LUND UNIVERSITY

CALLUM Linear Transmitter - Architecture and Circuit Analysis

Strandberg, Roland

2004

CALLUM Linear Transmitter

Architecture and Circuit Analysis

Roland Strandberg

Lund 2004

LUND INSTITUTE OF TECHNOLOGY
Lund University

Department of Electroscience
Lund University
P.O. Box 118
SE-221 00 LUND
SWEDEN

No. 50
ISSN 1402-8662

Produced using LaTeX Documentation System.
Printed in Sweden by *Tryckeriet i E-huset*, Lund.
December 2004.

Abstract

This doctoral dissertation presents a study of linear radio transmitters based on the combined analog locked loop universal modulator (CALLUM) approach. Linear architectures such as CALLUM are very attractive for power-efficient operations, since they have no fundamental limitations prohibiting a 100 % efficiency for all envelope levels, without sacrificing the potential of a linear transmission. This issue is becoming increasingly important in modern communication standards, where the information content is present not only in the signal phase, but in its amplitude as well. Such modulation schemes, while improving the data rate for a given signal bandwidth, pose tough demands on the linearity of the transmitter. As the power amplifier (PA) in the transmitter handles the largest signals and is the main power consumer in the radio, sufficient linearity should be achieved in conjunction with a high power efficiency of the PA, especially if the equipment is battery operated, as is the in case of mobile applications.

In this work, three different CALLUM architectures (i.e., CALLUM 1, CALLUM 1lin, and CALLUM 2) are studied in terms of loop gain, bandwidth, stability, and frequency compensation. A simplified baseband model of a general CALLUM is presented for efficient simulation of the system, and to gain knowledge about performance differences between the CALLUM derivatives. The investigated CALLUM architectures make use of Cartesian feedback, as it provides better matching between the I and Q signal paths than polar feedback.

From baseband simulations with different signal component generator (SCG) implementations, the spectral performance of CALLUM 1 and CALLUM 1lin for an EDGE modulated signal is significantly better than that of CALLUM 2, for a given maximal loop gain. It can be concluded that the radical simplifications leading to CALLUM 2 have severe effects on the spectral properties of the output signal, while the actual implementation of the SCG becomes much simpler than for example CALLUM 1. The effect of propagation delay in the feedback loop is also included in the model, and indeed this delay appears to be the limiting factor in the achievable closed-loop signal bandwidth. A lag-lead frequency compensation network is used to trade bandwidth for increased insensitivity to time delays. The frequency compensation is quite efficient for all CALLUM versions studied when they operate on a $3\pi/8$-shifted 8PSK, as acceptable delays may in this case become much larger without jeopardizing the stability. It is worth noting that CALLUM 1lin performs very well in terms of maximum acceptable time delay for a certain standard.

Implementations of the circuits for the CALLUM 2 architecture are presented together with simulation results of the fundamental blocks. A differential analog SCG realizing the control equations for CALLUM 2, and a variable-gain amplifier (VGA) were simulated together with other functional blocks to form a complete baseband-modeled CALLUM 2 transmitter. From the simulated spectral performances, the SCG and VGA implementations proved to be appropriate for an EDGE-modulated signal.

Contents

Preface

This thesis summarizes my academic work in the Analog and RF Design group at the department of Electroscience, Lund University, for the Ph.D. degree in Applied Electronics. The content of the thesis is divided into two parts, where the first part starts with a wide approach to linear amplification, with focus on linear transmitter architectures and CALLUM in particular. The second part contains five research papers. A more detailed outline of the thesis is as follows.

Chapter 1 motivates the need for linear and power efficient amplifiers, and takes a historical glance at electronic communications. Interesting to note is that the distortion of the amplifier has been a limiting factor from the very beginning. The fundamental relation on information capacity given by Shannon is discussed briefly, followed by a presentation of the test signals used in this thesis. These are the TETRA, EDGE, and W-CDMA standards, each an example of narrow, normal, and wide bandwidth, respectively.

Chapter 2 deals with various techniques to achieve linear amplification. Linear power amplifiers are tricky to build as they need to be not only linear but also power efficient. Often a linear transmitter with reasonable efficiency is realized by applying linearization to a power efficient but nonlinear power amplifier. The different techniques to achieve linear amplification are divided into two groups: 1) includes methods where the power amplifier is fed with a signal having a varying amplitude, 2) schemes where the power amplifier is fed with a constant-envelope phase-modulated signal. This second group is known as linear transmitter architecture, and special attention is devoted to the LINC and the CALLUM concept.

Chapter 3 discusses the linear amplification with nonlinear components (LINC) transmitter architecture. Both the LINC and the CALLUM concept is promising as no fundamental limitation prevents 100 % efficiency for all signal levels. The signal decomposition, that is performed by the signal component separator, into two constant-envelope phase-modulated signals is presented mathematically. The spectral regrowth that appears inside the LINC transmitter is demonstrated for two suitable modulations, i.e., $3\pi/8$-shifted 8PSK and $\pi/4$-shifted QPSK. The LINC concept promises high efficiency, but is achievable only with ideal power amplifiers and signal combiner. The efficiency issue is discussed for a number of possible combiners.

Chapter 4 covers the combined analog locked loop universal modulator (CALLUM) transmitter architecture, which in its nature is closely related to LINC, as they share the fundamental idea with the signal decomposition into two constant-envelope phase-modulated signals for highly efficient amplification, without sacrificing the potential for linear transfer. The control equations implemented

in the signal component generator for the most common CALLUM versions are derived. A baseband model is presented for efficient simulation of the system, and it also opened up for calculation of the loop gain based on linearized control equations. The loop gain expressions are presented in 3D contour plots for CALLUM 1, CALLUM 1lin, and CALLUM 2, respectively. The information available in the loop gain plots predicts the performance of the CALLUM derivative, and also supports the investigation of system stability and the related design of frequency compensation. This part of the chapter overlaps Paper III and Paper V to some extent, and is reprinted here to make the design flow easier to follow.

In the absence of loop time delay and when maximum bandwidth is desired, the frequency compensation best suited for the job is the phantom zero. However, the time delay will indeed be a limiting factor of the achievable closed-loop bandwidth, and the lag-lead compensation is used to trade bandwidth for increased robustness against time delay. System simulations, based on the baseband model, are presented to differentiate the expected performance from three versions, i.e., CALLUM 1, CALLUM 1lin, and CALLUM 2. From these simulations it is clear that CALLUM 2 has a performance handicap compared to the other two, but it proved to be adequate for the modulation used in EDGE. This chapter is closed by an overarching comparison of the CALLUM derivatives.

Chapter 5 contains implementation aspects of the CALLUM transmitter architecture on a circuit description level. Many of the circuits that together constitute the CALLUM transmitter are presented down to transistor level together with key simulation results. A differential analog SCG implementing the control equations for CALLUM 2 is presented, which together with a variable-gain amplifier (VGA) is an important part of the signal processing of the control signals to the VCOs. The oscillators have a special layout of the symmetrical inductor for reduced mutual coupling between the VCOs. The design also covers the implementation of a phase-locked loop based on a phase-frequency detector for synchronization of the free-running frequency of the VCOs to the reference LO signal. A downconversion part is needed in the CALLUM architecture, and is here chosen as a direct downconversion topology with passive CMOS mixers for low 1/f-noise.

Chapter 6 manifests the conclusions drawn from the previous chapters, and it also contains brief summaries and assessments of the contributions by each included paper.

Appendix A summarizes the implementation of a CALLUM 2 transmitter architecture on silicon.

Included papers: The main contribution to the thesis is derived from the following publications:

R. Strandberg and J. Yuan, "Analysis and implementation of a semi-integrated buck converter with static feedback control," *Midwest Symposium on Circuits and Systems*, vol. 2, pp. 934–937, 8–11 Aug. 2000.

R. Strandberg, P. Andreani, and L. Sundström, "Bandwidth considerations for a CALLUM transmitter architecture," *IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 4, pp. 25–28, 26–29 May 2002.

R. Strandberg, P. Andreani, and L. Sundström, "Spectrum emission considerations for baseband-modeled CALLUM architectures," *IEEE Transactions on Microwave Theory and Techniques*, Accepted for publication.

R. Strandberg, P. Andreani, and L. Sundström, "Implementation of the signal component generator of a CALLUM 2 transmitter architecture in CMOS technology," *Norchip Conference*, pp. 183–186, 8–9 Nov. 2004.

R. Strandberg and J. Piper, "Analytical expression of the efficiency of phantom zero compensation applied on negative-feedback amplifiers," *Norchip Conference*, pp. 87–90, 8–9 Nov. 2004.

The author's interest for feedback systems has resulted in development of a course, *Advanced Analog Design*. The structured design flow for negative-feedback amplifiers was expanded and applied on the CALLUM architecture. Substrate noise and low-noise digital logic has also been studied as part of the work at the department. The following articles and documents contain both overlapping and complementary material, but are not considered as part of this thesis.

R. Strandberg and J. Yuan, "Single input current-sensing differential logic (SCSDL)," *Norchip Conference*, pp. 328–333, 8–9 Nov. 1999.

R. Strandberg and J. Yuan, "Single input current-sensing differential logic (SCSDL)," *IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 1, pp. 764–767, 28–31 May 2000.

J. Piper, R. Strandberg, and F. Tillman, *Advanced Analog Design – Lecture Notes*. Department of Electroscience, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden, 2004.

J. Piper, R. Strandberg, and F. Tillman, *Advanced Analog Design – Exercises, Laboratory manual*. Department of Electroscience, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden, 2004.

Acknowledgments

During my time as a graduate student I have met many people that contributed in one way or another to this thesis, but to mention them all here is impractical and I restrict myself to those that have had substantial impact on this work.

I have been fortunate to be part of the Circuit Design group at the Dept. of Electroscience, Lund University. First, I would like to express my gratitude to my supervisors, Prof. Pietro Andreani and Adjunct Prof. Lars Sundström, for introducing me to the very interesting field of power amplification, in particular to linear ditto. When I was an undergraduate student, Prof. Andreani helped me take the first steps in analog integrated circuit design and he also introduced me to Prof. Jiren Yuan, who gave me the opportunity to start at the Dept. of Applied Electronics (now Dept. of Electroscience), Lund University, as a Ph.D. student in the Mixed Signal group. Prof. Sundström has encouraged me to really enjoy theory, research, and academia. He is also a *gold mine* when it comes down to specifications of radio communication systems, solving problems with RF measurements, Matlab programming, etc. After every discussion with Prof. Sundström I have something new to think about – indeed inspiring. I am also indebted to Dr. Thomas Mattsson for putting the right question at the right time during the analog meetings, accentuating weak points of the design, and for sharing some of his knowledge in VCO design.

I am grateful for the financial support of this project from the EXSITE program (Explorative System-Integrated Technologies), and the CCCD (Competence Center for Circuit Design), which both have funding provided by VINNOVA (Swedish Agency for Innovation Systems).

I would like to thank the people in the Analog and RF Design group for interesting discussions on transceivers. In particular, Niklas Troedsson for his assistance on the design of the on-chip inductors of the VCO. Fredrik Tillman for being my lunch mate, and for helping me sort my thoughts on matters in the daily life. Markus Törmänen for sharing my deep interest for hunting and target shooting. Martin Anderson, Martin Lantz, Gang Xu and Johan Piper for being good friends and my next door neighbors at the office.

Thanks also to the faculty, technical staff and administration at the department. In particular, to Erik Jonsson for his excellent UNIX support and for being a good friend. A special thanks goes to Agneta Ahlberg, Pia Bruhn, Stina Ahlenius, Britta Olsson, Lars Olsson, Bertil Larsson, Stefan Molund, Mats Ågren and Lars Hedenstjerna.

I will not forget my dear colleagues at the department (both present and former) for supporting and encouraging me in the daily work and for all interesting discussions in the coffee room on various topics. Finally, I would like to acknowledge all the support and love from my family and my friends.

Roland Strandberg

List of Acronyms

3GPP	3rd Generation Partnership Project
AC	Alternating Current
ACLR	Adjacent Channel Leakage power Ratio
ADC	Analog-to-Digital Converter
AHDL	Analog Hardware Description Language
AM	Amplitude Modulation
AS	Anti-Series
ASIC	Application-Specific Integrated Circuit
BJT	Bipolar Junction Transistor
CC	Common Collector
CD	Common Drain
CE	Common Emitter
CP	Characteristic Polynomial
CS	Common Source
CMOS	Complementary Metal Oxide Semiconductor
CALLUM	Combined Analog Locked Loop Universal Modulator
DC	Direct Current
DSP	Digital Signal Processor
EER	Envelope Elimination and Restoration
EDGE	Enhanced Data rates for GSM Evolution
ETSI	European Telecommunications Standardization Institute
EVM	Error Vector Magnitude
FET	Field Effect Transistor
FFT	Fast Fourier Transform
FM	Frequency Modulation
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HDL	Hardware Description Language

I	In-phase component
IP	Intellectual Property
ISI	Inter-Symbol-Interference
kbps	kilo bits per second
LINC	LInear amplification using Nonlinear Components
LNA	Low-Noise Amplifier
LP	Loop gain Poles
LUT	Look-Up Table
Mbps	Mega bits per second
MFM	Maximum Flat Magnitude
NFB	Negative Feedback
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PCB	Printed Circuit Board
PM	Phase Modulation
Q	Quality factor, or Quadrature component
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RF	Radio Frequency
RRC	Root Raised Cosine
SCG	Signal Component Generator, CALLUM
SCS	Signal Component Separator, LINC
SNR	Signal-to-Noise Ratio
THD	Total Harmonic Distortion
TETRA	Terrestrial Trunked Radio
UMTS	Universal Mobile Telecommunications System
VCO	Voltage Controlled Oscillator
VHDL	Very High-level Design Language
VLL	Vector Locked Loop
W-CDMA	Wideband Code Division Multiple Access

List of Symbols

$A\beta$	loop gain
A_t	transmittance: source-load relation
$A_{t\infty}$	asymptotic gain: source-load relation when $A\beta \to \infty$
B	(s^{-1}) system bandwidth
C_{ox}	oxide capacitance per area unity
e	marks electronic quantity (voltage, current or power)
e_ℓ	load signal (voltage or current)
e_s	source signal (voltage or current)
HD_n	n^{th} order harmonic distortion
f_c	(s^{-1}) center frequency
I_{KF}	(A) high current corner for forward beta, BJT
I_Q	(A) bias current
I_{S}	(A) saturation current, BJT
i_i	(A) incremental input signal current
i_o	(A) incremental output signal current
$\Im$	Imaginary part
IM_n	n^{th} order intermodulation distortion
j	imaginary unit $(j^2 = -1)$
k	$= 1.380658 \times 10^{-23}$ J/K, Boltzmann's constant
L	(m) channel length, FET
m_d	modulation depth
q	$= 1.60217733 \times 10^{-19}$ C, electronic charge
$\Re$	Real part
s	complex angular frequency, or Laplace transform variable
t_{ox}	(m) gate-oxide thickness, FET
V_{AF}	(V) forward Early voltage, BJT
V_{AR}	(V) reverse Early voltage, BJT
V_{BE}	(V) static base-emitter voltage

V_{CE}	(V) static collector-emitter voltage
V_{DS}	(V) static drain-source voltage
V_{GS}	(V) static gate-source voltage
V_{SB}	(V) static source-bulk voltage
V_T	=kT/q (V) thermal voltage, BJT
V_{th}	(V) threshold voltage, FET
V_Q	(V) bias voltage
v_i	(V) incremental input signal voltage
v_o	(V) incremental output signal voltage
W	(m) channel width, FET
β	feedback network parameter
β_F	forward current gain, BJT
γ	body effect coefficient, FET
ϵ_0	$= 8.854 \times 10^{-12}$ As/Vm, permittivity of vacuum
η	efficiency
λ	(V^{-1}) channel-length modulation parameter, FET
μ	(m^2/Vs) mobility of charge carrier
μ_0	$= 12.56637 \times 10^{-7}$ Vs/Am, permeability of vacuum
ν	output circuit loading network parameter
ξ	input circuit loading network parameter
ρ	direct feed-through network parameter
τ_F	(s) forward transit time, BJT

Transmitters

Linear Transmitters

Chapter 1

Introduction

1.1 Motivation

Modern communications standards rely on linear transceivers, where the designers building the power amplifiers have a demanding task to solve. As the acceptable error vector magnitude (EVM) for recently adopted modulation schemes is decreased compared to the acceptable EVM for GSM (Global System for Mobile communications), high priority is to find a power amplifier which fulfills both the power efficiency and the linearity specification.

The CALLUM architecture promises high efficiency without compromising on the possibility to achieve linear transfer. Yet this architecture is not as thoroughly investigated in terms of transmit spectrum, bandwidth and stability as other established linearization techniques, e.g., feedback, feedforward, predistortion. For proper selection of CALLUM version, architectural level performance differences in terms of spectrum emission and bandwidth are requested. An appropriate selection of CALLUM derivative depends on the standard targeted, e.g., TETRA, EDGE, NADC, W-CDMA, and by directly picking the appropriate CALLUM version the design process will be less time consuming.

The CALLUM derivatives have different spectral and bandwidth properties, and important parameters as loop gain, stability, frequency compensation and robustness against time delay should be investigated. A design approach is needed for the CALLUM system that connects to the design of negative-feedback amplifiers and phase-locked loops. The loop gain is the dominant factor for distortion reduction and shapes the spectrum emission from the transmitter. Equally important is a balanced frequency compensation, which is a prerequisite for system performance optimization in terms of bandwidth and stability, with or without presence of loop time delay. Guidelines for the designer help to shape the spectral performance of the transmitter according to the spectrum emission mask for the communications standard.

Previous implementations of the CALLUM architecture are base on discrete building blocks, whereas the aim in this work is a high degree of integration for reduced

loop delay, and to keep the number of low frequency loop poles to a minimum for improved performance in terms of distortion suppression and bandwidth capability.

1.2 Communications in the Mirror

Much of the work done in the field of electronics is due to the man's will to communicate. The very beginning of electronic communications started in 1825, as the British inventor William Sturgeon exhibited a device that laid the foundations for large-scale electronic communications: the electromagnet. A few years later, in 1830, an American, Joseph Henry, demonstrated the potential of Sturgeon's device for long distance communication by sending an electronic current over one mile of wire to activate an electromagnet which caused a bell to strike. Thus the electric telegraph was born. Samuel F.B. Morse successfully exploited Henry's invention commercially. While a professor of arts and design at New York University in 1835, Morse proved that signals could be transmitted by wire. He used pulses of current to deflect an electromagnet, which moved a marker to produce written codes on a strip of paper – the invention of the Morse Code.

Until 1877, all rapid long-distance communication depended upon the telegraph. That year, a rival technology developed that would again change the face of communication – the telephone. Two inventors Elisha Gray and Alexander Graham Bell both independently designed devices that could transmit speech electrically (the telephone). Both men rushed their respective designs to the patent office within hours of each other, Alexander Graham Bell patented his telephone first.

The telegraph and telephone are both wire-based electrical systems, and Alexander Graham Bell's success with the telephone came as a direct result of his attempts to improve the telegraph. The telegraph was basically limited to receiving and sending one message at a time. Bell's extensive knowledge of the nature of sound enabled him to conjecture the possibility of transmitting multiple messages over the same wire at the same time. His harmonic telegraph was based on the principle that several notes could be sent simultaneously along the same wire if the notes or signals differed in pitch, which today is a well defined and used technique[1].

At the turn of the century the telephone network was a passive device, as it had been since Bell's invention. Carbon microphone added energy from a battery to the weak acoustic signal from a speaker's voice, but once the wave entered the line it traveled to the receiver without further amplification. The telephone network reached its limits of transmission regarding distance due to the attenuation of the wire. By adding inductance at intervals along the wire, loading coils could decrease signal loss by a factor of three or four, and thus increase the maximum transmission distance proportionally. The long-distance transmission lines suffered from attenuation and distortion such the voice signal was barely understandable.

By adding repeaters as the solution to the problem of long-distance transmission the telephone network became an active network. By 1913 the vacuum tube was used

[1] *http://inventors.about.com*

to amplify signals in telephone repeaters, which made intercontinental communication possible. To meet the great demand, as telephony was becoming every man's property, much attention was directed at putting more conversation onto a single line. Modulation of several voice signals onto a high-frequency carrier was the most promising method. This new technique demanded linear amplifiers, otherwise intermodulation products would reduce the signal-to-noise ratio of the adjacent channels. The amount of amplification can be quite substantial and since each amplifier adds a little distortion, the speech can be garbled beyond recognition. The telephone network was at that point in the mid 1920's limited by the distortion of the amplifiers, and an extensive search for linear ditto started. Harold S. Black came up with two elegant solutions to the problem, the feedforward and the feedback technique (see also chapter 2).

At that time electricity in the wires became merely a carrier of messages, not a source of power, and hence opened the door to new ways of thinking about communications. The generalized system is capable of carrying any signal as a new currency: *information*. Today voice is a small but still very important part of the total amount of information communicated, and the later sophisticated communications systems put even higher demand on the linearity of the transmitter.

1.3 Electronic Signal Processing

The ability of an electronic system to transfer information per time unit is characterized by its channel capacity, C. Since the system is reciprocal, the word transferred can be exchanged for the word received. An upper bound for the capacity of a continuous channel was formulated by Shannon [1],

$$C = B \log \left(1 + \frac{S}{N} \right), \tag{1}$$

where B is the system bandwidth, S is the average signal power, and N is the average noise power.

It is the limited power handling capability, the limited bandwidth, and the noise contribution of the implementation of any signal processing function, that restricts its information transfer ability. The ratio between the desired signal power and the undesired noise power is referred to as the signal-to-noise ratio (SNR). From (1) it is clear that it is more favorable to increase the system bandwidth than the SNR to enhance the channel capacity. However, the available bandwidth is a scarce resource, and wise utilization is of major importance. For radio communication systems the bandwidth is in most cases limited to a narrow bandwidth and the demand on the channel capacity has to be reached by improving the SNR. The signal power is confined by the maximum power level that can be handled, and the noise ultimately by the thermal noise floor.

The noise power in (1) is the undesired fluctuations added to the information signal. The term noise in circuit theory is limited to noise stemming from stochastic

processes, which here will be called random noise. The sources of undesired fluctuation covered by the information theory also includes the unwanted signal dependent noise we normally refer to as distortion.

1.4 Signals

Signals are fluctuations on a quantity, and in electronics the signals are fluctuations of the energy carried by the electrons. As the electron has a fixed charge, q, the signal is discrete by nature, but on a macro scale the signal can be considered as continuous in amplitude and time. The signal fluctuations can be represented in time, frequency (or phase), and magnitude. The signals are used to carry information, which is a generalized dimensionless quantity that represents knowledge. Only those signals that carry information are desired, as other signals (noise) degrade the operation of the signal processing function.

1.4.1 Functions

The information is taken from a source and via a channel transfered to the recipient where the information is consumed. Accordingly, in any electronic information handling system three essential elements can be distinguished: 1) the acquisition of the information by a sensor, 2) the processing and transmission of the signal, and 3) the dissipation of the information in a transducer [2].

Amplification and filtering are among the most important functions in the electronic signal processing system. Amplification was in the early telephone network used to counteract the attenuation of the signal as it propagated through the wires, see section 1.2. As the first repeaters were included in the system the attenuation was no longer the main problem, rather distortion and random noise. The distortion depends on the signal magnitude and becomes increasingly troublesome at high signal power levels. These unwanted effects are part of the implementation of the amplifier, and the performance of the electronic signal processing function is limited by these errors [3].

1.5 Communication Standards

Every communication standard has its unique properties, and for any feedback system the channel bandwidth can be a true challenge to handle. Three different standards are presented in the next sections, each one a representative for narrow (TETRA), normal (EDGE) and wide (W-CDMA) bandwidth, respectively. Modulated signals for TETRA, EDGE and W-CDMA standard will be used as test signals in simulations.

1.5.1 Terrestrial Trunked Radio, TETRA

TETRA is a set of standards developed by the European Telecommunications Standardization Institute (ETSI) that describe a common mobile radio communication infrastructure throughout Europe. This infrastructure is targeted primarily at the mobile

radio needs of public safety groups (such as police and fire departments), utility companies, and other enterprises that provide voice and data communications services.

The people expects prompt assistance in emergency situations, and for the emergency services to be able to work efficiently together, a high performance radio network is a prerequisite. The digital radio network provides a number of advantages, e.g. all air-borne messages are encrypted, the network will appear to each agent (police, fire department, etc.), as if they have their own virtual private network. The almost instantaneous call set up times, enables the press to talk (PTT) capability. The PTT function allows simple simultaneous communications between all members of a working group, a bit like an old analog radio system, except with greater range. The TETRA standard features call set up times of less than half a second. By comparison, GSM has a set up time of over seven seconds.

When setting up a connection between two mobile phones (terminals, handsets) the normal mode is to use a basestation as repeater. TETRA also incorporates a direct mode that allows handsets to communicate directly with each other without the use of a repeater, just like walkie-talkies. An extension of direct mode operation (DMO) allows a standard mobile radio to act as a repeater, passing transmissions from one radio to another until reaching a basestation. This feature provides a temporary network extension into areas where there might not otherwise be coverage, e.g., inside buildings, basements and parking garages.

TETRA uses a technique called time division multiple access (TDMA) to squeeze four users into a single 25 kHz radio channel. In a TETRA system each user is assigned one of four timeslots (each timeslot is 14.167 ms). The TETRA standard defines data transfer rates in the range of 2.4 kbits/s to 28.8 kbits/s depending on the number of available timeslots, the security encryption level and the amount of error correction required. For maximum data rate all four timeslots are used, very much like in the general packet radio service (GPRS) in the GSM net.

Since the radio can usually only transmit during one timeslot, it can spend the remainder of the time receiving. By switching back and forth between transmitting and receiving 18 times every second, a TETRA radio user has the ability to talk and listen at the same time, just like a normal telephone call. This simultaneous communication is referred to as full-duplex. Most analog and even some digital systems limit the user to either speaking or listening at any particular time (called half-duplex).

The TETRA standard offers high spectral efficiency, and the carrier spacing is only 25 kHz (also the channel bandwidth). In Europe the frequency band 380 – 400 MHz is reserved for public safety and security use. The 410 – 430 MHz band in turn is reserved for commercial professional mobile radio use. Outside Europe, the 800 MHz band has become the dominant frequency band for TETRA. Several different transmitter power classes exist for the basestation: {0.6, 1, 1.6, 2.5, 4, 6.3, 10, 15, 25, 40} W, and for the mobile station: {1, 3, 10, 30} W.

The modulation in TETRA is a $\pi/4$-DQPSK. Differentially encoded quadrature phase shift keying (DQPSK) and $\pi/4$-shifted-DQPSK modulation are examples of modulation schemes taking advantage of a number of non-coherent detection schemes.

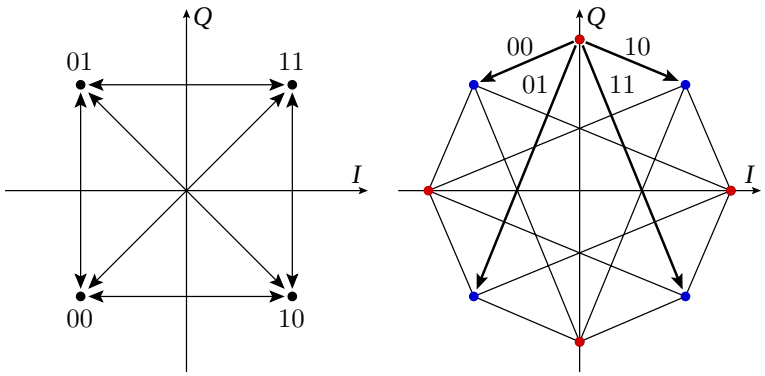

Figure 1: The constellation diagram of QPSK (left) and $\pi/4$-DQPSK (right).

Non-coherent detection is suited for a variety of applications. Phase referencing is a formidable task, and thus differential phase encoding and non-coherent detection strategies are of considerable interest. In mobile communication the received signal suffers from random phase fluctuations, due to shadow and multipath fading. This in turn makes coherent demodulation schemes rather unattractive and impractical. This is one of the reasons why ETSI decided to use non-coherent detection in the TETRA-project, and also the current North America Digital Cellular (NADC) standards specify $\pi/4$-shifted-DQPSK modulation for the digital mobile environment due to its high spectral efficiency and for the fact that it can be detected using a variation of the conventional analog FM receiver [4]. The constellation diagram of a QPSK and a $\pi/4$-DQPSK modulation is illustrated in Figure 1. The indicated trajectories will be smeared out when a square-root raise cosine filter is applied [5].

Considering a case when each symbol contains two bits. In a $\pi/4$-DQPSK constellation diagram the four possible symbols are mapped to one and only one of the following phases in the set $\Phi = \{\lambda, \lambda + \pi/2, \lambda + \pi, \lambda + 3\pi/2\}$, where λ is an arbitrary angle offset. Let us assume another set of possible phases $\overline{\Phi} = \{\lambda + \pi/4, \lambda + 3\pi/4, \lambda + 5\pi/4, \lambda + 7\pi/4\}$. The transmitted symbol can be one of totally eight different phases, and two consecutive transmitted symbols alternate between the sets Φ and $\overline{\Phi}$. The information carried by the input signal is transmitted by the phase shift of the radio signal relative the last received symbol.

1.5.2 Enhanced Data Rates for GSM Evolution, EDGE

In the second generation (2G) of the global system for mobile communication the data rate that GSM provides is not high enough for smooth operation of Internet services. The up and running second generation system has been expanded into 2.5G, which is a step in between 2G and 3G. Enhanced Data rates for GSM Evolution (EDGE) is one

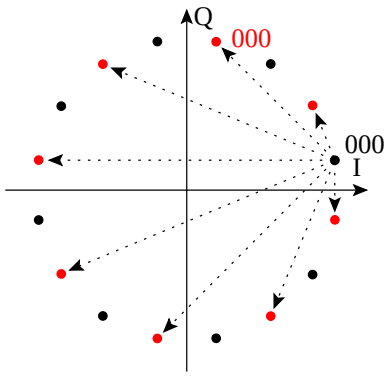

Figure 2: The $3\pi/8$-shifted 8PSK modulation scheme used in EDGE.

example of a 2.5G system.

The modulation used in EDGE is a linearized Gaussian $3\pi/8$-shifted 8PSK when the quality of the radio channel allows it. The eight point phase shift keying (8PSK) constellation diagram is rotated by $3\pi/8$ radians between each symbol, such that no point-to-point transition goes through the origin, as illustrated in Figure 2.

The $3\pi/8$ radians shift of the consellation diagram between each symbol reduces the ratio between the largest and smallest envelope of the transmitted signal. The $3\pi/8$-shifted 8PSK modulation is recognized by the hole in the center of the I-Q diagram, see Figure 3. The less variation in amplitude, the less nonlinearities of the power amplifier are excercised. This also means that a pure phase-modulation can be handled by grossly nonlinear amplifiers without ruining the information.

EDGE offers greater data rate while occupying the same bandwidth as the older GSM format. A significantly better spectrum efficiency (i.e., bits/Hz) is obtained as 3 bits are carried in each symbol. The modulating symbol rate for both EDGE and GSM is $1/T_s = 1625/6$ ksymb/s (i.e., approximately 270.833 ksymb/s) contained in a 200 kHz channel bandwidth. The GSM is specified to handle a data rate of 9.6 kbits/s, whereas EDGE theoretically has three times higher data rate. In EDGE the data points are represented by Dirac pulses, which excite a linear pulse shaping filter. This filter is a linearized gaussian minimum shift keying (GMSK) pulse. More information regarding filtering of the data in EDGE is found at ETSI[2].

