_d \end{aligned} \quad (2.119)$$

Near convergence, the expectations of the cross-terms in the previous equation will be zero because v_m and v_d are decorrelated. If the signal cancellation loop is locked, ϵ_α will be zero and the first term of equation (2.119) will vanish. The remainder is simply the negative of equation (2.116). Hence, it appears that covariance of v_o and v_e is a good approximation of the output power gradient

$$\sigma_{oe}^2 = -\frac{\partial P_o}{\partial \beta} \quad (2.120)$$

Should the first loop not be exactly adjusted, the second loop may converge to a value of β other than unity. This observation is based on equation (2.119). Notice that the covariance is a function of not only the distortion power, but the PA output power as well. The combination produces a masking effect because the signal power is many orders of magnitude larger than the distortion power. If the signal loop is not perfectly adjusted, it will make it difficult to detect changes in the distortion power level. The problem can be reduced by inserting a notch-filter at the input to the v_o side of the adaptation circuitry to reduce the detected signal power at the feedforward output.

Since the expectation cannot be calculated exactly, it is again necessary to approximate the gradient by using a statistical measure

$$D_\beta = v_o v_e^* \quad (2.121)$$

The circuit implementation of the distortion loop adaptation algorithm is shown in Figure 2.17. It consists of a complex mixer to form D_β , an integrator to cancel out some of its self noise and a gain element K_β to control the time constant of adaptation.

In view of the comments made above, there is also a filter to reduce the undesirable masking effect.

Figure 2.17 Adaptation circuit for distortion cancellation.

2.5.3 Coefficient accuracy requirements

The accuracy to which the cancellation paths can be adjusted will ultimately decide the level of distortion suppression obtainable from the feedforward linearizer. Since intermodulation distortion is the predominant product of the amplifier's nonlinear character, its power relative to the main signal is an accurate gauge of the overall level of distortion suppression. Based on this observation, a set of definitions will be introduced which relate the required coefficient accuracy to obtain desired reductions in IM power. It should be noted that although these definitions were developed by J. Cavers in his work on the gradient feedforward [19], they are independent of the adaptation technique. The first definition measures the ratio of IM power to signal power at the amplifier output:

$$IMSR_a = \frac{P_d}{|\alpha_0|^2 P_m} \quad (2.122)$$

The second definition is of greater importance because it is the true indicator of the effectiveness of the linearizer. It measures the amount of distortion at the feed-

forward output. If it is assumed that the signal cancellation coefficient α is exactly matched to α_0 , then from equation (2.112) one obtains

$$v_o = \alpha_0 v_m + (1 - \beta) v_d \quad (2.123)$$

Taking the ratio of the distortion power to signal power, the second definition is

$$IMSR_o = \frac{|1 - \beta|^2 P_d}{|\alpha_0|^2 P_m} = |\epsilon_\beta|^2 IMSR_a \quad (2.124)$$

By inspection, it appears that β must be accurate to within 1% in order to lower the IM power at the output by 40 dB. This is being optimistic because it assumes that the adaptation of α and β are independent. When α and β are coupled, as is the case in the stochastic gradient method, a misadjustment in α causes β to converge to a value other than one. If equation (2.119) is set to zero with $\epsilon_\alpha \neq 0$, this value is shown to be dependent on the accuracy to which the signal cancellation loop can be adjusted.

$$\beta = \frac{P_d - \epsilon_\alpha^* |\alpha_0|^2 P_m}{P_d + |\epsilon_\alpha|^2 |\alpha_0|^2 P_m} \quad (2.125)$$

Substituting equation (2.122), the β that results from imperfect signal cancellation can be expressed in terms of $IMSR_a$

$$\beta = \frac{IMSR_a - \epsilon_\alpha^*}{IMSR_a + |\epsilon_\alpha|^2} \approx 1 - \frac{\epsilon_\alpha^*}{IMSR_a} \quad (2.126)$$

Where the approximation holds if ϵ_α is small. Combining this with the definition of equation (2.124), a more realistic indication of the accuracy requirement for α is obtained

$$|\epsilon_\alpha| \approx \sqrt{IMSR_a IMSR_o} \quad (2.127)$$

As an example, if the PA output $IMSR_a = -20$ dB is to be reduced to a final intermodulation level of $IMSR_o = -40$ dB, then α must be adjusted to within 0.1 % of α_0 . At the beginning of section 2.4, it was stated that all linearizers would be analysed for the case of two closely spaced input tones before due consideration was given to signals with non-negligible bandwidth. Throughout sections 2.5.1 and 2.5.2, this state-

ment was observed. Now, the frequency dependence has been re-introduced by means of the coefficient accuracy requirements. If the PA is assumed to have a perfectly flat amplitude response and is properly delay matched, then the accuracy requirement of 0.1 % can be interpreted as the gain flatness requirement for α . This is a very stringent requirement.

In section 2.5.2, it was indicated that the adaptation of the distortion cancellation loop contained a bias term that would cause β to converge to a value other than unity. This explains why the accuracy of α as stated in equation (2.127) is so crucial to obtaining good IM suppression. However, if a rectangular band rejection filter is placed in front of the power detector used to sample the output signal, the masking effect can be reduced and the level of IM suppression improved. In [21], J. Cavers presents an analysis of the adaptation of β including such a filter. The salient result is a modified version of equation (2.127)

$$|\epsilon_\alpha| \approx \frac{1}{3A_m} \sqrt{IMSR_a IMSR_o} \quad (2.128)$$

where A_m is the stopband attenuation. For example, if A_m is -40 dB, the α coefficient accuracy requirement is relaxed from 0.1 % to 3.3 %. Although this remains a stringent result, it will be shown in Chapter 6, that it is possible to achieve over a bandwidth of 500 MHz.

2.5.4 Delay mismatch and bandwidth

Up to this point, the study of the feedforward linearizer has presupposed that there exists negligible delay difference between the upper and lower branches. Over a narrow band, this is a perfectly reasonable assumption. This is less likely for the case of a broadband application where a small difference in group delay may translate to large mismatches at the band-edges. To illustrate the effects of a delay mismatch, assume that the feedforward is operating with α and β adjusted to their optimum values at the centre frequency. Next, suppose that there exists a delay mismatch τ_α in the reference branch of the signal cancellation circuit and a delay mismatch of τ_β in the reference branch of the dis-

tortion cancellation circuit. Using equations (2.97) to (2.101), the complex envelope of the linearizer output signal becomes

$$\begin{aligned} v_0(t) &= v_a(t - \tau_\beta) - v_e(t) \\ &= v_m(t - \tau_\beta) + v_d(t - \tau_\beta) - [v_m(t) + v_d(t) - v_m(t - \tau_\alpha)] \end{aligned} \quad (2.129)$$

Taking its Fourier transform,

$$\begin{aligned} V_o(f) &= [e^{-j\pi f \tau_\beta} + e^{-j\pi f \tau_\alpha} - 1] V_m(f) + [e^{-j\pi f \tau_\beta} - 1] V_d(f) \\ V_o(f) &\approx V_m(f) - (j2\pi f \tau_\beta) V_d(f) \end{aligned} \quad (2.130)$$

The approximation applies provided that $f\tau_\alpha \ll 1$ and indicates that the output contains the desired signal $V_m(f)$ and some distortion. At $f = 0$, the band centre, the distortion term is zero. However, towards the band edges it will grow, depending on the accuracy to which the delay of the upper and lower branches are matched. For example, if a minimum of 30 dB distortion suppression is desired across the entire bandwidth, then the product $2\pi f \tau_\beta$ must be less than 0.032. Assuming that the distortion bandwidth B is roughly the same as that of the multicarrier signal, then the delay mismatch product $B\tau_\beta$ cannot exceed 0.5 %. If $B = 2$ GHz, τ_β cannot exceed 2.5 ps, a very stringent requirement.

2.6 Broadband feedforward implementation considerations

One of the main disadvantages of adaptive feedforward is its complexity compared to other linearization techniques. This section will expand upon this aspect by presenting a full block diagram of a feedforward linearizer for use over the 27.35 GHz to 29.35 GHz range. The bandwidth and nominal frequency level raises three main concerns: adapting for variable path delays in the branches, implementing the signal and distortion loop adaptation circuitry, and the requirement for low-loss RF components with flat in-band responses. The main source for this section is U.S. patent 5,489,875 which was filed by J. Cavers and dated February, 1996 [26].

In the previous section, it was established that accurate delay matching is essential to realize significant IM suppression. The example cited extremely stringent

requirements for this matching, requirements which would be impossible to meet simply by using measured lengths of delay line. In order to achieve the required degree of accuracy, it is necessary to add a variable delay to the cancellation branches. This variable delay control can be achieved using two vector modulators and a length of delay line τ_1 as shown in Figure 2.18. Appropriate adjustment of the vector modulators allow for the interpolation of the delayed signal in the upper branch and the undelayed signal in the lower branch. Hence, the overall delay can range from zero to the delay of the upper branch. Delays above this range remain possible, but with reduced accuracy [26].

The adaptation procedure for the variable delay is identical in principle to that discussed in sections 2.5.1 and 2.5.2. However, there are now two complex coefficients that need to be adapted, α_1 and α_2 . The trick is to decompose the gradient signals into two components with bases given by the delayed reference and undelayed reference signals, $v_m(t)$ and $v_m(t - \tau_1)$. Consider the statistical gradients of the signal cancellation loop:

$$\begin{aligned} D_{\alpha_1} &= v_e(t) v_m(t)^* \\ &= \alpha_0 v_m(t - \tau_0) v_m(t)^* + v_d(t) v_m(t)^* \\ &\quad - \alpha_1 |v_m(t)|^2 - \alpha_2 v_m(t - \tau_1) v_m(t)^* \end{aligned} \quad (2.131)$$

$$\begin{aligned} D_{\alpha_2} &= v_e(t) v_m(t - \tau_1)^* \\ &= \alpha_0 v_m(t - \tau_0) v_m(t - \tau_1)^* + v_d(t) v_m(t - \tau_1)^* \\ &\quad - \alpha_1 v_m(t) v_m(t - \tau_1)^* - \alpha_2 |v_m(t - \tau_1)|^2 \end{aligned} \quad (2.132)$$

At this point, it is important to note that the adaptation of the parameters α_1 and α_2 is coupled. This coupling may induce oscillations or cause the system to convergence to a non-optimal state. The problem is circumvented by decorrelating the two reference signals $v_m(t)$ and $v_m(t - \tau_1)$ prior to computing the statistical gradients D_{α_1} and D_{α_2} . This can be achieved by passing them through the matrix transform Q .

Figure 2.18 Signal cancellation loop with adaptive delay.

$$\begin{bmatrix} c_1(t) \\ c_2(t) \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} v_m(t) \\ v_m(t - \tau_1) \end{bmatrix} \quad (2.133)$$

If this matrix is chosen appropriately, it will decouple the correlated $v_m(t)$ and $v_m(t - \tau_1)$ processes and produce the uncorrelated $c_1(t)$ and $c_2(t)$ processes. Selection of the components of the Q matrix is guided by a technique known as the Karhunen-Loeve expansion. The orthogonalization of random processes is described in many books on probability and random variables and won't be discussed in any further detail here. Once the independent gradients are calculated, they are reverted to the basis defined by the vectors $v_m(t)$ and $v_m(t - \tau_1)$ using Q^{-1} . This configuration is shown in Figure 2.18.

Another complication associated with broadband feedforward linearization is the implementation of the adaptation circuitry. Not only must it be able to calculate D_α

and D_β rapidly, but it must be able to do so over a wide bandwidth. Since most broadband commercial wireless services will include multiple carriers, the adaptation circuitry should have the capacity to “notch-out” the distortion bands for the carrier cancellation and the carrier bands for the distortion cancellation. If this is not done properly, then the masking effect described in section 2.5.3 will cause β to converge to an incorrect value. Although less noticeable in the carrier cancellation loop, a small bias in the α coefficient will exist if the distortion components of v_e are not removed.

This problem is mitigated by calculating a partial gradient as opposed to a full gradient. The partial gradient is a simple application of Parseval’s theorem which states that the energy of a signal can be calculated either in the time domain or in the frequency domain.

$$\int_{-\infty}^{\infty} v^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |V(\omega)|^2 d\omega \quad (2.134)$$

The right-hand side can be separated into a series of integrals over limited bandwidths. For the signal cancellation loop, all integrals over bands that do not correspond to a modulated carrier are considered to be zero. Hence, the adaptation signal for α , when taken over a sum of partial bands appears as

$$\alpha = \frac{K_\alpha}{\pi} \sum_{k=1}^N \left(\int_{\omega_k - 2\pi B_k}^{\omega_k + 2\pi B_k} V_e(\omega) V_m(\omega)^* d\omega \right) \quad (2.135)$$

where the ω_k ’s and B_k ’s represent the centre frequencies and bandwidths of the channels in the multicarrier signal. A similar expression can be devised for the distortion cancellation loop. By excluding the signal bands in the latter case, the masking of the distortion by the signals is effectively eliminated.

Figure 2.19 shows schematically what a wideband partial gradient adaptive feedforward linearizer might look like. The first, and most obvious difference between this schematic and all the previous feedforward representations is the location of the signal

loop variable gain, phase, and delay adjuster. To date, it has always been located in the lower branch whereas in Figure 2.19 it is placed immediately before the PA. This alternative placement produces a negligible difference in the optimal values of α_1 and α_2 [7]. At the same time, any inadvertent distortion generated by the first set of vector modulators will be removed at the output.

Below the RF circuit are two partial gradient controllers which each contain three RF downconverters (RFM1, RFM2, RFM3) and an RF oscillator (RFLO1) to drive them. The purpose of these elements is to shift the 28 GHz signal down to an IF frequency which can be handled by cheaper low-frequency circuits. Following filtering (BPF1, BPF2, and BPF3), there is a second set of mixers (M4, M5, M6) and a second oscillator (LO2). By stepping the frequency of LO2, it is possible to selectively downconvert each carrier or distortion band one after the other to a range suitable for sampling. After the second down-conversion, a second bank of filters (BPF4, BPF5, and BPF6) will

Figure 2.19 Feedforward linearizer with adaptive delay and partial gradient.

ensure that only the narrow bandwidth of interest passes on to the analog to digital converters (ADC1, ADC2, and ADC3).

The sampled signals are stored as floating point numbers in a buffer of a digital signal processor (DSP) which will perform all subsequent adaptation operations. The first step is to convert the delayed and undelayed inputs to the uncorrelated processes c_1 and c_2 . Next, the converted inputs are multiplied by the sampled error signal and integrated over a finite time interval to remove some of the gradient signal self-noise. This process will be repeated until the oscillator LO2 has stepped to each carrier or distortion band. Finally, the results of each multiplication and integration will be summed, converted to the original coordinate axes, and output as a level to drive the vector modulators.

Unfortunately, this scheme will not accomodate true dynamic channel allocation without some modification. Basically, the use of the common IF bandpass filters BPF4, BPF5, and BPF6 requires that all carriers be of exactly the same bandwidth. To achieve true dynamism, these filters must be adaptable. One solution is for them to have bandwidth as wide as the highest data rate signal. Following sampling, the variability can then be implemented by changing the coefficients of a set of digital filters within the DSP block.

2.7 Summary

The intent of this section was to impress upon the reader the causes and classes of amplifier non-linearity. Subsequent to this, several alternative linearization strategies were considered and finally, feedforward was settled on due to its inherent capability of handling wide bandwidth signals. This was followed by a more thorough investigation of the operation of an adaptive feedforward linearizer. Section 2.6 gives a complete impression as to the level of complexity required to realize a wideband feedforward at Ka-band and it is intended to raise some important questions.

The first question concerns the large number of MMIC chips that are required to execute the variable gain, phase, and delay tuning, and the number of mixers and oscillators to downconvert the RF signal. Each one of these chips is expensive, and is required to have a bandwidth which can handle from 500 MHz to 2 GHz. According to Figure 2.19, the total number of MMICs required is 12: 4 vector modulators, 2 power amplifiers, 6 downconverters and one oscillator (the RFLO1 is assumed shared between both the signal and distortion cancellation branches). This remains optimistic because it assumes that vector modulators are single MMICs. The second question concerns the consumption of power. All of these MMICs, plus the additional low frequency RF mixers, oscillator (LO2), analog to digital converters, and digital to analog converters will draw power and thus drive down the efficiency of the overall system. In fact, the loss of efficiency is most certainly higher than if a large number of corporately fed PAs were simply backed off. Hence, efficiency is really not the issue with this linearizer. The third, and final question concerns the large number of splitters and combiners in the RF power path. Each one of these will introduce conductor and dielectric losses which will degrade the total output power. In all likelihood, it will be necessary to add further preamplifiers before the PA and the AA in order to restore some signal power. This will just add to the loss in efficiency of the overall circuit.

It would be of considerable value if further study were conducted to assess the overall viability of the feedforward project. Following a careful examination of power consumption, cost, and linearity, it may turn out that, rather than linearize a single PA simply adding a large number of backed-off PAs in parallel may provide comparable or possibly even superior performance.

CHAPTER 3: Vector modulator architectures

3.1 Introduction

The principal RF design concern of this thesis is a vector modulator. Its function is to electronically vary the amplitude and phase of a signal over a continuous range of states. The response is limited only by a maximum transmission amplitude. This chapter will introduce two possible configurations for such a device: phase split-amplitude adjust-combine, and phase shift-amplitude adjust. It will also explore some of the possibilities for realizing the amplitude control function, review the available active device technologies, and decide which one of the two vector modulator configurations will be used.

3.2 Phase split, amplitude adjust, and combine

The first approach to developing a vector modulator is to use hybrid couplers to divide the input signal into equal-amplitude quadrature components. Using a cascade of one branchline coupler and two rat-race couplers, it is possible to define a set of orthogonal vectors in a two-dimensional phase space. By varying the amplitude of these vectors and combining them in phase, the resultant signal can assume any desired ampli-

Figure 3.1 Schematic of a phase split, amplitude adjust and combine vector modulator.

tude and phase (subject to upper and lower amplitude bounds). This configuration is shown in Figure 3.1. Assuming an input signal of the form

$$S_i = A(t) \sin[\omega t + \phi(t)] \quad (3.1)$$

The couplers will split the signal into four components as follows

$$v_{ik}(t) = \frac{cA(t)}{4} \sin\left[\omega t + \phi(t) + \frac{\pi}{2}(5 - 2k)\right] \quad k = 1, 2 \quad (3.2)$$

$$v_{ik}(t) = \frac{cA(t)}{4} \sin[\omega t + \phi(t) + \pi(k - 3)] \quad k = 3, 4 \quad (3.3)$$

where c represents any losses incurred during the splitting process. Each one of these “vectors” is amplified or attenuated by an amount specified by the G_k ’s and summed using a corporate combiner of three Wilkinson couplers to yield

$$\begin{aligned} v_0(t) &= K_3 A(t) \sin[\omega t + \phi(t)] + K_2 A(t) \sin[\omega t + \phi(t) + \pi/2] \\ &\quad + K_4 A(t) \sin[\omega t + \phi(t) + \pi] + K_1 A(t) \sin[\omega t + \phi(t) + 3\pi/2] \\ &= (K_3 - K_4) \sin[\omega t + \phi(t)] + (K_2 - K_1) A(t) \sin[\omega t + \phi(t) + \pi/2] \\ &= KA(t) \sin[\omega t + \phi(t) + \theta] \end{aligned} \quad (3.4)$$

where

$$\begin{aligned} K &= \sqrt{(K_3 - K_4)^2 + (K_2 - K_1)^2} \\ \theta &= \text{atan}\left[\frac{K_2 - K_1}{K_3 - K_4}\right] \end{aligned} \quad (3.5)$$

The K_k ’s include the losses c and h of the combiners and the gain of the amplifiers, i.e. $K_k = chG_k$.