In comparison with GSM a significant change is made in the RF amplifier section. With the EDGE signal, the amplifier must faithfully reproduce the constellation at a high RF power level without distorting the signal in any significant manner. The nonlinearities of the reproduced signal are separated into two different figures of merit. The constellation distortions of both magnitude and phase are measured using the error vector magnitude (EVM) method (specified at a 7% limit for an entire basesta-

[2] *www.etsi.org*

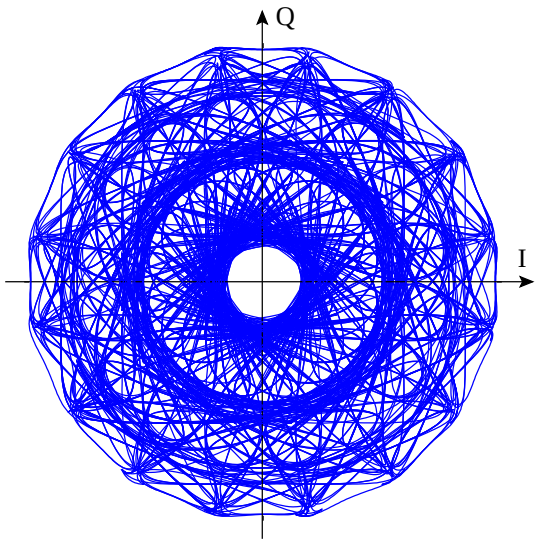

Figure 3: The I-Q diagram for a $3\pi/8$-shifted 8PSK modulated EDGE signal of approximately 1600 symbols.

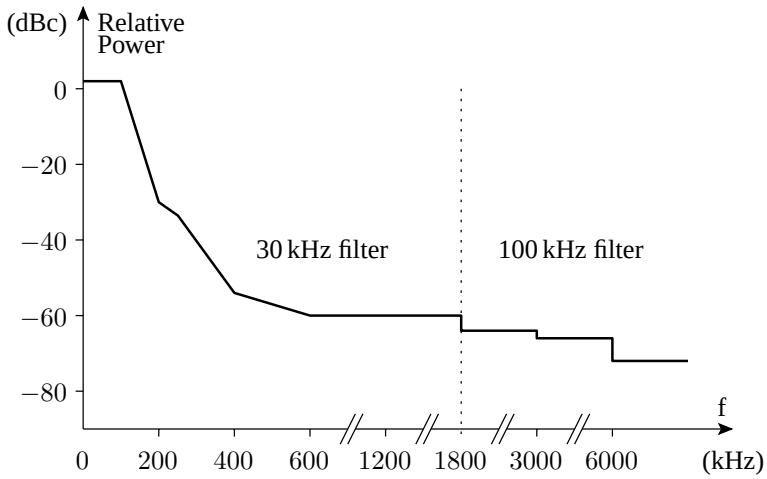

Figure 4: The spectrum mask for a mobile station operating on a modulated signal according to the EDGE standard ($3\pi/8$-shifted 8PSK assumed). The filter and video bandwidth are given in the figure.

tion). The additional spectral energy produced is measured by comparing the power ratio produced in the intended frequency channel to the ratio of undesired power splattered into other frequencies, which is called the adjacent channel leakage power ratio (ACLR). The adjacent channel power is caused by: 1) continuous modulation spectrum and wideband noise, 2) switching transient spectrum (the power ramping up and down). The spectrum emission mask for a mobile station with output power ≤ 33 dBm in the 900 MHz range due to 8PSK modulation is shown in Figure 4. This spectrum mask is the same as for GSM 900 except for the -60 dBc at 400 kHz offset in GSM (GMSK modulation). For further information, see 3GPP[3] TS 05.05.

An advantage for the EDGE system is that it shares the radio equipment and the data core of a GSM/GPRS network. Thereby the cost for this extension is small compared to setting up a completely new system as 3G W-CDMA. An advanced EDGE implementation allows for the multiplexing of EDGE and GPRS services on the same radio timeslots, enabling EDGE capabilities without the need for additional RF spectrum.

1.5.3 Wideband Code-Division Multiple-Access, W-CDMA

The demand for high-speed mobile data communications is growing rapidly. However, the capacity of existing GSM networks is not sufficient to provide the data rates required for Internet access or video streaming. W-CDMA or UMTS, as it is called throughout Europe, is a standard which has been developed to accommodate higher data rates of up to 2 Mbps. Even though W-CDMA is focusing on high data rates it still supports simple features like a plain voice call or sending of SMS. The modulation used in W-CDMA is hybrid phase shift keying (HPSK) with a pulse shaping root raised cosine filter with roll-off, $\alpha = 0.22$. For HPSK, the probability of a zero-crossing is limited to every other chip, and is therefore reduced by half compared to a rectangular QPSK. The probability of 0° phase shift transitions is also half as high. All this improves the peak-to-average power ration of the signal by approximately 1 – 1.5 dB[4]. The chip rate is 3.84 Mchips/s and the channel spacing 5 MHz. W-CDMA is a code division multiple access system, which means that the available frequency channel is broken down by different code sequences that are multiplied by the user signals of the individual subscribers. Multiple subscribers transmit on the same frequency and at the same time. The closest frequencies outside the transmit band, 2.515 – 3.485 MHz, are filtered using a 30 kHz filter. Further out a 1 MHz filter is used in the band 4 – 12 MHz[5].

In the domain of digital communications, the demands for linearity of transmit and receive systems are constantly increasing. The first indications of this were the stronger requirements imposed on power amplifiers and receiver front ends with the introduction of digital modulation schemes in the field of satellite communications.

[3] *www.3gpp.org*

[4] *http://we.home.agilent.com*, application note: AN 1335.

[5] The details about the spectrum emission mask requirements are found in TS 25.101 at *www.3gpp.org*.

The linearity requirements were further increased by such modulations methods as 16QAM, 64QAM or even higher, which cause strong amplitude fluctuations in the transmitted signal. As an additional complication, the permitted mean phase errors and amplitude fluctuations are smaller than with lower-order modulation. In the case of QPSK, the use of offset-QPSK at least solved the problem with linearity partially, as small signal levels are being avoided within the modulation signal.

High Speed Downlink Packet Access, HSDPA

Improvements and enhancements are being made to the W-CDMA telecommunications system. Called high speed downlink packet access (HSDPA) the new technology promises to increase the download data rate five fold. The introduction of this technology has come about as a result of the need to drive down costs as well as increasing the achievable data rates. Current trends show the volume of packet switched data rising and overtaking the more traditional circuit switched traffic. By adopting a packet based approach to the delivery of digital content as well as IP based person to person digitized voice, a single session can be used for multiple purposes and this can be used to improve revenues. With this approach in mind the use of HSDPA is a key element in providing the user with a better service as a result of increased capacity and usage for the service providers.

One of the keys to the operation of HSDPA is the use of an additional form of modulation. Originally W-CDMA used only HPSK as the modulation scheme. However, under the new system 16QAM is also used when the link is sufficiently robust. 16QAM can carry a higher data rate, but is less resilient to noise. The robustness of the channel and its suitability to use 16QAM instead of HPSK is determined by analyzing information about the channel. When using the new scheme it will be possible to achieve peak data rates of 10 Mbps within the 3.84 MHz channel bandwidth offered the W-CDMA signal. Users will gain higher data speeds as well as shorter service response times and better availability of services. However, new mobile designs are needed to handle the increased data throughput rates. Nevertheless, the advantages of HSDPA mean that it will be widely used as networks are upgraded and new phones introduced.

Chapter 2

Linear Amplification Techniques

The use of a linear modulation scheme in a mobile environment requires a linear transmitter and a linear power amplifier in particular. Otherwise the spectral properties of the modulated signal will be deteriorated due to intermodulation distortion (IMD). The IMD causes the spectrum to rise in adjacent channels and can cause interference for users of these neighboring channels. As one of the most power consuming part in a wireless handset, the power amplifier must be efficient not to degrade talk time. A standard solution is to back-off the power amplifier to obtain a high degree of linearity, but it also reduces the efficiency to unacceptable levels in mobile equipment due to the limited battery capacity [6]. In fact, it is well known that linearity and efficiency trade with each other.

To build a linear transmitter with reasonable efficiency we can apply linearization to a power efficient but nonlinear power amplifier. The distortion generated by the power amplifier has to be suppressed/cancelled by an additional circuitry. This group of techniques includes methods where the power amplifier is fed with a signal having a varying amplitude. Thus distortion is generated and it can be cancelled either by isolating the distortion from the information signal and subtract it from the amplifier output (feedforward), or the input signal to the power amplifier is shaped such the overall transfer is linear (predistortion or negative-feedback).

There is a walk-around to the problem by using a signal decomposition that does not exercise the nonlinearities of the power amplifier. These schemes have in common that the power amplifier is fed with a constant-envelope phase-modulated signal, and are known as linear transmitter architectures, e.g., LINC and CALLUM. Another scheme splits the signal into its polar components, and only the phase information is carried by the input signal to the amplifier as in envelope elimination and restoration (EER).

2.1 Overview of Linearization Techniques

When faced with the problem of building a linear power amplifier either the approach will be to linearize the nonlinear amplifier or use a linear architecture. These two groups can further be divided into more specific methods, as illustrated in Figure 5.

Predistortion is conceptually simple, where two complementary nonlinearities are cascaded to get a linear transfer. The nonlinearity has to be known in advance to be

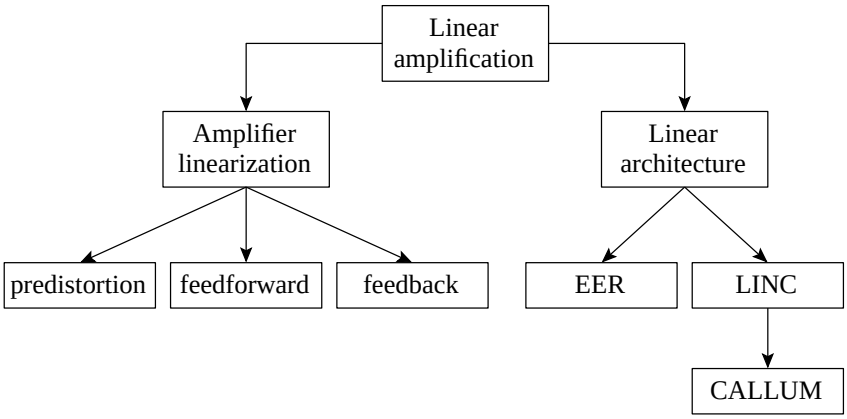

Figure 5: Overview of techniques to accomplish linear power amplification.

able to cancel it, and predistortion can potentially be broadband as no feedback is used. Indeed, there is some kind of feedback for adaptation of the predistorter via an algorithm that can track slow changes of the characteristic such as aging, load variation, temperature and transmit power.

Feedforward is another open-loop solution which relies on successful isolation of the error and requires high degree of phase (delay) and amplitude balance along the parallel signal paths. The bandwidth over which the feedforward provides significant linearity improvements depends in part on the bandwidth over which the group delay of the individual amplifiers may be tracked accurately by realizable time-delay elements [7].

Feedback is probably the best known technique to reduce distortion. As the nonlinearity must not be known in advance it can be applied on almost any system. On the downside are the stability hazard associated with negative-feedback, and the loop time delay that efficiently puts an upper limit on the loop bandwidth. Direct RF feedback is abandoned in favor of modulation feedback, which is the only sensible solution for large systems.

Envelope elimination and restoration is a linear transmitter architecture which uses a modulated intermediate frequency input signal. This signal is split into its polar components, magnitude and phase. The phase-modulated signal is fed to a power amplifier, which has a modulated supply voltage to form the linear transfer. It has the sensitivity to delay mismatch between the two signal paths in common with the feedforward technique. The power efficiency is potentially high as the drain voltage is adjusted continuously to produce the desired output level.

LINC relies on a divide-and-conquer approach, where the modulated signal is decomposed into two constant-envelope phasors, each amplified by a separate amplifier and combined at the summing port to form the output. As LINC is an open-loop

system the performance depends strongly on the degree of matching between the signal paths. Integrated design has pushed the performance, but most designs contain circuitry for adjustment of the gain and phase imbalance.

CALLUM is based on the LINC concept with negative-feedback and is conceptually interesting as it, in theory, promises 100 % efficiency for all input levels. As a matter of fact, the type of power combiner at the output affects the efficiency to a large extent. The linearizing effect of the feedback reduces the sensitivity to gain/phase imbalance, and opens up for use of new types of power combiners with potentially higher efficiency than for the hybrid combiner. More efficient combiners allow PA interaction with load pulling and increased distortion. The amount of loop gain suppresses any distortion component by approximately the same amount. In general, CALLUM shares both good and bad properties of the feedback technique.

2.2 Predistortion

Predistortion is one straightforward solution to achieve linear transfer from the nonlinear power amplifier by using a preceding predistorter. The predistorter contains an inverse of the function describing the transfer of the power amplifier. The result of the cascade is ideally a linear transfer. In Figure 6 is the predistortion transmitter illustrated in its simplest form as an open loop system.

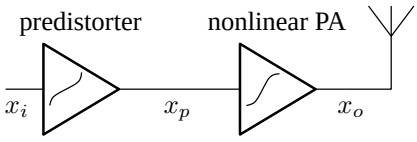

Figure 6: Basic idea of predistortion.

2.2.1 Analog Predistortion

The nonlinear transfer of the amplifier can be expressed by a polynomial of infinite order. However, for calculations more terms than necessary should be avoided due to rapidly increasing complexity. An adequate order is when the truncated polynomial approximates the nonlinear function with a precision such that the residual part alone does not give rise to more distortion than the specification demands, including some design margin. For weak nonlinearities a third order predistorter can be enough to suppress the distortion as wanted. For higher demands on distortion suppression and/or with more nonlinear amplifiers, e.g., when forced into (deep) saturation, the solution is to increase the order of the polynomial. This requires a more advanced adaptation algorithm as more coefficients have to be adjusted. Already by using a 5^{th} order polynomial Westesson et al. [8] reported more than 30 dB and 10 dB suppression of

the IM_3 and IM_5, respectively, of a Mini-Circuits MAR-1 amplifier forced into saturation. The predistorter was implemented in CMOS technology and worked at baseband or IF. The measurements were performed at 200 MHz center frequency and the amplifier was exercised with a two-tone test. Rahkonen et al. achieved similar distortion suppression with a 5^{th} order polynomial implemented in a 0.8 μm BiCMOS process, intended for baseband or IF applications with a modulation bandwidth of 40 MHz [9].

If we let the transfer of the amplifier to be represented by the truncated complex polynomial of third order the output signal of the amplifier can be written as,

$$x_o(x_p) = \alpha_1 x_p + \alpha_3 |x_p|^2 x_p, \tag{2}$$

where α_i is the complex-valued coefficient of i^{th} order. The third order predistorter is illustrated at a functional level in Figure 7, and the output from the predistorter is given by,

$$x_p(x_i) = x_i + (c_{3i} + jc_{3q})|x_i|^2 x_i. \tag{3}$$

Combining (2) and (3) yields,

$$x_o(x_p(x_i)) = \alpha_1 x_i + (\alpha_1(c_{3i} + jc_{3q}) + \alpha_3)|x_i|^2 x_i + o(|x_i|^4 x_i). \tag{4}$$

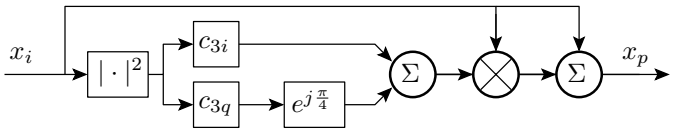

Figure 7: Third order complex-valued predistorter.

Hence, by choosing the coefficients of the predistortion polynomial properly the third order distortion component can be cancelled at the expense of moving nonlinearities to higher order terms. For a properly selected polynomial order these higher order nonlinearities are much smaller than the distortion terms cancelled [8]. As the coefficients of the predistorter power series are complex-valued both AM-to-AM and AM-to-PM distortion are handled. In (4) is a Cartesian representation used for the complex-valued coefficients, but equally efficient would a polar representation be. However, implementation aspects can prove one representation to be more suitable than another.

A low-order polynomial is only capable of cancelling weak nonlinearities. For grossly nonlinear amplifiers more general schemes have been developed that are based on DSP techniques and look-up tables. With a look-up table with enough entries it is possible to represent an almost arbitrary nonlinearity [5].

2.2.2 Digital Predistortion

The area of digital predistortion is too wide to be covered in full. Therefore, only a few examples of possible solutions of digital predistortion are shown here. The digital predistorter is recognized by a digital signal processor (DSP), and a memory look-up table (LUT). The DSP is part of the baseband signal generation and controls the access of the LUT as well as it computes updates of the table words via an adaptation algorithm.

Mapping Predistortion

A brute force solution of the predistorter implementation is to use a very large two-dimensional table. The input signal to the table is represented by its Cartesian components and the output is mapped via the coefficients in the table to generate the Cartesian output of the predistorter. With sufficient resolution of the LUT any distortion or deterministic error can be cancelled, as long as it is memoryless. Even other sources of distortion can be corrected, e.g., misalignment and nonlinearities of the quadrature modulator. The main drawback with this direct mapping is the size of the LUT [10]. With the same performance the number of words in the LUT can be reduced at least by a factor of ten by using the DSP to calculate interpolated values from the words stored in the LUT [11]. In general, the predistortion block uses information about the amplifier nonlinearities stored in the LUT to predistort the baseband signal. The adaptation algorithm adjusts the values in the LUT such the difference between the baseband input and the modulation feedback signal is minimized. The exact implementation depends on the algorithm chosen.

Complex Gain Predistortion

The major drawback with mapping predistortion is the size of the two-dimensional table. In fact, it is not the size in a geometric meaning that is the main obstacle, rather the long adaptation time required to update all these words. However, if the predistorter is restricted to only correct for nonlinearities of the power amplifier, then a complex-valued one-dimensional table will do since the amplifier characteristic is a function of the input amplitude only [5][6]. Complex gain predistortion employs a LUT indexed solely by a function of the input signal envelope and this allows complex gain predistortion to map the signal amplitude related distortion (AM-to-AM and AM-to-PM conversion) of the RF amplifier with much smaller LUT than mapping predistortion [11]. Another important parameter is the adaptation time, which is significantly reduced for the reduced LUT. A complex gain predistorter with complex-valued gain factors given in Cartesian form is illustrated in Figure 8.

[6] Again the amplifier is assumed not to suffer from any memory effects.

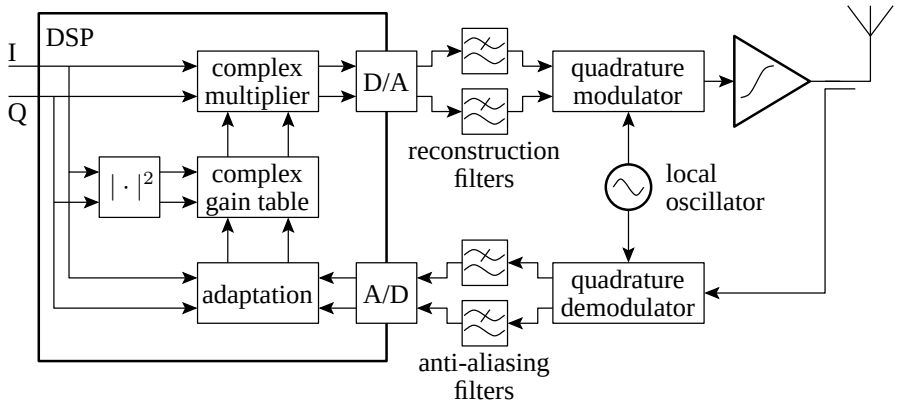

Figure 8: Complex gain table predistorter.

2.2.3 Adaptation of the Predistortion Circuit

In general, adaptation is only required for drifts in the characteristic of the amplifier. These drifts are mainly caused by changes in temperature, supply voltage variations, aging of devices and switching between channels [12]. Most of these are slowly varying processes and thus do not require fast adaptation (or equivalently, large bandwidth of the adaptation circuit). The adaptation algorithm is capable of automatic adjustment of the predistorter without any hand-tuning, making it a more universal technique.

Many linearization techniques, whether Cartesian modulation feedback or predistortion, rely on comparison of the complex-value output of the amplifier with the desired modulated signal. This requires a full feedback path, including a highly linear quadrature demodulator and a delay compensator, which add to the cost and complexity of the circuit. In the strive for more efficient adaptation schemes Stapleton et al. presented in [12] a technique based on iterative adjustment of the predistorter parameters for a 5^{th} order analog complex-valued polynomial to minimize the out-of-band power. The technique operates at IF, with a simple feedback path without any instability problems as for Cartesian modulation feedback. Since Stapleton et al. presented the technique in the early 1990's it has been refined and in [13] experimental results demonstrate an adjacent channel power ratio reduction of 11 dB for W-CDMA, at a carrier frequency of 2140 MHz.

The adaptation of the predistorter is a large and active research field. After chosen a suitable technique (analog or digital) for the predistorter a clever adaptation algorithm can enhance the versatility of the predistorter by reducing the adaptation time to acceptable levels. The adaptation can be made either in a separate training sequence or in parallel with the regular transmission of information.

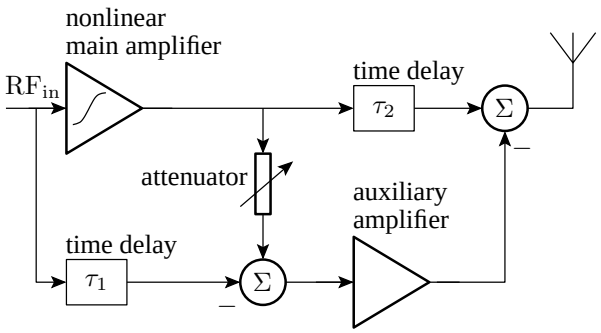

Figure 9: Block schematic of the feedforward system.

2.3 Feedforward

The first scheme H. S. Black invented for reducing amplifier distortion was feedforward [14]. Negative feedback [15] was another technique Black invented (described in section 2.4), but the feedback technique is restricted to relatively narrow bandwidths and is limited by conditional stability and finite distortion suppression as the loop gain is finite. However, the feedforward technique is unconditionally stable[7] and can, in theory, completely eliminate the distortion [5]. In particular, feedforward has been successfully applied to a number of RF amplifier linearization problems and has provided linearization over many tens of megahertz at 900 MHz center frequency [17]. The feedforward technique has been widely used in linearization of power amplifiers in wireless communication systems, e.g, for highly linear amplifiers for multichannel basestation transmitters.

In Figure 9 the block diagram of a typical feedforward system is shown. A two-way power divider splits the RF input into two paths; one signal goes to the main amplifier and the other goes through a delay line, τ_1, compensating for the delay of the main amplifier. The delayed and undistorted input signal is compared with a properly attenuated sample of the main signal in a 180° combiner. This type of combiner works as a subtraction, giving zero output if the main amplifier has no amplitude or phase distortion. If there is any amplitude or phase distortion in the main amplifier, the combiner produces a nonzero error signal.

The auxiliary amplifier has a power handling capacity about ten times less than of the main amplifier. The auxiliary amplifier brings this error signal back to the original level at which the output signal was taken and is recombined with the output after the main output has been delayed, τ_2, compensating for the delay in the auxiliary

[7] Actually Kenington et al. analyzed the possibility of instability in a feedforward loop in [16]. Instability is possible either in the error-detecting loop or the error-cancellation loop. Large loop gain in combination with insufficient reverse isolation of the power combiner is the main cause for oscillation in the feedforward system.

amplifier. The combiner at the output of the feedforward system can be a directional coupler. Although, equal amplitude and opposite phase are sufficient for cancellation at a single frequency, perfect broadband cancellation occurs only when the signals have equal amplitude, 180° phase difference and equal delay [18].

This cancellation is achieved without the possibility of instability, which would be associated with high-gain feedback systems, and without their bandwidth constraints [19]. To obtain cancellation of the nonlinearities the delays, τ_1 and τ_2, and the attenuator of the feedforward system are the parameters to be adjusted. In basestations using feedforward, bulky and expensive coaxial cables in the order of 20 feet in length are utilized to provide about 25 ns of time delay. Not only does this take a lot of room, it also adds to the cost. This is a major handicap towards the development of miniature basestations using feedforward linear power amplifiers [18]. The demands on the delay elements are high and to achieve at least 25 dB of distortion cancellation, as would be required in mobile radio applications, an amplitude error better than 0.5 dB and a phase error better than 0.5° would need to be achieved [19]. For high frequency applications it is evident that the performance of this scheme is not only dependent on the amplitude match but also on the phase/delay match along the parallel signal arms to obtain perfect signal and distortion cancellation, respectively. In practice fixed delays can be used, but still they need to have the right value. Therefore, a lot of work has been done in the field of adjustable delay elements, and both theoretical and practical studies of delay and amplitude imbalances [18–21].

2.4 Feedback

One of the most basic and important functions in any radio communication system (or processing of information in general) is amplification. Associated with the implementation of the mathematical function, numerous parasitics of the amplifier appear, e.g., power consumption, noise, and distortion. In the beginning of the twentieth century the telephone industry tried to extend the communication distances and amplifiers were needed to compensate for transmission line attenuation. The problem was not insufficient amplification, rather the problem was distortion. The main solution at the time was to guarantee *small-signal* operation of the amplifier, which today can be thought of as exaggerated *back-off*. The valid input amplitude was limited to only a fraction of the available dynamic range, thereby a more linear operation could be achieved. The major drawbacks with this solution are its (very) low efficiency and excessive heat dissipation as 100 W power amplifier processes milliwatt signals [7].

Back in the nineteen-twenties the graduate of Worcester Polytechnic named Harold S. Black struggled to improve the Bell System's new open-wire telephone system. Also for this system to operate properly the major obstacle was distortion, which Black solved after much time of figuring and persistent search the idea of negative-feedback (NFB) amplifier came to him in a flash [15]. He realized that the nonlinearity of such amplifiers would be suppressed by the amount of loop gain applied. Ever since has the NFB technique been applied to build high-performance amplifiers [22], and the

technique has proven to be one of the most frequently used in other control systems. The potential risk for instability can be controlled with correct frequency compensation, and full understanding of frequency compensation and available compensation strategies is of utmost importance when dealing with any feedback system. Hendrik W. Bode's and Harry Nyquist's development of mathematical methods for designing the frequency behavior to specified tolerances led to the precise, stable, reliable circuitry that made the amplifier a precision component. The properties of negative feedback are treated thoroughly in many textbooks of which the following are strongly recommended [22–25].

2.4.1 The Asymptotic-Gain Model

Black's feedback model is one of the best known, and it describes the idea of applying feedback in a simple way. This model treats signals at a signal processing level, which is quite far from the electrical system. The asymptotic-gain model more accurately represents the implementation of the electronic signal processing function that the system constitutes [3]. This is why the asymptotic-gain model is used instead of Black's feedback model.

An electronic signal can be represented either in voltage or current. The interaction between voltage and current is known as impedance (or its inverse, admittance). A model that represents an electronic implementation of a feedback system should also cover the interaction between voltage and current. Examples of important features the model has to cover are input circuit loading, output circuit loading, and direct feedthrough effects. All these effects are undesirable, but cannot be ignored for correct treatment of the circuit. The model that covers these different effects are shown in Figure 10.

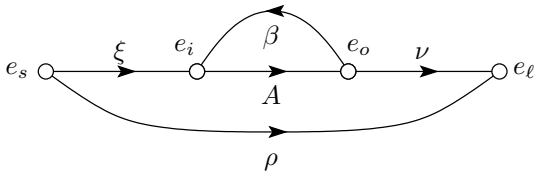

Figure 10: Signal-flow graph describing the asymptotic-gain model.

The transmission matrix of Figure 10 is an alternative way of representing the signal-flow,

$$\begin{bmatrix} \xi & \beta \\ \rho & \nu \end{bmatrix} \begin{bmatrix} e_s \\ e_o \end{bmatrix} = \begin{bmatrix} e_i \\ e_\ell \end{bmatrix}, \qquad e_o = Ae_i. \tag{5}$$

The amplification is represented by the parameter A, and is called the reference variable. The four network parameters $\xi,\ \nu,\ \beta$, and ρ correspond to the input and output

circuit loading, feedback, and direct feed-through, respectively. The network parameters are calculated according to,

$$\beta = \left.\frac{e_i}{e_o}\right|_{e_s=0}, \qquad \xi = \left.\frac{e_i}{e_s}\right|_{e_o=0}, \qquad \nu = \left.\frac{e_\ell}{e_o}\right|_{e_s=0}, \qquad \rho = \left.\frac{e_\ell}{e_s}\right|_{e_o=0}. \quad (6)$$

Frequently a direct transfer exists between the input and output caused by various kinds of parasitic coupling. This is modeled via the direct transfer factor ρ, i.e. the gain when the loop gain is zero.

From Figure 10 it can be derived that the closed-loop transfer from the source to the load becomes,

$$A_t = \frac{e_\ell}{e_s} = \rho + \xi\nu\frac{A}{1 - A\beta}, \quad (7)$$

where the term $A\beta$ is called the loop gain. This product has major influence on the distortion, the stability, and the bandwidth properties of the feedback system. The relation $1 - A\beta$ frequently appears when the properties of a NFB amplifier are investigated. Bode gave this relation a name of its own: the return difference, $F = 1 - A\beta$. When the magnitude of the loop gain is high it is very close to the magnitude of the return difference [3].

The closed-loop transfer A_t approaches the asymptotic gain as the magnitude of the loop gain goes to infinity,

$$A_{t\infty} = \lim_{|A\beta|\to\infty} A_t = \rho - \frac{\xi\nu}{\beta}. \quad (8)$$

The asymptotic gain is the desired transfer and it can be identified in the transmittance equation (7). The transmittance is given by,

$$A_t = A_{t\infty}\frac{-A\beta}{1 - A\beta} + \frac{\rho}{1 - A\beta}. \quad (9)$$

These results are found from the asymptotic-gain model, but in practical amplifiers the direct feed-through effect should be made negligible, i.e. $|\rho| \ll |\xi\nu A/(1 - A\beta)|$, and is simply ignored. The transmittance in (9) can be simplified to the form which is very well known,

$$A_t = A_{t\infty}\frac{-A\beta}{1 - A\beta}. \quad (10)$$

The design of a negative feedback amplifier is, with the help of the asymptotic-gain model, reduced into two successive steps. Initially the feedback network is designed to achieve the desired transfer $A_{t\infty}$ followed by implementation of the active part to achieve sufficiently large loop gain $A\beta$ [3]. The analogy for other types of feedback systems is striking, and these guidelines can be used there as well.

2.4.2 Direct and Modulation Feedback

The standard type of feedback in NFB amplifier design is *direct* feedback, meaning there is no frequency translation within the loop. The direct feedback is the simplest and most straightforward type of feedback, and it works accurately for frequencies where the magnitude of the loop gain is sufficiently high, normally up to a few decades below the cut-off frequency, f_T, of the transistor. As the gain of the active device drops off at higher frequencies and the poles of the system potentially giving stability problems the direct feedback loses its advantage. By taking the system to its frequency limit the loop gain is zero and the system works as an open-loop system without any distortion reduction. Application of direct feedback in a larger systems with numerous blocks within the loop is very often hampered by its limited bandwidth due to the loop time delay, which directly degrades the phase of the system.

An efficient RF power amplifier is more or less tuned to the carrier frequency and the harmonic distortion that appears at multiples of the carrier frequency is assumed to be sufficiently suppressed by the tuned circuit, otherwise external filters can be used to suppress distortion outside the transmit band. Only the intermodulation distortion is important and this is a function of the modulation signal solely. Therefore, it is possible to apply modulation feedback instead of direct feedback of the RF signal. In modulation feedback the loop bandwidth for such system is relatively small compared to direct feedback. The reduced bandwidth opens up for high loop gain, even though within the loop the signal is translated to RF, amplified and translated back again to low frequency [5]. The complexity is increased significantly compared to the direct feedback structure. This leads to increased loop time delay, which is one of the primary limiting factors of modulation feedback [26].

2.4.3 Cartesian Modulation Feedback

Cartesian feedback is a common technique for linearizing RF power amplifiers which carry signals having amplitude variations. AM-to-AM and AM-to-PM amplifier errors are approximately reduced by the magnitude of the loop gain [27]. The Cartesian feedback transmitter technique was first proposed by Petrovic [28] in the early 1980's and was later refined to form the direct-conversion architecture shown in Figure 11, where I and Q represent the complex-valued input baseband signals. The output of the transmitter is sensed and fed back via a demodulator and compared with the input signal to obtain the error signal. The error signal is fed to the loop filter followed by upconversion in a quadrature modulator before it reaches the nonlinear power amplifier.

The Cartesian modulation feedback transmitter technique is studied thoroughly both in theory and successfully in practical experiments [26, 29–34]. In the experiments the transmitter operates on a carrier frequency from a few MHz to 2.4 GHz. Carrara et al. reported in [35] a Cartesian feedback circuit that achieves a linearization bandwidth as high as 60 MHz and an intermodulation improvement of up to 24 dB when tested on a W-CDMA 5 MHz input signal with a carrier frequency of 1.95 GHz.

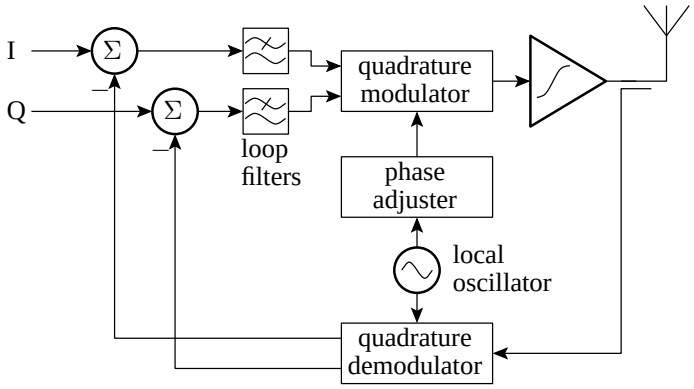

Figure 11: Block schematic of the Cartesian feedback system.

This feedback technique is considered to work only on narrowband signals, but Carrara et al. have proven its usage for rather wideband signals.

In practice a phase adjuster is required to adjust the phase automatically to preserve the stability. The phase adjuster can be placed in the system as illustrated in Figure 11 to differentiate the local oscillator phase between the modulator and the demodulator. This little feature of the schematic drastically increases the complexity of the technique. Automatic phase adjustment is a large and important research topic in this field.

2.4.4 Phase Adjustment

Feedback is a well known technique for linearizing nonlinear systems. In case of loop delay in the system, instability can result unless the operational bandwidth is limited. Unfortunately the stability of modulation feedback schemes can be affected by the absolute RF phase shift around the feedback loop which changes with carrier frequency, amplifier device, and environmental fluctuations, such as supply voltage, temperature and antenna loading. In common with other closed feedback loops this technique is only conditionally stable. Incorrect phase adjustment causes the loop phase margin to be reduced. This can cause peaking in the closed-loop frequency response, which in turn can increase the out-of-band noise floor, normally $1-10$ MHz from the carrier frequency [36]. Rising of the out-of-band noise floor is the last warning prior instability. These phenomena have been experimentally observed and are analyzed in [37].

Published methods for controlling the phase adjuster include minimizing the open-loop phase difference between the input and the feedback signal [38], and using look-up table of stored correction values obtained from a special test at manufacture [39]. Many of the phase adjusters reported to date have been discrete implementations and

not always been totally automatic. The calibration is sometimes done in a special test transmission or with manual trimming. This problem has hampered fully monolithic integration of the Cartesian feedback system. However, Dawson et al. have in [40] solved the synchronous downconversion problem with an analog phase alignment regulator integrated in a 0.25 μm CMOS process, and the paper also accurately motivates the need for a phase adjuster.

After the phase is adjusted near optimum it is time for frequency compensation of the system. The frequency response of linear feedback systems can be shaped and their performance enhanced by means of phase-lag, lead-lag and lag-lead compensators by producing the required amounts of gain and phase in the critical regions of the systems stability margins. The main limitation of the Cartesian transmitter is its linearization bandwidth. Boloorian et al. proved in [41] that the bandwidth of the Cartesian feedback stucture can be further improved by efficient frequency compensation of the loop. Experimental results reveal a factor of 2 increase of the linearization bandwidth. The amount of bandwidth increase depends on the actual system. However, the effect of frequency compensation is clearly there.

The research on the Cartesian modulation feedback technique has produced several circuits with special functions. In an attempt to improve the efficiency at low power levels a dynamic biasing scheme was proposed by Briffa et al. in [42]. Another interesting derivative of the Cartesian feedback was presented by Johansson et al. for multi-carrier applications [43].

2.4.5 Polar Modulation Feedback

The polar modulation feedback system is depicted in Figure 12 and was proposed by Petrovic et al. [44]. It is closely related to the envelope elimination and restoration (EER) technique (see section 2.5.1) as it completely avoids the nonlinear characteristic of the main amplifier. The input signal is at an intermediate frequency (IF) and the input signal is split up in its polar components, amplitude and phase, and compared with their respective counterparts of the amplifier output. The phase is controlled in a PLL structure, which basically contains limiter, phase detector, charge pump, loop filter, VCO and amplifier. The VCO feeds the power amplifier with a constant-envelope phase-modulated signal. The amplitude control is constituted from a regular negative-feedback structure, where the amplitude error signal modulates the collector voltage of the power amplifier. In common with all other feedback structures, it is applicable to any form of modulation, but is restricted to narrowband systems due to the bandwidth expansion that is associated with the polar representation of the signal. McCune et al. reported that both modulation paths (amplitude and phase) must be designed to support these wider bandwidths, which for EDGE are 1 MHz for the magnitude component and 2 MHz for the phase component [45].

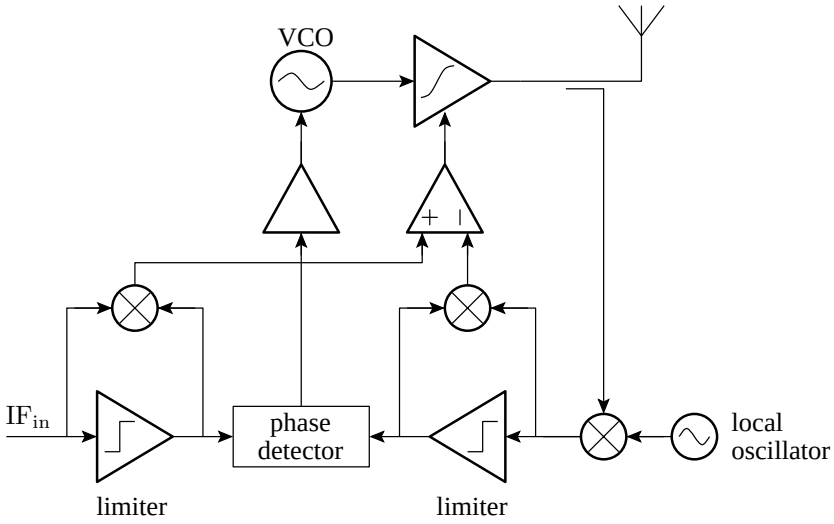

Figure 12: Block schematic of the polar modulation feedback system.

2.5 Linear Transmitter Architectures

Highly efficient linear power amplification can be achieved by linearizing a nonlinear but very efficient PA. Examples of established linearization techniques we have seen so far are predistortion, feedforward, and feedback. Alternatively, an intrinsically linear transmitter architecture can be adopted. The linear transmitter architecture is recognized by completely avoiding to exercise the nonlinear characteristic of the amplifier, and a few successful solutions exist. The envelope elimination and restoration transmitter splits the input signal in its phase and amplitude component (see section 2.5.1). Both LINC and CALLUM are based on a kind of divide-and-conquer approach, which first transforms an amplitude- and phase-modulated signal into two constant-envelope phase-modulated signals, and then recombine them after they have been power amplified (see section 2.5.2 and 2.5.3).

2.5.1 Envelope Elimination and Restoration

Already in 1952 Leonard R. Kahn introduced the single-sideband transmission by envelope elimination and restoration (EER) technique [46]. The principle of the described method is shown in Figure 13. A modulated intermediate frequency signal is split into its polar components, magnitude and phase, by means of an envelope detector and a limiter. Amplification is accomplished by a process in which the phase-modulation component of the modulated signal is amplified by means of a Class-C amplifier to achieve amplitude discrimination (limiter behavior), and the IF signal is translated to RF with a mixer. The input signal to the power amplifier should be large

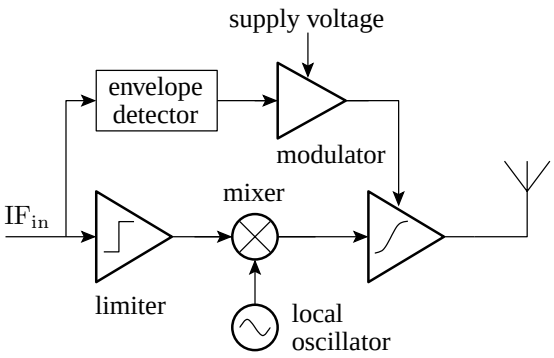

Figure 13: Envelope elimination and restoration block diagram.

enough to saturate the amplifier, thereby achieving high efficiency. The amplitude envelope is restored at the final amplifier by modulating the supply voltage.

Kahn compared a linear transmitter with the EER technique and found the average efficiency to be 47.1 % for the linear transmitter and approximately 69 % for the EER system when operating on a two-tone test signal [47]. The efficiency improvement produced by the EER technique is substantial. Since Kahn's initial experiments this linear transmitter architecture has been refined, implemented in an integrated CMOS technology [48], and applied on other types of modulations with improved power efficiency of the amplifier [49–53]. An efficiency of about 60 % for all signal amplitudes with a variety of signals, including single-side band (SSB), AM, and FM was reported by Raab et al. in [54]. For signals with high peak-to-average ratios, the average efficiency of a EER technique transmitter can be three to four times that of a traditional linear RF power amplifier [55].

The linearity of a EER transmitter does not depend upon the linearity of the RF power amplifier, in the same way as pulse-width modulation does not depend upon the linearity of individual transistors. Consequently, the intermodulation distortion (IMD) characteristic of an EER transmitter primarily depends upon parameters such as the bandwidth of the envelope modulator and the difference in delay between the envelope and phase signals when they are recombined in the final amplifier. Since the two paths, amplitude and phase, are inherently different in nature the delay matching error between the two paths is a big issue. The performance degradation of such mismatch has been studied in [55, 56].

2.5.2 Linear Amplification with Nonlinear Components, LINC

Substantial attention has been devoted to increase the efficiency of the power amplifier. At the same time it has been increasingly important to achieve a certain degree of linearity when operating on modulation schemes containing a fluctuating amplitude.

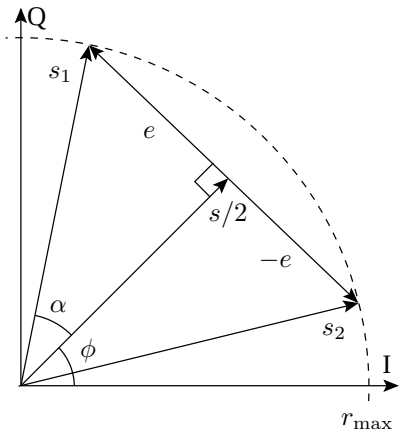

Figure 14: Component separation in the LINC transmitter.

One promising technique is the LINC architecture, which can utilize highly efficient power amplifiers without suffering from their nonlinear characteristic independent of applied modulation at the input. Linear amplification with nonlinear components is shortened LINC, and is an architecture that D.C. Cox [57] rediscovered in the mid 1970's, probably after studying the original work done already in the mid 1930's by H. Chireix [58]. At that time LINC was an unknown acronym, but the out-phasing technique was introduced to overcome increasing problems with cost and power efficiency of high power AM-broadcast transmitters [59]. The solution suggested by Cox was suitable for modulation schemes exhibiting both amplitude and phase variations [5].

The idea behind the LINC architecture came from a simple observation in the I-Q diagram where a vector can be divided into two vectors of equal magnitude but different phase, see Figure 14. The baseband signal is assumed to be both amplitude- and phase-modulated, and the complex-valued signal is represented by the vector $s(t)$. The two new signals, $s_1(t)$ and $s_2(t)$, have a constant envelope equal to $r_{\max}$ and they carry, of course, only information in their phase. e is an error vector that together with s can be used to create s_1 and s_2. The equations for the LINC transmitter are saved for chapter 3.

The basic LINC transmitter architecture is illustrated in Figure 15 at block level. The non-constant envelope input signal is resolved to two outphased constant-envelope signals, which are individually amplified by highly efficient, but grossly nonlinear amplifiers. A linearly amplified replica of the input signal is formed by the vector summation of the outputs from the power amplifiers. The LINC scheme has the potential of achieving both high power efficiency and high linearity at the same time [60].

The creation of the constant-envelope signals, $s_1(t)$ and $s_2(t)$, is nontrivial and

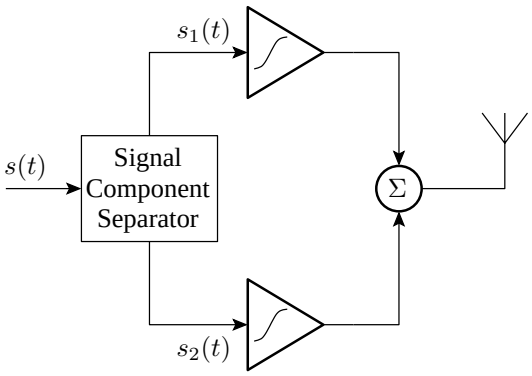

Figure 15: The LINC transmitter principle.

a pure analog implementation is a true challenge. Indeed, various analog implementations of the signal component separator (SCS) working at 200 MHz intermediate frequency have been presented by Shi et al. [61–64]. Test results of these experimental chips showed that spurious levels around -50 dBc could be obtained with a $\pi/4$-shifted DQPSK modulated North American Digital Cellular (NADC) signal. The power consumption of the chips, which use external power amplifiers, were in the 100 mW range.

The implementation of the SCS has been one of the major obstacles with the LINC concept, where an earlier paper [65] suggested a completely analog solution using the phase-modulation method. The description of the constant-envelope vectors includes a $\arccos$ (or $\arcsin$) function, which is hard to implement in an analog circuit. Today, the evolution of DSP techniques makes it possible to implement the SCS all in software using the in-phase and quadrature-phase method [66]. A block diagram of a DSP-driven SCS realization is illustrated in Figure 16, where e is the error vector shown in Figure 14. However, the bandwidth of the constant-envelope phasors are significantly higher than that of the modulated input signal. Therefore the DSP and D/A converters need to operate with sampling rates at least 15 – 20 times the bandwidth of the input signal [60, 67]. As can be understood this has significant impact on the power consumption of the DSP and D/A converters, as it is roughly proportional to the clock frequency.

Furthermore, the phasors generated at baseband by the DSP must be upconverted by a quadrature modulator, which suffers from gain and phase imbalance between the I and Q component. As LINC is an open-loop system these imbalances degrade performance and normally baseband compensation of this misalignment has to be done to get good performance out of the LINC transmitter. The spectral sensitivity to quadrature modulator misalignment has been studied in [68].

One major disadvantage with LINC is the extremely tight tolerance on the match-

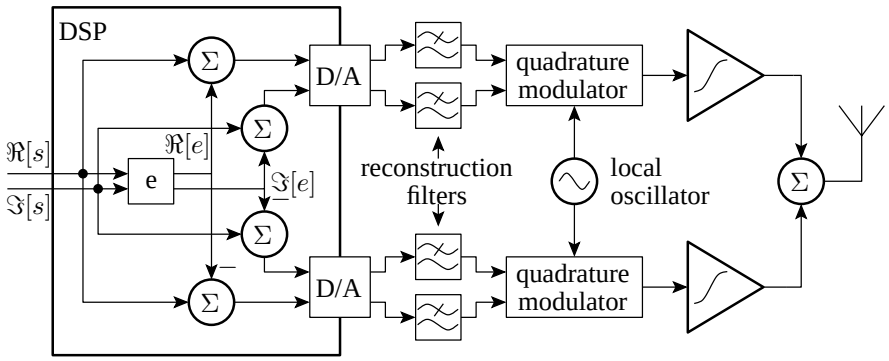

Figure 16: The LINC transmitter with a DSP-based SCS operating at baseband.

ing of the two paths. The out-of-band spectrum is created by incomplete cancellation of the quadrature signals. Typical requirements for most practical applications are approximately $0.1-0.5\,\text{dB}$ gain matching or $0.4^\circ - 2^\circ$ in-phase matching according to Zhang et al. [69]. The gain and phase imbalance in the LINC transmitter has been studied theoretically [70, 71] and a few proposed techniques to minimize these imbalances are found in [69, 72–75].

2.5.3 Combined Analog Locked Loop Universal Modulator

The linear transmitter architecture that goes under the acronym CALLUM was first proposed by Bateman [76]. CALLUM is related to the LINC technique in that the two constant-envelope phasors from the outputs of the power amplifiers are combined to produce the output signal to the antenna. The basic CALLUM architecture is illustrated in Figure 17 at block level. The in-phase and quadrature-phase component of the complex-valued baseband signal, $s(t)$, is the input signals to the system. The feedback signal is the baseband equivalent of the transmitter output, which is obtained from a quadrature demodulator. The signal component generator (SCG) contains a set of control equations that generate the control signals to the voltage controlled oscillators (VCOs) from the input and fed back signals. The output from the VCOs are the constant-envelope phase-modulated signals as in the LINC transmitter. The CALLUM architecture is studied in more detail in chapter 4, and there it will be clear that many of the numerous derivatives presented in the open literature originates from the CALLUM 1 structure and only differ from the chosen set of control equations in the SCG.

By nature distortion is deterministic opposed to noise. If the distortion characteristic is mapped it can be reduced by for example pre- or post-distortion, but changes of the nonlinearty over time call for calibration or adaptation of such systems. Therefore, a more universal approach is to use negative-feedback. CALLUM is an expansion of

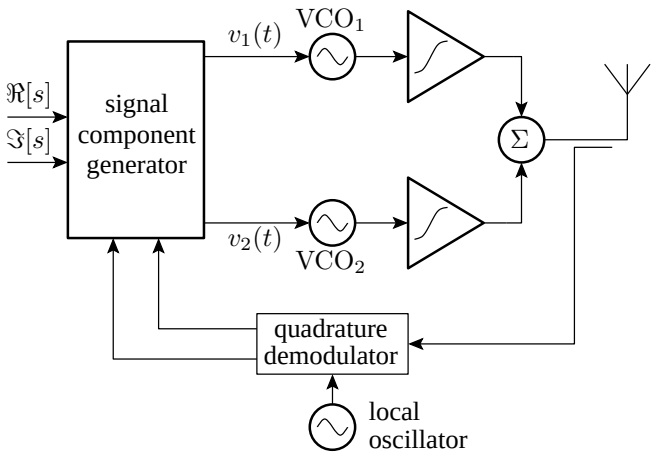

Figure 17: The CALLUM transmitter principle.

the LINC concept, since it is based on a LINC core placed in a feedback loop. Feedback as every other concept has its pros and cons. Feedback is associated with high performance systems, a universal way to reduce distortion no matter its origin, stability hazards, and limited to narrowband signals. In the CALLUM system the stability issue is delicate as the VCO in the loop degrades the phase by 90° and the inevitable loop time delay efficiently limits the bandwidth of the system to only cope with relatively narrowband signals. In addition to these two fundamental limitations comes the influence of the loop poles on the stability properties.

A few features of the CALLUM architecture are that the selected channel frequency is set by the local oscillator in the feedback path, and frequency upconversion is accomplished within the topology by the two loop VCOs. For integrated design, building blocks containing inductors are not very popular in the industry due to the associated chip area the inductors occupy. The VCO also needs an area consuming capacitor bank (normally switched for coarse tuning) and a varactor (for fine tuning). As CALLUM has two VCOs this can be considered as a drawback. The modulation feedback (Cartesian or polar) of CALLUM leads to a baseband (or IF) implementation of the SCG, which simplifies the design (or reduces the design time), opens up for high performance circuits with little or negligible influence on the stability properties, and low power consumption compared to a direct RF solution.

Several variations on the original CALLUM architecture, usually referred to as CALLUM 1, have appeared in different papers. Chan and Bateman [77] reviewed no less than six: CALLUM 1, 2, 3, 4, the vector locked loop (VLL), and higher-order. The various CALLUM implementations were investigated in terms of stability in the I-Q plane and speed estimations performed by step response analysis. The theoretical studies were supplemented with measurements on physical implementations of the

CALLUM architecture built from discrete blocks. Several CALLUM designs were simulated in [78], where their RF output spectra for a TETRA-like baseband modulation were presented, together with actual measurements for CALLUM 2. The work was an important contribution to the understanding of the properties of the different CALLUM versions.

Reported Measurement Results

Bateman reported in the beginning of the 1990's the first working CALLUM implementation [76]. The CALLUM version under test was based on a Cartesian feedback structure (referred in the literature as CALLUM 3), which in its most basic form only can maintain lock when the signal at the summing port at the outputs is within $\pm 90^\circ$ of the local oscillator phase. For the system to be stable in more than one quadrant a sign-switching matrix was inserted prior the VCOs. The system suffers from discontinuities in the signal, which will push the loop out of lock and rise the far out spectrum. However, the system could handle a 30 kHz channel bandwidth with intermodulation distortion suppression in excess of 55 dBc for an offset-QPSK modulation with roll-off factor $\alpha = 0.30$ at 159 MHz center frequency. Measurements on the same switching matrix CALLUM version Chan et al. reported an ACLR of approximately -50 dB [77, 79]. In the late 1990's Jennings et al. published measurement results for CALLUM 1, based on a DSP-driven SCG [80]. In response to the 5 kbaud $\pi/4$-DQPSK modulated signal at 233 MHz center frequency, the adjacent channel power is in the range -50 to -55 dBc, whereas the 18 kbaud signal response is slightly worse at approximately -45 to -50 dBc. The performance degradation for more wideband signals can be expected as the loop gain drops off at high frequencies. Jennings et al. also produced a working CALLUM 2 prototype at a center frequency of 220 – 240 MHz. The applied test signal was a TETRA signal, which employs $\pi/4$-DQPSK modulation at a data rate of 18 kbaud with roll-off factor $\alpha = 0.35$. The adjacent channel is located 25 kHz off center and the adjacent channel power was approximately -50 dBc [78, 81, 82].

Chapter 3

Linear Amplification with Nonlinear Components

The LINC and CALLUM transmitter architectures share the fundamental signal processing core, as the input signal is manipulated to form two constant-envelope signals fed to the power amplifiers. In this chapter we focus on LINC, and the control equations for generation of these phasors. Both the LINC and the CALLUM concepts are promising as no fundamental limitation prevents 100 % efficiency for all signal levels. Many other linearization techniques using conventional amplifier classes have be proposed as solutions, but all suffer a fundamental limit to the theoretical efficiency obtainable. This is set by the efficiency characteristic of the particular class and could typically be 75 – 100 % at peak envelope power (PEP), but usually falls considerably at lower envelope levels, tending to 0 % at zero output level [66]. In practice, it is a challenge to achieve ideal performance, as there is still today a rather large gap between the signal graph and the electrical circuit implementation.

3.1 LINC and the Signal Decomposition

The LINC transmitter is illustrated in Figure 18, where $s(t)$ is the amplitude- and phase-modulated input signal to the system. Since no frequency translation is performed by the circuit in Figure 18 the input signal must be at the carrier frequency. It is possible to put in mixers for frequency translation. Then the input signal can be at IF, or by using quadrature modulators in the forward paths preceding the power amplifiers, the input signal can be at baseband. A realization of the SCS at RF really puts hard constraints on the processing speed of the circuit and the power consumption will probably be high. A baseband solution, on the other hand, not only needs quadrature modulators but also introduces more error sources, e.g., I-Q misalignment. This error must be corrected for optimum performance, and the overhead circuitry is substantial. Which way to go is not definite, but the trend is to include the SCS in the digital baseband signal processing circuit together with correction for modulator misalignment and other imperfections.

The input signal is decomposed into two signals, $s_1(t)$ and $s_2(t)$, by a projection of $s(t)$ on the circle with radius $r_{\max}$ according to Figure 19. These phasors have constant-envelope but varying phase, and for maximum output signal the phasors are

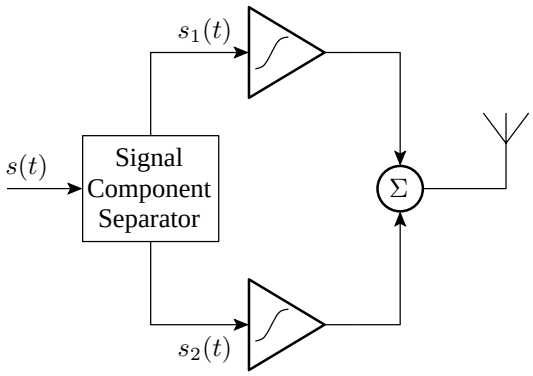

Figure 18: The block schematic of a LINC transmitter.

aligned with each other. To produce a zero output the phasors are in anti-phase. Note that for zero output the phase can take any value as long as the phasors are opposed. This undetermined state should be avoided as it gives numerical problems in terms of 'divide by zero'. The LINC transmitter can process a constant-envelope phase-modulation without direct efficiency penalties even if the transmitter is intended for linear modulations. Therefore, the coexistence of the GSM and the EDGE communication systems potentially could be operated by the same transmitter.

3.2 Signal Component Separator

The signal component separator (SCS) is the heart of the architecture, and the algorithm implemented in the SCS will differ depending on the kind of signal representation chosen, as will be shown in section 3.2.1 and 3.2.2. The transmit spectrum depends on the input spectrum, and any divergence from the input spectrum originates from vector errors of the constant-envelope signals. These errors can basically arise from inaccurate calculations in the SCS, or by gain and phase imbalance between the paths.

3.2.1 Phase-Modulation Method

The information signal applied at the input of the SCS is given by,

$$s(t) = r(t)\cos(2\pi f_c t + \phi(t)), \tag{11}$$

where $r(t)$ is the amplitude modulation. The length of the vector is $r(t) = |s(t)|$, and f_c is the carrier frequency. $\phi(t)$ denotes the phase-modulation part of the signal. The input signal is divided into two new signals according to,

$$s(t) = s_1(t) + s_2(t). \tag{12}$$

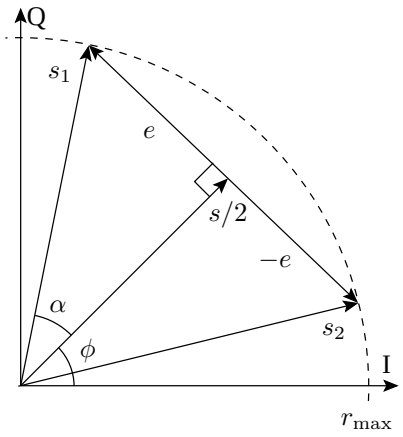

Figure 19: Component separation in the LINC transmitter.

From Figure 19 it can be seen that the two constant-envelope signals are described by,

$$\begin{aligned} s_1(t) &= r_{\max}\cos(2\pi f_c t + \phi(t) + \alpha(t)) \\ s_2(t) &= r_{\max}\cos(2\pi f_c t + \phi(t) - \alpha(t)). \end{aligned} \tag{13}$$

Vector addition yields,

$$\begin{aligned} s_1(t) + s_2(t) &= r_{\max}\left\{\cos(2\pi f_c t + \phi(t) + \alpha(t)) + \cos(2\pi f_c t + \phi(t) - \alpha(t))\right\} \\ &= r_{\max}\left\{\cos(\beta(t) + \alpha(t)) + \cos(\beta(t) - \alpha(t))\right\} \\ &= r_{\max}\left\{\cos(\beta(t))\cos(\alpha(t)) + \cos(\beta(t))\cos(\alpha(t))\right\} \\ &= 2r_{\max}\cos(2\pi f_c t + \phi(t))\cos(\alpha(t)). \end{aligned} \tag{14}$$

Identification of $s(t) = r(t)\cos(2\pi f_c t + \phi(t))$ and (14) gives the angle relation,

$$\alpha(t) = \arccos\left(\frac{r(t)}{2r_{\max}}\right). \tag{15}$$

Early implementations of the LINC transmitter were based on this phase-modulation method, which includes an $\arccos$ (or $\arcsin$) function [65, 83]. Although the LINC concept has been in existence since the out-phasing technique in the 1930's, the generation of the constant-envelope signals has always been inaccurate as the $\arccos$ phase term is difficult to implement using analog techniques. Much attention was directed towards the implementation of the $\arccos$ function, and the circuit complexity hampered the acceptance of the LINC concept. Modern DSP technology has made a significantly more accurate component separator possible, and the linearity performance of the transmitter is now dominated by the RF gain and phase match between the two amplified components [66].