The most significant disadvantage to this configuration is the inherent loss introduced by splitting the signal. For example, if $K_3 = K_2 = 1$ and $K_4 = K_1 = 0$, the insertion loss is 6 dB. This configuration throws away half the power even when the amplitude response is maximized. The problem can be remedied by reducing the number of orthogonal paths from four to two. However, in order to achieve the full phase range, bi-phase amplitude controllers, such as that shown in Figure 3.2, are required. Basically, the

Figure 3.2 Block diagram of a bi-phase amplitude controller.

single-pole double-throw (SPDT) switches direct the signal into a phase shifted or a non-phase shifted path. Hence, the amplitude of this device can range from $-K$ to $+K$ with the exception of states very close to zero transmission.

3.3 Phase shift and amplitude adjust

The second approach involves cascading a 0 to 360° analog phase shifter and a variable gain element (see Figure 3.3). Unlike the four path vector modulator of the previous section, this configuration does not suffer from any inherent insertion loss penalties other than those introduced by the phase shifter.

Figure 3.3 Schematic of a phase shift and amplitude adjust vector modulator.

The analog phase shifter can be realized using a 3 dB quadrature coupler with reflective terminations as shown in Figure 3.4. A signal entering at the input is split into two equal amplitude quadrature components and reflected to the output. The overall insertion phase of the device is controlled by the reactance of a pair of varactor diodes. Thus, by adjusting the bias supplied to the two varactors, a range of insertion phases is possible. In general, to obtain a full 360° phase shift, it is necessary to cascade at least two individual stages. This topology has been successfully implemented at X and Ku-band [27] and is definitely adaptable for Ka-band. Furthermore, it has demonstrated octave bandwidths, very low insertion loss, and excellent input return loss.

Figure 3.4 Analog phase shifter topology.

3.4 Amplitude control

A vector modulator requires an analog amplitude control element in order to adjust the output signal level. The controller can be active or passive and can be realized in a number of different technologies.

For the first vector modulator configuration, an active amplitude controller is essential if some of the loss generated by the phase split and combine operations is to be restored. The most intuitive approach is to design a standard small-signal microwave amplifier with gain that varies as a function of applied bias. The variable gain amplifier (VGA), as it is called, can be realized at Ka-band using either HBT or PHEMT technology. However, certain measures must be taken to flatten the inherent transistor gain roll-off and to reduce the input and output return loss. The latter problem arises because it is impossible to match the transistor exactly at all bias states.

If the second vector modulator configuration is used, then it is possible to use a passive amplitude controller — a variable attenuator. The main advantage to this approach is greater bandwidth and flatter amplitude response. There are two common attenuator configurations as shown in Figure 3.5: the T-network and the π -network. The

Figure 3.5 Variable attenuators, (a) T-network and (b) π -network.

variable resistance can be created using either a PIN diode or a field effect transistor (FET). The attenuation achieved by either of these schemes is given by

$$K = \frac{R_2}{R_1 + R_2 + Z_0} \quad (3.6)$$

and is derived by simple network analysis. Unfortunately, it is very difficult to suppress the input and output return loss in these configurations. When changing the attenuation, the values of R_2 and R_1 must be adjusted by very specific amounts or the terminal impedances will change and cause a mismatch. Another disadvantage to these attenuators is the requirement for two control inputs to adjust the two separate variable resistances.

By comparison, an attenuator based on the quadrature-hybrid reflection topology is much simpler. Referring to Figure 3.6, the device appears almost identical to

Figure 3.6 Variable attenuator, reflection type.

the analog phase shifter discussed in the previous section. However, instead of reflective varactor diode terminations, the attenuator uses variable resistors in the form of cold FETs. To achieve the variable resistance, the FET is biased in the triode or linear operating region.

3.5 Review of active device technologies

For Ka-band MMIC applications, there are several possible device technologies available to the designer. The most prominent among them are based on the GaAs pseudomorphic high electron mobility transistor (PHEMT) and the GaAs heterojunction bipolar transistor (HBT). Both technologies are capable of delivering sufficient gain and power to permit the development of high performance Ka-band RF-circuits such as mixers, amplifiers, and oscillators. This capability is due to a number of factors, the most notable being the intrinsic electrical benefits offered by GaAs itself. These benefits include high electron mobility, semi-insulating substrates for low parasitics, and monolithic integrated circuit compatibility. Furthermore, PHEMT and HBT devices are fabricated using advanced epitaxial growth techniques which enable device designers to insert deliberate band-gap discontinuities called heterojunctions. These structures can dramatically increase both the speed and gain of these devices.

The field is not completely limited to GaAs as the last ten years of MMIC development have also seen the entry of InP PHEMTs and HBTs, and more recently, the SiGe HBT. The InP devices are, without a doubt, the fastest available with useable gain extending up to 213 GHz [28]. Unfortunately, the processing is prohibitively expensive, complicated, and low yield. Nonetheless, the unrivalled material properties of InP make it attractive for high frequency and ultra-broadband applications.

The SiGe HBT, on the other hand, is a more recent technology which comes in two varieties. The first, which is being investigated in Canada by Nortel Technologies and SiGe Microsystems, is based on the existing CMOS silicon-based semiconduc-

tor techniques. Its main advantage is that it uses high-volume, low cost Si fabrication techniques. However, the technology is limited to applications below 20 GHz and uses lossy conductive substrates. The second variety is being developed in Europe by Daimler-Benz and aims to compete directly with the GaAs HBT at frequencies up to 77 GHz. It has evolved along slightly different lines from the North American technology as it uses molecular beam epitaxy and high resistivity Si substrates. The disadvantage with this technology is that there are currently no significant cost-savings relative to GaAs.

Although five separate device technologies have been reviewed, the SiGe and InP alternatives must be discarded because they are either too expensive or inaccessible at this point in time. This means that the field is reduced to one of the two high-speed GaAs technologies. A GaAs PHEMT process can produce Schottky varactors, variable resistors, and gain blocks for variable gain amplifiers. Hence, it could easily accommodate either vector modulator configuration. The GaAs HBT process, on the other hand, is more limited. Firstly, it cannot generate an effective variable resistance as the slope in the measured DC-IV curves does not vary significantly with changes in applied base bias. Secondly, although a variable capacitance is possible using the base-collector junction, the process tolerances tend to be rather poor [33]. This presents a problem for the analog reflection topology because it requires near-identical terminations at both of the reflecting ports in order to obtain good input/output match. Thus, the HBT is best suited to the first vector modulator configuration with its combination of passive structures and variable gain amplifiers.

3.6 Configuration selection

In the previous four sections, the possible vector modulator configurations have been outlined and some concepts for the amplitude controller have been presented. In addition, the GaAs PHEMT process appears to be the most amenable to the application at hand. However, the final decision must account for non-technical factors as well. When

work on this thesis project began, the fabrication schedule indicated that a Nortel HBT run was the most imminent. Moreover, there were no immediate plans for a PHEMT run.

Following consideration of the facilities and time available, the vector modulator for this thesis must be of the phase split, amplitude adjust, and combine variety as described in section 3.2. According to the arguments of sections 3.4 and 3.5, the amplitude controller must be a variable gain amplifier.

3.7 Summary

The purpose of this chapter is to assert that a vector modulator can be designed in more than one way. Having done this, the next aim was to explain the rationale for choosing the first configuration over the second.

CHAPTER 4: Variable gain amplifier design

4.1 Introduction

The key component required to realize the vector modulator of section 3.2 is an MMIC variable gain amplifier (VGA). The purpose of this chapter is to introduce the design and characteristics of an HBT variable gain amplifier intended to operate at 28 GHz. The chip area available is 1.25 x 2.5 mm. The bandwidth is chosen to range from 27.35 to 29.35 GHz and will bound all possible bands likely to be auctioned by Industry Canada for 28 GHz LMCS use. The first sections of this chapter lay the ground work by reviewing the theory of small-signal microwave amplifier design and the operation of the heterojunction bipolar transistor. This is followed by a series of sections which document the issues and methods specific to variable gain amplifier design. The chapter ends with an assessment of the final design in terms of its simulated and measured performance.

4.2 Fundamental amplifier design techniques

Transistor amplifiers can be designed according to three main specifications: maximum gain, low noise, and high power. The choice of mode of operation will decide the goals of the input, output and/or interstage matching networks. The variable gain amplifier developed here is a maximum gain amplifier. It consists of two gain stages or blocks, both of which are simultaneously conjugately matched. According to microwave transistor amplifier design theory, this will yield the maximum possible gain. This section will briefly examine some of the essential theory required to develop an amplifier of this type.

Figure 3.1 illustrates the fundamental design for a two-stage maximum gain amplifier matched to 50 Ω source and load impedances. It consists of four distinct

segments: transistor gain blocks, input matching network, interstage matching network and output matching network.

Figure 4.1 A two-stage transistor amplifier with matching networks.

Each block is completely characterized by its N -port scattering or S matrix. As shown in equation (4.1) and Figure 4.2, it relates the incident and reflected waves at all N ports.

$$\begin{bmatrix} V_1^- \\ V_2^- \\ \vdots \\ V_N^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1N} \\ S_{21} & & & \vdots \\ \vdots & & & \\ S_{N1} & \dots & & S_{NN} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ \vdots \\ V_N^+ \end{bmatrix} \quad (4.1)$$

where V_n^+ is the amplitude of the wave incident on port n , and V_n^- is the amplitude of the voltage wave reflected from port n . These wave amplitudes assume that all terminating impedances are equal. In other words, the coefficients of the scattering matrix are identical to the S parameters that would be measured on a network analyser with common $50 \, \Omega$

Figure 4.2 An N -port network characterized by S parameters.

source and load terminations. Removing this assumption, a more general set of wave amplitudes is defined as

$$\begin{aligned} a_n &= V_n^+ / \sqrt{Z_{0n}} \\ b_n &= V_n^- / \sqrt{Z_{0n}} \end{aligned} \quad (4.2)$$

Here, a_n represents an incident wave at the n th port, and b_n represents a reflected wave from the n th port. The n th port is characterized by characteristic impedance Z_{0n} instead of 50Ω . Using these amplitudes, the generalized scattering matrix becomes

$$[b] = [S'] [a] \quad (4.3)$$

with elements given by

$$S'_{ij} = \left. \frac{b_i}{a_j} \right|_{a_k = 0 \text{ for } k \neq j}, \quad S'_{ii} = \left. \frac{b_i}{a_i} \right|_{a_k = 0 \text{ for } k \neq i} \quad (4.4)$$

The transistor gain blocks shown in Figure 4.1 are two-port networks and can be characterized by a 2×2 scattering matrix as defined by equations (4.3) and (4.4). In the design of microwave amplifiers, the most useful measure of the power amplification is the transducer power gain, G . Referring to Figure 4.3, the power gain is defined as the ratio of the average power P_L delivered to the load Z_L when the maximum power available from the generator is P_a . With no assumptions regarding the source and load impedances, the power gain of the 2 port is defined as

$$G = \frac{P_L}{P_a} = \frac{(Re Z_L) |I_2|^2}{|V_S|^2 / 4 (Re Z_S)} \quad (4.5)$$

Figure 4.3 A 2-port network excited by a generator at the input.

It can also be shown that $G = |S_{21}|^2$ for a two-port network. Now, it is desirable to adapt equation (4.5) to use the S parameters defined in equation (4.1). This is accomplished using a set of equations which relate S' to S [12].

$$S' = A^{-1}(S - \Gamma^{T*})(I_2 - \Gamma S)^{-1}A^{T*} \quad (4.6)$$

where

$$\begin{aligned} A &= \frac{1}{2}Z_0^{-1}R'^{-1}(I_2 - \Gamma^{T*}) \\ \Gamma &= (Z' - Z)(Z' + Z)^{-1} \\ Z' &= \begin{bmatrix} Z_S & 0 \\ 0 & Z_L \end{bmatrix}, \quad Z = Z_0 I_2 \\ R' &= \begin{bmatrix} \frac{1}{2}(ReZ_S)^{-1/2} & 0 \\ 0 & \frac{1}{2}(ReZ_L)^{-1/2} \end{bmatrix}, \quad R = Z_0 I_2 \end{aligned} \quad (4.7)$$

Substituting equation (4.6) and (4.7) into (4.5) yields the transducers power gain in terms of its non-generalized scattering parameters

$$G = |S_{21}|^2 = |S_{21}|^2 \frac{(1 - |\Gamma_S|^2)(1 - |\Gamma_L|^2)}{|(1 - \Gamma_S S_{11})(1 - \Gamma_L S_{22}) - \Gamma_S \Gamma_L S_{12} S_{21}|^2} \quad (4.8)$$

To offset the use of the non-generalized scattering parameters while accounting for the mismatch between the transducer ports and the external 50 Ω terminations, the reflection coefficients Γ_S and Γ_L now appear. These two quantities are defined in Figure 4.4 and are related to the source and load impedances as follows

$$\Gamma_S = (Z_S - Z_0)/(Z_S + Z_0) \quad (4.9)$$

$$\Gamma_L = (Z_L - Z_0)/(Z_L + Z_0) \quad (4.10)$$

Figure 4.4 2 port network showing source and load reflection coefficients.

The term Z_0 in the above equations is the reference impedance; $50\ \Omega$ in most instances. Γ_s and Γ_L are very important because they can be controlled. By inserting the appropriate matching networks between the generator and the input and the load and the output, it is possible to present optimal reflection coefficients such that G is maximized. With reference to Figure 4.4, this condition is met when $\Gamma_{IN} = \Gamma_s^*$ and $\Gamma_{OUT} = \Gamma_L^*$ as it implies that there are no reflections at the input or output of the gain block or transducer. The transducer is now said to be simultaneously conjugately matched.

If S_{12} of the transducer is not equal to zero, then the selection of the optimal load and source impedances becomes a coupled problem. That is to say, there is a unique pair of load and source reflection coefficients that will simultaneously satisfy the requirements for maximum G . This pair is determined by solving the following pair of coupled equations

$$\Gamma_{IN} = \Gamma_s^* = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - \Gamma_L S_{22}} \quad (4.11)$$

$$\Gamma_{OUT} = \Gamma_L^* = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - \Gamma_s S_{11}} \quad (4.12)$$

Γ_{IN} represents the reflection coefficient seen looking into the input of the transducer when the output is presented with Γ_L . Similarly, Γ_{OUT} represents the reflection coefficient seen looking into the output of the transducer when the input is presented with Γ_s . Solving this pair of equations simultaneously yields the following optimal source and load reflection coefficients, Γ_{SM} and Γ_{LM}

$$\Gamma_{SM} = (C_1^*[B_1 \pm (B_1^2 - 4|C_1|^2)^{1/2}]) / (2|C_1|^2) \quad (4.13)$$

$$\Gamma_{LM} = (C_2^*[B_2 \pm (B_2^2 - 4|C_2|^2)^{1/2}]) / (2|C_2|^2) \quad (4.14)$$

where

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 \quad (4.15)$$

$$B_2 = 1 - |S_{11}|^2 + |S_{22}|^2 - |\Delta|^2 \quad (4.16)$$

$$C_1 = S_{11} - \Delta S_{22}^* \quad (4.17)$$

$$C_2 = S_{22} - \Delta S_{11}^* \quad (4.18)$$

$$\Delta = S_{11}S_{22} - S_{12}S_{21} \quad (4.19)$$

The minus sign in equations (4.13) and (4.14) is applied when $B_{1,2} > 0$ and the plus sign when $B_{1,2} < 0$. Using these optimal reflection coefficients, the maximum available gain can be obtained by substituting into equation (4.8).

Competing with the requirement for maximum gain is the stipulation that the amplifier be stable, or in other words, not oscillate. A transducer is said to be stable if $|\Gamma_{IN}| < 1$ and $|\Gamma_{OUT}| < 1$ which in turn implies that the real components of Z_{IN} and Z_{OUT} always be positive. This stability is unconditional so long as the two-port meets the following conditions

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |S_{11}S_{22} - S_{12}S_{21}|^2}{2|S_{12}S_{21}|} > 1 \quad (4.20)$$

$$|S_{11}| < 1, \quad |S_{22}| < 1$$

K is referred to as the stability factor and is derived from equations (4.13) and (4.14). For a more thorough development of this parameter, see Appendix I of Tri T. Ha [12]. As long as these conditions are met from DC to the highest possible frequency where the device exhibits gain (f_{max}), then it will be unconditionally stable. However, in the case of the variable gain amplifier, it will be shown that there is a region wherein the device is potentially unstable ($K < 1$). In such an instance, it is imperative that the region of instability be

bounded. This can be accomplished by determining all possible values of Γ_S and Γ_L that result in $|\Gamma_{IN}|, |\Gamma_{OUT}| > 1$.

As it turns out, the boundaries for source and load stability can be plotted as circles in polar coordinates. All points on these circles correspond to the case where $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$. The centre points and radii for these circles are determined by manipulating equations (4.11) and (4.12) and are given as

$$\begin{aligned} C_L &= \frac{S_{11}\Delta^* - S_{22}^*}{|\Delta|^2 - |S_{22}|^2}, & C_S &= \frac{S_{22}\Delta^* - S_{11}^*}{|\Delta|^2 - |S_{11}|^2} \\ r_L &= \left| \frac{S_{12}S_{21}}{|\Delta|^2 - |S_{22}|^2} \right|, & r_S &= \left| \frac{S_{12}S_{21}}{|\Delta|^2 - |S_{11}|^2} \right| \end{aligned} \quad (4.21)$$

The most useful means of interpreting these boundaries is to plot them on a Smith chart. Figure 4.5 shows such a plot for the load stability circle. If only passive loads are considered and $|S_{11}| < 1$, then the portion of the Smith chart that lies inside the circle represents choices of Γ_L that result in $|\Gamma_{IN}| > 1$. Occasionally, the load stability circle will enclose the origin. In this case, all load reflection coefficients within the circle will be stable. Should $|S_{11}| > 1$, then the opposite of these two instances is true. A similar set of rules applies to the source stability circle which is a plot of all values of the source reflection coefficient, Γ_S , that give $|\Gamma_{OUT}| = 1$. For example, when $|S_{22}| < 1$ all points within the source stability circle are stable if the circle encloses the origin. If the circle does not include the origin, then all source reflection coefficients outside the circle are stable.

For the case of the variable gain amplifier, there exists a band of frequencies for which the gain blocks become potentially unstable (i.e. $K < 1$). More specifically, $|\Gamma_{OUT}| > 1$ between 16 and 19 GHz when the Γ_S for the second stage is chosen for maximum gain at 28 GHz. Although these frequencies fall outside the operating band, there is a possibility that a transient signal may exist for a short period of time which contains harmonics at these frequencies. Owing to the instability, the transistor will amplify these har-

Figure 4.5 Load stability circle plotted on the Smith chart.

monics until they swamp out the desired performance. The issue of stability and the variable gain amplifier will be addressed again in sections 4.4.2, 4.6.2, and 4.6.4.

As illustrated in Figure 4.1, the amplifier consists of two gain stages or blocks. This configuration necessitates the inclusion of an interstage matching network. Unlike the input and output matching networks which transform the gain stage impedances to $50\ \Omega$, the interstage matching network must transform Γ_{IN2} seen looking into the second stage to Γ_{LM1} for the first stage. This raises yet another issue because the Γ_{IN2} depends on the reflection coefficient presented to the output of the second stage. Similarly, Γ_{LM1} depends on the source reflection coefficient presented to the input of the first stage. The approach espoused in this work is to simply design the interstage matching network

using Γ_{IN2} and Γ_{LM1} . This assumes that the input of stage one can be presented with exactly Γ_{SM1} and that the output of stage two can be presented with exactly Γ_{LM2} .

The interstage matching network serves another useful purpose and that is to flatten the gain. When operating a transistor at high frequency, the gain rolls-off at approximately 6 dB/octave. Hence, over an appreciable bandwidth, even if the input and output matching networks could present the optimal reflection coefficients at each frequency, the gain would still roll-off. The interstage matching network can be used to offset this problem. This is accomplished in two ways: introducing deliberate mismatch at the band-edges and/or designing a sloped interstage passband response. If the slope of this response increases by 6 dB/octave, then it stands to reason that it will counteract the transistor roll-off. Unfortunately, these compensation techniques result in a decrease in the overall gain of the amplifier. These ideas are put into practise in section 4.5.3.

4.3 GaAs HBT description and modelling

This section will provide a description of the active device used for the variable gain amplifier. It is divided into two subsections. The first is qualitative in nature and describes the main attributes of the GaAs HBT. It is followed by a subsection which details the characteristics and components of the HBT's non-linear model.

4.3.1 Operation and description of the GaAs HBT

The GaAs HBT operates according to the same general principles as the Si npn-BJT. That is to say, it is a bipolar device consisting of a p-doped base layer sandwiched between n-doped emitter and collector layers. However, several features of the GaAs HBT enable it to operate at much higher frequencies than its Si BJT counterpart.

The figure of merit used for comparing the high-frequency performance of comparable bipolar transistors is the maximum frequency of oscillation, f_{max} . This measure corresponds to the frequency at which the unilateral gain of the common-emitter connected transistor equals unity. Unilateral gain U is obtained by adjusting a lossless

feedback network until $S_{12} = 0$ and then conjugately matching both ports. However, this measurement is virtually impossible to make directly on GaAs HBTs owing to the high frequencies involved. Normally, it is obtained through extrapolation or calculated from the following formula

$$f_{max} = \left(\frac{f_T}{8\pi R_b C_{bc}} \right)^{1/2} \quad (4.22)$$

Here, R_b is the base resistance, C_{bc} is the base-collector depletion capacitance, and f_T is the gain-bandwidth product or cut-off frequency. All three parameters are subject to change based on material properties and device structure. However, in comparing the GaAs HBT to the Si BJT, only f_T and R_b need be considered to explain the HBT's superior high-frequency performance.

GaAs, as a material, has a much higher mobility than Si for equivalent electric fields. By definition, mobility μ equals the slope of the velocity versus electric field curve at a specified value of the electric field. At lower field levels, as shown in Figure 4.6, the GaAs velocity-electric field characteristic is a great deal steeper than that of Si. However, care must be taken to avoid operating the GaAs device at field levels beyond 4×10^3 V/cm as the mobility becomes negative and the electrons will travel opposite to the desired direction. If this precaution is properly observed, then electrons will traverse the emitter-base depletion region, the base region and the base-collector depletion region much more quickly in a GaAs HBT than in a Si BJT. The time required to cross these regions within the device is given as the emitter to collector transit time, τ_{ec} . Typical tran-

Figure 4.6 Steady-state velocity-field characteristics for electrons in GaAs and Si.

sit times for a Si BJTs are of the order of 12 picoseconds while those for GaAs HBTs are closer to 4 picoseconds [29].

The transit time of a device is inversely proportional to the gain-bandwidth product: $f_T = (2\pi\tau_{ec})^{-1}$. This is the frequency at which the common-emitter *current gain* is reduced to unity. Since the GaAs HBT has a much lower emitter-to-collector transit time than the Si BJT, it has a higher gain-bandwidth product.

To appreciate why the base resistance of equation (4.22) is superior in the HBT, the device structure must be considered in greater detail. Shown in Figure 4.7 is a biased npn transistor with its associated energy-band diagram. Electrons are injected into the base at the reverse-biased emitter and diffuse through the base to the collector junction. Once at this point, the electrons are swept across the base-collector junction by the potential drop. These charge transfers form the emitter and collector currents I_{en} and I_{cn} . As the electrons diffuse across the p-doped base, there is a possibility that some will recom-

Figure 4.7 NPN transistor with associated energy-band diagram.

bine with holes. This phenomena is represented in Figure 4.7 by the use of a “*” for the recombination event and the currents I_{rm} and I_{rp} . Finally, holes can leak from the base into the emitter and from the collector into the base. The latter is considered insignificant while the former is represented by the current I_{ep} .

There exist two common measures of HBT or BJT device performance.

The first is the static common-emitter current gain

$$\beta = \frac{I_C}{I_B} = \frac{I_{cn}}{I_{ep} + I_{en} - I_{cn}} = \frac{I_{en} - I_r}{I_{ep} + I_r} \quad (4.23)$$

where substitutions have been made for the collector and base current. Notice also that the difference in the emitter injection current I_{en} and the collected current I_{cn} is simply the recombination current $I_r = I_{rp} = I_{rn}$. The second figure of merit is the emitter efficiency. It is intended to indicate what proportion of the emitter current is due to the electrons entering the base

$$\gamma = \frac{I_{en}}{I_{ep} + I_{en}} \quad (4.24)$$

These simple relationships are very useful for comparing similar devices. Clearly, to increase both the current gain and the emitter efficiency, it is desirable to design a device with the smallest possible I_{ep} . Numerous textbooks develop expressions for these currents based on the carrier distributions at the edges of the depletion regions. For the case of a simple planar junction, the two currents of interest as far as equations (4.23) and (4.24) are concerned are approximated by the following [29].

$$I_{ep} \approx \frac{qAD_E}{L_E} p_{eo} e^{-\frac{qV_p}{kT}} \quad (4.25)$$

$$I_{en} \approx \frac{qAD_B}{W} n_{bo} e^{-\frac{qV_n}{kT}} \quad (4.26)$$

The crucial terms in these expressions are the equilibrium minority electron carrier density in the base n_{bo} , the equilibrium minority hole carrier density in the emitter p_{eo} , the potential difference seen by the base electrons V_n , and the potential difference seen by the emitter holes V_p . The remaining terms are: the fundamental unit of charge q , the cross-sectional area of the base-emitter junction A , the diffusion coefficients in the emitter and base D_E and D_B , and the average diffusion lengths charge carriers travel before recombining L_E and W . Substituting equations (4.25) and (4.26) into (4.23) and (4.24) yields

$$\beta = \frac{n_{bo} v_{nb}}{p_{eo} v_{pe}} e^{\frac{q}{kT}(V_p - V_n)} = \frac{n_{bo} v_{nb}}{p_{eo} v_{pe}} e^{\frac{\Delta E_s}{kT}} \approx \frac{N_E v_{nb}}{N_B v_{pe}} e^{\frac{\Delta E_s}{kT}} \quad (4.27)$$

$$\gamma = \frac{\beta}{1 + \beta} \quad (4.28)$$

where $D_E/L_E = v_{nb}$ and $D_B/W = v_{pe}$, the average velocities of the electrons in the base and the holes in the emitter.

In a homojunction, the bandgap is constant throughout the device. This means that $\Delta E_g = 0$ and β is a function of the relative doping of the emitter to the base. To obtain perfect emitter efficiency, γ must approach unity and β must be large. In terms of equation (4.27), this requirement can only be met if $N_E \gg N_B$. In practise, the emitter can be doped no more than one or two orders of magnitude greater than the base. On the other hand, if the base is doped too lightly, then its resistance R_b will increase. According to equation (4.22), this will reduce the maximum frequency of oscillation.

The heterojunction in an HBT provides another degree of freedom with which to improve device characteristics. A heterojunction is formed when two materials with different bandgap energies but similar lattice constants are joined. For GaAs HBTs, the junction consists of a GaAs base and either an AlGaAs or GaInP emitter. Careful scrutiny of Figure 4.7 will reveal that the emitter band gap is greater than the base band gap by an amount $\Delta E_g = q(V_p - V_n)$. From equation (4.27), it is immediately obvious that the magnitude of the gap energy has a significant affect on β and consequently the emitter efficiency as well. By maximizing the band gap difference, it is possible to relax the doping restrictions on the base relative to the emitter. By allowing the base doping to increase, the base resistance drops and consequently f_{max} increases.

In summary, the GaAs HBT is a device which operates in a similar manner to a Si BJT. Due to the inherently superior mobility of the electrons in GaAs, the base transit time is much lower and the gain-bandwidth product is larger. By forming a heterojunction, the emitter efficiency can be improved and the base resistance lowered. All of these factors contribute to increasing the maximum frequency of oscillation and make the GaAs HBT viable for MMIC applications at Ka-band. The only limitation of the GaAs HBT is

the poor thermal conductivity of the substrate. As will be shown in the following section, this is a significant issue involved in GaAs HBT operation and modelling. For further details on GaAs HBTs and their use in microwave integrated circuits, the reader is referred to the review papers by B. Bayraktaroglu [29] and B. Oyama [30].

4.3.2 Large-signal HBT device model

Nortel Technologies, the research division of Northern Telecom, has a mature GaAs HBT fabrication process with a gain-bandwidth product averaging about 60 GHz. The heterojunction, shown in Figure 4.8, consists of an n-doped GaInP emitter and a p-type GaAs base. GaInP is preferred over AlGaAs because it forms a better contact to the thin base layer, and hence provides higher yields and better reliability [32].

Figure 4.8 Nortel GaInP-GaAs HBT cross-section.

The Nortel HBT is a fairly complicated device and requires a sophisticated model to accurately represent its electrical behaviour. The Gummel-Poon (GP) model, which is designed for Si-bipolar transistors, provides the core of the Nortel large-signal model [33,34]. This core is depicted within the dashed box of Figure 4.9. The diodes D_{bf} , D_{br} , D_e , and D_c represent the electron-hole injections and space-charge recombination mechanisms of the two p-n junctions in the device. They simulate the observed base current. The intrinsic device resistances R_{bi} and R_{ci} model the small losses incurred as the

Figure 4.9 Standard BJT large-signal model (enclosing GP model).

base and collector currents spread out over the contact metallization area and enter the transistor proper. The capacitances C_{be} and C_{bc} are used to account for both excess charge storage during forward bias and the charge separation created by the p-n junction depletion regions. The current source I_{ce} is an amplified version of the base current.

Outside the dashed box of Figure 4.9 there are several additional elements. The leakage resistances R_{be} , R_{ce} and R_{bc} are very large and serve only to assist the simulator by providing a DC path when the collector-base and emitter-base junctions are reverse biased. The diode D_{bcx} and the capacitance C_{bcx} represent the extrinsic base-collector junction.

The GP model is called non-linear because, R_{bi} and R_{ci} excepted, most of the above parameters vary with the input signal level. Most of the parameters also contain specific temperature dependencies. For the case of the Si BJT, the operating temperature can be assumed constant. The greater thermal conductivity of the Si substrates permits the excess heat to be transported away from the emitter junction. This is not so with the GaAs HBT. As the device is operated at higher power, the junction temperature will increase.

Figure 4.10 Modified GP model for GaAs HBT.

To accommodate this phenomenon, a thermal subcircuit must be included as part of the model (see lower right corner of Figure 4.10). The circuit estimates the emitter-base junction temperature rise $T - T_0$ by multiplying the collector current I_c by the collector voltage V_{ce} . The thermal resistance R_{th} characterizes the equivalent thermal resistance of the GaAs substrate. When it is combined with the thermal capacitance C_{th} the resulting time constant controls the rate at which the temperature of the junction varies.

There are two significant thermal phenomena that are inherent to the GaAs HBT and not the Si BJT. The first can be isolated by applying a constant potential V_{be} to the device. As the temperature increases, the contact potential between the base and the emitter will appear to become smaller. This is because more electrons have the required thermal energy to overcome the emitter-base potential barrier. As a result, the number of electrons that are injected into the base will increase. The phenomena manifests itself as a rise in the collector current, and accordingly, an additional increase in the emitter junction

temperature. This effect will become ever more acute at higher applied bias and can lead to thermal run-away as is evident from Figure 4.11(a). Ultimately, the device will overheat and fail. The reduction in the base-emitter contact potential is modelled in Figure 4.10 by the temperature-controlled voltage source, V_{jt} .

Figure 4.11 DC I-V curves (a) vbe voltage source (b) current source at the base.

To isolate the second phenomena, the HBT must be biased with a constant DC current I_b . In this configuration, the collector current I_c is observed to roll-off (see Figure 4.11(b)) with increasing V_{ce} . This occurs because the rise in V_{ce} leads to an increase in junction temperature, the consequence of which is a larger base-emitter hole injection current I_{ep} . According to equations (4.23) and (4.24), growth of I_{ep} results in a reduction to both the emitter efficiency γ and the common emitter current gain β . Since I_b is held constant and $I_c \approx \beta I_b$, the collector current will also decrease in value. This process is modelled by a temperature controlled voltage source V_{pe} and a diode D_{ipe} .

The existence of the base-emitter heterojunction in the HBT warrants one additional modification to the standard GP model. Although deliberately ignored in Figure 4.7, there exists a tall, thin conduction band spike which hinders electrons as they cross from the emitter to the base. By carefully grading the dopants near the junction, it is possible to reduce it, but it cannot be completely eliminated. As a result, the electrons flowing from the emitter to the base must tunnel their way through. From elementary physics, the probability of electrons tunnelling through a barrier depends on the relative potential dif-

ference between the two sides and the height of the barrier itself. This effect is modelled in Figure 4.10 by a non-linear voltage source V_{eb} which varies according to the base-emitter terminal voltages. The relationship is empirical.

The modified Gummel-Poon model is designated as being both non-linear and large-signal. It will simulate all AM/AM, AM/PM and memory effects. Furthermore, it is valid when the applied RF voltages and/or currents are a significant fraction of the applied bias signals. To simulate non-linear circuits under large-signal conditions requires the use of special techniques such as: harmonic balance, time-domain analysis, or Volterra series. For this design, a combination of simulation techniques was used. The preliminary design of the gain blocks and matching networks was accomplished using the small-signal linear test bench of HP-EEsof Libra™. The linear test bench handles the large-signal model by measuring and storing its S parameters for -50 dBm input power. However, the final simulation results presented in section 4.6.1 make use of the HP-EEsof large-signal test bench which is based on the harmonic balance technique. This generates a more realistic prediction of the VGA performance and is a better basis for comparison with measured results.

4.4 Special design considerations

The variable gain amplifier has two mutually dependent criterion that require particular attention. It must exhibit the maximum possible variation in gain while simultaneously minimizing the degradation in the input and output match that occurs from having multiple bias conditions. The purpose of this section is to present the method used to make trade-offs between these two main criteria. In short, the quality of the final design depends on applying the correct amounts of DC and RF feedback. The techniques employed for this design make no claims regarding their efficiency and capability. In fact, the term “technique” is being used rather loosely here. More precisely, this section relates the salient details of how the VGA gain blocks came to be configured the way that they

are. The individual steps will be highlighted as best as possible. However, some cannot be fully justified because they are the result of adaptation and fine adjustment.

4.4.1 Bias network design

The approach to biasing the HBT in the variable gain application is very important and it will ultimately decide the range of gain achievable. The simplest, and most common method of injecting bias into any device is via a high-impedance quarter-wavelength short-circuited stub. According to basic transmission line theory, the impedance seen looking into the bias feed will be an open circuit. To prevent DC from leaking to ground, the feed-line is terminated by a large shunt capacitance.

The one drawback to this technique is its use of quarter wavelength lines. In addition to occupying large amounts of critical chip area, they are inherently narrow-band. This means that the RF will not see a perfect open at the entrance to the bias network across a very wide band of frequencies. An alternative approach is to replace the feed line with a choke inductor. This will present a large impedance to ground for the RF, but will still pass DC without hindrance. The Nortel HBT process readily accommodates the use of large spiral inductors, which unlike transmission lines, have limited distributed effects. For this reason, they are designated as lumped elements. Being less frequency sensitive, the bias network will not significantly affect the frequency response of the circuit. This approach is well suited to applications at Ka-band because large inductors are still relatively small in terms of the amount of chip real-estate they use up. The RF choke used for the variable gain amplifier is depicted in Figure 4.20.

Having established the bias insertion technique, it follows that the range of bias conditions selected must maximize the variation in the gain. Figure 4.12 shows a set of curves depicting the maximum gain of the HBT at 28 GHz as a function of the voltages V_{be} and V_{ce} . Clearly, the main variation in the maximum gain of the device occurs as a result of changes in the base-emitter bias. Hence, any variable bias scheme should include

Figure 4.12 HBT Maximum available gain plotted versus V_{be} and V_{ce} .

a mechanism to sweep the base bias. It should be pointed out that the range shown in Figure 4.12 is just below the base-emitter turn-on voltage of 1.4 V [33]. At greater values of V_{be} , the maximum gain is larger, but the variation is less dramatic. Furthermore, obtaining a respectable match across the entire bias range becomes impossible.

Examining Figure 4.12, a logical bias range is between the limits $1.28 \text{ V} < V_{be} < 1.32 \text{ V}$ and $0.5 \text{ V} < V_{ce} < 2.5 \text{ V}$. As it turns out, this offers between 0 and 6 dB gain. The simplest way to take advantage of this bias range is to use a voltage divider circuit as shown in Figure 4.13. If the DC resistances R_{be} and R_{ce} seen looking into the base and collector are assumed very large (which they are), then the base and collector terminals can be considered open circuits. By keeping V_b fixed at 1.3 V, a network of resistors can be chosen such that by sweeping V_c from 0.5 V to 2.5 V, V_{be} will increase from 1.28 V to 1.32 V. By applying the superposition principle, the following relationships can be readily established

$$V_{be} = \frac{R_b}{(R_{bc} + R_c) + R_b} V_c + \frac{(R_{bc} + R_c)}{(R_{bc} + R_c) + R_b} V_b \quad (4.29)$$

Figure 4.13 DC feedback bias loop.

$$V_{ce} = \frac{(R_{bc} + R_b)}{(R_{bc} + R_b) + R_c} V_c + \frac{R_c}{(R_{bc} + R_b) + R_c} V_b \approx V_c \quad (4.30)$$

The approximation in equation (4.30) is made because the value of R_{bc} must be very large to permit V_c to vary by volts and V_{be} to vary by tens of millivolts. Upon examining equation (4.29), it would appear that it has the form of a straight line with slope given by $R_b/(R_{bc} + R_c + R_b)$ and intercept given by $V_b(R_{bc} + R_c)/(R_{bc} + R_c + R_b)$. If the bounds of the desired bias range in Figure 4.12 are taken as two points on this line, the slope becomes 0.020 V/V and the intercept 1.27 V. With four variables in equation (4.29) and two fixed values, some latitude is possible in selecting specific parameters. The ones used here are $R_b = R_c = 50 \, \Omega$, $R_{bc} = 2400 \, \Omega$, and $V_b = 1.3 \, \text{V}$. The inclusion of R_c might seem somewhat unnecessary. From a bias perspective it is. However, its presence is useful as it can further isolate the collector power supply from the RF signal.