3.2.2 In-Phase and Quadrature-Phase Method

A general baseband representation of the band limited RF signal can be written as,

$$s(t) = r(t)e^{j\phi(t)}, \quad \text{where} \quad 0 < r(t) \leq 2r_{\max}. \tag{16}$$

This signal is split into two signals, $s_1(t)$ and $s_2(t)$, having constant amplitudes such that,

$$s(t) = s_1(t) + s_2(t), \tag{17}$$

where $|s_1| = |s_2| = r_{\max}$. The phasors, $s_1(t)$ and $s_2(t)$, can be calculated with the method described in [66], which is a generalization of the in-phase and quadrature-phase method. The vector representation of the signals in Figure 19 yields,

$$\begin{aligned} s_1(t) &= \frac{s(t)}{2} + e(t) \\ s_2(t) &= \frac{s(t)}{2} - e(t). \end{aligned} \tag{18}$$

The error signal e is a signal that is in quadrature with the input signal, $s(t)$, and is determined by,

$$\begin{aligned} e(t) &= s_1(t) - \frac{s(t)}{2} = \sqrt{\left(r_{\max}^2 - \frac{r^2(t)}{4}\right)} e^{j(\phi(t)+\frac{\pi}{2})} \\ &= j\frac{r(t)}{2}e^{j\phi(t)}\sqrt{\left(\frac{2r_{\max}}{r(t)}\right)^2 - 1} = j\frac{s(t)}{2}\sqrt{\left(\frac{2r_{\max}}{|s(t)|}\right)^2 - 1}. \end{aligned} \tag{19}$$

The quadrature signal, $e(t)$, is calculated once and stored in a two-dimensional LUT as calculation of $e(t)$ would require multiple clock cycles, which is an unacceptable speed penalty. $e(t)$ is added to $s(t)$ in one branch and subtracted from $s(t)$ in the other branch to obtain two constant envelope signals [84]. Now, the constant-envelope vectors can be formulated by combining (18) and (19).

$$\begin{aligned} s_1(t) &= \frac{s(t)}{2}\left(1 + j\sqrt{\left(\frac{2r_{\max}}{r(t)}\right)^2 - 1}\right) \\ s_2(t) &= \frac{s(t)}{2}\left(1 - j\sqrt{\left(\frac{2r_{\max}}{r(t)}\right)^2 - 1}\right) \end{aligned} \tag{20}$$

3.2.3 Spectral Regrowth

The linear modulation applied on $s(t)$ gives the input signal a well defined bandwidth, but as the input signal is decomposed into two constant-envelope phasors, $s_1(t)$ and

$s_2(t)$, the relative bandwidth is increased. The spectral regrowth is due to the transfer of the amplitude information to the phase, whereas extra phase modulation is added to the existing phase modulation. For signals as two-tone tests or modulations having zero crossings, the wideband nature of the signal components are due to instantaneous phase discontinuities that occur when the envelope passes through zero. Hetzel et al. realized that the two signal component phasors effectively swap over, i.e., $s_1(t)$ becomes $s_2(t)$ and vice versa, when the envelope of the two-tone input signal crosses through zero, it is possible to take account of these changes and swap the phasors back again [85].

From (20) it can be seen that a zero envelope gives a zero in the denominator. For the SCS of the LINC transmitter this might be a problem as the outputs are undefined, even if it is only for a short while as the signal passes through zero. An efficient remedy is to use modulation schemes that avoid the origin of the I-Q diagram. The scheme used in Enhanced Data rates for GSM Evolution (EDGE) is a $3\pi/8$-shifted 8PSK, and the $\pi/4$-shifted quadrature phase shift keying (QPSK) are examples of such modulations. In Figure 20 the EDGE modulated baseband signal $s(t)$ is shown together with the spectrum of the constant-envelope signal $s_1(t)$, for which the bandwidth is significantly increased. There is a coupling between the bandwidth expansion and the modulation depth, m_d, defined as the ratio between the largest and smallest envelope of the modulated signal. In general, the larger modulation depth, the larger bandwidth regrowth. For EDGE $m_d \simeq 7$, and for $\pi/4$-shifted QPSK modulation $m_d \simeq 10$, and the spectrum for $s_1(t)$ is relatively wider for $\pi/4$-shifted QPSK due to its higher modulation depth, as seen in Figure 21.

The instantaneous frequency component the system has to handle cannot be extracted directly from the spectrum graph in Figure 20. According to [86] the instantaneous frequency is, at first-order approximation, proportional to the modulation depth of the signal. For an EDGE modulated signal the instantaneous frequency is approximately three times higher than the inverse of the time between two consecutive symbols.

3.3 Efficiency Issues

An amplifier is most efficient in converting DC power to RF output power when operating in its saturation region for a constant-envelope input signal. Digital modulation schemes, such as M-ary PSK, have an efficiency advantage in their constant envelope. The frequency spectrum is limited by nature, and the increasing number of users of mobile handsets have switched the interest to linear digital modulation schemes like M-ary QAM and $\pi/4$-shifted QPSK, since they can transmit more data over a given bandwidth [87]. A way of combining the constant-envelope input signal and the linear modulation is to use a linear transmitter architecture as LINC, which offers high efficiency and high linearity at the same time. In practice 100 % efficiency is a very optimistic figure of merit, as the ideal efficiency number can only be approached with ideal power amplifiers and signal combining. Normally a power combiner with high

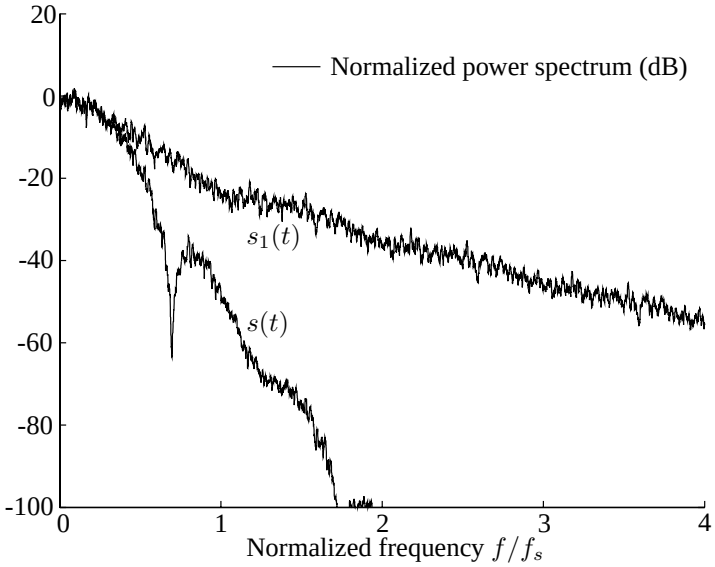

Figure 20: EDGE modulation with modulation depth $m_{\mathrm{d}} \simeq 7$.

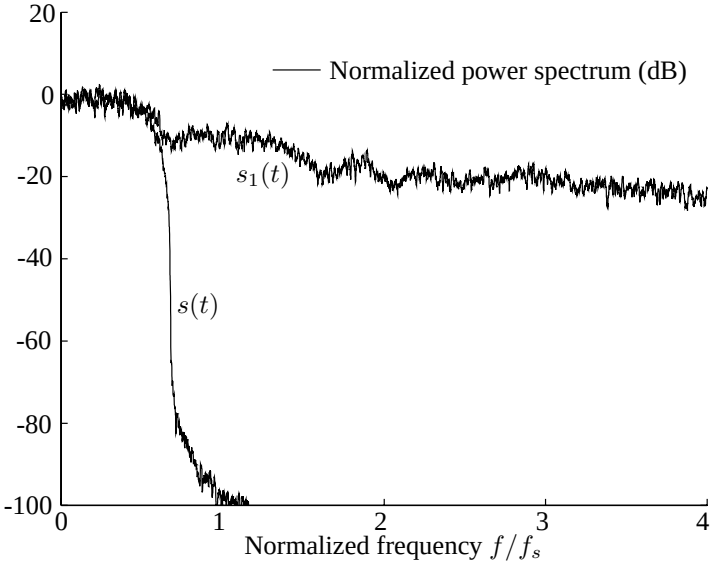

Figure 21: $\pi/4$-shifted QPSK modulation with modulation depth $m_{\mathrm{d}} \simeq 10$.

isolation is used, which poses the major obstacle to achieve high efficiency.

3.3.1 Power Combining

Linear modulation schemes combined with limited battery capacity impose primarily restrictions on the power consumption of a standard handset. The out-phasing technique has been proposed as one solution, since the two out-phased constant-envelope signals are amplified separately by two highly efficient power amplifiers and the outputs are added before the signal propagates to the antenna. This approach allows the power amplifiers to operate continuously at their peak power efficiency, still with a possibility for a highly linear output. One inherent drawback with this approach is the associated power loss in the combining network. When using a hybrid combiner, which is the standard combiner in a LINC transmitter, this loss can be significant [88]. Other possible power combining techniques exist, but suffer from incomplete isolation between the power amplifiers, resulting in increased out-of-band distortion.

Hybrid Combiner

The hybrid combiner has one summing port connected to the antenna and another subtracting port connected to a dummy load. The vector summation and subtraction at the output and the dummy port, respectively, are illustrated in Figure 22. The hybrid combiner directs the in-phase signal, $s(t)$, to the load, whereas the quadrature-phase signal, $\pm e(t)$, is fed to a dummy load.

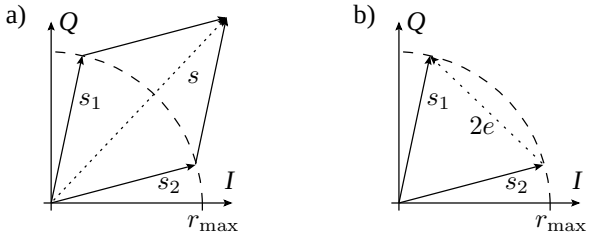

Figure 22: a) Vector summation giving the wanted signal $s(t)$ at the output summing port. b) Vector subtraction forms the error signal, which is directed and terminated at the difference port by a matched load.

One obstacle to get linear output from a LINC transmitter, next to gain/phase imbalance, is the use of a combiner whose input ports are coupled [87]. Therefore the hybrid combiner is most common in the LINC transmitter as it provides high isolation between the input ports, and thereby high linearity. The drawback with the hybrid combiner is that much of the efficiency inherent in the LINC transmitter topology is lost as the error vector, e, in Figure 19 is directed to a matched load at the difference

port. For large envelope levels the error vector is small and the efficiency high, but for small envelope levels the situation is the opposite.

To isolate the impact of the hybrid combiner on the efficiency of a LINC transmitter Sundström et al. [88] assumed the efficiency of the hybrid combiner and the power amplifier to be 100 %. From these simulations offset-QAM, QAM and $\pi/4$-shifted DQPSK perform best of the tested modulation methods with efficiencies in the 40 – 50 % range for roll-off factors $0.3 \leq \alpha \leq 0.5$. The simulations also show that the efficiency of 16-QAM is less than 25 % at best and 64-QAM even worse with $<$20 % efficiency.

The efficiency of the LINC transmitter depends on the modulation, and η_m is given by,

$$\eta_m = \frac{\overline{r}^2}{r_{\max}^2}, \tag{21}$$

where $\overline{r}^2$ is the square of the average amplitude of the baseband signal calculated from,

$$\overline{r} = \frac{1}{r_{\max}^2} \int_0^{r_{\max}} p_m(r) r^2 dr. \tag{22}$$

$p_m(r)$ is the probability density function describing the amplitude probability for the chosen modulation scheme. For high modulation efficiency the average amplitude should be high, e.g., PSK, but a compromise can be linear modulations with large $\overline{r}$.

A power re-use technique based on an AC/DC converter was proposed by Langridge et al. to increase the power efficiency when using a hybrid combiner [89]. The power at the difference port that normally is wasted as heat, now is recycled back to the voltage supply. The reported peak re-use efficiency was 63 % for the power converter at 1.96 GHz.

Wilkinson Combiner

An alternative to the four port hybrid combiner is the Wilkinson power combiner, illustrated in Figure 23a). It consists of two $\lambda/4$ transmission lines with a characteristic impedance of $\sqrt{2}Z_0$. The combiner/splitter has tree ports and can be used in either direction. In case of power combining the inputs are port 2 and 3, and the output is at port 1. When the inputs have equal phase and amplitude, no power is dissipated in the $2Z_0$ isolating resistor. When a mismatch occurs at port 2 or 3, the reflected signals split through the two transmission lines, travel back to the output, split again, and travel back to the input ports. The reflected wave returns to the input ports in two parts, each 180° out of phase from each other. The value of the isolating resistor $2Z_0$ was selected so that the two paths of the reflected wave have equal amplitude and perfect cancellation results [90]. The Wilkinson coupler has high isolation between the

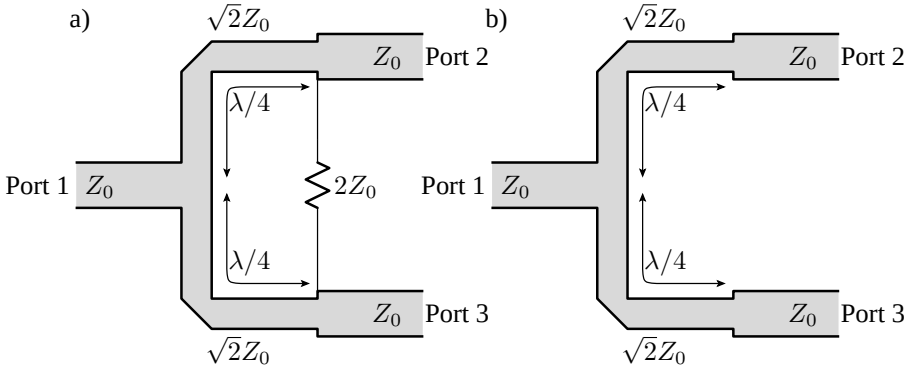

Figure 23: a) The Wilkinson coupler. b) The tee coupler.

input ports and the efficiency, in combination with the LINC concepts, is the same as for the hybrid combiner. These combiners are also called matched combiners, which provide a constant load impedance seen by the amplifiers. This results in a continuous maximum output power from the amplifiers and a constant amount of DC power taken from the supply.

Signal Combination with Power Amplifier Interaction

The more efficient combining techniques allow the power amplifiers to affect each other by coupling their outputs [91, 92]. A lossless combiner is a device that contains no internal power termination. The lossless tee is the three port Wilkinson combiner without the isolating resistor, allowing the nonlinear outputs of the high efficiency amplifiers to be coupled, see Figure 23b). This has the effect of modulating the effective load impedance presented to the power amplifiers, whereas the output power is a function of the load seen by the amplifier. This varying load will inevitable result in increased distortion. In order to avoid degrading the linearity of the system, the amplifiers should be designed to function as either voltage or current sources and, together with the combining network, give voltage or current summation at the output [6].

Based on [93] Shi et al. [6] investigated the efficiency property of a complementary-parallel common-source (CPCS) output stage of the power amplifier connected to a $\lambda/4$ transmission line for impedance transformation from low impedance at the amplifier output to high impedance at the summing node.

In general, a $\lambda/4$ transmission line corresponds to half a revolution in the Smith-chart, and the two power amplifiers with low output impedance are each connected to a $\lambda/4$ stripline. Figure 24 illustrates only one amplifier connected to the tuned load.

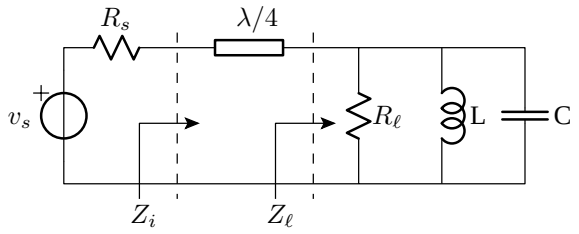

Figure 24: A $\lambda/4$ stripline as impedance transformer.

The load impedance is given by,

$$Z_\ell = \frac{sR_\ell L}{s^2 R_\ell LC + sL + R_\ell}. \tag{23}$$

The impedance Z_i can be found by using

$$Z_i = \frac{Z_0^2}{Z_\ell}, \tag{24}$$

where Z_0 is the characteristic impedance of the transmission line. In a specific case where $R_\ell = 50\,\Omega$, $L = 1\,\text{nH}$, $C = 31.3\,\text{pF}$ and $Z_0 = 50\,\Omega$ the impedance Z_i was calculated for a number of frequencies.

$$\begin{array}{lcl} Z_i(900\,\text{MHz}) & \simeq & 50 \\ Z_i(3 \cdot 900\,\text{MHz}) & \simeq & 50 + j1200 \\ Z_i(5 \cdot 900\,\text{MHz}) & \simeq & 50 + j2200 \end{array}$$

Only at the fundamental frequency there is active power to the load; at the 3^{rd} and 5^{th} harmonic mostly reactive energy pulsates in the wires.

Power Delivered to the Antenna

The antenna is designed to have nominal $R_\ell = 50\,\Omega$ input impedance over the frequency band. Suppose we want a linear amplifier delivering $1\,\text{W}$ into a $50\,\Omega$ resistive load at 900 MHz, and that a 3.3 V supply voltage is available, then the maximum power for a sinusoidal signal to the load without impedance transforming network is given by,

$$P_{\max} = \left(\frac{V_{\text{DD}}/2}{\sqrt{2}}\right)^2 \frac{1}{R_\ell} \simeq 0.03\,\text{W}. \tag{25}$$

An impedance transformation network is necessary to supply the demanded power to the load and the network has to be linear as no correction of errors is possible. The

maximum load resistance seen from the output of the PA is $R_{\max} = 1.36\,\Omega$. The characteristic impedance for the transmission line is then,

$$Z_o = \frac{V_{DD}/2}{\sqrt{2}}\sqrt{\frac{Z_\ell}{P_\ell}} = 8.25\,\Omega. \tag{26}$$

If the output from the PA is a square-wave and the wave is filtered with a tuned network in parallel with the load it becomes slightly different. The amplitude of the square-wave is $V_{\rm DD}/2$ and the fundamental component is $4/\pi$ times larger.

3.3.2 Efficiency Enhancement by Supply Voltage Adjustment

The transmitted power tends to decrease as the size of the cell in the cellular structure decreases. This trend comes from an increasing demand for bandwidth to the users of mobile terminals as new services arise such as WAP, video, and images. As discussed in section 3.3.1 the efficiency of the hybrid combiner is low for small envelopes of the information signal, $s(t)$, since the phasors $s_1(t)$ and $s_2(t)$ are almost out of phase, and thereby having a large error vector, e, which is terminated by the dummy load and wasted as heat.

The probability density function (PDF) of the transmit power for a handset (also called user equipment (UE)) operating on the W-CDMA standard is shown in Figure 25a). The curves are from simulations in urban and rural area scenarios, respectively, and the model includes both distance attenuation using Okamura-Hata propagation model and log-normal shadow fading. The maximum transmit power is 21 dBm and the median (when the cumulative distribution function (CDF) is 50 % in Figure 25b)) is $\simeq -3$ dBm, which is significantly below the maximum transmit power. Even though the peak output power is seldom used, the power amplifier has to be designed to handle this maximum power level, see Figure 26a). When the output power is controlled to a lower level by the basestation, the output amplitude must be reduced. Even for maximum amplitude at this reduced power level the constant-envelope phasors have a large angle in between, resulting in low efficiency of the LINC transmitter.

The supply voltage of the power amplifier determines $r_{\max}$ of the system. The value of $r_{\max}$ should be chosen as $r_{\max} = \max(r_i)/2$, meaning that the maximum input signal amplitude generates the maximum output signal for best power efficiency. The power amplifier is one of the most power consuming parts in the transmitter chain, and adaptation of the supply voltage would be of great interest to enhance power efficiency at lower transmit power levels. For lower output power $r_{\max}$ should be reduced to $r_{\rm opt}$ as illustrated in Figure 26b). The drain voltage of the power amplifier is controlled by a DC/DC converter to reduce the voltage to its optimal value. Global stability is one of the most appreciated features of a DC/DC converter, and it can be guaranteed with a certain type of nonlinear control. A semi-integrated DC/DC Buck converter (with external filter) was developed with this application in mind [94].

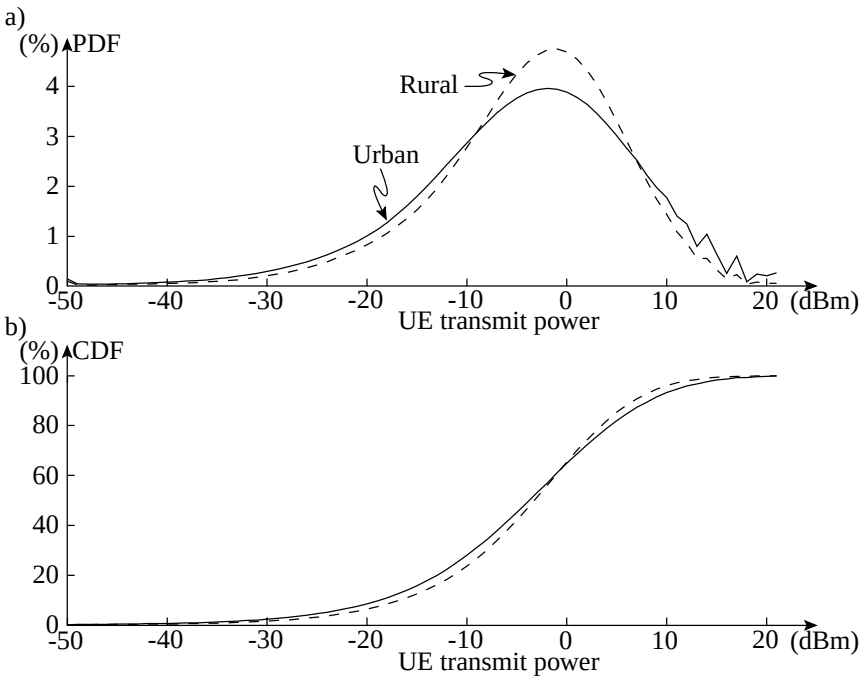

Figure 25: W-CDMA uplink transmit power by users per time slot level. a) The PDF for urban and rural scenario, respectively. b) Corresponding CDFs. Courtesy Muhammad Kazmi, Ericsson AB.

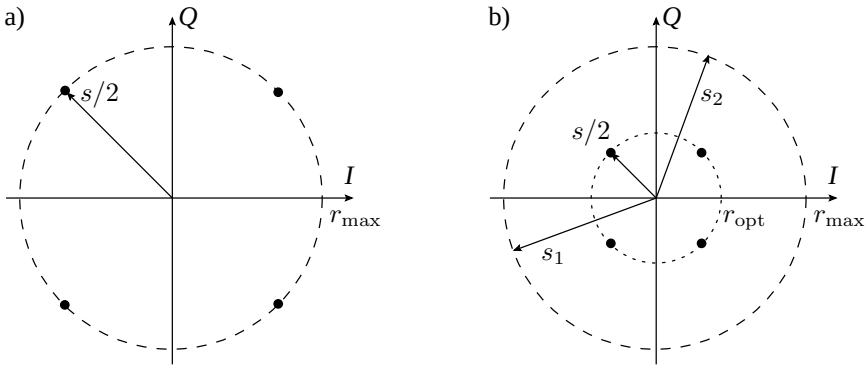

Figure 26: A QAM modulation at: a) maximum output power. b) reduced output power.

Chapter 4

The Combined Analog Locked Loop Universal Modulator

The CALLUM transmitter architecture is analyzed in this chapter that includes derivation of the control equations for the most common CALLUM topologies. The baseband modeling of the system is presented together with an investigation of the loop properties based on linearized control equations. The loop gain, bandwidth, stability, and frequency compensation are shown for the CALLUM transmitter. The modeling is expanded to include the most important parameter, the loop time delay, and how the presence of delay changes the frequency compensation strategy. System simulations, based on the baseband model, are presented to differentiate the expected performance from three versions, i.e., CALLUM 1, CALLUM 1lin, and CALLUM 2. From these simulations it is clear that CALLUM 2 is inferior compared to the other two. This chapter is closed by a comparison of the different CALLUM derivatives in terms of spectral performance, cost, and complexity.

4.1 CALLUM

CALLUM is in its nature closely related to LINC, as they share the fundamental idea with the signal decomposition into two constant-envelope phase-modulated signals for highly efficient amplification, without sacrificing the potential of linear transfer. The conceptual differences are that the architecture contains a frequency translation from baseband (or IF) to RF, and that CALLUM has a global feedback loop for error correction. CALLUM is a larger and to some extent a more complex system than LINC. However, the feedback makes the matching of the gain and phase between the paths less critical and thereby somewhat relaxes the design constraints in that respect. Also the signal component generation is less critical as long as the generation of the error signal is processed carefully. The signal component generator (SCG) does not generate the constant-envelope phasors directly, but delivers the correct control signals to the VCOs, which in turn generate these phasors. A critical part of the system is the downconversion of the RF output signal, as its linearity defines the maximum achievable linearity of the CALLUM architecture.

The input baseband signal, $s(t)$, is most often a linear digital modulation with information in both the amplitude and the phase of the signal. The baseband input and

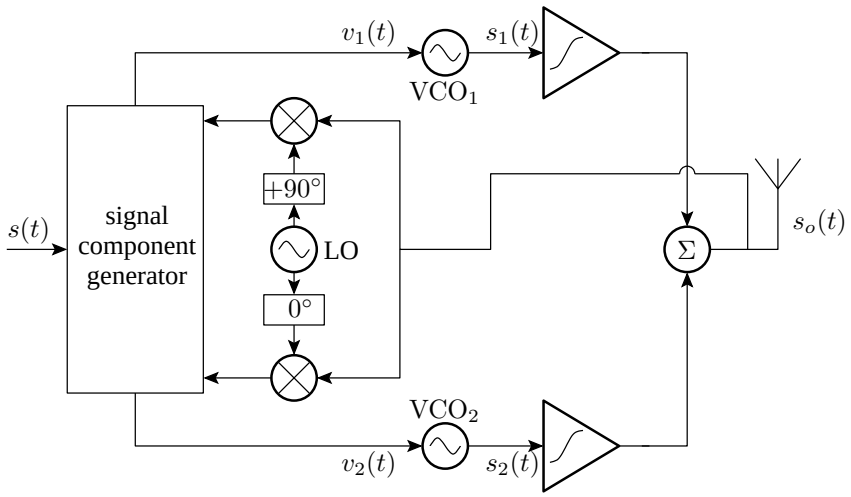

Figure 27: The CALLUM transmitter principle.

feedback signals are mathematically manipulated by the SCG, which outputs are the two control signals that are fed to the VCOs. The output from the VCO is a phase-modulated signal at RF, with a constant amplitude that drives the PA into saturation for high power efficiency. The input signals to the PAs, $s_1(t)$ and $s_2(t)$, are amplified and added together at the summing output port of the architecture. The vector addition of the outputs of the PAs should form an amplified replica of the input signal. A generic model of the CALLUM transmitter architecture is illustrated in Figure 27.

4.2 The Signal Component Generator

The signal component generator (SCG) is the heart of the architecture and synthesizes the control signals to the succeeding VCOs. The control signals from the SCG are based on a set of nonlinear equations, which contain the generation of the error signal as an important part. The error signal is the difference between the input signal, and the linearly attenuated and downconverted signal taken at the output of the transmitter. As in any feedback system the error signal should be equal to zero, which guarantees a linear amplification.

The derivation of the original set of control equations is based on an idealized open-loop system, and not from the closed-loop CALLUM architecture. The derivation of the constant-envelope signal is taken from the LINC concept. However, these equations will serve as the foundation for the control of the CALLUM system. Even though there are many similarities between the open-loop system of CALLUM and

LINC, the VCOs in the forward paths are a token for the CALLUM architecture. The derivation of the constant-envelope signals uses the phase-modulation method, see section 3.2.1, and only the most important equations will be repeated here. The constant-envelope signals are given by,

$$s_1(t) = r_{\max}\cos(2\pi f_c t + \phi(t) + \alpha(t)) \quad (27)$$
$$s_2(t) = r_{\max}\cos(2\pi f_c t + \phi(t) - \alpha(t)), \quad (28)$$

where the angle α is,

$$\alpha(t) = \arccos\left(\frac{r(t)}{2r_{\max}}\right). \quad (29)$$

The input control signal to the VCO is a voltage and the output is a frequency. Since the information at this point in the schematic is stored in the phase rather than in the frequency of the signal, the control signals to the VCOs cannot be applied directly. To exploit the sensitivity of the VCO it is wise to differentiate the phase with respect to time, $\dot{\phi}_{\mathrm{VCO}}$, which is equal to the angular velocity ω_{VCO}. The arguments of the sinusoidals in (27) consist of the carrier frequency and the information signal. The carrier frequency is generated by applying a suitable time-invariant voltage to the VCO, whereas the information is applied via the control voltage, $v_1(t)$. One illustrative way to understand the difference between the information signal and the carrier frequency is that the information signal changes the amplitude and phase of the signal vector in the I-Q diagram, as the carrier frequency rotates the complete coordinate system. The rotation of the I-Q diagram itself does not add any information, and is simply a translation of the information in frequency.

The VCO is characterized by its sensitivity, K_{VCO}, near its operating point, and the transfer of the VCO is given by, $\omega_{\mathrm{VCO}} = v_{\mathrm{VCO}}K_{\mathrm{VCO}}$. It can be concluded that by differentiating the phase of the information signal with respect to time, explicit expressions for the control equations to the VCOs are obtained as,

$$\begin{aligned}\omega_{\mathrm{VCO1,2}} &= \frac{d\phi(t)}{dt} \pm \frac{d\alpha(r(t))}{dt} = \frac{d\phi}{dt} \pm \frac{d\alpha}{dr}\cdot\frac{dr}{dt} \\ &= \dot{\phi}(t) \pm \frac{d\left(\arccos\left(\frac{r(t)}{2r_{\max}}\right)\right)}{dr}\cdot\frac{dr}{dt} = \dot{\phi}(t) \mp \frac{\dot{r}(t)}{\sqrt{4r_{\max}^2 - r(t)^2}} \quad (30)\end{aligned}$$

By applying the sensitivity relation of the VCO to (30) the time-varying control signals on polar form arise as,

$$v_1(t) = \frac{\dot{\phi}(t)}{K_{\mathrm{VCO}}} - \frac{\dot{r}(t)}{K_{\mathrm{VCO}}\sqrt{4r_{\max}^2 - r(t)^2}} \quad (31)$$
$$v_2(t) = \frac{\dot{\phi}(t)}{K_{\mathrm{VCO}}} + \frac{\dot{r}(t)}{K_{\mathrm{VCO}}\sqrt{4r_{\max}^2 - r(t)^2}}. \quad (32)$$

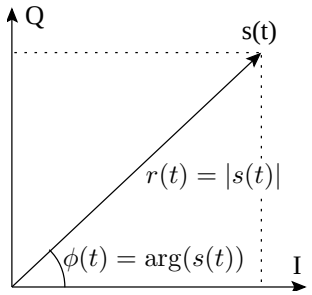

Figure 28: Signal representation in a Cartesian coordinate system.