If this network is connected to the transistor, the bias conditions can be summarized as a locus of points (see Figure 4.14). This locus is shown superimposed on the DC-IV curves to give an impression of the general bias vicinity. Using this biasing arrangement, the maximum available gain extends from 0 dB to 6 dB (as expected). Of

Figure 4.14 DC-IV curves with DC feedback bias points superimposed.

course, the maximum will not be attainable because it is impossible to present simultaneous conjugate matches to two amplifier stages across the entire bias range. To ensure the best overall input and output matches, it is necessary to design for midrange bias levels, thus sacrificing the gain at the upper end. In the following section, an RF feedback circuit is proposed to desensitize the input and output impedance to bias variations. Unfortunately, this will also reduce the maximum obtainable gain.

4.4.2 RF feedback design

Amplifiers employing negative feedback are recognized primarily for ease of matching and capacity to span wide bandwidths. The technique can be viewed as placing a variable element, the HBT, in parallel with a fixed element, the passive feedback network. Intuitively, the overall variation in the terminal characteristics should be reduced.

The gain blocks for the variable gain amplifier use a capacitively coupled, common-emitter configuration with parallel feedback. The selection of the feedback elements is important if the benefits outlined above are to be realized. Owing to constraints in

topology, most MMIC feedback amplifiers are limited to a series resistor and inductor. The resistor R_{fb} reduces the overall gain of the transistor block and the inductor L_{fb} provides some frequency dependence. At low frequencies, the inductor is a short, while at high frequencies its reactance will limit the gain reduction caused by the resistor. This should, in effect, increase the bandwidth of the gain block. The capacitor C_{fb} is large and serves only to prevent DC shorting of the collector to the emitter. The feed inductances L_{feed} are parasitic elements which represent the narrow feeds required to connect the HBT to the feedback network. The configuration is shown in Figure 4.15

Choosing values for R_{fb} and L_{fb} is the subject of numerous analyses and design techniques. They all calculate the terminal characteristics of the feedback gain block by summing device and feedback admittance matrices and converting the result to S parameters. Depending on the transistor and the practical limits imposed on the values of the lumped feedback elements, the S parameters can then be tailored to achieve perfect matching and unilateral transmission response ($S_{12} = 0$). The process of connecting S

Figure 4.15 The VGA gain block with RF feedback.

parameter response to feedback elements is aided by a graphical technique of which several discussed in references [35]-[39].

The trouble with these methods is that they use purely ideal lumped elements without any consideration of parasitic and distributed effects. This is fine for low frequency, but at Ka-band they cannot be discounted. Figure 4.15 shows a schematic of the first amplifier gain block as it appears to a linear circuit simulator such as HP-EEsof Libra™. It is comprised of seven segments: the HBT, the input feed, the output feed, the two choke inductors, the RF feedback network and the DC feedback network. Everything contained within the solid box, except for the HBT, has been electromagnetically simulated and includes all possible parasitic and distributed effects. The RF feedback network is located outside the box, and at this point consists of ideal lumped elements.

By varying the values of the feedback elements, it is now possible to observe the resulting variations in G_{MAX} , Γ_{SM} and Γ_{LM} using the linear circuit simulator. Although the numbers do not reflect the presence of the parasitics and distributed character within the feedback network, they are close enough to provide a basis for the selection of real elements. Figure 4.16 and Figure 4.17 show gain block parameters plotted as a function of the feedback elements for the first gain stage only. The resistor and the inductor are permitted to vary over the ranges of 40 Ω to 100 Ω and 0.2 nH to 1.2 nH respectively and the blocking capacitor is fixed at 1 pF. Furthermore, the applied bias is fixed at $V_c = 2$ V and the frequency at $f = 28$ GHz.

One of the greatest difficulties of designing a variable gain amplifier is the existence of multiple bias states. Since the bias varies, an optimal set of parameters at one bias may not be optimal at another bias state. The approach espoused in this thesis is to design the amplifier as if it were to be used exclusively for a nominal bias condition and to accept the degradation that occurs as it is varied. The collector bias $V_c = 2$ V represents the best compromise between high and low gain and ensures that the degradation is within acceptable limits.

Figure 4.16 G_{max} versus R_{fb} and L_{fb} for the first gain block.

Figure 4.17 Γ_{sm} and Γ_{lm} versus R_{fb} and L_{fb} .

Based on Figure 4.16 and Figure 4.17, the optimal feedback parameters are: $R_{fb} = 60 \, \Omega$ and $L_{fb} = 0.8 \, \text{nH}$. These would appear to yield less than optimal results: $\Gamma_{SM} = 0.16 \angle -67^\circ$, $\Gamma_{LM} = 0.73 \angle 10^\circ$, and $G_{MAX} = 3.5 \, \text{dB}$. This is because the feedback elements are selected with a view to including the shift that will occur when the ideal elements are implemented in microstrip. For example, when a $0.8 \, \text{nH}$ ideal inductor is replaced by an equivalent spiral inductor, there will be some associated parasitic shunt capacitance. For this particular inductor, the added parasitics will shift the reflection coefficients in such a way that the requirements for the input and output matches are eased. The electromagnetic modelling of the feedback elements is discussed at greater length in section 4.4.3.

The design of the second stage proceeds according to the same principles as the first. At $28 \, \text{GHz}$, the optimal feedback parameters turn out to be the same: $R_{fb} = 60 \, \Omega$ and $L_{fb} = 0.8 \, \text{nH}$. Unfortunately, these parameters also yield a gain block that is conditionally stable between 16 and $19 \, \text{GHz}$. After completing a preliminary design, it was discovered that the amplifier would be unstable at these frequencies. In order to match the input of the second gain stage at $28 \, \text{GHz}$, the amplifier output reflection coefficient becomes greater than one at $17 \, \text{GHz}$. There were three possible solutions to this dilemma: change the matching network, change the topology of the second gain stage, or stabilize the gain stage by adding additional feedback. The matching network could not be changed owing to space constraints. The gain block topology could not be changed without straining the output match requirements. However, It was possible to stabilize the second stage using a shunt feedback resistor. By trial and error, the gain stage was made unconditionally stable at $17 \, \text{GHz}$ and $V_c = 2.0 \, \text{V}$ by connecting a $300 \, \Omega$ resistor between the collector terminal and ground. This moved the source stability circle of the second stage just outside of the Smith chart for all bias states below $V_c = 2.5 \, \text{V}$. The stability problem is discussed further in section 4.6.2. The two final gain blocks, including all

feedback elements are shown in Figure 4.18. The 2 pF DC-blocking capacitor in the shunt feedback segment has no effect at either 17 or 28 GHz.

Figure 4.18 RF feedback elements for (a) first and (b) second gain block.

At 28 GHz and $V_c = 2$ V. The conjugate matching coefficients and maximum gain for these the two amplifier stages are as follows

Table 4.1 Conjugate matching coefficient and maximum gain for VGA gain blocks.

	Gain Block 1	Gain Block 2
Γ_{SM}	$0.27 \angle 146^\circ$	$0.45 \angle 129^\circ$
Γ_{LM}	$0.68 \angle 41^\circ$	$0.50 \angle 55^\circ$
G_{MAX}	3.26 dB	0.68 dB

Two points should be conceded here. Firstly, the inclusion of shunt feedback has virtually eliminated any gain from the second stage. The only rationale for retaining the stage is to ease the output matching requirements. Secondly, these numbers are definitely different from those selected directly from Figure 4.16 and Figure 4.17, even considering the comments made regarding the anticipated parasitics. The reason for this discrepancy is that the feedback networks have now been replaced by a single electromagnetically simulated block. Details regarding this simulation will be discussed at the end of the following section.

The methods applied in this section are by no means intended to be a general basis for variable gain amplifier design. Unfortunately, the need for accurate electro-

magnetic modelling of circuit elements is just not compatible with elegant optimization techniques. As a result, it is impossible to provide conclusive evidence that the final results shown in Table 4.1 are indeed optimal. Nonetheless, the remainder of chapter 4 will illustrate that they are sufficient to produce a respectable input/output match across a fairly broad band of frequencies and bias states.

4.4.3 Electromagnetic modelling of circuit elements

The Nortel HBT fabrication process contains a complete library of circuit elements including transmission lines, nichrome and implant resistors, metal-insulator-metal (MIM) capacitors, and spiral inductors. All of these structures can be formed using one or combinations of three metallization layers which are separated by layers of dielectric, either SiON or benzocyclobutene (BCB). The upper two metal layers, me2 and me1, can also be used to realize the whole range of transmission lines and related structures: Lange couplers, stubs, DC feed lines, Wilkinson power combiners, and quadrature hybrids. Interconnections between the various layers are made using $2\text{ }\mu\text{m} \times 2\text{ }\mu\text{m}$ metalized holes: posts for me1-me2 and vias for me1-shotcap.

Figure 4.19 shows planar and cross-sectional views of all of the lumped-circuit elements that are used in the implementation of the variable gain amplifier. By definition, the size of a lumped element is significantly smaller than the operating wavelength, and therefore it exhibits only minor phase shift with frequency. Thus, they have the advantage of smaller size and wideband frequency response. The biggest obstacle to designing Ka-band circuits containing lumped elements is modelling. At high frequency, the parasitic behaviour of lumped elements can have a significant affect on overall circuit

Figure 4.19 Nortel HBT interconnect layers and lumped elements.

performance, especially if they are used in sensitive matching networks. The term parasitic refers to small and often unforeseen effects that arise in circuits due to the presence of ground planes, proximity to other elements, and fringing fields. However, when properly modelled, these effects can be included in the desired response.

Sonnet™ Em version 4.0 (Em) is one of several possible electromagnetic analysis tools that is capable of calculating scattering parameters for 3-D planar geometries. It has been used extensively in the design of the variable gain amplifier to model every bend, junction, capacitor, inductor, and resistor in the RF path. Furthermore, all possible couplings between closely spaced elements have either been accommodated or discounted as insignificant. Since the electromagnetic simulator deals with geometry rather than specific circuit elements, it will predict parasitic behaviour.

The input to Em is a drawing file which defines the dielectric layers, metal patterns, ports, reference planes and enclosing rectangular metallic box. The analysis proceeds by subdividing the metallization patterns into small rectangular subsections and assigning an arbitrary current to each one. Next, the simulator evaluates the tangential

electric fields due to the individual currents and adjusts their magnitudes such that the weighted residual of the total tangential electric field goes to zero on all metallization. This process is repeated until there is no appreciable change from one iteration to the next. Once the corrected current distribution is ascertained, the N -port S parameters are easily determined.

Each portion of the variable gain amplifier takes three steps to design. The first step involves choosing network topologies and ideal element values using a Smith chart. With the ideal values as a guide, the next step is to substitute equivalent Libra microstrip models and to tune their parameters until the desired response is achieved. Unfortunately, most of the models are not guaranteed to be accurate at 28 GHz. At these frequencies, the required dimensions often violate the recommended parameter limits. Therefore, once a segment has been tuned to the desired response in Libra, the topology is redrawn exactly for Em and analysed. The resulting S -parameters typically differ from the Libra results by no more than 10% in magnitude and 5% in phase. By adding or subtracting short sections of high-impedance microstrip line at the input or output of the element, the phase error can usually be matched exactly over a 1 to 2 GHz bandwidth. Depending on the case, the magnitude can be corrected to within 5% of the Libra results.

The remainder of this subsection reviews several elements particular to the Nortel HBT process and presents comparisons between Libra models and Em simulated results. For accurate comparison, no attempt has been made to “tune” the responses. The first element is a 1.2 nH spiral inductor which is used as the RF choke in the variable gain amplifier. The drawing file used by Em for the analysis of this inductor is shown in Figure 4.20. The rectangular frame around the inductor corresponds to the metallic “box” which is used to define the simulation boundary conditions. If the box is small relative to a wavelength, yet large enough to keep all circuit elements well away from the boundaries, its effects on the results will be nil. To assist with the latter requirement, Em permits de-embedding of the transmission line near the box edges. Hence, the inductor in Figure 4.20

Figure 4.20 1.2 nH choke inductor.

contains long leads connecting it to ports 1 and 2. The long arrow on the left points to the location of the port 1 reference plane. Although not visible, the situation is similar for port 2. The simulated results now include only the me2 spiral and the me1 exit feed.

When swept from 10 to 36 GHz, the resonance frequency of this inductor (i.e., when $\text{ang}(S_{11}) = 0^\circ$) is predicted to be 30 GHz. By comparison, the equivalent lumped Libra MRIND model anticipates a resonance frequency of 32 GHz (as shown in Figure 4.21). Considering the simplicity of the Libra model, it is rather surprising to observe that these two simulators have generated remarkably similar results. The MRIND model is well-known and employs empirical expressions to transform physical dimensions into the values for the lumped element shown in Figure 4.22. As mentioned in section 4.4.2, a spiral inductor with a nominal inductance value of L also has the parasitics C_2 and C_3 . Thus, any set of Γ_{SM} and Γ_{LM} predicted in Figure 4.17 will be shifted when the ideal inductor is replaced by a spiral inductor. This does not imply that use of Figure 4.17 is unhelpful, it simply means that promising results must be checked using an equivalent

Figure 4.21 Comparison of spiral inductor models.

Figure 4.22 Lumped element model for a rectangular spiral inductor.

spiral inductor before they can be considered optimal. As it turns out, the parasitic shunt capacitances associated with a spiral inductor of nominal inductance 0.8 nH are helpful.

The Nortel HBT process is capable of fabricating two varieties of capacitor. Large capacitors, such as those used for DC blocks and RF grounds, consist of a 0.3 μm thick layer of SiON dielectric sandwiched between the metal layers me1 and shotcap. This arrangement provides approximately 142 pF/mm² which can easily reach 1 or 2 pF without using up too much chip real-estate. These capacitors are very accurately modelled owing to the fact that they are virtually invisible for RF signals with frequency 10 GHz or greater. The second variety is for small capacitances, which are rated at

Figure 4.23 50 fF me1-me2 capacitor.

22 pF/mm². These capacitors are formed using the 1.5 μm benzocyclobutene layer and the two interconnect metals: me1 and me2. The example in Figure 4.23 shows one such capacitor which has a nominal capacitance of 50 fF. Three separate sets of predictions are shown for the capacitor in Figure 4.24. The Em results are developed directly from Figure

Figure 4.24 Comparison of me1-me2 capacitor models.

4.23 and are compared against two Libra models: microstrip thin film capacitor (MTFC) and thin film capacitor (TFC). Clearly, all three results diverge significantly at higher frequency with the phase of S_{11} varying by as much as 15%. Unlike the inductor model, MTFC and TFC do not appear to match Em very well beyond 20 GHz. At this point, a decision must be made. Which results should be trusted? Well, definitely not the TFC model. Its circuit representation of a capacitor and resistor in series does not account for the parasitic shunt capacitance between the lower plate and the GaAs substrate. The MTFC model is certainly more realistic. Unlike the TFC model, it does include the effects of the substrate and the ground plane. However, it makes no accommodation for possible fringing effects at the edges of the plates, nor does it consider the discontinuities created by the leads at the input and the output. Since the Em input files represent the specifications of the final structures and accommodate all fringing, it has been decided that the Em results take precedence over the Libra models.

By far the biggest asset of the full-wave simulator is its ability to analyse coupling between multiple elements within a circuit. It offers the greatest insurance that a circuit will meet its design requirements. The simulated variable gain amplifier is composed entirely of electromagnetically analysed segments. All of the break-points between these segments are made in the transmission line portions or “leads” where the modes are almost TEM. Cascading several such segments together produces results that are virtually identical to a full simulation on the grouped structure. One of the critical segments, the feedback network, is shown in Figure 4.25 and includes the nominally 0.8 nH inductor, 1 pF capacitor, and 60 Ω resistor.

By using the Em simulator, all possible couplings between these closely spaced elements, their respective feeds, and the general topology have been allowed for in the final simulation. The final gain blocks consist of five individually Em simulated segments: the feedback network as shown in Figure 4.25, the input/output leads to the HBT, and the RF choke inductors. When fully assembled, the linear test bench in Libra is used

Figure 4.25 Complete feedback network for Em simulation.

to cascade all the parts, analyse the HBT model, and calculate the final conjugate matching coefficients. The results presented in Table 4.1 were generated using this approach.

4.5 Matching networks

Having established the required conjugate matching coefficients for the two gain stages and the network design methods, the final designs for the matching networks will now be reviewed. Note that all reflection coefficients in this section are for $V_c = 2.0$ V and $f = 28$ GHz. Figure 4.26 shows the complete amplifier with its matching networks. Naturally, the values of the elements in these networks are nominal only and do not include any parasitic quantities.

Figure 4.26 Variable gain amplifier schematic showing matching networks.

4.5.1 Input matching network

Of the three matching networks, the demands placed on the input section are the least stringent. In fact, with the conjugate match requirement of $0.27 \angle 146^\circ$, it can be easily achieved with only a few small elements as it is already *relatively* close to 50Ω . Figure 4.27 depicts the input matching network and marks off three separate measurement

Figure 4.27 Input matching network.

Figure 4.28 Input matching on the Smith chart.

planes. Starting from the right, Γ_{IN1} is the reflection coefficient seen looking into the input of the first stage. For simplicity, it is assumed that $\Gamma_{IN1} \approx \Gamma_{SM1}^*$. Strictly speaking, this is untrue because it is only possible if the output of the first stage is presented with its conjugately matched load, Γ_{LM1} . As shown in section 4.5.3, this requirement is not achieved. However, the result for the interstage match is sufficiently close that the assumption does not pose any serious problems.

The shift to $B - B'$ reflects the addition of a small 0.2 nH series inductance, provided by a narrow, folded transmission line and a small 55 fF shunt capacitance. The small shunt capacitance is difficult to realize because it can be easily dominated by the length of line required to connect the capacitor output to the ground pad. To avoid this problem, the capacitor is connected to the through-wafer via using a very low-impedance pad. The shift to $B - B'$ is plotted on Figure 4.28.

It is important to note that the line joining Γ_{IN1} and Γ_B is meaningless and is present for illustration purposes only. If the matching network was composed of purely

ideal elements, then the shift to $B - B'$ would be represented on the Smith chart by an upwards movement along a contour of constant resistance for the series inductance, followed by a downwards shift along a contour of constant conductance for the shunt capacitor. The movements would be depicted as lines representing the increasing inductive reactance and capacitive susceptance. However, for the structures shown in Figure 4.27, the presence of parasitics and mutual coupling blurs the exact shape of these movements. The joining lines have been “sketched in” simply to guide the reader from point to point.