The form of the representation, i.e., polar or Cartesian, of the control equations should be matched with the intended implementation. This will be discussed in more detail at the end of this section. The baseband signal can equally be represented on Cartesian form, see Figure 28, and the baseband signal can be decomposed according to,

$$s(t) = I(t) + jQ(t) \tag{33}$$
$$s_1(t) + s_2(t) = I_1(t) + jQ_1(t) + I_2(t) + jQ_2(t). \tag{34}$$

The in-phase (I) and quadrature-phase (Q) component are related to the polar representation as,

$$I(t) = r(t)\cos(\phi(t)) \tag{35}$$
$$Q(t) = r(t)\sin(\phi(t)). \tag{36}$$

Now, express the polar representation on its equivalent Cartesian form,

$$r(t) = \sqrt{I^2(t) + Q^2(t)} \tag{37}$$
$$\phi(t) = \arctan\left(\frac{Q(t)}{I(t)}\right). \tag{38}$$

Since the equations for the VCO control voltage already are derived in (31) and (32), these equations on Cartesian form are found via a change of basis. Before giving the complete control equation we start by deriving $\partial r/\partial t$, which is given by,

$$\begin{aligned}
\frac{\partial r}{\partial t} &= \frac{\partial r}{\partial I}\frac{dI}{dt} + \frac{\partial r}{\partial Q}\frac{dQ}{dt} \\
&= \frac{1}{2\sqrt{I^2(t)+Q^2(t)}}2I(t)\dot{I}(t) + \frac{1}{2\sqrt{I^2(t)+Q^2(t)}}2Q(t)\dot{Q}(t) \\
&= \frac{I(t)}{\sqrt{I^2(t)+Q^2(t)}}\dot{I}(t) + \frac{Q(t)}{\sqrt{I^2(t)+Q^2(t)}}\dot{Q}(t) \\
&= \frac{I(t)}{r(t)}\dot{I}(t) + \frac{Q(t)}{r(t)}\dot{Q}(t).
\end{aligned} \tag{39}$$

The expressions compact rather nicely and the phase time derivative, $\partial\phi/\partial t$, yields,

$$\begin{aligned}
\frac{\partial\phi}{\partial t} &= \frac{\partial\phi}{\partial I}\frac{dI}{dt} + \frac{\partial\phi}{\partial Q}\frac{dQ}{dt} \\
&= \frac{1}{1+\left(\dfrac{Q(t)}{I(t)}\right)^2}\left(-\frac{Q(t)}{I^2(t)}\right)\dot{I}(t) + \frac{1}{1+\left(\dfrac{Q(t)}{I(t)}\right)^2}\frac{1}{I(t)}\dot{Q}(t) \\
&= -\frac{Q(t)}{I^2(t)+Q^2(t)}\dot{I}(t) + \frac{I(t)}{I^2(t)+Q^2(t)}\dot{Q}(t) \\
&= -\frac{Q(t)}{r(t)^2}\dot{I}(t) + \frac{I(t)}{r(t)^2}\dot{Q}(t). \qquad (40)
\end{aligned}$$

By combining the results from (39) and (40) with the control equation in (31) the control voltage to VCO_1 is then given by,

$$\begin{aligned}
v_1(t) &= \frac{\dot{\phi}(t)}{K_{\text{VCO}}} - \frac{\dot{r}(t)}{K_{\text{VCO}}\sqrt{4r_{\max}^2 - r(t)^2}} \\
&= -\frac{Q(t)}{K_{\text{VCO}}r(t)^2}\dot{I}(t) + \frac{I(t)}{K_{\text{VCO}}r(t)^2}\dot{Q}(t) \\
&\quad -\frac{I(t)}{K_{\text{VCO}}r(t)\sqrt{4r_{\max}^2 - r(t)^2}}\dot{I}(t) - \frac{Q(t)}{K_{\text{VCO}}r(t)\sqrt{4r_{\max}^2 - r(t)^2}}\dot{Q}(t) \\
&= \dot{I}(t)\left(-\frac{Q(t)}{K_{\text{VCO}}r(t)^2} - \frac{I(t)}{K_{\text{VCO}}r(t)\sqrt{4r_{\max}^2 - r(t)^2}}\right) \\
&\quad +\dot{Q}(t)\left(\frac{I(t)}{K_{\text{VCO}}r(t)^2} - \frac{Q(t)}{K_{\text{VCO}}r(t)\sqrt{4r_{\max}^2 - r(t)^2}}\right). \qquad (41)
\end{aligned}$$

The sensitivity of the VCO is usually given in (Hz/V), and this unit is also used by the simulator. Collecting the control voltages for the VCOs, when using $2\pi K_{\text{VCO},f} = K_{\text{VCO}}$ yields,

$$\begin{aligned}
v_1(t) &= \frac{\dot{I}(t)}{2\pi K_{\text{VCO},f}}\left(-\frac{Q(t)}{r(t)^2} - \frac{I(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}}\right) \\
&\quad +\frac{\dot{Q}(t)}{2\pi K_{\text{VCO},f}}\left(\frac{I(t)}{r(t)^2} - \frac{Q(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}}\right) \qquad (42) \\
v_2(t) &= \frac{\dot{I}(t)}{2\pi K_{\text{VCO},f}}\left(-\frac{Q(t)}{r(t)^2} + \frac{I(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}}\right) \\
&\quad +\frac{\dot{Q}(t)}{2\pi K_{\text{VCO},f}}\left(\frac{I(t)}{r(t)^2} + \frac{Q(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}}\right). \qquad (43)
\end{aligned}$$

The appropriate control signals to the VCOs, $v_1(t)$ and $v_2(t)$, are either the polar representation given in equations (31) and (32), or the Cartesian form in (42) and (43). However, if the chosen set of control equations imply a polar feedback implementation, it should be noted that the amplitude and phase will be detected in two completely different paths, with different and unknown delays. When the fed back amplitude and phase information is not synchronized in time the spectral performance degrades similar to the EER technique (section 2.5.1) and the polar modulation feedback (section 2.4.5) with delay mismatch. By instead using a Cartesian representation, and integration on silicon, the nominally identical I and Q paths will have much better matched delays.

There is a complication in the control equations, namely the differentiation of the complex-valued input signal with respect to time. These time derivatives are difficult to implement in electronic circuits, (e.g., due to noise and stability hazard). An approximate solution to these time derivatives is presented next in section 4.3.

4.3 Approximation of the Control Signal

The open-loop control equations for the VCOs contain time derivatives. As mentioned before these are hard to implement and an approximation would be desired for the closed-loop control equations. The search for a closed expression of the control voltages, $v_1(t)$ and $v_2(t)$, starts by approximating the time derivative of the input signal with a difference ratio. The CALLUM system is assumed to only suffer from a small time delay, τ, around the loop. Then the time derivative is approximated by,

$$\dot{x}(t) = \lim_{\tau \to 0} \frac{x(t) - x(t-\tau)}{\tau}. \tag{44}$$

The I component of the input signal is denoted $I_i(t)$. The output signal $I_o(t-\tau)$ is somewhat delayed and the downconversion of the signal from the antenna can be considered as a linear attenuation with transfer, $1/A_{t\infty}$. The approximative time derivative of the in-phase signal is given by,

$$\dot{I}(t) \simeq \frac{I_i(t) - \dfrac{I_o(t-\tau)}{A_{t\infty}}}{\tau}. \tag{45}$$

By introducing a new parameter $k_1 = 1/\tau$ (Hz), the sought time derivative is approximated $\dot{I}(t)) \simeq k_1\{I_i(t) - I_o(t-\tau)/A_{t\infty}\}$. A reasonable assumption is that the delay around the loop is frequency independent for narrowband signals, meaning that k_1 is a constant gain factor.

4.3.1 Control Equations for CALLUM 1

In section 4.2 and 4.3 the preparations were made to formulate the original set of control equations for the CALLUM transmitter architecture. The original version is

referred to as CALLUM 1 and is based on the assumption that the system only suffers from a small time delay around the loop, and in all other aspects behave as an ideal system. The derived control equations (46) and (47) are recognized as the control law for CALLUM 1, and are based on the combination of (42), (43) and (45). As will be shown later this original set of equations are the foundation of numerous derivatives of the CALLUM linear transmitter.

$$\begin{aligned} v_{1,\mathrm{C1}} &= \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left\{ \left(I_i(t) - \frac{I_o(t-\tau)}{A_{t\infty}} \right) \left(-\frac{Q_i(t)}{r(t)^2} - \frac{I_i(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}} \right) \right. \\ &\quad \left. + \left(Q_i(t) - \frac{Q_o(t-\tau)}{A_{t\infty}} \right) \left(\frac{I_i(t)}{r(t)^2} - \frac{Q_i(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}} \right) \right\} \end{aligned} \tag{46}$$

$$\begin{aligned} v_{2,\mathrm{C1}} &= \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left\{ \left(I_i(t) - \frac{I_o(t-\tau)}{A_{t\infty}} \right) \left(-\frac{Q_i(t)}{r(t)^2} + \frac{I_i(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}} \right) \right. \\ &\quad \left. + \left(Q_i(t) - \frac{Q_o(t-\tau)}{A_{t\infty}} \right) \left(\frac{I_i(t)}{r(t)^2} + \frac{Q_i(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}} \right) \right\} \end{aligned} \tag{47}$$

Since the closed system in a real implementation includes blocks that are nonlinear, maximum allowable loop gain is wanted for optimum performance. Therefore, a new gain factor, g_1, is introduced that is used for adjusting the loop gain. Large loop gain will increase the signal tracking capability as well as the linearity of the closed-loop system.

The control law for CALLUM 1 includes a continuous amplitude scaling to allow a constant loop gain, whereas other derivatives will have varying loop gain over the valid complex-valued input range. The CALLUM 1 version is considered as the best performing derivative. The loop gain characteristic is discussed further in section 4.5.

Polar representation of CALLUM 1

The last part of this section contains a derivation of the CALLUM 1 control equation on polar form. Minor rewriting of (46) using $\omega_{\mathrm{VCO}} = \dot{\phi} = 2\pi f_{\mathrm{VCO}} = 2\pi K_{\mathrm{VCO},f} v_{\mathrm{VCO}}$ yields,

$$\begin{aligned} \dot{\phi}_{1,\mathrm{C1}} &= g_1 k_1 \left\{ \left(I_i - \frac{I_o}{A_{t\infty}} \right) \left(-\frac{Q_i}{r^2} - \frac{I_i}{r\sqrt{4r_{\max}^2 - r^2}} \right) \right. \\ &\quad \left. + \left(Q_i - \frac{Q_o}{A_{t\infty}} \right) \left(\frac{I_i}{r^2} - \frac{Q_i}{r\sqrt{4r_{\max}^2 - r^2}} \right) \right\}. \end{aligned} \tag{48}$$

Explicit notification of the variables' time dependency is omitted for the sake of clarity. The input vector is represented by (r, ϕ), whereas the downconverted output vector at the input of the SCG is given by (w, θ). The in-phase signal is given by, $I = r\cos\phi$,

and the quadrature signal, $Q = r\sin\phi$. The error signals are given by,

$$\Delta I = I_i - \frac{I_o}{A_{t\infty}} = r\cos\phi - w\cos\theta = r\cos\phi - w\cos(\phi - \Delta\phi) \quad (49)$$

$$\Delta Q = Q_i - \frac{Q_o}{A_{t\infty}} = r\sin\phi - w\sin\theta = r\sin\phi - w\sin(\phi - \Delta\phi), \quad (50)$$

where $\theta = \phi - \Delta\phi$. Now, apply the following variable substitution,

$$a = r\sqrt{4r_{\max}^2 - r^2} \quad (51)$$

$$b = r^2. \quad (52)$$

Revisiting (48) yields,

$$\begin{aligned}
\dot{\phi}_{1,\mathrm{C1}} &= g_1 k_1 \left\{ \left(I_i - \frac{I_o}{A_{t\infty}} \right) \left(-\frac{I_i}{a} - \frac{Q_i}{b} \right) + \left(Q_i - \frac{Q_o}{A_{t\infty}} \right) \left(\frac{I_i}{b} - \frac{Q_i}{a} \right) \right\} \\
&= g_1 k_1 \left\{ (r\cos\phi - w\cos(\phi - \Delta\phi)) \left(-\frac{r\cos\phi}{a} - \frac{r\sin\phi}{b} \right) \right. \\
&\quad \left. + (r\sin\phi - w\sin(\phi - \Delta\phi)) \left(\frac{r\cos\phi}{b} - \frac{r\sin\phi}{a} \right) \right\} \\
&= -g_1 k_1 r \sqrt{\frac{1}{a^2} + \frac{1}{b^2}} \left\{ (r\cos\phi - w\cos(\phi - \Delta\phi))\cos(\phi - \varphi_1) \right. \\
&\quad \left. -(r\sin\phi - w\sin(\phi - \Delta\phi))\cos(\phi - \varphi_2) \right\} \\
&= -g_1 k_1 r \sqrt{\frac{1}{a^2} + \frac{1}{b^2}} \\
&\quad \left\{ w[-\cos(\phi - \Delta\phi)\cos(\phi - \varphi_1) + \sin(\phi - \Delta\phi)\cos(\phi - \varphi_2)] \right. \\
&\quad \left. -r[-\cos\phi\cos(\phi - \varphi_1) + \sin\phi\cos(\phi - \varphi_2)] \right\}. \quad (53)
\end{aligned}$$

In order to further simplify the expression in (53), the following relations are used,

$$\begin{aligned}
\varphi_1 &= \arctan\frac{a}{b} \\
\varphi_2 &= \arctan\frac{-b}{a} = -\arctan\frac{b}{a}.
\end{aligned}$$

By studying a right-angled triangle it can be found that $-\arctan\frac{b}{a} = -\frac{\pi}{2} + \arctan\frac{a}{b}$ (which is equal to $\varphi_2 = -\frac{\pi}{2} + \varphi_1$). These relations are put in (53) and the equation is modified such the relation $\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$ appears. The same

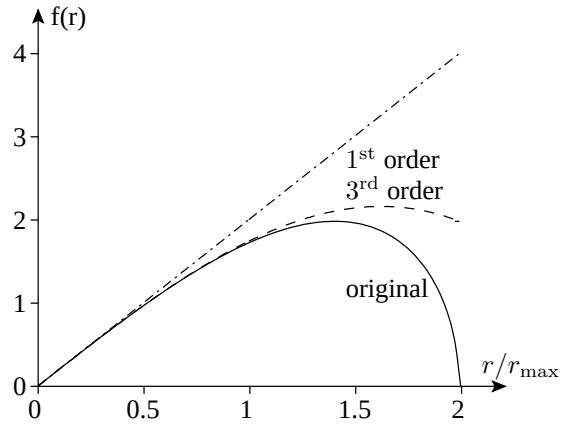

Figure 29: Comparison of a 1^{st} and 3^{rd} order approximation of the original square-root expression, $r(t)\sqrt{4r_{\max}^2 - r^2(t)}$.

calculation holds for $\dot{\phi}_2$ with minor changes.

$$\dot{\phi}_1 = g_1 k_1 r \sqrt{\frac{1}{a^2} + \frac{1}{b^2}} \left\{ w\cos(\Delta\phi - \varphi_1) - r\cos\varphi_1 \right\} \tag{54}$$

$$\dot{\phi}_2 = -g_1 k_1 r \sqrt{\frac{1}{a^2} + \frac{1}{b^2}} \left\{ w\cos(\Delta\phi + \varphi_1) - r\cos\varphi_1 \right\} \tag{55}$$

4.3.2 CALLUM 1 with Linearized Denominator

The control equations, (46) and (47), for CALLUM 1 contain summation, subtraction, multiplication, division and square-root operators. The square-root term is the most demanding computational challenge, as well as its complicated circuit implementation. The strive for simplifications of the realization of the SCG, the square-root term can be expressed by a Taylor series, and an approximation of the square-root term in the denominator taken for $r = 0$ yields,

$$r\sqrt{4r_{\max}^2 - r^2} = 2r_{\max}r - \frac{1}{4r_{\max}}r^3 + O(r^5). \tag{56}$$

The first and third order approximation are shown together with the original square-root expression in Figure 29. The matching of the first order approximation to the original square-root term is fairly good when $r \leq r_{\max}/2$, but beyond this point the curves diverge and even the sign of the derivative differs for large values of r. If the cost increase can be handled when adding another term in the Taylor expansion, the approximation gets significantly better.

The control equations for CALLUM 1 with linearized denominator, hereafter CALLUM 1lin, are based on a first order approximation of the square-root term. The denominator of the control equation has been simplified, but still r and r^2 have to be evaluated and a division performed. The control equations for CALLUM 1lin are given by,

$$\begin{aligned} v_{1,\mathrm{C1lin}} &= \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left\{ \left(I_i - \frac{I_o}{A_{t\infty}} \right) \left(-\frac{Q_i}{r^2} - \frac{I_i}{2r_{\max} r} \right) \right. \\ &\quad \left. + \left(Q_i - \frac{Q_o}{A_{t\infty}} \right) \left(\frac{I_i}{r^2} - \frac{Q_i}{2r_{\max} r} \right) \right\} \qquad (57) \\ v_{2,\mathrm{C1lin}} &= \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left\{ \left(I_i - \frac{I_o}{A_{t\infty}} \right) \left(-\frac{Q_i}{r^2} + \frac{I_i}{2r_{\max} r} \right) \right. \\ &\quad \left. + \left(Q_i - \frac{Q_o}{A_{t\infty}} \right) \left(\frac{I_i}{r^2} + \frac{Q_i}{2r_{\max} r} \right) \right\}. \qquad (58) \end{aligned}$$

4.3.3 CALLUM 2

For an all-analog implementation of the control equations even the complexity of CALLUM 1lin can be hard to cope with. It is possible to make further simplifications for easier SCG implementation at the expense of degraded spectral performance capability of the transmitter. Since the division is an expensive operation the denominator terms of the CALLUM 1 control equations are disregarded altogether in CALLUM 2. The control equations (59), (60) lack amplitude scaling and as a consequence will the loop gain vary significantly over the valid input range. The control equations for CALLUM 2 are given by,

$$\begin{aligned} v_{1,\mathrm{C2}} &= \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left\{ \left(I_i - \frac{I_o}{A_{t\infty}} \right) \left(-Q_i - I_i \right) \right. \\ &\quad \left. + \left(Q_i - \frac{Q_o}{A_{t\infty}} \right) \left(I_i - Q_i \right) \right\} \qquad (59) \\ v_{2,\mathrm{C2}} &= \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left\{ \left(I_i - \frac{I_o}{A_{t\infty}} \right) \left(-Q_i + I_i \right) \right. \\ &\quad \left. + \left(Q_i - \frac{Q_o}{A_{t\infty}} \right) \left(I_i + Q_i \right) \right\}. \qquad (60) \end{aligned}$$

4.3.4 CALLUM 3

In a single loop feedback amplifier the input signal to the active part is the error signal formed by the difference between the input and the feedback signal. An intuitive control equation would be to translate this knowledge from feedback amplifier design and let the error signal constitute the major part of the control equation. This idea was originally presented by Bateman [76], where each branch in the forward path either

produce the error signal of the in-phase or the quadrature-phase signal. CALLUM 3 can also be seen as a further simplified version of CALLUM 2, and the control equations for CALLUM 3 are given by,

$$v_{1,\mathrm{C3,orig}} = \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left(I_i - \frac{I_o}{A_{t\infty}} \right) \tag{61}$$

$$v_{2,\mathrm{C3,orig}} = \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left(Q_i - \frac{Q_o}{A_{t\infty}} \right). \tag{62}$$

A known problem with the proposed control equations in (61) and (62) is that the basic structure, as it first was presented, is only stable in one quadrant [76]. To overcome the stability problem a solution was found by including a switching matrix such the following control equations hold [76, 95, 96],

$$v_{1,\mathrm{C3}} = \mathrm{sign}[Q_i] \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left(I_i - \frac{I_o}{A_{t\infty}} \right) \tag{63}$$

$$v_{2,\mathrm{C3}} = \mathrm{sign}[I_i] \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left(Q_i - \frac{Q_o}{A_{t\infty}} \right). \tag{64}$$

In a feedback system, continuous signals vouch for a smooth processing of the signal. Hence discrete jumps of the control signals should be avoided. The sign function in the control equations for CALLUM 3 forces the control signals to the VCOs to change abruptly, which limits the versatility of CALLUM 3.

4.3.5 Vector-Locked Loop

The vector-locked loop (VLL) is related to LINC and CALLUM, as the basic idea of signal decomposition is the same. The VLL is a feedback system, which reduces some of the fundamental problems with the LINC approach, e.g. branch mismatch, but on the other hand unavoidably diminishes the attainable linearization bandwidth. The idea for the control equations are illustrated in Figure 30. As usual for this family of linear transmitter architectures, the input signal, s, is decomposed into two constant-envelope signals, s_1 and s_2, as shown in Figure 30a). The gray vectors in b) indicate the position before a change in amplitude was made, and it can be seen that an increase of the magnitude of the input signal decreases the phase of s_1 as the phase of s_2 increases. In c) the input signal phase is increased and the constant-envelope vectors change their phase equally. In d) the combination of b) and c) is shown, where s_1 changes only a little and s_2 quite a bit due to the combination of phase and amplitude change add together.

The VLL is based on a polar representation as the change of the vectors are directly coupled to the amplitude and phase of the input signal. The idea was first proposed by Senderowicz et al. [97], and further modifications were made by Bateman et al. [77, 98]. The VLL transmitter is illustrated in Figure 31. From Figure 30 the control

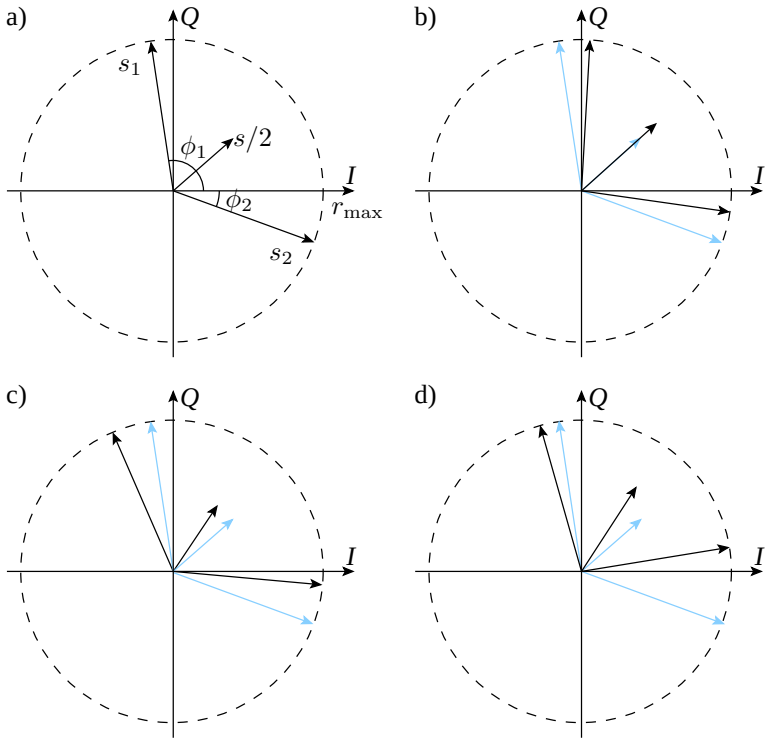

Figure 30: a) The input signal, s, and the constant envelope vectors, s_1 and s_2. b) An increase of input signal amplitude makes s_1 and s_2 to rotate in opposite direction. c) A change of the phase will equally change the phase of s_1 and s_2. d) A more realistic change during signal processing. The combination of amplitude and phase change is shown here.

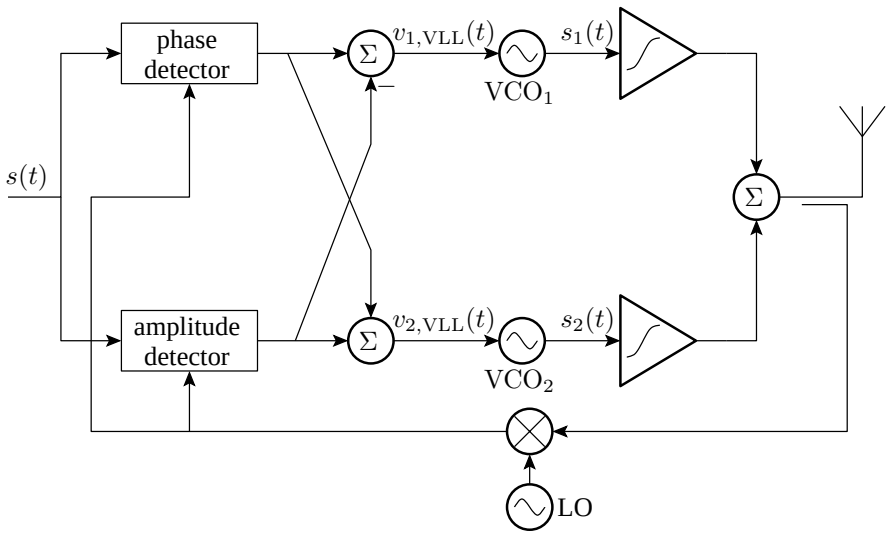

Figure 31: The vector-locked loop version of the linear transmitter principle.

equations for the VLL are formulated as,

$$v_{1,\mathrm{VLL}} = \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left(-\Delta r + \Delta\phi\right) \tag{65}$$

$$v_{2,\mathrm{VLL}} = \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left(\Delta r + \Delta\phi\right). \tag{66}$$

The loop forces the amplified output signal to accurately track the input signal provided that the loop gain is high enough. In the baseband frequency domain, the VLL is a low-pass circuit. The cut-off frequencies for the phase and magnitude are directly defined by the respective loop gains. For distortion introduced at the output, the loop forms a high-pass function. If a first-order approximation is used, the linearization capability at a fixed modulation frequency is increased by the same amount as the loop gain. These are basic properties of any feedback system, and the VLL is no exception [99].

For a certain linearization capability at a specific modulation bandwidth, the loop gain must be maximized. As the loop gain couples directly to the bandwidth of the loop, stability may be jeopardized in the presence of loop time delay. The loop delay is the most devastating issue in feedback linearization systems like VLL, and it should be effectively minimized [99]. It is generally known that in polar transmitters, where the information is transfered in the phase and magnitude of the signal, the corresponding system bandwidth must be considerably higher than for the baseband signal. Another problem is the nonlinear characteristic of the amplitude detector, which makes the

loop gain vary considerably with the input signal amplitude. Since the stability properties set the maximum allowed loop gain for the worst-case scenario, the system will work sub-optimal in large regions of the complex-valued input range. This, of course, effectively limits the linearization capability of the VLL. The system has to work at an IF as the phase detector cannot operate at DC. The phase must be defined relative to a known signal (reference), and one way to determine the phase is by measuring the time between a known signal and the information signal pass a reference level. By relating the time to the actual frequency, a measure of the phase can be achieved.

A prototype of the VLL transmitter built from off-the-shelf components by Kaunisto achieved 25 dB better distortion performance than the stand alone PA at the band edge of a 100 kHz $\pi/4$-shifted DQPSK modulation at 2.15 GHz center frequency.

4.4 Baseband Modeling of CALLUM

A baseband simulation is characterized by the absence of the carrier frequency. As a result, simulations based on a baseband model are far more efficient in terms of computational workload, but at the same time some information is discriminated. The dramatic reduction of the simulation time is the major drive behind the baseband model. In all CALLUM architectures an up- and down-conversion of the information signal are performed, which can be excluded in certain types of simulations. In this section we pave the way for deeper understanding of the investigation based on linearized control equations in section 4.5.

Since the carrier frequency in a baseband simulation is 0 Hz, the signal representation has to be in the complex-valued domain. A normal circuit simulator based on time domain analysis, e.g. Spectre, cannot handle complex-valued signals directly. This fact calls for a work-around. The carrier frequency, f_c, effectively only rotates the I-Q diagram and it can be removed without sacrificing the baseband information. The reference signal corresponding to 0° phase is aligned along the x-axis in an orthogonal xy-coordinate system. The real and imaginary part of the output signal from the VCO are represented by,

$$v_{\mathrm{real}} = 2r_{\max}\cos\left(2\pi\int_0^t K_{\mathrm{VCO},f}v_{in}(t)dt + \varrho\right) \tag{67}$$

$$v_{\mathrm{imag}} = 2r_{\max}\sin\left(2\pi\int_0^t K_{\mathrm{VCO},f}v_{in}(t)dt + \varrho\right), \tag{68}$$

where the initial angle is,

$$\varrho = \arccos\left(\frac{v_{in}(0)}{2r_{\max}}\right). \tag{69}$$

The value of the signal in present time depends on the initial condition and the control signal from the start. This representation is impractical why another definition of the

input signal, $s(t)$, and its relation to the constant envelope vectors, $s_1(t)$ and $s_2(t)$, are used.

$$\begin{aligned} s_1 &= \Re\{s_1\} + j\Im\{s_1\} \\ s_2 &= \Re\{s_2\} + j\Im\{s_2\} \\ s &= \Re\{s\} + j\Im\{s\} = s_1 + s_2. \end{aligned}$$

It is now easy to identify the I and Q components as,

$$I = \Re\{s_1\} + \Re\{s_2\} \quad (70)$$
$$Q = \Im\{s_1\} + \Im\{s_2\}. \quad (71)$$

Actually, from these equations the implementation of the baseband model of the system is made explicit. The baseband modeling of a generic CALLUM system is shown in Figure 32. The VCOs have a $\pi/2$ radians phase shift to accommodate the complex implementation. To guarantee this relative phase shift between the VCOs, they are implemented as mathematical functions in AHDL[8].

The baseband representation is of great importance (not only of academic interest) since the simulation time is reduced to only a fraction of the time it would take to do the same simulation with the RF frequency present. For example, complete bursts of transmitted data can be easily simulated with the baseband model. However, at a final system verification phase, all blocks have to be included and simulated at RF. The purpose with the final system simulations including the RF signal is to make sure the loop locks, and simple test signals are processed correctly. Even this limited simulation can be very time consuming.

4.5 Loop Properties Using Linearized Control Equations

Due to tighter demands on the error vector magnitude (EVM) for most future radio standards, highly linear power amplifiers are needed, especially in basestations. Today the combination of sufficiently backed-off amplifiers and a linearization scheme, such as feedforward or predistortion, is required to meet the linearity specifications [99]. In a feedback system the loop gain is the key parameter as it directly determines the amount of distortion suppression possible. The loop dynamics limit the maximum amount of loop gain, thereby the system itself (before applying the feedback) cannot be overly nonlinear if a certain spectrum emission mask should be fulfilled. The importance of the linearizing effect accomplished by the loop gain should not be underestimated, and as a consequence this section is devoted to loop gain calculations of the CALLUM transmitter.

[8] Analog Hardware Description Language is handled by many circuit simulators (Spectre in our case).

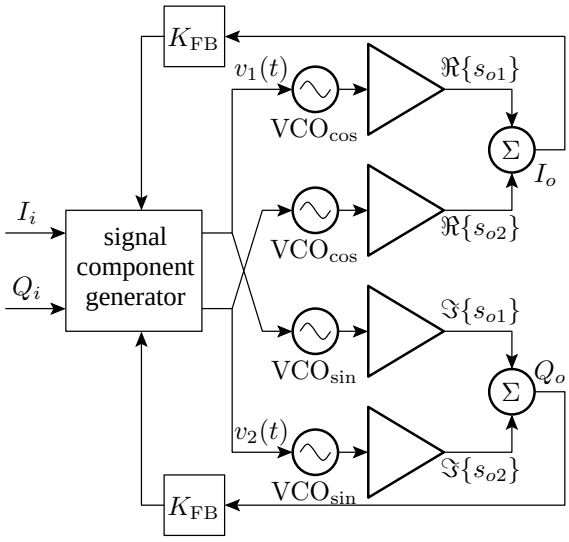

Figure 32: Baseband implementation of a general CALLUM system.

4.5.1 Signal Representation

The CALLUM system is governed by a set of nonlinear equations. The nonlinear description of the control equations, valid for any signal amplitude, gives little insight in terms of loop gain, bandwidth and stability. However, all these basic properties fall into place with a linear system. The nonlinear equations hamper the direct use of a very powerful mathematical tool, i.e., Laplace or Fourier transformation. The work-around is to linearize the equations prior Laplace transformation. The linearized equation approximates the actual behavior of the nonlinear equation in a limited region around a predetermined operating point.