Looking to the right at $B - B'$, the input network is about as close to matched as possible: $\Gamma_B = 0.06 \angle 170^\circ$ or -25 dB. The purpose of the shift to $A - A'$ is to provide the capacitive coupling to the circuit and is not intended to be part of the matching network. Since the amplifier does not operate at DC, the 1 pF capacitor at the input prevents any such leakage. The capacitor is sufficiently large that it has very little effect at 28 GHz; Γ_A is $0.03 \angle -156^\circ$.

4.5.2 Output matching network

The purpose of the output matching network is to transform $50 \, \Omega$, or $\Gamma_{50} = 0$ to $\Gamma_{LM2} = 0.5 \angle 53^\circ$. This task is accomplished mainly by a shunt inductor with a nominal inductance of 0.6 nH. When the network is realized, two narrow lengths of high-impedance transmission line and a microstrip taper must also be included to feed to

Figure 4.29 Output matching network.

the inductor correctly (see Figure 4.29). The first segment, enclosed between the output port and the $C - C'$ reference plane, introduces only a small change in the inductive reactance. As seen from Figure 4.30, $\Gamma_C = 0.09 \angle 82^\circ$ is still fairly close to the origin. The move from $C - C'$ to $B - B'$ includes the effects of two small series inductances, a large shunt inductance, a through-wafer via at port 3, and some parasitics associated with the slight coupling near the microstrip T junction. Clearly, these additional elements have an effect, because Γ_B and Γ_C are on wildly different conductance contours: $j\overline{B}_C = 0.96$ versus $j\overline{B}_B = 0.57$. Were there only an ideal 0.6 nH shunt inductor, the conductances would be equal.

Figure 4.30 Output matching on the Smith chart.

Once the $B - B'$ reference plane is reached, the output is still slightly mismatched; but this is in anticipation of the series capacitor and its feed lines. Like the one at the input port, this capacitor serves only as a DC block. Nonetheless, owing to the way in which it is connected, it does shift the response of the matching circuit. This effect is reflected in the shift from $\Gamma_B = 0.51 \angle 87^\circ$ to $\Gamma_A = 0.51 \angle 55^\circ$ on Figure 4.29 and is quite close to the desired value of Γ_{LM2} .

4.5.3 Interstage matching network

The salient points concerning the design of the interstage matching network were covered in section 4.2. When an amplifier is designed for maximum gain, the interstage matching network should not only present the conjugate matched impedances to the first and second stages, but attempt to counteract the inherent 6 dB/octave transistor gain roll-off. For simplicity, it has been assumed that both the output of the second stage and the input of the first stage are exactly matched. If this is not the case, then Γ_{LM1} and Γ_{SM2} are not the reflection coefficients for maximum RF coupling from stage 1 to stage 2. However, the designs shown in sections 4.5.1 and 4.5.2 are close enough to the optimal reflection coefficients that the assumption does not present a problem.

Figure 4.31 Interstage matching network.

It should be conceded that no concerted effort was made to design a matching network with a particular sloped response. Certainly, there are plenty of synthesis techniques which exist in the literature (the most noteworthy being that of Mellor et al. [40]). The trouble is that these techniques require large numbers of lumped elements. Owing to the amount of chip area available, there is not enough space to accommodate them. Moreover, the Em modelling of large numbers of elements demands longer simulation times and more design iterations. Given these limitations, the gain flatness achieved

by the present interstage matching network is simply the result of a trial and error process. The complete interstage matching network is illustrated in Figure 4.31.

The matching network consists of a DC block to isolate the base bias level of stage two from the collector bias level of stage one, a 0.4 nH spiral inductor, and a short length of folded inductive transmission line. Starting from the right, Γ_{IN2} is the conjugate of Γ_{SM2} and is shown plotted in Figure 4.32. At first glance, the shift from Γ_{IN2} to Γ_C appears incorrect for a series capacitor which usually moves the response downwards along a contour of constant resistance. The upwards response seen in Figure 4.32 is actually due to the capacitor's input and output feed lines which behave somewhat like small inductors. The shift from $\Gamma_C = 0.45\angle-153^\circ$ to $\Gamma_B = 0.78\angle-58^\circ$ corresponds to an increase in reactance without significant change in resistance; the normal response for an inductor. The final shift to $\Gamma_A = 0.8\angle-42^\circ$ is due to the folded microstrip line.

There is clearly a sizeable mismatch here because Γ_A and Γ_{LM1} differ substantially. By the trial and error process hinted at previously, a flatter response is

Figure 4.32 Interstage matching on the Smith chart.

obtained by deliberately introducing this mismatch. After having designed a matching network with a much closer match, the full amplifier response was noted. Then, a few sets of modifications were made to the interstage matching network including such things as: plus or minus a half turn of the inductor and varying the width and length of the folded microstrip line. After several variants, the final match was decided upon because it yielded slightly less gain for the overall amplifier response, but improved flatness.

4.6 Simulated and measured results for the variable gain amplifier

The criteria for a successfully designed variable gain amplifier is good matching, stable operation, and flat transmission response at all frequencies and bias points. Furthermore, the variation in gain with bias should be maximized. The final simulated results are based entirely on Em scattering parameter results, the Nortel non-linear HBT model, and Libra microstrip models for the DC bias networks. The large-signal test bench in Libra is used to calculate the S parameters of a network consisting of the aforementioned components for a given power level. The circuit schematic is shown in Figure 4.33. To avoid cluttering the diagram, the microstrip models used for the DC bias network have been replaced by ideal lumped elements.

Figure 4.33 Circuit simulator schematic of the variable gain amplifier.

The individual Em simulated portions of the circuit are shown enclosed in boxes. With the exception of the matching networks, each one represents a complete set of S parameters which range from 10 to 40 GHz. In order to realize the shunt feedback section of the second gain block, it is necessary to employ a special processing feature: the capacitor on via. This is simply an me1-shotcap capacitor which is located directly above a via hole. By invoking this design cell, a great deal of chip area is conserved. The capacitor portion of this structure is simulated using the Libra thin film capacitor model (TFC) which gives reasonable results for large values of capacitance. The resistor portion of the shunt feedback segment is included in the Em simulation of the stage two output feed. The only other structure not previously discussed is the via hole model used for making ground connections. The Nortel HBT process library has a simple inductor/resistor network which is used to characterize its behaviour. Since the via is $160\ \mu m \times 160\ \mu m$, it is often shared by several portions of the circuit. As a result, the model is also designed to predict the mutual coupling between separate leads.

4.6.1 Large-signal simulation

Referring to Figure 2.19, variable gain amplifiers may be located in the main RF path of the feedforward linearizer. Since it will be placed just before the power amplifier, it may have to be required to pass signals with amplitudes that exceed the small-signal or linear criteria. Thus, although the amplifier has been designed under the assumptions associated with small-signal analysis, a large-signal simulation provides a more accurate prediction of its capability.

The simulation is carried out using the large-signal test bench in Libra which employs the harmonic balance method to determine power-dependent S parameters. The non-linear circuit is separated into a linear sub-circuit and a non-linear sub-circuit. If there are N connections between the non-linear components and the linear components, then there will be N ports in the non-linear subcircuit. Each one of these ports connects

directly to the linear subcircuit. In addition to the N connections, the linear subcircuit has M ports for sources and loads. For the case of the variable gain amplifier, the HBTs form the non-linear portion, while everything else becomes the linear portion. At first glance, forming this partition is difficult to imagine. Nonetheless it is always possible to rearrange the nodes in such a way that separation is achieved. The simulator will then iteratively match the frequency domain solution of the linear sub-circuit to the time domain solution of the non-linear sub-circuit at each of the N mutual connections. Once the solution has converged, the voltages and currents are conveyed to the source and load points via the linear sub-circuit and converted to large-signal S parameters.

After some experimentation with the final design, it was determined that the simulated 1 dB compression point occurred at an input power of -5 dBm. To ensure “linear” operation, the input power was fixed at -7 dBm, slightly below this point. This will ensure that the harmonics of the fundamental carrier frequency are small and can be ignored. This power level is used both in this section and section 4.6.4 to enable the best possible comparison between simulated and measured results. It should also be noted that the collectors of both stages are biased at exactly the same level for all measurements meaning that $V_c = V_{c1} = V_{c2}$.

Figure 4.34 shows the input reflection coefficient (S_{11}) or input return loss as a function of the bias applied to the collectors of stages 1 and stage 2. Close inspection of the graph will reveal that the match is best at the centre frequency of 28 GHz for the case of $V_c = 1.0$ V. The worst case return loss is -14 dB from 24 to 32 GHz and corresponds to the case of $V_c = 2.7$ V.

The transmission coefficient (S_{21}) is depicted as a function of applied collector bias in Figure 4.35. As can be seen from the figure, the gain at 28 GHz extends from a minimum of -10 dB to a maximum of 8 dB indicating an 18 dB variation in gain. Compared to small-signal simulations, this is a range increase of approximately 5 dB. The gain

Figure 4.34 Simulated VGA input reflection coefficient.

Figure 4.35 Simulated VGA transmission coefficient.

Figure 4.36 Simulated VGA output reflection coefficient.

flatness is, at worst, -1 dB/GHz at 28 GHz. This sloped transmission characteristic will be discussed at greater length in section 6.4.

Compared to S_{11} , the output reflection coefficient in Figure 4.36 shows very little variation with bias. Hence, the curves for S_{22} almost completely overlap. At worst case, the output return loss is better than -11 dB from 26 to 30 GHz. The stability problem is also apparent in the figure. Notice that for $V_c \geq 1.9$ V, S_{22} is greater than unity between 17 and 20 GHz implying that the amplifier will oscillate. This condition is cause for concern because it means that, in simulation at least, the maximum gain will be restricted to values below 6 dB. In terms of the applied bias, V_c can be increased no more than 1.9 V.

4.6.2 Amplifier stability

As illustrated in section 4.6.1, the root of the stability problem is a band of frequencies centred around 18.5 GHz. In order to discuss the issue further, it is necessary

Figure 4.37 Source and load stability circles at 17 GHz and $V_c = 2.7$ V.

to revert to the small-signal behaviour of the amplifier. This will cause the oscillatory condition to shift from 18.5 GHz to 17 GHz and will alter the collector bias setting at which the oscillations begin from 1.9 V to 2.5 V. However, from a qualitative point of view, the problem remains the same. There is a band of frequencies below the design bandwidth that are unstable.

Before discussing the stability issue further, some definitions need to be elucidated. In Figure 4.37, there are four reflection coefficients, namely Γ_{S1} , Γ_{L1} , Γ_{S2} , and Γ_{L2} . With reference to Figure 4.33, Γ_{S1} is the reflection coefficient seen looking into the input matching network terminated by a 50 Ω source impedance. Γ_{L1} is the reflection coefficient seen looking into the interstage matching network terminated by the second stage, output matching network and 50 Ω load impedance. Naturally, this implies that Γ_{L1} will vary with applied bias. The same logic applies to Γ_{S2} and Γ_{L2} .

Referring to Figure 4.37, the stability problem is due to the fact that the interstage matching network presents a source reflection coefficient, Γ_{S2} , which is in the unstable region of the stage 2 source stability circle. Although this is shown for $V_c = 2.7$ V and $f = 17$ GHz, the same is true for all values of $V_c > 2.5$ V and for all frequencies between 16 and 19 GHz. This explains why all of the stabilization efforts have

been focused on the second stage, rather than the first. Prior to the addition of the $300\ \Omega$ resistor, the source stability circle occupied almost half of the Smith chart for $V_c > 1.4\text{ V}$ at the same range of frequencies. Thus, the amplifier was almost always unstable.

Examining the stage 1 stability circles and source/load reflection coefficients, it is clear that there is little potential for instability here. Since stage 1 and stage 2 have identical feedback circuits, and are biased at exactly the same level, there remains only one explanation for the difference in stability: topology. This difference was first announced in section 4.4.2, and it was briefly stated that the particular gain block configuration for stage 2 was necessary in order to avoid straining the requirements on the output matching network. Unfortunately, it is not possible to provide a particularly insightful explanation for this statement because there were a number of competing factors that lead to the final topology. It suffices to say that in the process of numerous design iterations, decisions were made regarding the matching circuits and the gain block configuration. If the stage 2 gain block had been identical to the first, the output and interstage matching networks would have become too big to fit in the chip area available.

4.6.3 MMIC layout

An MMIC layout is an exact representation of a circuit's features. That of the variable gain amplifier is shown in Figure 4.38. The chip is $1.25\text{ mm} \times 2.5\text{ mm}$. Naturally, the $6.5 \times 3\ \mu\text{m}$ HBTs appear as small dots at the ends of the input and output feeds. For close agreement with simulated results, all critical passive elements must be carefully reproduced in the layout. For the variable gain amplifier, this means that the matching networks and the feedback circuits should appear exactly as they do to the full-wave simulator, Em. There are some limits to this requirement. For example, the layout of the amplifier shows inductors which are mitered at the corners whereas the simulated ones are not. In practice, such features are far too small to have an appreciable affect on the Em simulation. They exist to prevent current crowding at corners which can damage the metalliza-

Figure 4.38 MMIC layout of the variable gain amplifier.

tion over the circuit's lifetime. The mitering does little to affect the RF behaviour of the inductors themselves.

There are other small features which, fortunately, do not need to be modelled carefully. Throughout the Em simulations, all connections between me1, me2, and shotcap were made using vertical sheet metallization. On the layout, these connections are made using small ($2\ \mu\text{m} \times 2\ \mu\text{m}$) metallized holes in the dielectrics. At Ka-band, such dimensions are easily considered short circuits. Thus, the vertical sheet metallization is an adequate approximation to the actual structure.

The layout of the DC bias networks is guided by the space available and not simulation results. The large choke inductors and shunt capacitors prevent any in-band signals from entering the power supplies. At lower frequencies, where the choke inductors no longer provide isolation, the wavelength becomes sufficiently large that the dimensions become unimportant. Nonetheless, it is still important to include the elements of the DC bias network in all simulations. There is always a possibility of some unexpected resonance between elements. Without realizing it, a designer can create a short-circuit somewhere or affect the active device in some way that it begins to oscillate.

For the variable gain amplifier, the first stage is biased at the lower left while the second stage is biased at the upper right. The bias signals are applied to four of the $100\text{ }\mu\text{m} \times 100\text{ }\mu\text{m}$ pads which ring the circuit in Figure 4.38. From each of the pads, the DC passes through $50\text{ }\Omega$ resistors which appear as small squares. The $2400\text{ }\Omega$ resistor, by comparison, is a long thin strip. The RF grounds are formed using 4 large 2 pF capacitors located at either end of the $2400\text{ }\Omega$ resistors. This configuration is ideal for on-wafer prober measurements as it offers plenty of space to bring the DC bias probes in on either side. However, having four DC bias points on such a small chip presents a challenge for the design of the vector modulator. The ramifications of this arrangement will be made clear in section 6.2.

There are nine separate ground connections for the amplifier. These grounds are made by etching via holes through the wafer and are represented on the layout by large squares. On top of each via is a $160\text{ }\mu\text{m} \times 160\text{ }\mu\text{m}$ area of metallization. For eight of the nine ground connections, the metallization consists of a thick pad layer sandwiched between relatively thin me2 and me1. All combined, the structure is strong enough to survive the aggressive etching required to bore through the wafer. The ninth via, adjacent to the HBT of the second stage, appears different from the other eight because it also includes a large DC blocking capacitor. The capacitor is formed between the lower me1 metallization and the shotcap metallization. Like the standard via, it contains a pad and me2 layer. The shotcap layer is considerably thinner than the combined me2-pad-me1 stack; hence the cap-on-via is rather delicate. It is expected that this will impact the circuit yield considerably. Nonetheless, the cap-on-via solves a major topological problem associated with the shunt feedback network.

The RF ports, at the left- and right-hand sides of the circuit, have the standard ground-signal-ground configuration for on-wafer testing using a probing station. This is an accurate testing method for circuits at Ka-band. Using on-wafer calibration standards, it is possible to place the measurement reference planes right at the input port and

output ports. The alternative testing technique is to mount an individual circuit in a testing jig. Since the MMIC is so small, it is not possible to attach a coaxial-to-microstrip launcher directly to the circuit. Usually, the chip is connected via bondwire to a short length of microstrip line on a ceramic substrate. The ceramic microstrip is then attached to the launcher. The best possible calibration for this arrangement cannot de-embed the bond wire. At 30 GHz, this will affect the measured results.

4.6.4 Measured results

The amplifier, as described in section 4.6.3 was returned from Nortel's GaAs foundry in November 1997. This subsection will review the on-wafer measurement results. These measurements consist of input/output return loss and gain as a function of the applied collector bias. Furthermore, in completing the discussion on stability, a spectrum is provided describing the amplifier output port when oscillations begin.

Before presenting the results, some comments are in order. Firstly, the stability problem was less severe than initially predicted. In section 4.6.2, it was determined that the source impedance presented to the second stage was in the unstable region of the associated source stability circle. For the small-signal case, this was true for all collector bias levels above 2.5 V and for all frequencies between 16 and 19 GHz. The large-signal simulation revealed similar behaviour, only the oscillations started up at a collector bias of 1.9 V and the band of affected frequencies shifted up by 1.5 GHz. As it turns out, the measurements reveal a much more stable amplifier than these simulated results anticipated. For an input power of -7 dBm, it is possible to apply a collector bias up to 2.7 V without inducing oscillatory behaviour.

All testing was carried out at CRC using the setup shown in Figure 4.39. The principal instrument is the HP8510C network analyser with its associated test set (HP8516A) and sweep source (HP8340B). The ports of the test set are connected to the amplifier via 36" rigid K-standard coaxial cables and a Cascade Microtek Probing Station

Figure 4.39 Test set-up for on-wafer VGA characterization.

model 1032. This provides measurement capability from 45 MHz to 40 GHz and calibration up to the input and output ports of the amplifier. The probe station itself is outfitted with four probe manipulators, two for RF insertion/measurement and two for bias insertion. The latter two have eight probe tips spaced at 100 μm intervals. This spacing corresponds exactly to the bias pad arrangement on the MMIC. Hence, by using probes three and seven, it is possible to supply V_{b1} , V_{c1} , V_{b2} , and V_{c2} to the amplifier. The probes are in turn connected to the bias supplies via ribbon cables and a feed box.

The measurement results are shown in Figures 4.40, 4.41, and 4.42 respectively. As in the simulation, the bases of both first and section stage are biased at 1.3 V. An effort was made to adjust this level for optimum gain and input/output return loss, but no improvement was observed. As in the simulated results, the collectors of both stages were biased at exactly the same level for all measurements meaning that $V_c = V_{c1} = V_{c2}$.

Figure 4.40 shows the measured input return loss (S_{11}) as a function of V_c . As in the simulated results of section 4.6.1, the match at the centre frequency of 28 GHz is best for the case of $V_c = 1.0\text{ V}$. However, there is a degradation of about 5 dB at

Figure 4.40 Measured VGA input reflection coefficient.

Figure 4.41 Measured VGA transmission coefficient.

this point over the equivalent result of section 4.6.1. At worst case, the return loss is -10 dB from 22 to 30 GHz and, as in the simulations, corresponds to the case of $V_c = 2.7$ V.