Again the signals are described by the phase-modulation method, see section 3.2.1, and only relevant equations used in the following derivation of the linearized system characteristics are repeated here. The signal representation relations used when studying the loop gain in section 4.5 are based on Figure 33, and the relations are given by,

$$s_1(t) = r_{\max}\{\cos(\phi+\alpha) + j\sin(\phi+\alpha)\} \quad (72)$$

$$s_2(t) = r_{\max}\{\cos(\phi-\alpha) + j\sin(\phi-\alpha)\}, \quad (73)$$

where

$$\alpha(t) = \arccos\left(\frac{r(t)}{2r_{\max}}\right) \quad (74)$$

$$r(t) = \sqrt{I^2+Q^2}. \quad (75)$$

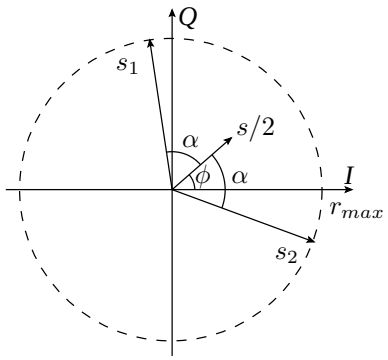

Figure 33: The definition of the angles used for the representation of the vectors.

The I and Q component can be identified from (72) as the real and imaginary part of s_1 and s_2, respectively. The I and Q component are defined by,

$$I = 2r_{\max}\cos(\phi)\cos(\alpha) \quad (76)$$

$$Q = 2r_{\max}\sin(\phi)\cos(\alpha), \quad (77)$$

where ϕ is,

$$\phi = \begin{cases} \arctan\left(\dfrac{Q}{I}\right) & if\ I \geq 0 \\ \arctan\left(\dfrac{Q}{I}\right) + \pi & if\ I < 0. \end{cases} \quad (78)$$

4.5.2 Loop Gain Calculations Based on Linearized Equations

In this section the baseband modeling of the CALLUM system will be unveiled. The results of the loop gain calculations will be displayed in the I-Q diagram for the most common derivatives of the CALLUM architecture, i.e., the continuous ones based on Cartesian feedback. Basically, it is only the control equations implemented in the SCG that differ between the CALLUM versions. Hence, the generic structure of the linearized baseband CALLUM model, displayed in Figure 34, is the same for all cases. As with phase-locked loops (PLLs), the relevant input and output signals are phases, while inside the loop different phase-to-voltage and voltage-to-phase conversions take place. The model in Figure 34 is roughly equivalent to half of the circuit in Figure 32, in that the cross-coupling between the I and Q part of the baseband equivalent has been neglected. This rather drastic simplification can be justified as follows: since the goal is to have a linear signal transfer, the required amount of loop gain is substantial,

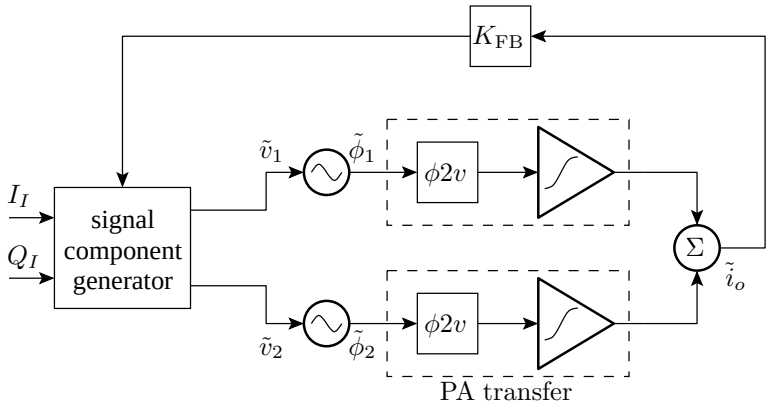

Figure 34: Reduced baseband model used for loop gain calculations.

and therefore the I (Q) component of the signal propagating through the Q (I) part of the system will be significantly attenuated[9]. This simplified model is only used for loop gain calculations. Closed-loop simulations are of course performed on the complete model of Figure 32.

The following notation is adopted to make the linearization procedure easy to follow: x_a is defined as the total signal, X_A its bias component, and $\tilde{x}_a$ its small-signal variation (i.e., $x_a = X_A + \tilde{x}_a$). We begin the linearization procedure by noting that the error signal for a constant input signal (I_I, Q_I) goes to zero as time goes to infinity, due to the presence of an integrator in the loop (it is well known from PLL theory that an oscillator behaves as an ideal integrator in the baseband PLL model [100]). At steady-state, the output signal will be equal to the input signal scaled by the asymptotic gain: $I_O = I_I A_{t\infty},\ Q_O = Q_I A_{t\infty}$. Now, in order to perform a loop gain analysis, all variations for the input signals are set to zero, whereas their bias components are fed to the system. The nonlinear system is linearized around this quiescent point, and the loop is opened at a suitable node (for instance, at the output of the signal adder) for open-loop calculations.

From the outputs of the VCOs to the output of the CALLUM architecture a phase-to-amplitude conversion takes place, indicated by $\phi 2v$ in Figure 34. The trigonometric function performed by the PA and the combiner is a projection of the output signal from the PAs onto the I-axis, as shown in Figure 35. The in-phase output signal is given by

$$I_o = \Re\{s_{o1}\} + \Re\{s_{o2}\} = K_{\mathrm{PA}} V_{\mathrm{VCO}} (\cos\phi_1 + \cos\phi_2), \tag{79}$$

[9]The I (Q) component is seen as a disturbance by the Q (I) part of the system, and as such is largely suppressed by it.

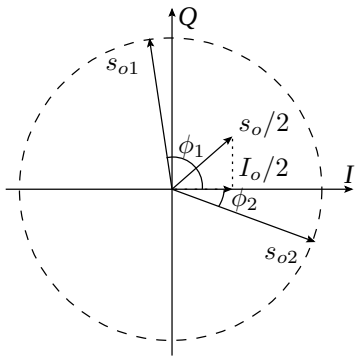

Figure 35: Projection of the output signal on the I-axis.

where $K_{\rm PA}$ is the transfer of the power amplifier (here considered as a constant), and $V_{\rm VCO}$ is the amplitude of the VCO signal. The baseband input signal level is determined by $r_{\max}$, which depends on the feedback gain, the PA gain, and the VCO gain ($r_{\max} = V_{\rm VCO}K_{\rm PA}K_{\rm FB}$). The transfer function for the VCO is given by $\tilde{\phi}_1 = K_{\rm VCO}\tilde{v}_1/s$, where s is the complex angular frequency. Straightforward linearization of (79) yields the output signal as

$$\tilde{i}_o = -K_{\rm PA}V_{\rm VCO}\left\{\sin(\Phi_1)\tilde{\phi}_1 + \sin(\Phi_2)\tilde{\phi}_2\right\}, \tag{80}$$

where $\Phi_1 = \Phi + \alpha$, $\Phi_2 = \Phi - \alpha$, and Φ and α are given by

$$\Phi = \arctan\left(\frac{Q_I}{I_I}\right) + n\pi \tag{81}$$

$$\alpha = \arccos\left(\frac{\sqrt{I_I^2 + Q_I^2}}{2r_{\max}}\right), \tag{82}$$

where $n = 0$ if $I_I \geq 0$, and $n = 1$ if $I_I < 0$. From the model in Figure 34 and using $V_{\rm VCO} = r_{\max}$, we can write the loop gain as,

$$A\beta = -\frac{K_{\rm FB}K_{\rm PA}K_{\rm VCO}r_{\max}}{s}\cdot\Big(\sin(\Phi_1)K_{\rm SCG,v_1} + \sin(\Phi_2)K_{\rm SCG,v_2}\Big), \tag{83}$$

where $K_{\rm SCG,v_1}$ ($K_{\rm SCG,v_2}$) is the SCG transfer relative to signal v_1 (v_2).

Comments on the Validity of the Model

When the system is tracking the input signal, a linearized model can be used when considering small changes in the in-phase or the quadrature-phase signal. For large

changes that occur during the acquisition process, only the governing set of nonlinear equations can accurately describe the system behavior [82]. In any CALLUM transmitter a signal variation on the in-phase signal will cause some unwanted signal on the quadrature-phase signal, which in turn will induce a signal on the in-phase signal. This loop goes on forever, but as there is attenuation the signal interaction decays rapidly. However, if the loop gain is assumed high, the interaction can be neglected. The model to investigate the loop gain assumes that there is no interaction between the signal paths.

This assumption is supported by simulations showing a small cross-coupling between the signal paths when the system has moderate to large loop gain. The applied signal variation on the I signal generates a much smaller variation on the Q signal (a few order of magnitude smaller). This small variation on the Q channel caused by the cross-coupling can be seen as a time-varying modulation of the bias, and will cause minor changes of the response. From simulations it was hardly detectable, unless entering the low loop gain region near the origin for the CALLUM 2 architecture.

4.5.3 Linearization of the Control Equations

In the expression for the loop gain (83) the transfers for the SCG, $K_{\mathrm{SCG,v_1}}$ and $K_{\mathrm{SCG,v_2}}$, have not been derived yet. To make the linearization easier to follow, we start with the control equations for CALLUM 2. With CALLUM 2 as a test vehicle the derivation of the linearized control equations for other CALLUM versions are straightforward.

CALLUM 2

The nonlinear control equations for CALLUM 2 are found in (59) and (60). Note that direct linearization of these equations include the effect of $K_{\mathrm{FB}} = 1/A_{t\infty}$, which is not part the SCG transfer. Therefore the result should be divided by K_{FB} in order to include the effect of the gain in the feedback path only once in the loop gain expression (83). The linearization of these equations results in,

$$\begin{aligned}
\tilde{v}_{1,\mathrm{C2}} &= \frac{g_1 k_1}{K_{\mathrm{VCO}}}\Big\{(-2I_I + K_{\mathrm{FB}}(I_O - Q_O))\tilde{i}_i + (-2Q_I + K_{\mathrm{FB}}(I_O + Q_O))\tilde{q}_i \\
&\quad + K_{\mathrm{FB}}(I_I + Q_I)\tilde{i}_o + K_{\mathrm{FB}}(-I_I + Q_I)\tilde{q}_o\Big\} \\
\tilde{v}_{2,\mathrm{C2}} &= \frac{g_1 k_1}{K_{\mathrm{VCO}}}\Big\{(2I_I - K_{\mathrm{FB}}(I_O + Q_O))\tilde{i}_i + (2Q_I + K_{\mathrm{FB}}(I_O - Q_O))\tilde{q}_i \\
&\quad + K_{\mathrm{FB}}(-I_I + Q_I)\tilde{i}_o + K_{\mathrm{FB}}(-I_I - Q_I)\tilde{q}_o\Big\}.
\end{aligned}$$

As indicated in Figure 34, only a small variation in the output signal, $\tilde{i}_o$, is considered. By investigating the properties of the system by varying a single variable at the time,

more insight is gained compared to a brute force multi-variable approach. Thus, the linearized transfers of the SCG are calculated in presence of the following signals: $\tilde{i}_i = 0$, $\tilde{q}_i = 0$, $\tilde{q}_0 = 0$, $Q_O = Q_I A_{t\infty}$, $I_O = I_I A_{t\infty}$.

$$K_{\mathrm{SCG,v_1,C2}} = \frac{g_1 k_1}{K_{\mathrm{VCO}}}(Q_I + I_I) \tag{84}$$

$$K_{\mathrm{SCG,v_2,C2}} = \frac{g_1 k_1}{K_{\mathrm{VCO}}}(Q_I - I_I) \tag{85}$$

The transfers of the SCG for CALLUM 1 and CALLUM 1lin are found in an analogous way. These results are reported below.

CALLUM 1

With the same assumptions as above the transfers of the SCG for CALLUM 1 are given by,

$$K_{\mathrm{SCG,v_1,C1}} = \frac{g_1 k_1}{K_{\mathrm{VCO}}}\left(\frac{Q_I}{r^2} + \frac{I_I}{r\sqrt{4r_{\max}^2 - r^2}}\right) \tag{86}$$

$$K_{\mathrm{SCG,v_2,C1}} = \frac{g_1 k_1}{K_{\mathrm{VCO}}}\left(\frac{Q_I}{r^2} - \frac{I_I}{r\sqrt{4r_{\max}^2 - r^2}}\right). \tag{87}$$

Here r is the amplitude of the signal taken in the operating point. Hence given by,

$$r = \sqrt{I_I^2 + Q_I^2}. \tag{88}$$

CALLUM 1 with linearized denominator

The transfer for the SCG for CALLUM 1 with linearized denominator is reported in analogy with the results for CALLUM 1 above.

$$K_{\mathrm{SCG,v_1,C1lin}} = \frac{g_1 k_1}{K_{\mathrm{VCO}}}\left(\frac{Q_I}{r^2} + \frac{I_I}{2r_{\max} r}\right) \tag{89}$$

$$K_{\mathrm{SCG,v_2,C1lin}} = \frac{g_1 k_1}{K_{\mathrm{VCO}}}\left(\frac{Q_I}{r^2} - \frac{I_I}{2r_{\max} r}\right) \tag{90}$$

4.5.4 Loop Gain Characteristics

All information to express the loop gain for any of the CALLUM versions is now available. The loop gain is found by inserting the transfers of the SCG in the generic expression for the loop gain (83). The fairly complicated loop gain expression prevents a direct interpretation; hence, the absolute value of the loop gain is plotted in the IQ-diagram, which gives a quick indication of the expected performance of the chosen implementation.

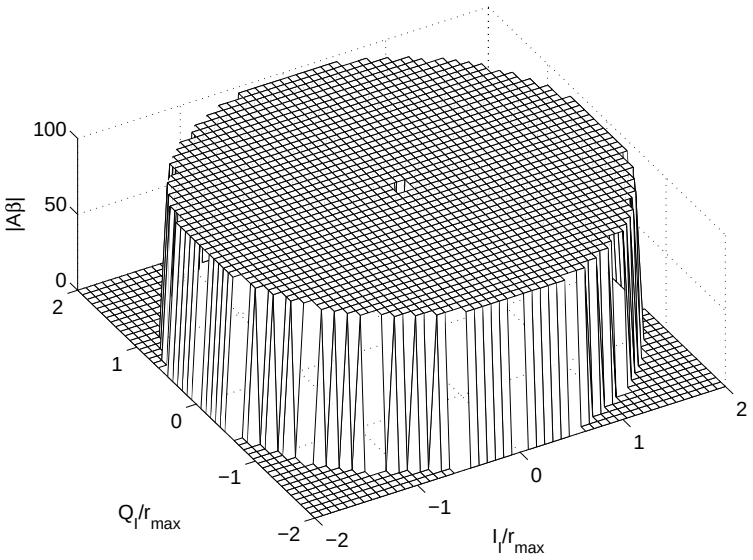

Figure 36: The loop gain for a CALLUM 1 system implementation. The loop gain refers to the in-phase component variation plotted at a fix frequency. The loop gain is set to zero outside the valid input range.

Loop Gain for CALLUM 1

The complete loop gain expression is given by,

$$A\beta_{\mathrm{C1}} = -\frac{K_{\mathrm{FB}}K_{\mathrm{PA}}K_{\mathrm{VCO}}r_{\max}}{s}\cdot\frac{g_1k_1}{K_{\mathrm{VCO}}}\left\{\sin(\Phi_1)\left(\frac{Q_I}{r^2}+\frac{I_I}{r\sqrt{4r_{\max}^2-r^2}}\right)\right.$$
$$\left.+\sin(\Phi_2)\left(\frac{Q_I}{r^2}-\frac{I_I}{r\sqrt{4r_{\max}^2-r^2}}\right)\right\}. \qquad (91)$$

It is possible to cancel the term K_{VCO} in (91), but when implementing the SCG on silicon the term g_1k_1/K_{VCO} is treated as a constant gain. Cancelling would be misleading and non-intuitive as the loop gain would appear as independent of the VCO gain, which it is not.

The equation describing the loop gain is not easy to interpret. Figure 36 shows the magnitude of the loop gain for CALLUM 1 for the in-phase component, plotted at a fix frequency. The maximum value of the loop gain is set to 100 to facilitate comparison between the CALLUM versions. Due to the loop gain normalization, the chosen frequency is arbitrary.

The loop gain for CALLUM 1 is constant over the whole valid input range $0 < r_i \leq 2r_{\max}$ in the IQ-plane, which is not obvious from the loop gain expression. This

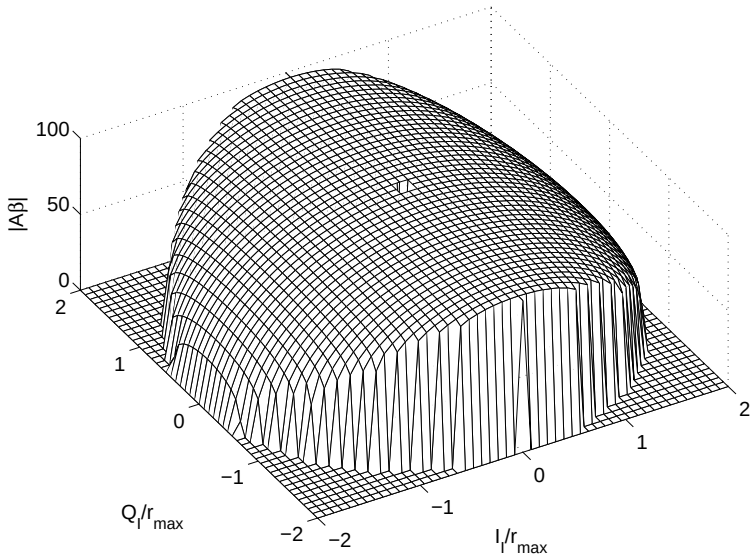

Figure 37: Loop gain for CALLUM 1 with linearized denominator.

is unique for the original CALLUM 1 control law, as it has a continuous amplitude normalization in the loop gain expression. In general, the loop gain is not defined for $r = 0$, and it shows a symmetry within each quadrant. For the quadrature-phase component variation, the loop gain will show the same pattern, rotated by $\pm 90^\circ$. The same information is contained in the in-phase and quadrature-phase plots. In the following we will not make this distinction.

Loop Gain for CALLUM 1 with Linearized Denominator

By using a first-order approximation of $r\sqrt{4r_{\max}^2 - r^2} = 2r_{\max}r$ the loop gain will be slightly different for large values of r as the difference between the original term and the approximation is noticeable, see Figure 29. The loop gain expression for CALLUM 1lin is given by,

$$\begin{aligned} A\beta_{\mathrm{C1lin}}(s) &= -\frac{K_{\mathrm{FB}}K_{\mathrm{PA}}K_{\mathrm{VCO}}r_{\max}}{s} \cdot \frac{g_1 k_1}{K_{\mathrm{VCO}}} \left\{ \sin(\Phi_1) \left(\frac{Q_I}{r^2} + \frac{I_I}{2r_{\max}r} \right) \right. \\ &\quad \left. + \sin(\Phi_2) \left(\frac{Q_I}{r^2} - \frac{I_I}{2r_{\max}r} \right) \right\}. \end{aligned} \tag{92}$$

This first-order approximation of the square-root term is expected to be accurate when $I_I \ll r_{\max}$. Thus, for small values of r the loop gain is expected to be similar as for CALLUM 1. The 3D loop gain plot for CALLUM 1lin is shown in Figure 37,

where the loop gain clearly drops for large values of I_I. A more accurate approximation, i.e., a higher order Taylor expansion, would reduce the drop in loop gain at the expense of increased circuit complexity.

Loop Gain for CALLUM 2

Both CALLUM 1 and CALLUM 1lin have high implementation costs, due to the relatively complicated mathematical operations (square root and/or division) that have to be performed. The strive for reduced complexity resulted in CALLUM 2, where the price to be payed is small loop gain (or lack thereof) at low signal levels, and a strongly varying loop gain within the amplitude interval of interest $0 < r_i \leq 2r_{\max}$, as shown in Figure 38. The valid input range is limited at the lower end by the bandwidth of the closed-loop system [86], and at the upper end by signal handling capability, i.e., by the onset of clipping. The loop gain for CALLUM 2 is given by,

$$\begin{aligned} A\beta_{C2}(s) &= -\frac{K_{\mathrm{FB}}K_{\mathrm{PA}}K_{\mathrm{VCO}}r_{\max}}{s} \cdot \frac{g_1 k_1}{K_{\mathrm{VCO}}} \\ &\quad \left\{ \sin(\Phi_1)(Q_I + I_I) + \sin(\Phi_2)(Q_I - I_I) \right\}. \end{aligned} \tag{93}$$

The loop gain for each point in the IQ-plane is shown in Figure 38. The information available in the 3D contour plots of the loop gain allows the investigation of system stability and the related design of frequency compensation. A prediction where the loop gain takes its largest value in the IQ-plane will serve as the point selected for compensation purposes, since this is the worst-case scenario in feedback systems. Clearly, in all CALLUM versions maximum loop gain is achieved when $Q_I = 2r_{\max}$. Thus, frequency compensation must be performed in presence of maximum envelope for the input signal.

4.5.5 Comparison of Loop Gain Characteristics

The loop gain information is important when designing the frequency compensation, and the worst-case scenario for the feedback system is achieved when $Q_I = 2r_{\max}$. Therefore, frequency compensation is best performed at almost maximum input envelope signal. Figure 39 shows a cut along the I- and Q-axis, respectively, for the investigated CALLUM derivatives. Noticeable is the low loop gain at low and intermediate envelopes for CALLUM 2.

CALLUM 1 Third-Order Approximation

So far CALLUM 1, CALLUM 1lin, and CALLUM 2 have been investigated in terms of loop gain characteristics in the IQ-plane. Which of the derivatives to choose for the implementation is basically a trade off between performance and cost. The complicated analog implementation of the control equations for CALLUM 1 can be exchanged for the simplified control equations (e.g., CALLUM 1lin or CALLUM 2) at

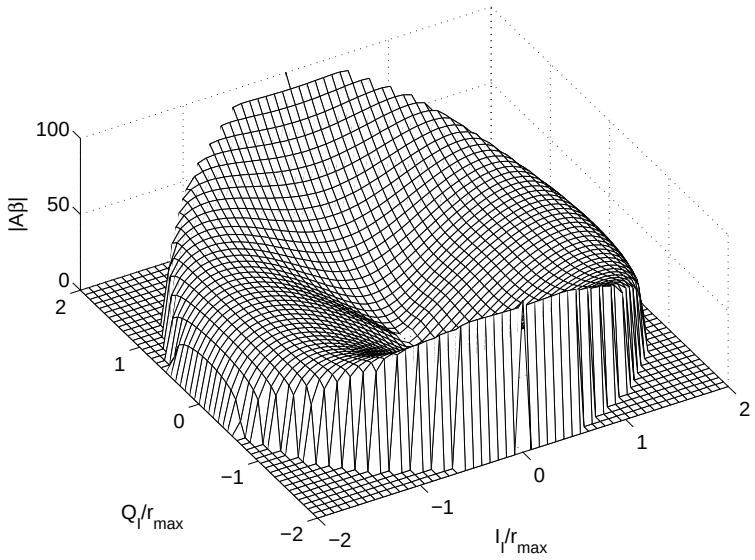

Figure 38: Loop gain 3D plot for CALLUM 2. The region with low loop gain for low envelope signal is one of the main drawbacks for CALLUM 2 compared to CALLUM 1.

the expense of reduced loop gain in some regions. From a designers point of view, it is interesting to find a control equation that makes the loop gain as homogeneous as possible over the IQ-plane, still with reasonable complexity. So far we have seen that CALLUM 1lin is a good compromise between performance and cost.

Now, assume that the design budget allows a control equation with slightly increased complexity compared to CALLUM 1lin. The linearized denominator term in (56) can be truncated after the third order term to better approximate the square-root expression, see Figure 29. As it is a refinement of CALLUM 1lin, a loop gain characteristic in between the original CALLUM 1 and its linearized successor is expected. By taking a closer look at Figure 29 the loop gain characteristic is expected to only differ for relatively large values of I_I, since the difference between the first- and third-order approximation is significant for large values of I_I. The square-root expression is not monotonous and cannot by any means be approximated over a large interval by a linear term. It turns out that the shape of the loop gain plot for CALLUM 1 third-order approximation is hardly distinguishable form Figure 37, but a cut of the loop gain contour plot along the I-axis for CALLUM 1 third-order approximation shows the difference compared to CALLUM 1lin, see Figure 40.

The loop gain improvement over CALLUM 1lin is significant for large values of I_I, but if the refined control equation pays off in noticeably higher performance is not

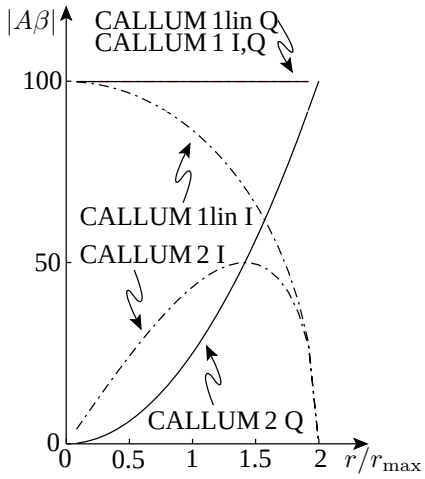

Figure 39: Loop gain comparison between three CALLUM implementations, i.e. CALLUM 1, CALLUM 1lin, and CALLUM 2. In the graph Q refers to the loop gain along the positive Q-axis in the IQ-plane.

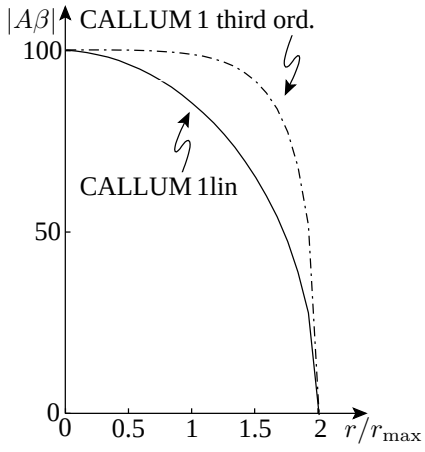

Figure 40: Loop gain comparison between CALLUM 1lin and CALLUM 1 with a third-order approximation of the square-root term.

completely clear. The complexity for CALLUM 1 third-order is somewhat reduced compare to CALLUM 1, but both these derivatives are most likely to be implemented in a digital SCG with pre-calculated control signals stored in a two-dimensional LUT. In that case the complexity can be considered the same.

4.5.6 Frequency Dependency of the Loop Gain

The loop gain 3D contour plots presented (e.g., Figure 36) are plotted at a fix frequency and normalized to have a peak loop gain magnitude of 100. However, for every point in the IQ-plane, the loop gain has a frequency dependency that contains very important information for the stability properties of the system. There are numerous ways to define the design space for the stability properties of the feedback system, where a step response with restricted over-shoot (time domain), a Bode plot without amplitude peaking (frequency domain), or direct pole-zero placement can be mentioned as commonly used design goals.

The frequency compensation is best performed in the Laplace domain, where the poles and zeros for the loop and the system can be identified. From the rules for root locus it is found that large loop gain is hazardous for the stability of the closed-loop system, and the worst-case for the system should be used when designing the frequency compensation. Consequently a set of input signals that results in high loop gain should be chosen. To assist the selection of a suitable set of input signals the 3D contour loop gain plots are useful (e.g., Figure 37). Remember that the loop gain plots are only for small in-phase variations. Normally a transmitted sequence of data will also exercise the quadrature component and the loop gain for a quadrature-phase variation is found by rotating the loop gain plot by $\pm 90^\circ$. With that in mind, a bias of the input signal close to maximum is appropriate for stability investigation and frequency compensation.

The VCO in the loop gives rise to a loop pole in the origin, since its transfer in the Laplace domain is $\phi(s) = vK_{\mathrm{VCO}}/s$. Unfortunately, this is not the only loop pole of the system. Significant contribution to the number of loop poles comes from both the SCG and the variable-gain amplifier (VGA) in the loop. Another very important phase degradation comes from a somewhat more diffuse mechanism, namely from the overall time delay around the loop. The delay depends on the physical size of the loop as well as the electrical delay of each component. Figure 41 shows a typical loop gain frequency dependency with one loop pole in the origin and a double pole at -100 Mrad/s. The pole placement maps well with the behavior seen from simulations of a CALLUM 2 system modeled at high level using a mixture of ideal blocks and transistors implementing various analog functions. The results from such simulations are principally interesting and will guide the designer through the complete implementation of a CALLUM transmitter.

Later on we will come back to the effect of loop delay (see section 4.9), but leave this very interesting topic for now. In chapter 5 the CALLUM 2 system is implemented on silicon and is investigated more thoroughly.

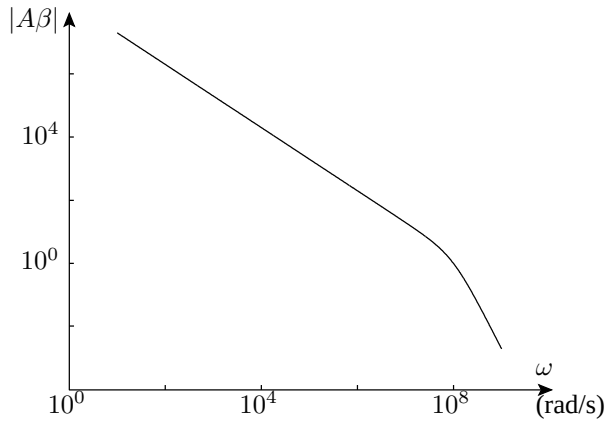

Figure 41: Loop gain plot with a loop pole in the origin from the VCO, and the SCG is assumed to contribute with one pole and the VGA with another, both placed at -100 Mrad/s.

4.6 The Asymptotic-Gain Model Applied on CALLUM

The asymptotic-gain model in Figure 10 accurately represents the implementation of the electronic signal processing function that the CALLUM system constitutes [25]. The input circuit loading effect, i.e., the attenuation and/or the signal domain transformation from voltage to current or vice versa. All this is put in the network parameter ξ. At the other end of the system the output circuit loading effect is represented by ν. The load signal is not possible to detect directly at the antenna, see Figure 27. Therefore, ν will mainly contain some attenuation of the output signal.

When modeling CALLUM the direct feed-through term ρ is set to zero, and if we for the sake of simplicity disregard the attenuation at both the input and output, i.e., $\xi = \nu = 1$, the asymptotic-gain is determined by the feedback factor β. The feedback path contains frequency translation by mixing and filtering together with other signal conditioning blocks. The attenuation in the feedback path is equal to the asymptotic-gain, A_{t_∞}.

The loop gain can be written on the form $A\beta = A\beta(0)\frac{N_{A\beta}}{D_{A\beta}}$, where the DC loop gain is extracted from the expression. The zeros of the loop gain are the roots of the numerator, $N_{A\beta} = 0$, and analogously the loop poles are the roots of the denominator, $D_{A\beta} = 0$. The denominator is written on the form,

$$D_{A\beta} = \left(1 - \frac{s}{p_1}\right)\left(1 - \frac{s}{p_2}\right)\cdots\left(1 - \frac{s}{p_n}\right), \tag{94}$$

where p_1 through p_n are the loop poles.