At 28 GHz, the maximum gain achievable from the amplifier is 4 dB for $V_c = 2.70$ V. By comparison, the minimum gain is -12 dB for the collector setting of $V_c = 0.37$ V. Clearly, this differs from the simulated results which predicted maximum and minimum gain of 8 dB and -10 dB respectively. The 4 dB reduction in gain is attributed to two factors. Firstly, the simulated results of section 4.6.1 assume lossless metal and dielectric. This measure was necessary in order to reduce the Sonnet Em simulation time. Synthesizing all aspects of the circuit with lossy materials was deemed impractical. Secondly, the input/output/interstage matches do not present exactly the same impedances as in the simulation. This is partly due to the use of perfect materials and the small errors introduced by the HBT model. The gain roll-off predicted by the large signal simulation is roughly commensurate with the measured result. From Figure 4.41, the gain slope in the vicinity of 28 GHz is at worst -1 dB/GHz.

Figure 4.42 Measured VGA output reflection coefficient.

Figure 4.43 VGA spectrum at the output port for $V_c = 2.9$ V.

The measured output reflection coefficient of Figure 4.42 shows only a modest variation with bias somewhat akin to the simulated results. However, the band of frequencies with $S_{22} < -10$ dB is quite narrow and extends only from 27 to 29 GHz. Fortunately the band is centred around 28 GHz as hoped. Compared to Figure 4.36, the stability problem is far less evident. S_{22} is less than unity across all frequencies below $V_c < 2.8$ V. Hence, stability is less of a concern than the design process has indicated.

To induce oscillations in the amplifier, the collector bias on stages 1 and 2 had to be increased beyond 2.7 V. The result of such an increase can be observed by connecting a spectrum analyser (HP8564E) at the output port with the input terminated by $50\ \Omega$. As shown in Figure 4.43, a large number of peaks appear indicating the existence of DC-RF power conversion with zero input signal. This confirms the existence of the instability and demonstrates the importance of careful design verification. It is essential

that all frequencies up to and beyond the design bandwidth be considered potentially oscillatory.

4.7 Summary

In this chapter, the steps involved in the design of a 28 GHz HBT variable gain amplifier circuit have been presented. All portions of the circuit have been simulated using Sonnet™ Em version 4.0 in order to obtain the best possible prediction of its behaviour across a bandwidth extending from 10 to 40 GHz. During the design process, a region of potential instability was discovered in the neighbourhood of 17 GHz. Its impact on the final design was reduced by adding a shunt feedback resistor at the output of the second stage. The final amplifier design was laid out according to the Nortel foundry rules, submitted for fabrication, and tested at CRC. The measured results compared well with those predicted by the simulations. The discrepancies were attributed to assumptions made in the simulations such as the use of perfect conductors and lossless dielectrics. Another possible source of variation between measured and simulated results is the HBT model itself.

In terms of the bandwidth required for LMCS, 27.35 to 29.35 GHz, the amplifier is not quite adequate to cover the full range. The measured output return loss increases quite dramatically towards the band edges and the gain is not sufficiently flat to ensure complete wideband distortion cancellation. However, given the dimensions of the final chip and the simplicity of the matching networks, a bandwidth half as wide as the LMCS specification is quite reasonable for the first design iteration.

CHAPTER 5: Passive component design

5.1 Introduction

As illustrated in Figure 3.1, the vector modulator includes a set of three input couplers and three Wilkinson power combiners. This section will review the basic principles and design techniques used to develop them.

Being electrically large, it is prohibitively expensive and wasteful to realize passive components in GaAs. A cheaper alternative is to use .010" alumina ceramic ($\epsilon_r = 9.9$). However, at 28 GHz, alumina circuits tend to be very lossy, not only because of their large size, but because the dielectric is less insulating than GaAs. Moreover, the lower dielectric constant and greater thickness of the alumina substrate means that a .010" (254 μm) wide 50 Ω line is almost 4 times wider than its .0028" (71 μm) equivalent on GaAs. This discrepancy in width will invariably make interconnections between alumina and GaAs circuits slightly reflective.

The phasesplitter and combiner are both designed by a three-step strategy similar to that used for the elements of the variable gain amplifier. As a starting point, the individual couplers are simulated on Libra using microstrip models for transmission lines, bends, resistors, and junctions. The critical dimensions of the circuits are initialized according to first principles and then modified using the optimization tools. This compensates for the non-ideal nature of the microstrip elements. Once satisfactory results are obtained, the phasesplitter and combiner are verified on a full-wave electromagnetic simulator. This will account for all possible coupling between circuit elements and provide the most realistic expectation of overall circuit behaviour.

For particularly large components, such as the phasesplitter and combiner, Sonnet Em is not the most desirable simulator. The electrical box required to enclose the circuit is essentially a waveguide. Once it becomes large enough, the cutoff frequency of the first mode will fall below the band of interest. At all frequencies above the cutoff point,

the results are corrupted by parasitic box modes and are no longer reliable. HP Momentum™, an alternative electromagnetic simulator, does not suffer from these problems. As its name suggests, it is based on the method of moments. This method meshes the structure into rectangles and triangles and represents each one as a circuit node. The interrelationship between these nodes is then described in terms of an impedance matrix, the elements of which depend upon the size, shape, and relative location of each discrete portion of the circuit.

After completing the three-step strategy, the results are acceptable, but often not optimal. The latter two-steps of Libra optimization followed by Momentum simulation are repeated several more times. In the following two sections, the dimensions of the finished circuits are presented along with simulated and measured results. The measurement setup consists of the HP8510C, HP8516A, and HP8340B connected in the same manner as depicted in Figure 4.39. It does not include the probe station. The chapter ends with a review of the photomasks used to pattern metal and resistor layers on the alumina substrates. These photomasks are used for fabricating the passive components for both the final vector modulator of Chapter 6 and the test pieces used in this chapter.

5.2 Phasesplitter design

5.2.1 Description

The term phasesplitter is used in this thesis to refer to a device which splits a signal into four orthogonal phase components. The first segment, a branch line coupler, splits the signal into equal amplitude components which are 90° apart. The two components are then further divided into equal amplitude antiphase signals using rat race hybrids. The complete schematic of these components and the equivalent layout are shown in Figure 5.1.

The branch line coupler, located at the far left of Figure 5.1, is composed of four quarter wavelength transmission line segments. The power split at the right-hand

Figure 5.1 Input phasesplitter schematic and layout.

ports is dictated by the relative impedance of the shunt to the series transmission lines. Assuming perfect junctions and lossless transmission lines, an impedance ratio of 1 to $\sqrt{2}$ will yield an even amplitude division between the two output ports. Furthermore, if the shunt transmission line has impedance Z_0 , then all ports will be matched to the characteristic impedance Z_0 at the centre frequency and the fourth port will be isolated.

Perfect amplitude balance is not possible in practise because of transmission line losses and parasitic effects associated with the T-junctions. Although the balance can be partially restored by adjusting the dimensions of the series and shunt lines, such modifications tend to degrade the input match, isolation and phase response. Using the optimization tools, these competing factors are automatically traded-off to yield the best compromise.

Next to the branch line coupler in Figure 5.1 are two rat race couplers: circular rings with four ports spaced at quarter wavelength intervals. The impedance of the ring is $\sqrt{2}Z_0$ and its circumference is $3\lambda/2$ at the centre frequency. This will ensure a good match, an even power split at the two outputs, and high isolation at the fourth port.

As can be seen from the implemented design in Figure 5.1, the “ring” is actually a rectangle. Albeit convenient for connections to the four amplifiers of the vector modulator, the bend in the $3\lambda/4$ wavelength section greatly disturbs the amplitude balance between the two outputs. As before, optimization on the linear circuit simulator can be applied to ensure the best possible balance between matching, phase and amplitude responses. For more details on branchline and rat-race couplers, the reader is referred to R. Collin [41].

5.2.2 Simulation and measurement

Figure 5.2 and Figure 5.3 compare the final Momentum simulated results to measured results and are based on the dimensions given in the layout of Figure 5.1. The amplitude imbalance is quite clear on the 0.6 diameter polar plots of Figure 5.2. It is about 1 dB and arises primarily from the branch line coupler. Although efforts were made to reduce it, it could not be eliminated without affecting the input and output matches.

The phase response from 27 to 29 GHz is quite satisfactory for the simulated results. The four vectors of the constellation are no more than 3 or 4 degrees away from being perfectly orthogonal. This performance is repeated across the entire band

Figure 5.2 Phasesplitter transmission response (a) simulated, (b) measured.

Figure 5.3 Simulated and measured phasesplitter matching results.

(plotted with filled symbols). By comparison, the measured results diverge at the band edges to as much as 15 degrees away from perfect orthogonality. Furthermore, the amplitude imbalance is much more pronounced and the transmission response considerably less flat. To illustrate this characteristic, the increment between symbols in Figure 5.2(b) is 200 MHz as opposed to 1 GHz in Figure 5.2(a). Finally, it should be noted that there is a ninety degree phase shift between the simulated and measured responses. This problem can be attributed to the calibration of the network analyser, a topic which is discussed in section 5.4.

The return loss results of Figure 5.3 appear quite promising in simulation, but are barely acceptable when measured. In simulation, all reflection coefficients are below -16 dB across a 4 GHz bandwidth. The measurements indicate no worse than -6 dB over the same frequencies, a level which threatens to drive up the VSWR of the final vector modulator to a considerable degree. For the sake of avoiding numerous and repetitive discussions, the other output reflection coefficients S_{33} , S_{44} and S_{55} are not presented. It

suffices to say that they exhibit approximately the same behaviour as S_{22} . These are better than -17 dB in simulation and better than -11 dB in measurement from 26 to 30 GHz. The isolation, which is also not plotted, is better than -15 dB between any combination of the output ports across the entire bandwidth in both simulation and measurement.

It is quite clear that the HP momentum simulations diverge considerably from the measured phasesplitter. As mentioned previously, the method of calibration is partially to blame. However, even this does not fully explain the observed variation between simulation and measurement. Another possibility is the coax-to-microstrip launchers. If two are not exactly alike, then their effects will not be nulled by the calibration.

5.3 Wilkinson power combiner design

5.3.1 Description

Once the vectors have been amplified, they are summed using the output power combiner. The combining is performed in two stages by three separate 3-dB

Figure 5.4 Output combiner schematic and layout.

Wilkinson 2-way power combiners. The schematic representation and final layout of the two-stage combiner is presented in Figure 5.4. It is designed to add all of the input signals in phase while simultaneously maintaining good input/output matches and high isolation between input ports.

The basic 3-dB Wilkinson power combiner consists of two quarter-wave-length sections with characteristic impedances of $\sqrt{2}Z_0$ and a $2Z_0$ Ω shunt resistor. The role of the resistor is to prevent a signal injected at one input from appearing at another, thus isolating the two ports. Provided that the ports are matched, the current through the resistor should be zero at the centre frequency.

The power combiner was designed using the same three-step approach as the phasesplitter. During the first step, a single stage was found to exhibit very poor matching at the inputs port using dimensions based on the first principles. As is often the case when applying theory to practical circuits, the bends and junctions must be accommodated before satisfactory results can be obtained. To compensate, the input and output ports are not matched to the system impedance of $Z_0 = 50 \Omega$. Furthermore, the resistor is not exactly $2Z_0$ but rather $1.27Z_0$. The new port impedances are realised by inserting lengths of non-50 Ω transmission line between the ports and the quarter wavelength sections. From the layout in Figure 5.4, it is possible to see that the input ports are preceded by wider segments of transmission line while a narrower line is connected at the output. Armed with more satisfactory results, the entire combiner, with dimensions as shown in the layout, is analysed using Momentum and then measured.

5.3.2 Simulation and measurement

The Libra and Momentum simulators yield very similar results. The transmission coefficients and the reflection coefficients are plotted in Figure 5.5. Here, there is clear evidence of the disadvantage of using Alumina over GaAs. The insertion loss is -10 dB. By misadjusting the resistor, the match at the ports is much improved, but it has also

Figure 5.5 Simulated and measured combiner transmission results.

added to the poor insertion loss. In an effort to improve matching, the currents in the resistor are no longer zero at the centre frequency. Nonetheless, the resistor still performs its role of suppressing the coupling between the input ports as shown by S_{32} in Figure 5.5. The worst results are for the two input ports adjacent to each other: two and three, and four and five. The measurements indicate -17 dB isolation. Isolation between the top two ports and the bottom two ports stands at about -25 dB.

The matching results are shown in Figure 5.6, and like the phasesplitter there is little or no comparison between the simulated and measured results as the curves simply do not overlap in any sensible fashion. Based on the figure, the input/output return loss is almost better than -10 dB across the 4 GHz bandwidth. This level is quite acceptable. Finally, no further reflection coefficients are plotted because they do not provide any

Figure 5.6 Simulated and measured combiner matching results.

further insight into the combiner's performance. Any discrepancies between the four input reflection coefficients are measurement or calibration related.

Like the phasesplitter, the simulated and measured results bear little resemblance. Once again, the calibration technique as described in section 5.4 is only partially to blame. The coax-to-microstrip launchers may also introduce some discrepancies.

5.4 Passive element photomasks

All of the passive components for the vector modulator are fabricated at the Communications Research Centre on a single 2" x 2" x .010" polished alumina substrate. The CRC fabrication facility requires two separate light field photomasks in order to separately define the metal and the resistor features. The masks are made of patterned chrome on low expansion coefficient (LEC) glass. In lightfield lithography, windows are etched in the chrome where features are to appear on the final circuit. Hence, when looking at the

mask, the pattern is clear and the background opaque. The composite drawing, which includes both the metal and resistors layers, is shown in Figure 5.7.

With a 2" square substrate, there is plenty of space to include two full copies of the phasesplitter and the combiner. The first pair are used to obtain the measurement results of sections 5.2.2 and 5.3.2 while the second pair are mounted in the vector modulator test jig. These pairs are identified as the "test pieces" and the "vector modulator components" respectively. In order to make measurements, it is necessary to fabricate a set of calibration standards. The TRL technique requires the three structures located at the top of Figure 5.7: a through (T), reflect (R), and line (L).

The phasesplitter and the combiners look slightly different from those of Figure 5.1 and Figure 5.4. The differences are included purely for calibration purposes.

Figure 5.7 2" x 2" photomask for fabrication of passive elements.

Figure 5.8 Phasesplitter/Combiner with TRL calibration standards.

For the test pieces, extra lengths of transmission line have been added at all ports. This length is exactly equal to that of the reflect standard and half that of the through standard. The network analyser has an internal program which will use stored measurements of the three standards to remove the effect of the extra line from all subsequent measurements. It may seem somewhat odd to add extra lengths of line to the circuits only to subtract them. However, the calibration standards will be mounted on brass blocks with the same coax-to-microstrip K-launchers as the test pieces (see Figure 5.8). When the calibration is performed, it will now include the slight impedance mismatch introduced by the launcher.

It should also be noted that some of the extra transmission line lengths on the test pieces are folded, unlike the straight calibration standards. In addition, the lengths of line leading to the ports emerging at the top and the bottom appear shorter than those emerging at the sides. The explanation for this has to do with the K-launchers used for the transition from coaxial line to microstrip. As shown in Figure 5.8, they are big, and as a consequence, the ports must be spaced out. Clearly, if the upper and lower ports were stretched out to the same length as the ones on the side, there would not have been enough space on the 2" x 2" substrate. As a result, they are cut-off and the remainders are located

in the centre of the mask. These *line extensions* are attached using wirebonds when the test pieces are mounted on their test-jigs. These ad-hoc solutions definitely had a significant effect on the calibrations and subsequent measurements as the bends and wirebonds will not have been accounted for. This problem is cited in both sections 5.2.2 and 5.3.2 as one of the principal causes for discrepancy between simulated and measured results. Unfortunately, the situation is unavoidable.

The vector modulator components also include the calibration lengths, but they will be retained on one side only. When fully assembled, the TRL calibration de-embeds the launchers and extra line at the input and output. The extra lengths of line at the outputs of the phasesplitter and the inputs of the combiner will be sawed off along the saw lines indicated.

The only additional features of note are test resistors, locations marks for dowel pins, metal pads for vias, and alignment marks. After all of the metal and resist is deposited and patterned, the entire substrate is baked. The resistivity of the Ti:W resistor layer is controlled by accelerated aging through baking. The test resistors are monitored throughout the baking process so that they can be adjusted to $50 \Omega/\text{square}$. The location marks are used for drilling .062" diameter holes through the substrate. The test jigs which will support the substrates have dowel pins located at exactly the same positions. This will ensure that the substrate is properly aligned on the jig and will prevent slippage while the epoxy is being cured. Each phasesplitter has three .040" x .040" pads. These form the top contacts for via holes which are ultrasonically drilled from the backside of the substrate. Metal contact is formed by sputtering metal onto the backside and into the hole. The last

features are alignment marks which appear as groups of squares. These ensure that during processing the resistors are properly aligned relative to the patterned metallization.

5.5 Summary

In this chapter, the passive components of the vector modulator were described, simulated, and measured. The photomask used to prepare the patterned alumina circuits was discussed with particular emphasis on the calibration arrangements. In terms of the results presented in this chapter, there are two principal sources of discrepancy between the simulated and measured results. The calibration strategy was less than ideal, and the coax-to-microstrip launchers used on the TRL standards may not have behaved exactly the same way as those on the final test pieces.

CHAPTER 6: Vector modulator assembly and test

6.1 Introduction

At this point, all of the components necessary to build a vector modulator have been reviewed. The purpose of this chapter is to describe the assembly of the vector modulator prototype, review the measured performance, and draw some conclusions regarding its suitability for the feedforward application.

6.2 Jig design

Since the vector modulator is comprised of several separate components, a special jig must be manufactured to support all of them. From the previous chapters, the vector modulator consists of GaAs variable gain amplifiers and alumina ceramic portions to split and combine the signal. Furthermore, the four amplifiers require a total of sixteen bias lines in order to provide DC power. A jig which will accommodate all of these needs is shown in Figure 6.1. This subsection will review some of its features.

The jig is essentially a block of brass .250" thick with a .006" tall ridge. The purpose of the ridge is to support the amplifiers such that they are level with the phasplitter and the combiner. Next to the amplifiers, on top of the ridge, are 32 chip capacitors. Sixteen of these are located immediately adjacent to the sixteen bias points of the four amplifiers. They are also wire-bonded to these bias points. Immediately behind and wire-bonded to the first set of capacitors, is a second set of capacitors. The second set is then connected to sixteen 4-40 filtercons which are threaded into two vertical .125" thick brass plates at either side of the jig. All combined, the capacitors and filtercons provide 2000 pF of capacitance to ground. At very low frequencies (i.e. MHz), the spiral inductors which isolate the RF portions of the VGAs from the DC bias networks become short circuits and the 2 pF on-chip shunt capacitors become open circuits. If the VGAs should oscillate at low frequency, the existing MMIC circuit components can do little to prevent

Figure 6.1 Vector modulator in test jig.

it. The off-chip capacitors are much larger in size and in dielectric constant and are able to short out the active devices over a much wider band of frequencies, including the low frequency range.

This DC biasing arrangement may seem somewhat complicated. Why so many separate chip capacitors? The answer is lack of space between adjacent base and collector bias points. In order to maximize the amount of shunt capacitance with the available space, it was necessary to employ multiple chip capacitors rather than larger single ones. Hence, the biasing arrangement which makes on-wafer measurements simpler, is rather inconvenient when the VGAs are mounted in a jig.

The phasesplitter and combiner are located on either side of the ridge and are level with the amplifiers. They are aligned on the jig by means of four .062" dowel pins. The holes which accept these dowels are positioned such that they align exactly with the location marks which are printed on the alumina substrates (see mask of Figure 5.7). To ensure a good electrical contact to the brass block, the substrates will be fixed into place using conducting epoxy.

RF connections are made using Directorate of Antennas and Integrated

Electronics (DAIE) K-launchers. These items, custom made by CRC, consist of a block of brass with a tapped hole drilled through them for a K-standard coaxial connector. When the launcher is screwed into the brass block, the centre pin of the connector will emerge flush against the microstrip line. The launcher has demonstrated an S_{11} of better than -25 dB up to 40 GHz.

6.3 Measurements

A vector modulator is a difficult device to characterize, primarily because when operating as designed, it can synthesize any desired transmission response. Thus, any test performed on the device must be repeated over a multitude of different states. For this section, a pattern of transmission states is generated for 28 GHz. Once the bias supplies to the vector modulator are correctly adjusted, the transmission response and matching results are simulated from 26 to 30 GHz.

The measurements are carried out using a test setup similar to that described by Figure 4.39 with the bias connections as shown in Figure 6.2 and Figure 6.3.

Figure 6.2 Schematic of vector modulator.

Figure 6.3 Vector modulator test setup.

However, like the phasesplitter and combiner measurements, the probing station is not used. Since the vector modulator is mounted in a jig, connections are made to the HP8516A using rigid cables.

Based on the amplifier measurements of section 4.6.4, it appeared that stability was not the issue it was originally made out to be. Hence, it is possible to biasing all collectors at the same level provided they are not increased beyond 2.70 V. Figure 6.3 shows how the various bias supplies are connected to the vector modulator. The eight bases of the mounted devices are all biased at $V_b = 1.3$ V by a single common power supply. Owing to the large amount of shunt capacitance and series inductance for each bias point, there is no fear of any coupling between amplifiers. The eight base bias points are denoted by b_{mn} where m is the amplifier identifier and n denotes first or second stage. The eight collector bias points are similarly identified and their respective power supplies are denoted V_{cn} where n identifies the amplifier.

Figure 6.4 shows 16 transmission states of the vector modulator at 28 GHz and for an input power of -3 dBm. This pattern contains several important cases which

should give an accurate overall measure of the circuit's behaviour. It also demonstrates the range of states achievable by the device. The collector biases required to achieve these states are identified as b_1 to b_{16} . These are summarized in the following table:

Table 6.1 Bias states for the vector modulator.

	V_{c1} (V)	V_{c2} (V)	V_{c3} (V)	V_{c4} (V)
b_1	0.406	0.595	0.403	2.700
b_2	0.406	1.450	0.403	2.204
b_3	0.406	1.798	0.403	1.603
b_4	0.406	0.540	1.807	0.601
b_5	0.706	0.391	1.860	0.400
b_6	1.555	0.401	1.502	0.400
b_7	1.822	0.404	0.458	0.400
b_8	1.822	0.405	0.400	1.403
b_9	0.396	0.516	0.561	1.150
b_{10}	0.391	0.518	1.103	1.323
b_{11}	0.534	0.550	0.400	0.400
b_{12}	0.585	0.516	0.480	0.741
b_{13}	0.394	0.811	0.931	0.400
b_{14}	0.394	0.478	1.192	0.660
b_{15}	0.704	0.401	0.499	0.400
b_{16}	0.493	0.400	0.400	1.243

The resulting transmission responses are divided into a set of three rings with nominal insertion losses of -20 dB, -25 dB, and -35 dB respectively. Thus, the polar chart of Figure 6.4 has an outer radius of 0.12 or -18.5 dB. The eight points farthest from the centre (b_1 to b_8) represent two amplifiers turned off and two operating at high gain. The imbalance in the phasesplitter now becomes quite evident as amplifiers 1 and 2 need

Figure 6.4 Vector modulator transmission response at 28 GHz.

not be turned on very much before they swamp out amplifiers 3 and 4. The inner four bias points (b_9 to b_{12}) are intended to illustrate the modulator's response for low bias levels where exact cancellation of antiphase channels is required. The bias points b_{13} to b_{16} rep-

Figure 6.5 Vector modulator transmission response.

resent the case of two orthogonal amplifiers turned off and two operating at mid-range gain.

The magnitudes of the transmission responses are plotted in Figure 6.5 as a function of frequency (26 to 30 GHz). In order to avoid cluttering the figure with curves, only three representative responses are given: b_1 , b_{10} , and b_{13} . These correspond to the three main groupings discussed above. Across the 27.35 to 29.35 GHz band, the high and mid-range gain responses b_1 and b_{13} vary by no more than ± 2 dB about the nominal level at 28 GHz. When inexact cancellation between antiphase amplifiers occurs, the gain flatness deteriorates even further to approximately ± 3 dB as exhibited by b_{10} . Over a 1 GHz bandwidth, the variation in gain drops by a half.

The input matching results are presented in Figure 6.6 as a band of curves. The match is better than -15 dB from 26 to 30 GHz for the three representative bias states. Surprisingly, the variation in input match with bias is very slight. When compared to the results of Figure 4.40, the corresponding vector modulator characteristic shows considera-

Figure 6.6 Vector modulator input return loss.

Figure 6.7 Vector modulator output return loss.

ble improvement. Clearly, the mismatch at the outputs of the phasesplitter has in some way compensated for the slight mismatch at the amplifier input.

The output match is completely constant with respect to the amplifier gain states as indicated by Figure 6.7. Since the output match of the variable gain amplifier is bias independent, it stands to reason that the vector modulator's output match will also be constant. From the above figure, the output return loss is better than -12 dB from 26 GHz to 28.4 GHz. This is not quite as good as desired as the output VSWR will be too high for proper operation in the 28.35 GHz to 29.35 GHz range.

6.4 Comments

Based on the results presented in section 6.3, it is possible to assess the capability of the phasesplit, amplitude adjust, and combine vector modulator configuration. This section will review the causes of the vector modulator gain variation and comment on its suitability as a component in a wideband feedforward linearizer at 28 GHz. It

will also examine the potential for the alternative phase shift and amplitude adjust configuration and propose a modification to the adaptive feedforward.

Clearly, it would appear highly ambitious to expect this vector modulator configuration to operate over the full 2 GHz bandwidth. At the very best, the gain variation is ± 2 dB. To explain why this variation occurs, recall the simple derivation in section 3.2. It sought to illustrate that the overall amplitude and phase response could be controlled by the gains of a set of four amplifiers. At that time, however, the gains were assumed real and their frequency dependence was ignored. If these two assumptions are dropped, then the amplifiers' transfer functions become

$$H_k(\omega) = K_k(\omega)e^{-j\omega\tau_k} \quad (6.1)$$

Where for the k th amplifier, $K_k(\omega)$ is approximately equivalent to $|S_{21}|$ as given by the curves in Figure 4.35 and τ_k represents group delay. If the input signal has a spectrum given by $V_{IN}(\omega) = |V_{IN}(\omega)|e^{j\theta(\omega)}$, then the output spectrum must be

$$\begin{aligned} V_{OUT}(\omega) = & \frac{1}{4}|V_{IN}(\omega)|e^{j\theta(\omega)}K_3(\omega)e^{-j\omega\tau_3} - \frac{1}{4}|V_{IN}(\omega)|e^{j\theta(\omega)}K_4(\omega)e^{-j(\omega\tau_4 - \pi)} \\ & + \frac{1}{4}|V_{IN}(\omega)|e^{j\theta(\omega)}K_2(\omega)e^{-j(\omega\tau_2 - \pi/2)} - \frac{1}{4}|V_{IN}(\omega)|e^{j\theta(\omega)}K_1(\omega)e^{-j(\omega\tau_1 - 3\pi/2)} \end{aligned} \quad (6.2)$$

Rearranging,

$$\begin{aligned} V_{OUT}(\omega) = & \frac{1}{4}|V_{IN}(\omega)|e^{j\theta(\omega)}([K_3(\omega)e^{-j\omega\Delta\tau_b} - K_4(\omega)]e^{-j\omega\tau_4} \\ & + [K_2(\omega)e^{-j\omega\Delta\tau_i} - K_1(\omega)]e^{-j(\omega\tau_1 - \pi/2)}) \end{aligned} \quad (6.3)$$

Here, $\Delta\tau_i = \tau_2 - \tau_1$ and $\Delta\tau_b = \tau_3 - \tau_4$ are the differences in group delay for the two pairs of antiphased amplifiers. Any losses and frequency dependence arising from the couplers have been absorbed by the K_k 's. For $V_{OUT}(\omega)$ to be an undistorted, yet phase shifted and amplitude adjusted replica of $V_{IN}(\omega)$, the inphase and quadrature components of equation (6.3) must be frequency insensitive. Hence,

$$K_3(\omega)e^{-j\omega\Delta\tau_h} - K_4(\omega) = K_{inphase} \quad (6.4)$$

$$K_2(\omega)e^{-j\omega\Delta\tau_i} - K_1(\omega) = K_{quadrature} \quad (6.5)$$

This result will only occur if two conditions are met. Firstly, the group delay must be identical for antiphase pairs of amplifiers for any applied bias. Secondly, as the gain is changed, it must vary by a constant amount at each frequency in the band. Moreover, this amount must be the same for each amplifier. Of these two conditions, only the first one is met. The second one is not, and hence some frequency variation in the overall vector modulator transmission characteristic results.

From Figure 6.5, it is evident that the high cancellation states suffer from even greater variation in transmission response. This cannot be explained purely on the basis of equation (6.3) because it makes an implicit assumption which conceals a problem. It presupposes that the phasesplitter generates perfectly orthogonal vectors. From section 5.2.2, this is known to be untrue. It is stated that the four vectors of the constellation vary by as much as 15 degrees from perfect orthogonality. This is quite a significant deviation and it has a profound effect when perfect cancellation is desired. If the difference between two equi-amplitude sinusoids is calculated where one is phase shifted, the result is a ripple. For the case of a modest 4 degree phase error, the resulting ripple can cause variations in the amplitude response by as much as ± 5 dB.

The absence of a flat gain response will certainly pose problems when the vector modulator is inserted in the feedforward linearizer. In section 2.5.3, coefficient accuracy requirements were stated. From the discussion following equation (2.128), the key requirement for -40 dB IM suppression was for the α coefficient to be adjustable to within 3.3 %. The calculation assumed that a -40 dB rectangular band rejection filter was placed in front of the power detector used to sample the output signal. Based on the measured results of section 6.3, the smallest gain variation was observed to be ± 2 dB over a 2 GHz bandwidth and approximately ± 1 dB over the more modest bandwidth of 27.35 GHz

to 28.35 GHz. Assuming a nominal attenuation of 20 dB, the coefficient misadjustment due to the gain variation is approximately 20% for the 2 GHz bandwidth and 10% for the 1 GHz bandwidth. This is far too high to achieve any distortion suppression. The latter case might be sufficient to suppress IM by -10 dB from $IMSR_a = -20$ dB to $IMSR_0 = -30$ dB.

Clearly, the vector modulator cannot accommodate such a wide band. This should come as no surprise as none of the components have so far demonstrated the required flatness or constancy. It is perhaps more insightful to work backwards. Using equation (2.127), assume that the modest -20 dB IM suppression is still desired. If the masking filter is included and the amplifier itself generates -20 dB IM levels, then the required coefficient accuracy is the 3.3% stated above. For the sake of argument, suppose that the two branches of the signal cancellation loop yield a nominal -30 dB insertion loss leading up to the subtractor. A 3.3% difference means that they could be no more than ± 0.3 dB apart. This sort of tolerance is still unrealistic but could possibly be met over the band of 27.35 GHz to 27.85 GHz assuming that the low-gain states of the vector modulator could be avoided. Of course, this prediction does not take into consideration the stringent delay matching requirements discussed in section 2.5.4.

It should be conceded that, of the two vector modulator architectures considered in chapter 3, this one is definitely inferior. Among its disadvantages is the inherent narrowband nature of the phasesplitter and combiner. As discussed in this section, the vectors must be very close to orthogonal in order to make use of the full dynamic range. Secondly, the gain of the amplifiers must change evenly at all frequencies or the difference signals in the orthogonal branches will not have flat characteristics. The trouble is, both of these properties are difficult to design for at Ka-band. The dependence on full-wave simulators for accurate modelling is simply not compatible with extensive optimization.

By comparison, the phase shift amplitude adjust configuration does not require the subtraction of nearly equal signals. Hence, there will be no ripple at the lower

will have suffered roughly the same roll-off characteristics. This will provide another degree of freedom for adjustment, while compensating for the non-ideal transmission characteristics of the vector modulators. Unfortunately, such an arrangement will increase the complexity of the adaptation circuitry and add to the overall cost of the linearizer. The configuration is shown, with adaptive delay matching in Figure 6.8.

6.5 Summary

This chapter has sought to characterize and explain the behaviour of a phase split, amplitude adjust and combine vector modulator. A complete jig design was reviewed and implemented. The addition of large amounts of shunt capacitance at all bias points suppresses all low-frequency oscillations and reduces the number of required power supplies from sixteen to five. Using the prototype vector modulator, a set of measurements were presented to give an idea of its performance. Based on these measurements, it was possible to make an assessment of the overall viability of the adaptive feedforward concept at Ka-band. It is concluded that the vector modulator, as designed, could function adequately over a 500 MHz bandwidth and could theoretically offer 20 dB IM suppression. A modification to the original design has been proposed to reduce the sensitivity of the feedforward to the vector modulator's non-ideal transmission characteristic.

CHAPTER 7: Conclusion

7.1 Summary

Wireless communications systems of the future, such as LMCS, WLAN, and Advanced SATCOM are being designed to support a large variety of audio, data, and video services. Due to frequency crowding at the lower end of the radio spectrum and the demand for large bandwidths, these systems must operate above 26 GHz. They must also be efficient in their use of spectral resources and resistant to channel fading.

These requirements have far-reaching implications for the overall system design and have prompted developers to investigate advanced transceiver architectures and signalling formats. Among other things, it is intended that the new systems use linear modulation formats and be capable of dynamic channel allocation. To accommodate both of these features simultaneously, all power amplification must be strictly linear and the signal combining technique adaptable to changes in the number, locations, and bandwidth of individual channels. The proposed solution is to combine all signals at low power and then amplify the sum with a single power amplifier.

The principal thrust of this thesis was to investigate adaptive feedforward linearization as a means of ensuring the linearity of the power amplifier. It was motivated by a CITR LMCS project which exemplifies many of the concerns listed above. The thrust was divided into two major goals. Firstly, a study was undertaken to review PA non-linearity, distortion, and linearization techniques. The intent was to provide sufficient background to appreciate a wideband adaptive feedforward solution based on the work of J. Cavers. The second goal was to present a prototype for the key RF component of the feedforward, a Ka-band vector modulator. Having done so, it was possible to compare the RF requirements of the linearizer to obtainable results. The contents of this thesis, simulation results, measurements and key findings will now be reviewed.

All power amplifiers are non-linear to some extent. The non-linearities manifest themselves in a variety of different behaviours: harmonic distortion, intermodulation distortion, phase distortion, linear distortion, and memory effects. Feedforward linearization is a technique which can, in theory, offer a means of suppressing all forms of distortion. As indicated in section 2.5, the level of distortion suppression depends entirely on the RF components and their ability to adjust the gain, phase, and delay correctly. Nonetheless, even if the cancellation can be effected across the entire band, external circuitry must be capable of adapting the RF parameters appropriately. The gradient feedforward linearizer has a robust and rapid converge algorithm which is proven reliable at lower frequencies [7].

In chapter 3, two types of vector modulator were introduced and compared. Owing to the particularities of the two architectures, and given the foundry services currently on offer, it was decided that the phase shift-amplitude adjust configuration could not be pursued. The alternative configuration was intuitively simple. It consisted of splitting a Ka-band signal into four phase states, amplifying them by different amounts, and then recombining them. The resultant is a summation of the four “vectors.”

The main component of the vector modulator of this thesis is a GaAs HBT variable gain amplifier. Anticipating considerable insertion loss from the splitting and combining circuits, a variable gain amplifier appeared more attractive than a variable attenuator. Its bandwidth and gain variation were maximized using both RF and DC feedback. The use of RF shunt feedback on the output stage simultaneously stabilized the amplifier, simplified the output match, and flattened the transmission response. However, its use also resulted in a loss of gain.

Throughout the VGA design process, Sonnet™ Em version 4.0 was used to model spiral inductors, MIM capacitors, and junctions. In some instances, entire portions of the circuit were simulated in order to account for mutual coupling between closely spaced features. The final simulation results for the amplifier anticipated 8 dB maximum

gain, -10 dB minimum gain, better than -14 dB input return loss from 24 to 32 GHz, and better than -11 dB output return loss from 26 to 30 GHz. All simulated results were based on -7 dBm input power. Following fabrication, the amplifier was tested using an on-wafer probing station and the following performance was observed: maximum gain of 4 dB, minimum gain of -12 dB, input return loss better than -10 dB from 22 to 30 GHz, and output return loss better than -10 dB from 27 to 29 GHz. As in the simulations, the input power was set at -7 dBm which is just below the gain compression point.

The phasesplitter and the 4-to-1 Wilkinson power combiner were designed and implemented on .010" alumina ceramic. The former consists of a branch-line coupler whose outputs are connected directly to a pair of rat-race couplers. The combined structure splits the signal into four components which are equal in amplitude and mutually orthogonal. When simulated on HP Momentum, the magnitudes of the vectors vary by no more than 1 dB and the phases are orthogonal to within 3 to 4 degrees between 27 and 29 GHz. The nominal insertion loss at 28 GHz is approximately -7.5 dB. The structure also exhibits better than -17 dB simulated return loss at all ports. The 4-to-1 power combiner is composed of three individual Wilkinson couplers and sums the phase shifted amplifier outputs. Based on HP Momentum simulation results, the structure exhibits -10 dB insertion loss, -20 dB return loss, and -15 dB isolation at 28 GHz. The amplitude variation is less than ± 0.5 dB.

When these structures are fabricated and tested, the results were far less optimistic than HP Momentum anticipated. The phase response of the phasesplitter is as much as 15 degrees away from orthogonality a mere 1 GHz from the centre frequency. The input return loss is much worse than predicted and stands at -6 dB from 26 to 30 GHz. The output return loss, although less severe, is still 6 dB worse at -11 dB between 26 and 30 GHz. The combiner shows similar discrepancies in the matching results with input return loss below -8 dB and output return loss below -20 dB. Simply stating these numbers does little to advance the concerns regarding the accuracy of the simulator. Plotting meas-

ured and simulated results on the same graphs reveals absolutely no similarity in responses. Although this is partly due to the calibration of the network analyser, one is still forced to conclude that HP momentum is not ideally suited to simulating complicated structures at 28 GHz.