If the asymptotic-gain is frequency independent, then the system poles are given

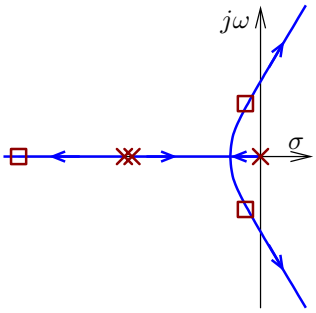

Figure 42: Root locus for a fictitious CALLUM 2 with three loop poles in the dominant group. The poles are positioned as follows: one in the origin and two poles at -100 Mrad/s.

by the roots of $D_{A\beta} - A\beta(0)N_{A\beta} = 0$. The exact positions of the system poles for higher order systems are hard or even impossible to determine by hand-calculations. However, to understand the frequency response the exact positions of the system poles are of minor interest, whereas the trajectory of the system poles are of major interest. The trajectory along which the system poles move with varying loop gain is better known as the root locus, see Figure 42. The figure shows that the desired frequency properties (maximum flat magnitude (MFM) or real system poles) are not possible to achieve without compensation. When the magnitude of the loop gain is zero there is no effect of the feedback and the system poles are positioned right on-top of the loop poles. When the loop gain is increased the system poles start to move along the root locus. For the third order system (all poles) shown in Figure 42 the system poles will penetrate the right half plane (RHP) as the loop gain is large enough. This is equivalent to an unstable system.

To get the desired stability properties, the root locus has to be altered by either changing the starting points of the loop poles or by changing the shape of the root locus. Several efficient frequency compensation techniques exist, e.g., phantom zero, pole split (e.g., local capacitive feedback, pole-zero cancellation), and local resistive feedback. Both resistive broad-banding and capacitive loading are compensation methods that should be handled with care, since their influence on either loop gain or bandwidth is significant [3, 22, 101].

A typical amplitude plot for the open- and closed-loop transfer before compensation is shown in Figure 43, assuming the same loop pole pattern as in Figure 42. The peaking in amplitude clearly shows that frequency compensation is needed.

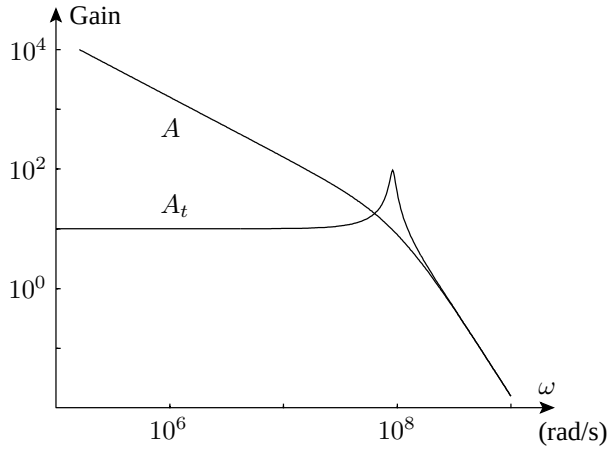

Figure 43: Open- (A) and closed-loop transfer (A_t) for CALLUM 2 without frequency compensation applied.

4.7 Bandwidth Estimation of a CALLUM Transmitter

Estimation of the bandwidth is highly desired, since the bandwidth has to accommodate the signal fed to the system. Depending on the modulation of the input signal it has a certain bandwidth, and due to the mapping of the linear modulation on two phase-modulated signals the effective bandwidth is increased, see section 4.11. The internal amplitude discrimination that appears within the CALLUM architecture results in increased bandwidth of the signals.

The characteristic polynomial (CP) of the closed-loop transfer, A_t, with an all-pole transfer (no zeros allowed) is given by,

$$\mathrm{CP}_{\mathrm{closed-loop}} = s^n - s^{n-1}\sum_{i=1}^{n} p_i + \cdots + (1 - A\beta(0))\prod_{i=1}^{n} p_i. \qquad (95)$$

The frequency behavior of the system is described by its CP and by changing the CP the system can be compensated.

Assume the bandwidth of the system should be maximized, meaning Butterworth position of the system poles if peaking in amplitude is not allowed. Butterworth position is also known as maximum flat magnitude (MFM) transfer. The poles are equidistantly placed on a half-circle in the left half plane (LHP) with the real axis as a line of symmetry [101]. The radius of the semi-circle equals the bandwidth of the transfer function. The characteristic polynomial for the system poles in Butterworth position

is given by,

$$\mathrm{CP}_{\mathrm{system}} = s^n - s^{n-1}\sum_{i=1}^{n} p_i^{'} + \cdots + \prod_{i=1}^{n} p_i^{'}. \tag{96}$$

The system poles are primed in order not to be interchanged with the loop poles. If the system poles can be positioned for MFM transfer the maximal bandwidth can be estimated by identification of the s^0 term of (95) and (96). The bandwidth estimation is also known as the loop gain poles (LP) product for obvious reasons.

$$\omega_0^n = \left| (1 - A\beta(0)) \prod_{i=1}^{n} p_i \right| \tag{97}$$

The bandwidth estimation is only valid for all-pole systems, and constitutes an estimation of the upper bound for the achievable bandwidth.

When dealing with a CALLUM transmitter the loop pole in the origin makes it incorrect to write $A\beta(s) = A\beta(0)\frac{N_{A\beta}}{D_{A\beta}}$, since $\lim_{s\to 0} A\beta = \infty$. In (97) there is also a conflict between infinite DC loop gain and a loop pole in the origin. Usually the information is restricted to a certain frequency band with a band-pass characteristic. When the lower bound of the information band, $\omega_{\mathrm{low}} \neq 0$, a possibility opens to pass this difficulty. Instead of having the pole in the origin, it can be assumed to be positioned at low frequency, still significantly below ω_{low}. Then a finite DC loop gain also exists. With this minor change the representation $A\beta = A\beta(0)\frac{N_{A\beta}}{D_{A\beta}}$ makes sense again and the maximum achievable bandwidth can be estimated.

For a realistic bandwidth estimation it must be possible to move the system poles to desired positions by means of frequency compensation. Only these poles that can be moved into Butterworth position, when the loop is closed, are called the dominant loop pole set. Let us start with all loop poles and estimate the MFM bandwidth. If $\Sigma p_i < \Sigma p_i^{'}$, then at least one non-dominant loop pole is included. Remove the most negative loop pole from the loop pole set and recalculate until $\Sigma p_i \geq \Sigma p_i^{'}$. The remaining loop poles make the dominant loop pole set.

Now apply the knowledge about the loop pole set on the CALLUM system. The MFM bandwidth is now straightforward to estimate,

$$\omega_0 = |(1 - A\beta(0))p_1 p_2 p_3|^{\frac{1}{3}} \simeq 100\,\mathrm{Mrad/s}. \tag{98}$$

The loop poles ($p_1,\ p_2, \ldots$) are the same as in Figure 44. When applying the check for the dominant loop pole set one of the high frequency poles belongs to the non-dominant group, since all established frequency compensation techniques only can make the loop pole sum more negative. The MFM bandwidth is estimated to, $\omega_0 = |(1 - A\beta(0))p_1 p_2|^{\frac{1}{2}} \simeq 100\,\mathrm{Mrad/s}$. This estimation is reasonable when inspecting Figure 43.

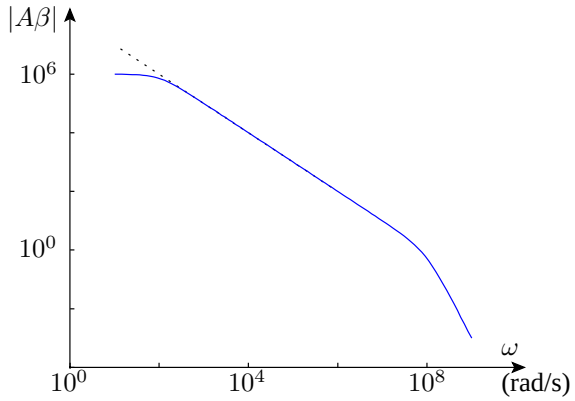

Figure 44: The pole from the VCO is moved to -100 rad/s, which can be assumed to be well below the lower bound of the information band. The SCG contributes with one pole and the VGA with another both placed at -100 Mrad/s. The DC loop gain is now well defined, $|A\beta(0)| = 10^6$.

4.8 Frequency Compensation of the CALLUM System

Feedback systems have several benefits, e.g., the reduced sensitivity of component parameter variations often in conjunction with matching, and the inherent property of reducing the nonlinearity can be mentioned as important features. The CALLUM architecture is to be considered as a large system including numerous block inside the loop. Not only does this fact involve a large number of loop poles, but also introduces a significant delay. The delay can be seen as an unfavorable phase shift, which limits the maximum possible loop gain for stable operation. In a completely integrated realization, where the drivers, power amplifiers and power combiner are on-chip the possibility for higher loop gain is opened. High loop gain feedback systems can handle more wideband input signals than a low loop gain system, for the same set of loop poles.

The model used for the system is simplified and a safety margin has to be used. The estimated bandwidth assumes the system poles to be in Butterworth position, but for increased robustness that is needed to withstand parameter variations, such as loop gain fluctuation, component spread, temperature, aging, unknown delay and other non-modeled parasitics the system poles should be placed on the real axis. As mentioned before one loop pole, which does not belong to the dominant loop pole set, must be cancelled since also non-dominant loop poles can affect the stability properties. A zero is put on top of the loop pole to be cancelled. However, the zero must be a phantom zero implemented in β, if the system should remain all-pole. The possible positions to implement this zero is at the input, the output or in the feedback network. This phantom zero is not really part of the compensation, but makes the bandwidth

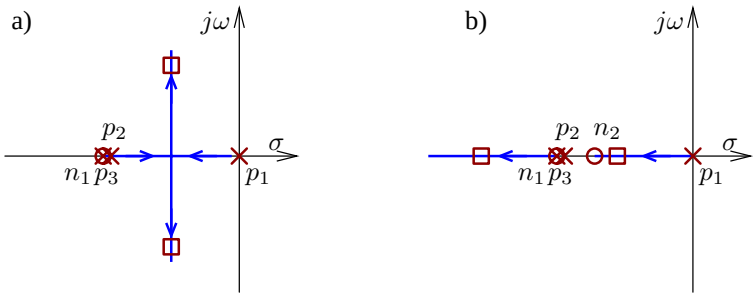

Figure 45: a) Elimination of the loop pole not part of the dominant loop pole set by a phantom zero. b) Further compensation for robustness by applying a second phantom zero. The phantom zeros are placed at -100 Mrad/s and -70 Mrad/s. Only real system poles are possible for this pole-zero map.

estimation valid. After this cancelling the system order is reduced by one, and in this example a second order system remains.

Frequency compensation is basically to match the CP for the uncompensated system with the CP for a system having its poles at the desired position. By making these CPs equal the closed-loop system will get the desired properties. If we take a closer look at the CP in (95) the constant term is already matched when choosing the Butterworth transfer characteristic. For other types, e.g., Bessel, the bandwidth will be reduced. The other terms, except s^n, have to be matched, which normally means that $n-1$ compensations have to be applied.

As mentioned above a phantom zero is placed to cancel one of the loop poles. The frequency happens to coincide with the peak in Figure 43. If no further compensation is applied the root locus allows complex system poles, which is undesired when designing for robustness. Another zero can be implemented at the second pole or at lower frequency. The maximal practical bandwidth for this type of compensation strategy is basically limited by the position of the second phantom zero. The frequency compensation strategy is illustrated in Figure 45.

4.8.1 Simulated Frequency Response for CALLUM 2

To get valuable information from the simulator regarding the stability properties, the amplitude plot from an AC analysis is needed. Before an AC simulation can be performed a stable DC solution of the system is required, but for a CALLUM system there is no such solution. This might sound strange at first, but when trying to answer the question 'what is the operating point of a VCO?' it probably becomes clearer. Since there is a VCO in the loop of any CALLUM system an ordinary DC simulation will not do, as for a simple DC simulation the VCO acts as a blocker. The output from the VCO is formed by the *input signal in time*. To find a stable DC solution a transient

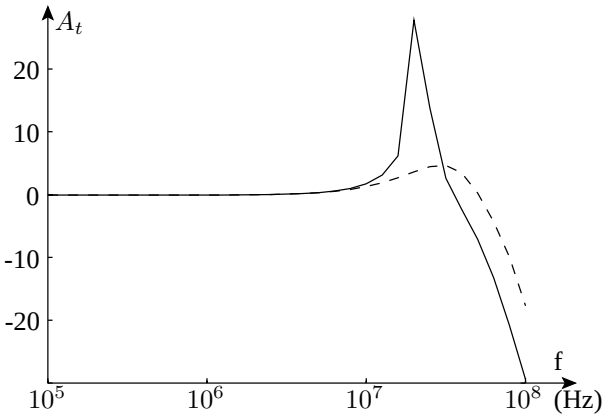

Figure 46: Transmittance plot of a CALLUM 2 feedback system before and after first order compensation using a phantom zero.

simulation is needed. However, there is a package in Cadence to cope with these kind of systems as they are common in radio electronics.

The simulator was controlled in such way that the closed-loop transfer could be extracted. Basically it was done by a periodic-steady-state (PSS) simulation, where the system was forced into a certain operating point in the IQ-plane. The PSS was followed by a periodic AC (PAC) simulation to produce the illustrative closed-loop transfer plot.

A CALLUM 2 system based on the baseband model in Figure 32 with the SCG implemented by CMOS transistors and a first order low-pass filter in the feedback network has been simulated with PSS and PAC. The bias point was chosen in agreement with the result from Figure 38, which for this system configuration corresponds to $I_I = 0\,\text{V}$, $Q_I = 100\,\text{mV}$. In Figure 46 the source-load transfer (transmittance) is shown for the uncompensated system at the selected quiescent point. The figure also shows the transfer when a phantom zero is applied. As discussed above, a single compensation of a system with three loop poles is not enough, as it allows complex system poles and consequently some peaking remains. Anyway, it is a large step in the right direction. For correct compensation another phantom zero should be implemented. The quiescent point in the IQ-plane maps to a specific loop gain, which in turn sets the bandwidth. The reduced bandwidth for low envelope signals is expected for CALLUM 2 and quantitatively proven by simulations.

A desired situation is shown in the root locus, see Figure 42, where all loop poles are positioned on the negative real axis. The loop poles are positioned along the negative real axis if there is no local feedback inside the global loop or any resonance circuits. For the closed-loop the system poles start to move along the root locus and the distance they travel is primarily determined by the magnitude of the loop gain.

Large loop gain forces the system poles far out, not unusually the system poles have much larger imaginary- than real-part, i.e., low damping factor, or even forced into the right half plane (instability).

4.8.2 Phantom Zero Compensation

To control the system poles into their desired position we need two phantom zeros, as stated in the compensation strategy outlined in 4.8. Phantom zero compensation is the one preferred, as it is the most orthogonal compensation method to the other design steps [25].

Efficiency of the Phantom Zero

A phantom zero compensation is often easy and straightforward to implement and considered as the most favorable compensation technique available. A very important aspect is the efficiency of the phantom zero [102], but this is very seldom seen or discussed in the open literature. Let us first define the efficiency, δ, of the phantom zero as,

$$\delta \equiv \frac{p_{\rm ph}}{n_{\rm ph}}, \tag{99}$$

where $p_{\rm ph}$ is the parasitic pole that appears when a phantom zero, $n_{\rm ph}$, is introduced. There is a saying *'no zero without a pole'*, which makes sense since every component known to date has limited bandwidth. This unwanted pole limits the efficiency of the applied phantom zero compensation. The effect of the efficiency is captured by any circuit simulator, but only gives vague guidelines to the designer, e.g., how to improve the efficiency. It is possible to find a simple analytical expression describing the efficiency and the result is useful for the designer.

From the pole-zero map it is straightforward to calculate the efficiency of the phantom zero (if the associated parasitic pole can be identified) using (99). Unfortunately most circuit simulators can produce neither the pole-zero map nor the root locus. In Figure 47 a second order system is compensated with a zero. The loop poles are located at $p_1 = -1\,\mathrm{krad/s}$, $p_2 = -2\,\mathrm{krad/s}$, and the zero, $n_{\rm ph} = -15\,\mathrm{krad/s}$. For the closed-loop system to have MFM transfer one of the complex system poles should be located at the intersection between the half-circle indicating the MFM bandwidth, ω_0, and the 45° line. The system poles should be at $p'_{1,2} = -\omega_0(1 \pm j)/\sqrt{2}$. The zero is shown in Figure 47 for the sake of clarity even though the zero must be a phantom zero for MFM characteristics, and the position of the parasitic pole, $p_{\rm ph}$, depends on the efficiency and is only indicated in the figure. The numbers in Figure 47 indicate the efficiency of the zero, and for $\delta = 1$ the pole $p_{\rm ph}$ is right on top of $n_{\rm ph}$ and the root locus is basically a straight line parallel with the $j\omega$-axis that intersects the real axis at $\sigma = -1.5\,\mathrm{krad/s}$. When the efficiency increases the root locus bends more and more into the left-half-plane, and for $\delta = \infty$ the familiar circle is achieved and the MFM bandwidth, $\omega_0 = 19\,\mathrm{krad/s}$, can be fulfilled.

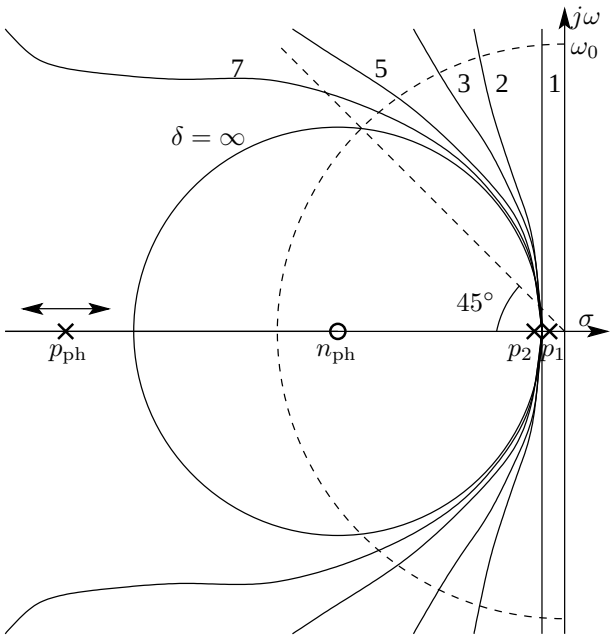

Figure 47: Root locus for the second order system compensated with a phantom zero versus efficiency.

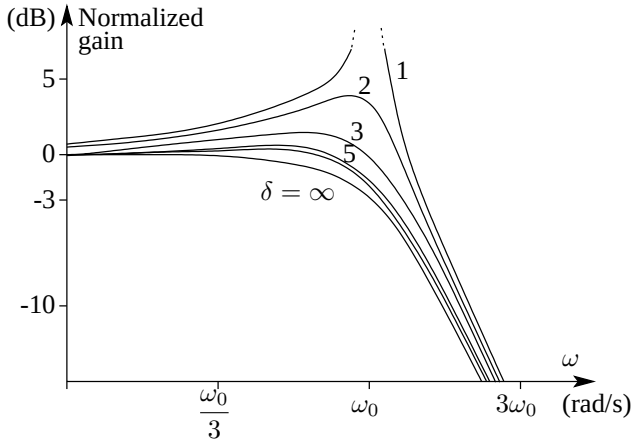

Figure 48: The amplitude plot of the second order system compensated with a phantom zero versus efficiency.

Table 1: The efficiency of the phantom zero and its impact on the circuit, where *high* is most desired.

Efficiency				Impact on root locus
1	$\leq$	δ	< 3	low
3	$\leq$	δ	< 7	moderate
7	$\leq$	δ		high

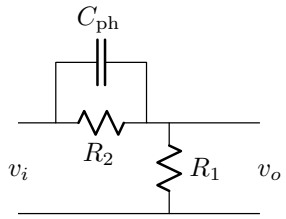

Figure 49: Implementation of a first order phantom zero in the resistive feedback network.

Frequency compensation is all about controlling the location of the system poles to their desired position. From Figure 47 it is clear that only with a phantom zero with infinite efficiency the estimated MFM bandwidth, ω_0, can be achieved. For non-ideal efficiency the system will have a damping factor $\zeta < 1/\sqrt{2}$, and as a consequence peaking will appear in the amplitude plot of the closed-loop system, see Figure 48.

We see that a good approximation of the ideal efficiency case is achieved for $\delta \geq 7$ and acceptable compensation is achieved for efficiency values as low as 5. Figure 47 is an example when the phantom zero is well separated from the loop poles (a factor of ten or more). As long as the phantom zero is well separated from the loop poles the variation of the root locus is very small. The guideline for the impact on the root locus versus efficiency is based upon the observation in the root locus, and given in Table 1.

Let us return to the CALLUM transmitter and investigate a common feedback network in terms of efficiency, see Figure 49. The network is assumed to have ideal drive and loading conditions. The capacitor, $C_{\rm ph}$, in Figure 49 implements the phantom zero, and the position of the zero is given by,

$$n_{\rm ph} = -\frac{1}{R_2 C_{\rm ph}}. \tag{100}$$

The unwanted (parasitic) pole appears at,

$$p_{\rm ph} = \frac{R_1 + R_2}{R_1} n_{\rm ph} = \delta n_{\rm ph}. \tag{101}$$

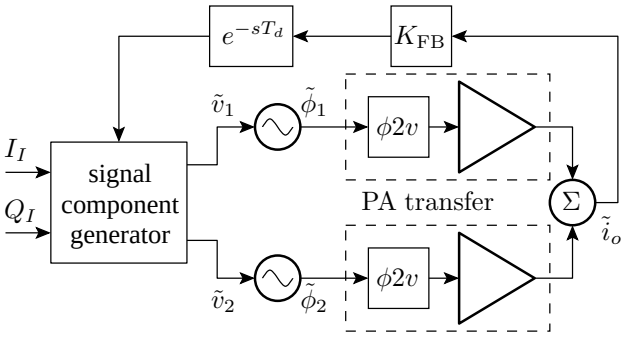

Figure 50: Model used for loop gain calculations including delay.

The efficiency is here determined by the relation between the resistor, R_1 and R_2. Depending on the efficiency of the phantom zero the compensation will be successful or not, as stated in Table 1.

4.9 Loop Properties with Delay

In section 4.8 it was mentioned that the system poles ultimately should be place on the real axis due to the extra negative phase shift from the loop delay and phase shift introduced by other high frequency poles and zeros [37]. Time delays between inputs and outputs are very common in industrial processes, e.g., economical, biological, and engineering systems. Time delays are also used to compensate for model reduction where high-order systems are represented by low-order models with delay. The delay depends on the physical size of the loop as well as the electrical delay of each component. Any delay in measurement equipment, in controller action, in actuator operation, in computer computation, and the like, is called transport delay or dead time. The delay always reduces the stability of a system and limits the achievable response time of the system. The introduction of the delay in the model used for loop gain calculations is shown in Figure 50. The loop gain results found in section 4.5.4 can handle time delay by adding the term e^{-sT_d}, and will not be repeated here. T_d is the time delay in seconds.

The investigated CALLUM 2 system with a loop pole at 0 Hz and a double pole at -100 Mrad/s was simulated in Matlab, where the described system was compensated with two phantom zeros place at -100 Mrad/s, and -200 Mrad/s, respectively. The closed-loop transfer is shown in Figure 51, and it looks nice for small loop delay, but as the delay is increased the amplitude curve gets its characteristic peak and beyond that an unstable system results.

A comment on the amplitude plot in Figure 51: The amplitude plot will look the same for a certain pole configuration as when the poles are mirrored in the imaginary axis. The amplitude plot only reacts to the distance of the poles and zeros to the imag-

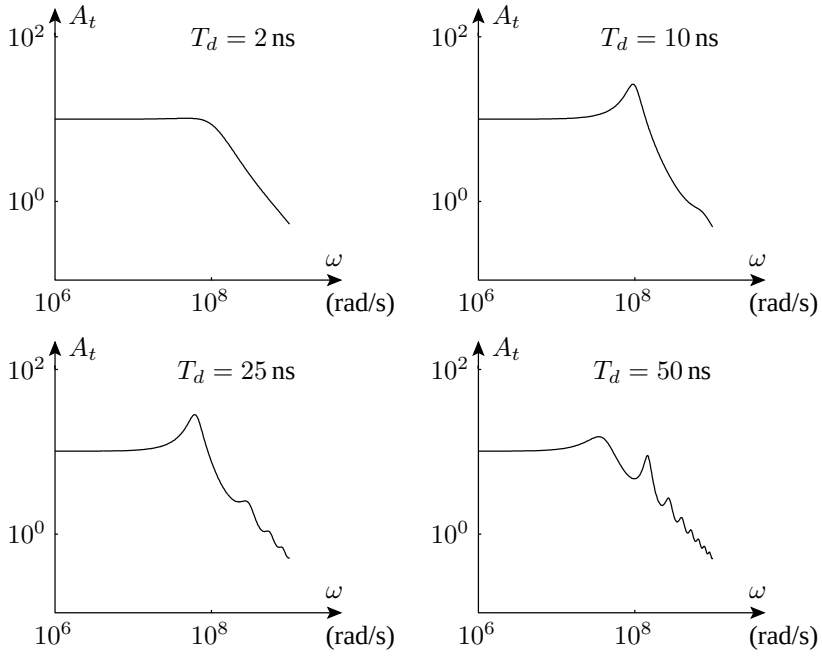

Figure 51: The closed-loop transfer affected by various loop delays. A delay of $T_d = 2\,\text{ns}$ results is stable system, but already for $T_d = 10\,\text{ns}$ the system poles are located close to the imaginary axis (stable or unstable not possible to say from the amplitude plot). For a relatively large delay, 25–50 ns, the graphs represent an unstable system.

inary axis, and the stability information is discriminated. Yet, the stability information is possible to achieve from the phase plot.

4.9.1 Non-Minimum-Phase System

The amplitude and phase curves of the transfer of a dynamic system are connected to each other. For a given amplitude curve it is possible to find a phase curve under which the actual phase curve is located. The phase curve of a minimum-phase system coincide with the predicted phase curve. Furthermore, it can be shown that a necessary and sufficient condition for a minimum-phase system is that all poles and zeros are located in the left half plane [103]. A system with a right half plane zero is an example of a non-minimum-phase system, and if it is possible, a non-minimum-phase system should be avoided in feedback structures, due to its additional phase degradation.

A constant time delay corresponds to a linear change in phase angle with frequency, and therefore a network having a linear phase variation with frequency would act as a constant time delay. Another important property of a time delay is that the amplitude of the signal is not changed. The time delay corresponds to e^{-sT_d} in the Laplace transform domain, and a general system function can always be factored into the product of a minimum-phase system function and an all-pass system function. The all-pass function is given by $q(-s)/q(s)$. From this representation it becomes clear that the all-pass function must possess a zero in the right half plane at ω_x for every pole in the left half plane at $-\omega_x$. Thus, the poles and zeros are images with respect to the origin and, since the functions are real, with respect to the imaginary axis [104]. The time delay expressed as e^{-sT_d} should be written on the form $q(-s)/q(s)$ in order to be an all-pass filter. The gain of the time delay block is always unity, $\left|e^{-sT_d}\right| = 1$, which characterizes an all-pass transfer. The time delay can be expressed by the series,

$$e^{-sT_d} = \frac{e^{\frac{-sT_d}{2}}}{e^{\frac{sT_d}{2}}} = \frac{1 - \frac{sT_d}{2} + \frac{s^2T_d^2}{8} + \cdots + \frac{\left(-\frac{sT_d}{2}\right)^k}{k!}}{1 + \frac{sT_d}{2} + \frac{s^2T_d^2}{8} + \cdots + \frac{\left(\frac{sT_d}{2}\right)^k}{k!}}. \tag{102}$$

A crude but very common approximation is a first order lag filter given by,

$$e^{-sT_d} \simeq \frac{1}{1 + sT_d}. \tag{103}$$

Obviously this is not an all-pass filter, but its justification is simplicity and that it accurately describes the phase down to approximately $-20°$. The pole is positioned at $p = -\frac{1}{T_d}$. This model will do for less critical applications.

The series expansion in (102) is better known as the Pade approximation. When referring to a specific number of terms for a Pade approximation we may use Pade (m,n), where m (n) is the number of zeros (poles). For all-pass magnitude characteristic and better phase approximation (down to approximately $-40°$) use Pade (1,1), which is

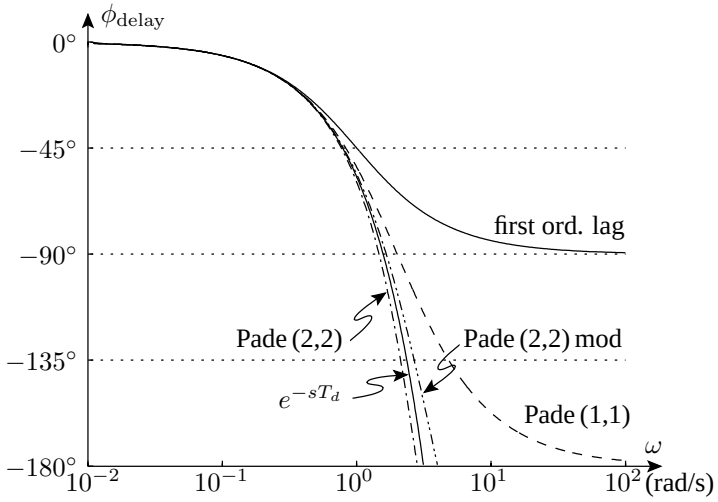

Figure 52: The phase characteristics for different approximations of $e^{-T_d s}$. The time delay is set to $T_d = 1\,\text{s}$ for simplicity.

given by,

$$e^{-sT_d} \simeq \frac{1 - \frac{sT_d}{2}}{1 + \frac{sT_d}{2}}. \tag{104}$$

The zero (pole) is positioned at $n = \frac{2}{T_d}$ ($p = -\frac{2}{T_d}$). When the time delay is very troublesome in the design of the feedback control loop an even more complex approximation is used, i.e., Pade (2,2) in (105), the modified Pade (2,2) in (106), or even higher order approximations.

$$e^{-sT_d} \simeq \frac{1 - \frac{sT_d}{2} + \frac{s^2T_d^2}{8}}{1 + \frac{sT_d}{2} + \frac{s^2T_d^2}{8}} \tag{105}$$

$$e^{-sT_d} \simeq \frac{2 - sT_d + \frac{s^2T_d^2}{8}}{2 + sT_d + \frac{s^2T_d^2}{8}} \tag{106}$$

The Pade (2,2) has a complex pair of poles (and zeros) located at $p_{1,2} = -\frac{2(1\pm j)}{T_d}$, ($n_{1,2} = \frac{2(1\pm j)}{T_d}$), whereas the modified Pade (2,2) has double poles (and zeros) at $p_{1,2_{\text{mod}}} = -\frac{4}{T_d}$, ($n_{1,2_{\text{mod}}} = \frac{4}{T_d}$). The phase response for different approximations of the time delay is shown in Figure 52. Remember all approximations have all-pass magnitude characteristic except for the first order lag, because it is a low-pass filter.

Table 2: Guideline for selection of time delay approximation.

$36\,\omega_0$	$\leq$	$1/T_d$			$\rightarrow$	Ignore the time delay entirely
$15\,\omega_0$	$\leq$	$1/T_d$	$<$	$36\,\omega_0$	$\rightarrow$	First order lag (Pade (0,1))
$6\,\omega_0$	$\leq$	$1/T_d$	$<$	$15\,\omega_0$	$\rightarrow$	Pade (1,1) approximation
$3\,\omega_0$	$\leq$	$1/T_d$	$<$	$6\,\omega_0$	$\rightarrow$	Pade (2,2) approximation
		$1/T_d$	$<$	$3\,\omega_0$	$\rightarrow$	Higher order Pade approximation and verification of closed-loop performance

There is always a trade-off between effort, visibility, and computational effort, on one hand and accurate prediction on the other. In order to find the appropriate approximation a table is set up. The Table 2 is an outline of which approximation to choose, based on the relation between the time delay, T_d, and the closed-loop bandwidth, ω_0. From Figure 52 the largest phase degradation, ϕ_{delay}, can be found for each approximation representing the phase characteristic satisfactory. The upper limit of the interval in the table is determined as $\frac{360^\circ}{\phi_{\text{delay}}}\omega_0$. Since the table is only a recommendation, it is good practice to verify the approximation used by computer simulation or measurement.