The vector modulator is measured at sixteen separate bias states as a demonstration of its capabilities. The gain was adjusted between -20 dB and -35 dB. At high gain, the transmission response varied by a maximum of ± 2 dB across the band of frequencies from 27.35 to 29.35 GHz. At higher attenuation, the response was less flat and generally unusable. This was attributed to the non-orthogonality of the input vectors at the band edges. The phase response of the vector modulator is completely unrestricted. That is to say, by careful adjustment of the gain of the four amplifiers, any phase from 0 to 360° is possible. The input and output return loss of the vector modulator are better than -12 dB across the 27.35 to 28.35 GHz band and for all possible bias states.

Based on the measured results, an assessment was made to establish how well the vector modulator would work in a feedforward linearizer. As it turns out, the bandwidth is limited to an absolute maximum of 500 MHz. Assuming that the adaptive delay could be implemented with the required accuracy, it is speculated that the linearizer could reduce intermodulation levels from -20 dB to -40 dB for the 27.35 to 27.85 GHz frequency block. It is conceded that these results are somewhat disappointing. However, given the extremely tight coefficient accuracy requirements and the inherent limitations of the vector modulator architecture, this performance should not come as any great surprise.

7.2 Future work

In all likelihood, this thesis has raised more questions than it has answered. For example, Figure 2.19 should serve to explain the extent and scale of a wideband feedforward linearizer. However, the overall feasibility of such a complicated and expensive circuit remains largely unanswered. When does it become more practical to use an array of

corporately fed backed-off class A amplifiers? How many can be arranged in this fashion? What are the cost trade-offs?

Even if it does turn out to be practical to realize the full feedforward depicted in Figure 2.19, the overall layout of the downconversion chains needs to be designed. This will include bandpass filters, downconverters, oscillators, and routing networks. The right combination of MMIC and hybrid technology is crucial if circuit losses and financial costs are to be minimized.

There remains several outstanding issues with regards to the vector modulator. It was clear right from the beginning that the phasesplit, amplitude adjust, and combine configuration suffered from some inherent problems. It would make sense to attempt the alternative phase shift and amplitude adjust architecture using PHEMT's as the reflective terminations. Based on examples at L and X band, this approach holds a great deal of promise and could potentially offer a much wider bandwidth and greater overall gain flatness.

Finally, the question of adaptive delay matching has been left unresolved. It has constantly been assumed that it is possible to synthesize group delay along with variable gain and phase. A demonstration of the principles and a prototype at Ka-band are logical extensions of the work on the vector modulator.

7.3 Final comments

Broadband wireless communications operating at or above the millimetric wave boundary are a reality. First generation WLAN, LMCS and broadband satellite services are commercially available. Nonetheless, these systems are only the beginning. The second generation systems must be more adaptable, spectrally efficient, and cheaper. Dynamic channel allocation and high-efficiency linear modulation schemes are two of many considerations which hold the promise of bringing these systems to fruition. However, the question of cost and implementation is far from answered. Complex systems at

Ka-band are certainly realizable, but they are expensive. GaAs circuits, which dominate RF applications at this frequency, cannot be manufactured to the scale of Si due to processing limitations and low yield. Alternative strategies, such as hybrid ceramic and GaAs circuits can reduce costs assuming modest production quantities, but at Ka-band they generally suffer a loss of performance.

In many ways, the vector modulator described in this thesis epitomizes the issues that must be surmounted in order to realize second generation systems. It uses both GaAs and alumina to control signals at Ka-band. As the work contained in this thesis indicates, this is a tricky proposition. Combiners, splitters, and transmission lines in microstrip are extremely lossy and frequency sensitive. The parasitic behaviour of junctions, capacitors, and inductors are significant and force the designer to simulate each component in excruciating detail: nothing can be taken for granted. Fabrication and assembly also must meet very demanding tolerances.

The front end of a 28 GHz broadband receiver faces many of the same challenges as the vector modulator, only on a much greater scale. The feedforward network alone requires some 20 GaAs components and at least a dozen splitters/combiners. Based on the vector modulator design, this level of complexity and the potential for extremely high loss, suggests that alternative approaches should be investigated.

REFERENCES

- [1] P.B. Kenington, R.J. Wilkinson, and J.D. Marvill, "A multi-carrier amplifier for future mobile communications systems," *Proceedings of IEE 6th International Conference on Mobile Radio and Personal Communications*, Coventry, United Kingdom, pp. 151-156, December 1991
- [2] A.K. Johnson and R. Myer, "Linear amplifier combiner," *Proceedings of IEEE Vehicular Technology Conference*, Tampa, Florida, June 1987, pp. 421-423.
- [3] C. Hopwood, "LMCS: the dawning of a new technology," *Wireless Telecom*, vol. 14, no. 4, December 1996, pp. 19-21.
- [4] CITR, "Draft Research Proposal 1996-98: Broadband Wireless Communications for Local Multipoint Communication Services (LMCS)," pp. 1-6, April 1996.
- [5] Y. Akaiwa and Y. Nagata, "A linear modulation scheme for spectrum efficient digital mobile telephone systems," *Proceedings of the International Conference on Digital and Mobile Radio Communications*, Venice, pp. 218-226, 1987
- [6] C.-L. Lu and K. Feher, "Noncoherent detection of $\pi/4$ -QPSK systems in a CCI-AWGN combined interference environment," *Proceedings of IEEE Vehicular Technology Conference*, San Francisco, June 1994, pp. 83-94.
- [7] S.J. Grant, "A DSP controlled adaptive feedforward power amplifier linearizer," M.A.Sc. Thesis, School of Engineering Science, Simon Fraser University, July 1996.
- [8] A.S. Sedra and K.C. Smith, Microelectronic circuits, Saunders College Publishing, pp. 646-674, 1991.
- [9] J.L.B. Walker, High-power GaAs FET amplifiers, Artech House, Boston, pp. 5-30, 114-118, 1993.
- [10] S. Sychaleun, "GaAs/AlGaAs HBT modelling and its application to a 14 GHz class A medium power amplifier," M. Eng. Thesis, Department of Electronics, January 1996.
- [11] P.H. Ladbrooke, MMIC Design: GaAs FETs and HEMTs, Artech House, Boston, pp. 69-83, 1989.
- [12] Tri. T. Ha, Solid-state microwave amplifier design, John Wiley & Sons, pp. 203-209, 1981.
- [13] G.L. Heiter, "Characterization of nonlinearities in microwave devices and systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-21, no. 12, pp. 797-805, December 1973.
- [14] H.S. Black, "Inventing the negative feedback amplifier," *IEEE Spectrum*, pp. 55-60, December 1977.
- [15] D.J.H. Maclean, Broadband feedback amplifiers, Research Studies Press, pp. 1-10, 1982.

- [16] M. Briffa and M. Faulkner, "Stability analysis of cartesian feedback linearization for amplifiers with weak nonlinearities," *IEE Proceedings on Communications*, Vol. 143, No. 4, pp 212-18, August 1996.
- [17] M. Johansson, T. Mattsson, L. Sundstrom, and M. Faulkner, "Linearization of Multi-Carrier Amplifiers," *Proceedings of IEEE Vehicular Technology Conference*, North Jersey, pp. 684-87, May 1993.
- [18] S.P. Stapleton and F.C. Costescu, "An adaptive predistorter for a power amplifier based on adjacent channel emissions," *IEEE Transactions on Vehicular Technology*, Vol. 41, No. 1, pp. 49-54, February 1992.
- [19] J.K. Cavers, "Adaptation behaviour of a feedforward amplifier linearizer," *IEEE Transactions on Vehicular Technology*, Vol. 44, No. 1, pp. 31-39, February 1995.
- [20] J.K. Cavers, "Adaptive feedforward linearizer for rf power amplifiers," U.S. Patent 5,489,875, February 6, 1996.
- [21] J.K. Cavers, "Adaptation behaviour of a feedforward amplifier linearizer," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1., pp. 31-40, February 1995.
- [22] S.J. Grant and J.K. Cavers, "A DSP controlled adaptive feedforward amplifier linearizer," *Proceedings of IEEE Vehicular Technology Conference*, 1996, pp. 788-792.
- [23] S. Narahashi and T. Nojima, "Extremely low-distortion multi-carrier amplifier - self-adjusting feed-forward (SAFF) amplifier," *Proceedings of IEEE International Communications Conference*, 1991, pp. 1485-1490.
- [24] R.E.J. Gerard, "Feedforward Amplifiers," U.S. Patent 4,412,185, October 25, 1983.
- [25] R.E. Myer, "Feed forward linear amplifier," U.S. Patent 4,885,551, December 5, 1986.
- [26] J.K. Cavers, "Adaptive feedforward linearizer for rf power amplifiers," U.S. Patent 5,489,875, February 6, 1996.
- [27] S. Lucyszyn and I.D. Robertson, "Synthesis techniques for high performance octave bandwidth 180° analog phase shifters," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-40, no. 4, pp. 731-9, March 1992.
- [28] P.T. Greiling, "InP based microwave/millimetrewave technology for military and commercial markets," *26th European Microwave Conference*, Prague, pp. 743-747, September, 1996.
- [29] B. Bayraktaroglu, "GaAs HBT's for microwave integrated circuits," *Proceedings of the IEEE*, vol. 81, no. 12, pp 1762-1785, December 1993.
- [30] B.K. Oyama, B.P. Wong, "GaAs HBT's for Analog Circuits," *Proceedings of the IEEE*, vol. 42, no. 12, pp. 1744-1761, December 1993.
- [31] P. Vizmuller, RF design guide: Systems, Circuits, and Equations, Artech House, Boston, pp. 64-68, 1995.

- [32] I.D. Robertson, et al., MMIC Design, *The Institute of Electrical Engineers*. London, pp. 36-50, 1995.
- [33] 5C20, HBT Design Handbook, *Unpublished internal document - Nortel Technologies*, pp. 1-1 to 1-18. January 1996.
- [34] Q.M. Zhang, H. Hu, J. Sitch, R.K. Surridge, J. M. Xu, "A new large signal HBT model," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-44, no. 11, pp. 2001-9, November 1996.
- [35] K.B. Niclas, "GaAs MESFET feedback amplifiers: design considerations and characteristics," *Microwave Journal*, , pp. 39-48, March 1980.
- [36] D.J. Ahlgren, W.H. Ku, "Gain-bandwidth properties of a class of matched feedback amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-35, no. 4, pp. 361-9, April 1987.
- [37] K.B. Niclas, W.T. Wilser, R.B. Gold, W.R. Hitchens "The matched feedback amplifier: ultrawide-band microwave amplification with GaAs MESFET's," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-28, no. 4, pp. 285-94, April 1980.
- [38] A.M. Pavio, "A network modelling and design method for a 2-18 GHz feedback amplifier," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-30, no. 12, pp. 2212-16, December 1982.
- [39] F. Perez and V.A. Ortega, "A graphical method for the design of feedback networks for microwave transistor amplifiers: theory and applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-29, no. 10, pp. 1019-27, October 1981.
- [40] D. Mellor and J. Linvill, "Synthesis of interstage networks of prescribed gain versus frequency slopes," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-23, no. 12, pp. 1013-20, December 1975.
- [41] R.E. Collin, Foundations for microwave engineering: second edition, McGraw-Hill, Inc., New York, pp. 432-3, 437-50, 1992.
- [42] S. Lucyszyn, I.D. Robertson, and A.H. Aghvami, "Negative group delay synthesizer," *Electronics Letters*, vol. 29, no. 9, pp. 798-800, April 1993.
- [43] S. Lucyszyn, and I.D. Robertson, "Analog reflection topology building blocks for adaptive microwave signal processing applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-43, no. 3, pp. 601-611, March 1995.

APPENDIX I: Fourier Expansions

This appendix contains three separate calculations of various Fourier series coefficients used in section 2.2. The first two are for the DC and first harmonic terms of the class AB amplifier output current waveforms of section 2.2.2. The third applies to the first harmonic of the saturated output current waveform of section 2.2.3.

Derivation of a_0 for the class AB amplifier output current waveform

From equation (2.11)

$$\begin{aligned}
 a_0 &= \frac{2}{T} \int_0^{T/2} i_{ds}(t) dt \\
 &= \frac{2}{T} \left[\int_0^{t_0} (xI_{MAX} + (1-x)I_{MAX}\cos(\omega t)) dt \right] \quad \text{where } t_0 = \frac{1}{\omega} \cos^{-1}\left(-\frac{x}{1-x}\right) \\
 &= \frac{2}{T} I_{MAX} \left[x \int_0^{t_0} dt + (1-x) \int_0^{t_0} \cos(\omega t) dt \right] \\
 &= \frac{2}{T} I_{MAX} \left[xt_0 + \frac{(1-x)}{\omega} \sin(\omega t_0) \right]
 \end{aligned}$$

Substituting for t_0

$$a_0 = I_{MAX} \left[\frac{x}{\pi} \cos^{-1}\left(-\frac{x}{1-x}\right) + \frac{(1-x)}{\pi} \sin\left(\cos^{-1}\left(-\frac{x}{1-x}\right)\right) \right] \quad (\text{A.1})$$

Using the unit circle and the right angle triangle contained therein

$$\begin{aligned}
 \sin\left(\cos^{-1}\left(-\frac{x}{1-x}\right)\right) &= \sqrt{1 - \left(-\frac{x}{1-x}\right)^2} \\
 &= \frac{\sqrt{1-2x}}{1-x}
 \end{aligned}$$

(A.2)

Substitute equation (A.2) into (A.1) to obtain the result of equation (2.13)

$$\begin{aligned} a_0 &= I_{MAX} \left[\frac{x}{\pi} \cos^{-1} \left(-\frac{x}{1-x} \right) + \frac{(1-x)}{\pi} \left(\frac{\sqrt{1-2x}}{1-x} \right) \right] \\ &= \frac{I_{MAX}}{\pi} \left[\sqrt{1-2x} + x \cos^{-1} \left(-\frac{x}{1-x} \right) \right] \end{aligned}$$

Derivation of a_1 for the class AB amplifier output current waveform

From equation (2.12)

$$\begin{aligned} a_1 &= \frac{4}{T} \int_0^{T/2} i_{ds}(t) \cos(\omega t) dt \\ &= \frac{4}{T} \int_0^{t_0} (x I_{MAX} + (1-x) I_{MAX} \cos(\omega t)) \cos(\omega t) dt \\ &= \frac{4 I_{MAX}}{T} \left[\int_0^{t_0} x \cos(\omega t) dt + \int_0^{t_0} (1-x) \cos^2(\omega t) dt \right] \\ &= \frac{4 I_{MAX}}{T} \left[x \left(\frac{\sin(\omega t)}{\omega} \right) \Big|_0^{t_0} + (1-x) \left(\frac{t}{2} + \frac{\sin(2\omega t)}{4\omega} \right) \Big|_0^{t_0} \right] \\ &= I_{MAX} \left[\frac{4x}{T\omega} \sin(\omega t_0) + \frac{4}{T} (1-x) \left(\frac{t_0}{2} + \frac{\sin(2\omega t_0)}{4\omega} \right) \right] \\ &= I_{MAX} \left[\frac{2x}{\pi} \sin(\omega t_0) + \frac{2}{T} (1-x) t_0 + \frac{(1-x)}{2\pi} \sin(2\omega t_0) \right] \end{aligned}$$

Applying the trigonometric identity $\sin 2\theta = 2 \sin(\theta) \cos(\theta)$

$$a_1 = I_{MAX} \left[\frac{2x}{\pi} \sin(\omega t_0) + \frac{2}{T} (1-x) t_0 + \frac{(1-x)}{\pi} \sin(\omega t_0) \cos(\omega t_0) \right] \quad (\text{A.3})$$

Substituting for the value of t_0

$$\begin{aligned} a_1 &= I_{MAX} \left[\frac{2x}{\pi} \sin \left(\cos^{-1} \left(-\frac{x}{1-x} \right) \right) + \frac{(1-x)}{\pi} \cos^{-1} \left(-\frac{x}{1-x} \right) \right. \\ &\quad \left. + \frac{(1-x)}{\pi} \sin \left(\cos^{-1} \left(-\frac{x}{1-x} \right) \right) \cos \left(\cos^{-1} \left(-\frac{x}{1-x} \right) \right) \right] \end{aligned} \quad (\text{A.4})$$

Substituting equation (A.2) into (A.4) yields

$$a_1 = I_{MAX} \left[\frac{2x\sqrt{1-2x}}{\pi(1-x)} + \frac{(1-x)}{\pi} \cos^{-1} \left(-\frac{x}{1-x} \right) + \frac{(1-x)\sqrt{1-2x}}{\pi(1-x)} - \frac{x}{1-x} \right] \quad (\text{A.5})$$

Simplifying (A.5) results in equation (2.14)

$$a_1 = \frac{I_{MAX}}{\pi} \left[\frac{x\sqrt{1-2x}}{1-x} + (1-x) \cos^{-1} \left(-\frac{x}{1-x} \right) \right]$$

Derivation of a_1 for the overdriven class A output current waveform

Starting at equation (2.28)

$$\begin{aligned} a_1 &= \frac{4}{T} \int_0^{T/2} I_{ds}(t) \cos(\omega t) dt \quad \text{where } t_1 \\ &= \frac{4}{T} \left[\int_0^{t_1} \frac{I_{MAX}}{2} \cos(\omega t) dt + \int_{t_1}^{T/2-t_1} \hat{I} \cos(\omega t) \cdot \cos(\omega t) dt \right. \\ &\quad \left. + \int_{T/2-t_1}^{T/2} \frac{I_{MAX}}{2} \cos(\omega t) dt \right] \end{aligned}$$

By symmetry, the first and third integrals will exactly cancel each other

$$\begin{aligned} a_1 &= \frac{4}{T} \left[\int_{t_1}^{T/2-t_1} \hat{I} \cos^2(\omega t) dt \right] \\ &= \frac{4}{T} \left[\hat{I} \left\{ \frac{t}{2} + \frac{\sin(2\omega t)}{4\omega} \right\} \right]_{t_1}^{T/2-t_1} \\ &= \frac{4\hat{I}}{T} \left[\frac{T/2-t_1}{2} + \frac{\sin(2\omega(T/2-t_1))}{4\omega} - \frac{t_1}{2} - \frac{\sin(2\omega t_1)}{4\omega} \right] \\ &= \frac{4\hat{I}}{T} \left[\frac{T}{4} - t_1 - \frac{\sin(2\omega t_1)}{2\omega} \right] \end{aligned} \quad (\text{A.6})$$

Substituting for t_1

$$a_1 = \hat{l} \left[1 - \frac{4}{T} \frac{1}{\omega} \cos^{-1}(u) - \frac{4}{T} \frac{1}{2\omega} \sin(2\cos^{-1}(u)) \right]$$

Simplifying, equation (2.29) is obtained

$$a_1 = \hat{l} \left[1 - \frac{2}{\pi} \cos^{-1}(u) - \frac{1}{\pi} \sin(2\cos^{-1}(u)) \right]$$

IMAGE EVALUATION TEST TARGET (QA-3)

APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc.. All Rights Reserved