4.9.2 Generalization of Time Delay

The group delay is defined as the effective time delay as a function of frequency of a network. A commonly encountered representation of filter phase response is called the group delay, defined by

$$T_d(\omega) = -\frac{\partial \phi_{\text{delay}}(\omega)}{\partial \omega}. \tag{107}$$

For linear phase response, i.e., $\phi_{\text{delay}} = -T_d\omega$ for some constant T_d, the group delay and the phase delay are identical, and each may be interpreted as time delay (equal to T_d). If the phase response is nonlinear, then the relative phases of the sinusoidal signal components are generally altered by the filter. A nonlinear phase response causes a smearing of transients. Another term for this type of phase distortion is phase dispersion.

Thus, the name *group delay* for $T_d(\omega)$ refers to the fact that it specifies the delay experienced by a narrowband group of sinusoidal components, which have frequencies within a narrow frequency interval around ω. The width of this interval is limited to that over which $T_d(\omega)$ is approximately constant. The effective time delay is often measured by finding the negative slope of the phase response in the frequency range of interest. The rate of change of the total phase shift with respect to angular frequency, $\frac{\partial \phi_{\text{delay}}(\omega)}{\partial \omega}$, through a device or transmission medium.

4.9.3 Bandwidth Limitation Due to Time Delay

Let us investigate the effect of the time delay on a CALLUM transmitter by isolating the problem. Assume a CALLUM system at block level with no loop poles except for the one originating from the VCO, which this alone contributes -90°.

Often the phase margin is used as a figure of merit for the stability properties of an amplifier. For desired system behavior both in time and frequency, the system poles should be in Butterworth position. The damping factor for a second order Butterworth is $\zeta = 1/\sqrt{2}$ and the phase margin is around 65° [105, 106]. This means that without any influence of other poles the time delay must not degrade the phase more than 25°. The phase change due to time delay is given by $-\omega T_d$, thereby the maximum loop bandwidth can be determined, if the desired phase margin and delay are known.

$$\omega_{0,\max} = \frac{90^\circ - \phi_m}{180^\circ} \cdot \frac{\pi}{T_d} \tag{108}$$

Equation (108) gives the opportunity for two interpretations. First, with a given phase margin and delay the maximum bandwidth of the loop can be determined. Second, for a given standard (e.g., EDGE, W-CDMA) the maximum allowed time delay can be determined for a given phase margin. Unfortunately, it is not straightforward to determine the loop bandwidth necessary to process an input signal with known bandwidth since the bandwidth of the input signal is normally increased within the loop, see also section 4.11.

The maximum loop bandwidth for a system with MFM transfer and a time delay of $T_d = 50\,\text{ns}$ is limited to,

$$\omega_{0,\max} = \frac{90^\circ - 65^\circ}{180^\circ} \cdot \frac{\pi}{50 \cdot 10^{-9}} \simeq 8.7\,\text{Mrad/s}. \tag{109}$$

In practice the bandwidth of the feedback system mainly will be set by the dead-time. Therefore the loop delay is a very important issue. From the implications caused by the time delay a quantitative estimation of the loop time delay becomes important and is discussed in section 4.10. When estimating the maximum loop bandwidth all loop poles, but the one from the VCO, is disregarded. The frequency compensation now becomes a bit different compared to how it was described earlier. The approach is basically to make the model (or assumption) valid instead of optimizing for maximum bandwidth. This implies reduction of high frequency parasitics by bandwidth limitation. One way is to use a lag-lead loop filter, which from simulation was found to be effective on the closed-loop system.

The root locus for the CALLUM system with $p_1 = 0$, $p_2 = -90\,\text{Mrad/s}$, $p_3 = -100\,\text{Mrad/s}$, $n_{\text{ph1}} = -100\,\text{Mrad/s}$ and a time delay of $T_d = 50\,\text{ns}$ approximated by the modified Pade (2,2) is shown in Figure 53. The simulated and predicted loop bandwidth match very well for system poles in Butterworth position. From the root locus a potential risk for high frequency peaking in amplitude can be predicted caused by the outer pair of complex poles. By limiting the bandwidth of the system the

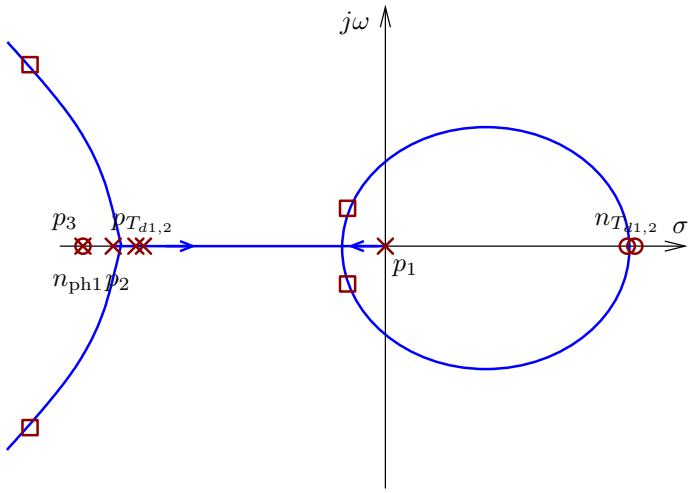

Figure 53: The root locus for a CALLUM 2 with a phantom zero compensation and a time delay of $T_d = 50\,\mathrm{ns}$. The delay is represented by the modified Pade (2,2) approximation.

potential peaking can be suppressed. The design of the compensating lag-lead loop filter is presented in next section.

4.9.4 Lag-Lead Compensation

The interpretation of a lag-lead compensation is that the amplitude of the loop gain should be reduced at high frequency in such manner that the phase still is above $\phi_m - 180°$ when the magnitude of the loop gain passes unity. The lag-lead compensation is a pole-zero pair with a transfer given by,

$$H_{\mathrm{ll}} = \frac{1 + \dfrac{s}{n_{\mathrm{ll}}}}{1 + \dfrac{s}{p_{\mathrm{ll}}}}, \tag{110}$$

where the pole is positioned at $p_{\mathrm{ll}}E = n_{\mathrm{ll}}$, and $E > 1$ for lag-lead behavior. Even though a rule of thumb discriminates a lot of information, it can in some cases accelerate the design process. A sensible placement of the zero is $n_{\mathrm{ll}} = 0.1\omega_0$, where ω_0 is the loop bandwidth [103]. The lag-lead compensation alters the shape of the loop gain characteristic, as illustrated in Figure 54, from which a number of conclusions can be drawn. The compensation will reduce the bandwidth, and the loop gain at intermediate and high frequencies is reduced. This reduction of global loop gain can either be

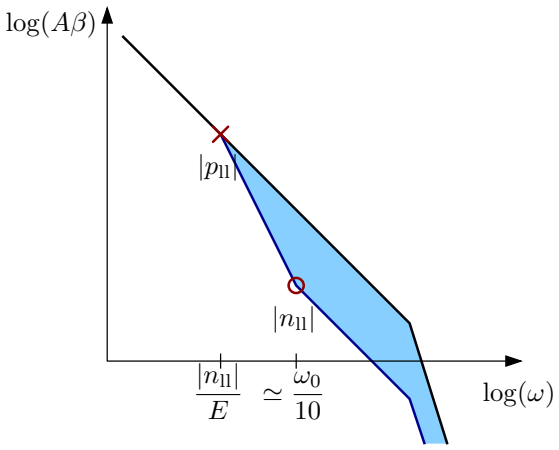

Figure 54: The loop gain, $A\beta$, before and after lag-lead compensation.

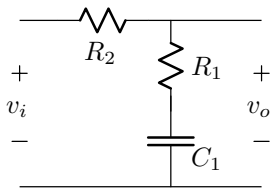

Figure 55: Passive implementation of a lag-lead loop filter.

wasted or used locally to linearize a nonlinear block. By using global feedback locally the distortion performance can at best be preserved, more often it is reduced [3].

One common passive implementation of a lag-lead filter is shown in Figure 55. With ideal drive and load conditions the zero and pole are given by,

$$n_{ll} = -\frac{1}{R_1 C_1} \tag{111}$$

$$p_{ll} = -\frac{1}{(R_1 + R_2)\, C_1}. \tag{112}$$

The distance between the pole and zero is given by $E = (R_1 + R_2)/R_1$. The zero-pole ratio E is best determined from system simulations in order to maximize the performance.

4.10 Estimation of Dead Time in the Loop

The transport delay dominates the phase characteristics at higher frequencies and gives a maximum limit to the value of the loop gain and the linearizing bandwidth. The transport delay should effectively be minimized, but before trying to minimize it, the origin of the delay in a transmitter chain must be identified. The delay is here divided in two subgroups, wave propagation and charge transport mechanism through devices.

4.10.1 Delay from Wave Propagation

The first group contains the delay resulting from the physical distance the signal has to travel. The speed of the electro-magnetic wave is limited by the speed of light in vacuum, c_0, and in any media, such as a coaxial cable, the actual velocity will be smaller. Often the velocity in the media is 70 – 90 % of c_0, which results in approximately 4 ns time delay per meter coaxial cable. As an example, in a feedforward power amplifier a 20 feet (1 feet = 0.3048 m) coaxial cable was used to accommodate the needed 25 ns delay in the error signal path, which is equivalent to a delay of 4.1 ns/m [18]. For a test circuit the total physical length of the loop will be around one meter, and the time delay from wave propagation is expected to be approximately 4 ns.

4.10.2 Delay from Charge Transport Mechanism Through Devices

The delay resulting from finite charge transport velocity can be viewed at either block level (e.g., PA, mixer, VGA, filter) or at device level (e.g., transistor, diode). Due to high f_T of the transistors they are considered as instantaneous, and their contribution to the total delay can be neglected. The transmitter is most often tuned to the operating frequency. This impedance matching gives maximum power to the load and additional far out filtering of the transmitted spectrum. The time delay originating from this filter is significant, especially when using high Q filters.

Time Delay from a Parallel Resonance Circuit

The parallel resonance circuit gives rise to a varying group delay, and by using the tools from section 4.9.2 the time delay can be found as shown below.

The impedance of a parallel RLC-circuit is given by,

$$G(j\omega) = \frac{sRL}{s^2RLC + sL + R}. \tag{113}$$

The characteristic polynomial is rewritten on a form containing the center frequency, ω_c, and the quality factor (Q-value).

$$s^2 + s\frac{1}{RC} + \frac{1}{LC} = s^2 + s\frac{\omega_c}{Q} + \omega_c^2 \tag{114}$$

The RLC-circuit has one zero in the origin, and the complex pole pair (if $2Q > 1$) is given by,

$$p_{1,2} = -\frac{\omega_c}{2Q} \pm j\omega_c\sqrt{1 - \frac{1}{(2Q)^2}}. \tag{115}$$

If Q is reasonably large the poles are approximately located at,

$$p_{1,2} = -\frac{\omega_c}{2Q} \pm j\omega_c = -a \pm jb. \tag{116}$$

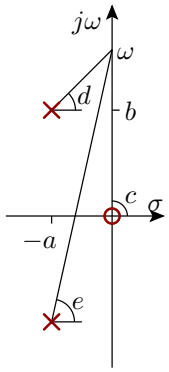

Figure 56: Pole-zero map of a parallel RLC-circuit.

The pole-zero map of the RLC filter is shown in Figure 56. From this map it is straightforward to find the phase expression. The argument, $\arg G(j\omega) = c - d - e$, has an convenient geometric interpretation, see Figure 56.

$$\arg G(j\omega) = \frac{\pi}{2} - \arctan\left(\frac{\omega - b}{a}\right) - \arctan\left(\frac{\omega + b}{a}\right) \tag{117}$$

From the phase information the group delay defined by (107) can be calculated for the parallel resonance circuit. The group delay looks like a Gaussian distribution centered at ω_c. The maximum time delay can be assumed to coincide with ω_c for any $Q > 1$. The time delay at ω_c is given by,

$$T_{d,\max} = \frac{1}{a}\left(1 + \frac{1}{1 + (4Q)^2}\right) \simeq \frac{1}{a} = \frac{2Q}{\omega_c} = \frac{2}{\mathrm{B}}, \tag{118}$$

where the relation from (116) has been used and that Q is sufficiently large to justify the approximation. B is the -3 dB bandwidth of the resonance circuit. This compact

formula can be use to investigate the additional time delay of the tuned filter and matching networks. For example, assume $Q = 5$ and $\omega_c = 2\pi \cdot 900$ Mrad/s to fit, e.g., the EDGE transmit band. The maximum time delay from the tuned filter $T_{d,\max} \simeq 1.8$ ns add to the total delay in the loop.

4.11 Spectrum Widening of Internal Signals in CALLUM

The output signal, $s_o(t)$, should have the same spectrum as the input signal, $s_i(t)$. The constant-envelope signals, $s_1(t)$ and $s_2(t)$, have proven to be significantly more wideband than the input signal, see section 3.2.3. These constant-envelope vectors are produced by the VCOs, which get their control signals, $v_1(t)$ and $v_2(t)$, from the SCG. In the CALLUM architecture, the VCOs are connected in front of the power amplifiers to fulfill two major tasks, i.e., to perform the frequency shift of the information, and to provide the power amplifiers with the appropriate control signals. Unfortunately, the necessary control signal for the VCO is more wideband as the VCO behaves as an integrator for the information signal, which means that the spectrum of the input signal is attenuated accordingly. Since the spectrum of the constant-envelope signals is fix for a certain modulation, the control signals to the VCOs have to adjust. This spectral widening that appears internally in the system puts increased constraints on the individual building blocks.

The spectrum of the information signal, $s(t)$, is shown in Figure 57 and 58, together with the signals $s_1(t)$ and $v_1(t)$ for EDGE and $\pi/4$-shifted QPSK modulations, respectively. The bandwidth of the constant-envelope signal in the loop is apparently dependent on the modulation depth of the signal, and the control signal to the VCO is spectrally even wider than the constant envelope signal, $s_1(t)$. Since the loop gain is decreasing at higher frequencies, large spectral regrowth limits the performance of the CALLUM architecture.

In order to find an approximate formula describing the instantaneous frequency of the constant envelope signal in a CALLUM system, we consider the constellation diagram for the modulation, see Figure 59. The trajectory has to move from one data-point to another in the modulation diagram in exactly the time between two consecutive symbols, T_s. The distance between the points a and b is denoted r_{ab}. For a signal envelope near $r_{\min}$ the angular velocity of $s_1(t)$ (and $s_2(t)$) is almost the same as for $s(t)$. A coarse approximation is to say that the signal moves at a constant speed from a to b on the time, T_s, between two symbols. Then the maximum instantaneous frequency of the vector $s_1(t)$ (or $s_2(t)$) is,

$$f_{\max} = \frac{r_{ab}}{2\pi T_s r_{\min}} \leq \frac{m_d}{\pi T_s}. \tag{119}$$

$r_{\min}$ equals the smallest output amplitude for the transmitted sequence, and the modulation depth, m_d, of a signal is defined as the ratio between the maximum and minimum amplitude. For $\pi/4$-shifted QPSK modulation ($m_d \simeq 10$) equation (119) predicts $f_{\max} \simeq 2.9/T_s$, and simulations give $f_{\max} \simeq 3/T_s$. Similarly, for an EDGE

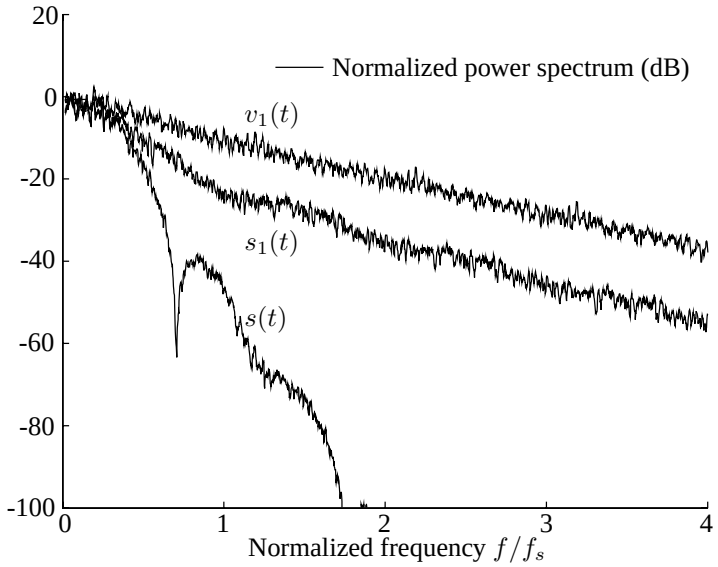

Figure 57: EDGE modulation with modulation depth $m_{\mathrm{d}} \simeq 7$.

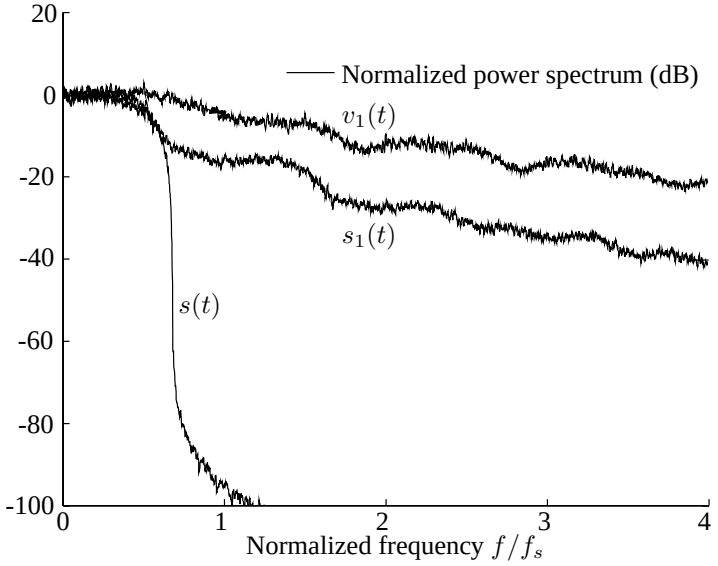

Figure 58: $\pi/4$-shifted QPSK modulation with modulation depth $m_{\mathrm{d}} \simeq 10$.

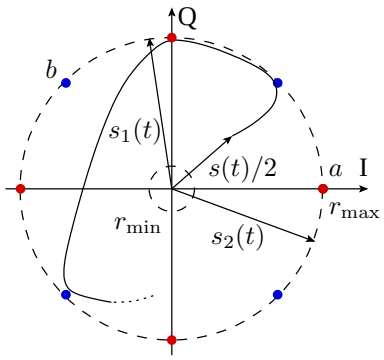

Figure 59: A $\pi/4$-shifted QPSK modulation scheme showing a stylized trajectory of the modulated signal, $s(t)$, and the constant envelope signals, $s_1(t)$ and $s_2(t)$.

modulated signal ($m_d \simeq 7$), the predicted $f_{\max}$ is $2.2/T_s$, and the simulated $f_{\max}$ is $1.4/T_s$. Most importantly, (119) shows that the instantaneous frequency is proportional to the modulation depth of the information signal.

4.12 Spectral Performance

The three CALLUM versions under test, that is, CALLUM 1, CALLUM 1lin, and CALLUM 2, have been modeled in a circuit simulator (spectre/spectreRF) at baseband. The schematic views were built at the block level using AHDL as description language. From the resulting simulations, the relative ranking of the different implementations, in terms of the amount of loop gain needed to correctly process a signal with a given modulation, and of the maximum acceptable loop delay, is presented in the next sections.

4.12.1 CALLUM and Spectrum Emission Mask for EDGE

The first issue investigated is how the implementation of the SCG affects the performance of the system when operating on an EDGE modulated signal. The loop gain plots in the IQ-plane (Figures 36 – 38) provide the information to adjust the peak magnitude of the loop gain for each CALLUM version to a constant value. This allows easy comparison of the spectral properties for each implementation. The simulated spectra are based on a random data sequence containing 33 ksymbols, using a $3\pi/8$-shifted 8PSK modulation and filtering according to the EDGE standard. The maximum peak amplitude of the input signal was chosen to be 98% of its valid input range ($2r_{\max}$), in order to avoid the problem of having a zero at the denominator

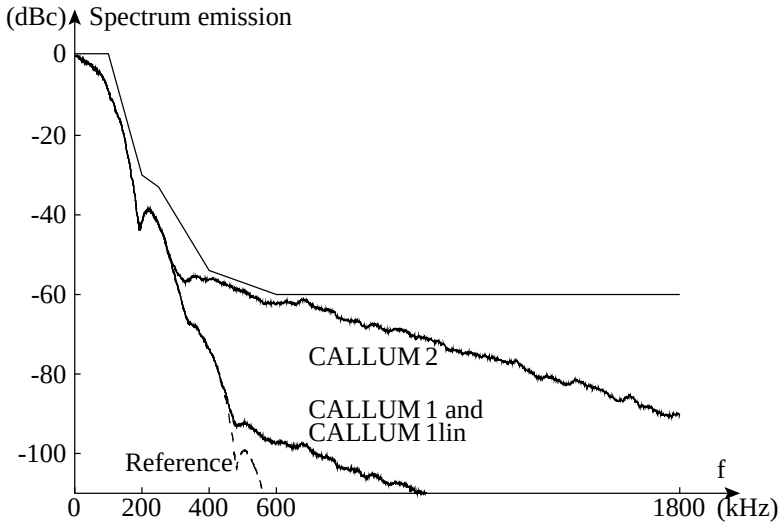

Figure 60: The spectrum emission mask for mobile handset equipment for EDGE transmit spectrum together with the normalized input and output signals from different implementations of CALLUM.

of the control equations in for example CALLUM 1. The simulated spectrum is processed through a 30 kHz filter bandwidth and normalized to the power in 30 kHz of the carrier. The technical specifications covering GSM/EDGE radio transmission can be found in TS 05.05 at *3GPP* and *ETSI*[10].

The EDGE spectrum emission mask for mobile stations in the GSM 900 band is plotted in Figure 60, together with the spectrum of transmitted signal for each CALLUM version. The input signal serves as a reference, from which it is possible to detect any spectrum degradation. The magnitude of the loop gain was adjusted in such way that CALLUM 2 barely passed the spectrum emission mask. The gap to CALLUM 1 and CALLUM 1lin is significant, and it can be concluded that dropping the denominator term in the control equations for CALLUM 2 has severe effects on the spectral properties. The low loop gain region for low and medium signal amplitudes relative to $r_{\max}$ (Figure 38) has really a strong impact on the performance of CALLUM 2. The spectral performances of CALLUM 1 and CALLUM 1lin, on the other hand, are much better as a result of a more leveled loop gain in the IQ-plane for these CALLUM version. It can be noted that CALLUM 1lin approximates the original CALLUM 1 very closely, and for the modulation tested they perform equally well.

[10] See *www.3gpp.org* and *www.etsi.org* for the complete standard specifications.

4.12.2 CALLUM and Spectrum Emission Mask for W-CDMA

In section 4.12.1 the peak loop gain was adjusted such that CALLUM 2 just passed the spectrum emission mask test for EDGE. The same simulation setting, apart from a loop gain scaling, was used also on a W-CDMA signal. Since the chip rate of a W-CDMA signal is 3.84 Mchips/s, compared to only 270.833 ksymb/s for EDGE, a bandwidth normalization was done. This normalization removes the effect of the larger bandwidth of the W-CDMA signal to enable a direct comparison with EDGE and various CALLUM configurations. This was accomplished by increasing the loop gain by 14 times, which results in a 14 times larger loop bandwidth for this first order system transfer.

Figure 61 shows the simulated output spectra for the CALLUM versions together with the spectrum of the input signal. The output power spectra obtained from the simulator were filtered before they were compared to the mask according to the standard TS 25.101, in which the details are found. The out-of-channel emission is specified relative the root raised cosine (RRC) filtered mean power of the modulated carrier. Close to the carrier (2.5 – 3.5 MHz) a 30 kHz measurement bandwidth is used, while at higher offsets (3.5 – 12.5 MHz) the bandwidth becomes 1 MHz. Again, CALLUM 2 barely passes the spectrum emission mask test. A second test was to calculate the adjacent channel leakage power ratio (ACLR) for CALLUM 2. The ACLR is the ratio of the RRC filtered mean power centered on the assigned channel frequency, to the RRC filtered mean power centered on an adjacent channel frequency. In the closest neighbor channel the space to the spectrum emission mask is fairly large, and the ACLR is more than 10 dB better than the specification demands. In the second neighbor channel, however, located ± 10 MHz relative the carrier, the spectrum almost touches the mask and the ACLR is calculated to 43.7 dB, which is more or less at the lower limit of the specification ($\geq$ 43 dB).

It can be suspected from the shape of the spectrum of CALLUM 2, due to its white-like character, that some sort of narrow spikes are superimposed on the signal in the time domain. In fact, it has been found from transient simulations that CALLUM 2 loses lock now and then, always in conjunction with small amplitudes of the input signal. Turning again to Figure 38, it is clear that the loop gain is low for small to medium signal amplitudes, and it is well known that the bandwidth of a system having only a single loop pole is proportional to the loop gain. Thus, the loop bandwidth decreases for small signal amplitudes, with the result that the loop cannot follow the signal in some (rare) cases. After the loop is unlocked, the process of lock acquisition starts almost immediately. Lock acquisition is a nonlinear and fast process (compared to the modulated signal), which produces glitch-like disturbances in the output signal, hence the white-noise-like floor in Figure 61. This unfavorable behavior of CALLUM 2 makes it very unsuitable for operations on a W-CDMA signal. In principle, it is true that it is always possible to increase the loop gain to get improved performance; however, the unavoidable presence of parasitic phenomena such as time delay and high frequency poles set strong limitations to this brute-force approach, as

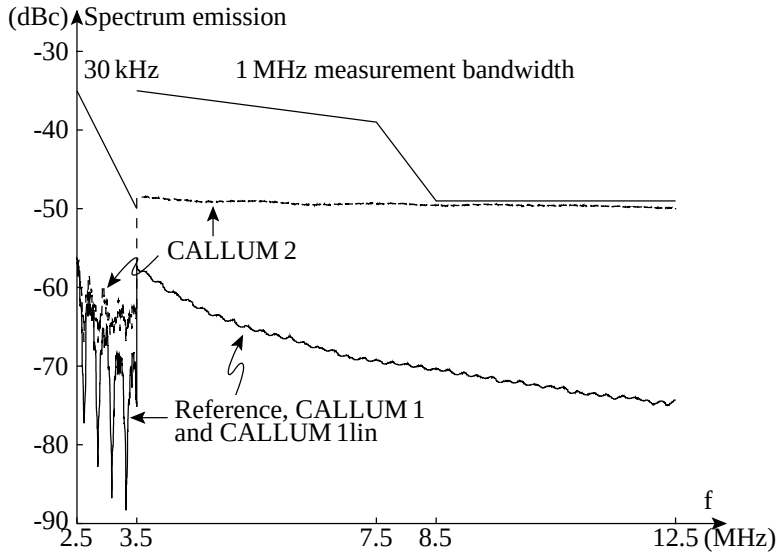

Figure 61: The spectrum emission mask for user equipment (UE) for W-CDMA transmit spectrum together with the reference and outputs for different implementations of CALLUM. Only the part of the spectrum covered by the spectrum emission mask requirement is plotted.

it increases the risk of instability, due to the large loop bandwidth. Hence, it is the limited bandwidth of the CALLUM 2 system that makes its application on a W-CDMA signal difficult. Further, the modulation used in W-CDMA, the so-called hybrid phase shift keying (HPSK), also known as orthogonal-complex quadrature phase shift keying (OCQPSK), allows zero crossings, and in any CALLUM architecture zero crossings give rise to a spectrally very wide signal within the loop, see section 4.11. We can conclude that building a linear transmitter for W-CDMA based on CALLUM 2 is a real challenge in terms of the bandwidth needed.

The performance for both CALLUM 1 and its linearized sibling is strikingly good, as it tracks the reference perfectly. Actually, it would be possible to degrade the spectral performance and gain in design robustness. In section 4.13.1 robustness will be investigated in terms of the maximum acceptable loop delay, and in section 4.13.2 a bandwidth reduction technique will be applied to further enhance the robustness against time delay.

4.13 Accounting for Time Delay in the Loop

A pure time delay is often used to compensate for a complexity reduction in the model of a given system. In this way, a high-order system can be represented by a low-order

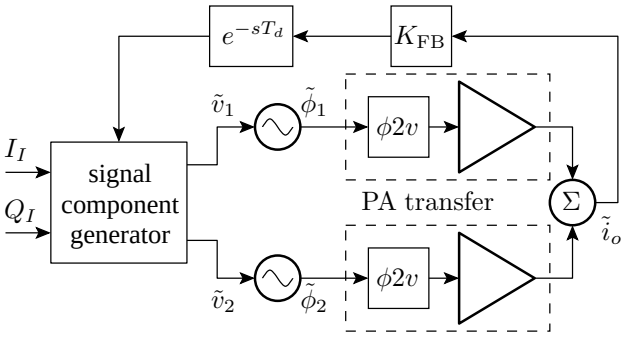

Figure 62: Model used for loop gain calculations including delay.

model, plus an appropriate time delay. A time delay always reduces the stability of a system, and limits its achievable response time. In the CALLUM architecture this delay originates from the fact that the PA output is filtered by resonance circuits (introducing a delay roughly proportional to their quality factor), and that the signal has to travel a physical distance both on chip and on the printed circuit board.

The effects of time delay on a CALLUM system are investigated here by isolating them from other sources; therefore, we will assume that the only pole present in the loop is the one due to the integrating action of the VCO. The introduction of the delay T_d in the model used for loop gain calculations is shown in Figure 62.

4.13.1 Acceptable Loop Delay in CALLUM

Since the time delay in the loop is expected to be the limiting factor for the loop bandwidth in most real-life implementations, we will examine its impact on CALLUM performances next.

The maximum acceptable time delay has been simulated for each CALLUM version and for the two standards already targeted – EDGE and W-CDMA. To create the necessary conditions for the simulations, the magnitude of the loop gain was adjusted, for each CALLUM version, to a level resulting in a close fit of the spectral emission, compared to the mask for the standard under test[11]. Thereafter, a time-delay was added, and the largest acceptable time delay for which the spectrum emission requirements were still fulfilled was noted in Table 3. For small time delays the spectral emission is not noticeably degraded, but as the delay increases, the far out power spectrum starts to increase (similar observations were reported in [37]). Further increments in time delay prevent the circuit from acquiring lock altogether.

The large difference between the maximum allowable time delay for CALLUM 1lin and CALLUM 2, respectively, is due to the fact that some fifty times lower peak loop

[11] It should be noted that in these simulations the loop gain varied between different CALLUM versions, while it was the same for all CALLUM versions in the simulations of section 4.12.

Table 3: Maximum acceptable loop delay to fulfill the spectrum emission mask.

	Maximum loop delay (ns)		
Standard	CALLUM 1	CALLUM 1 lin.	CALLUM 2
EDGE	11	48	1.0
W-CDMA	3.4	4.5	$<$0.1

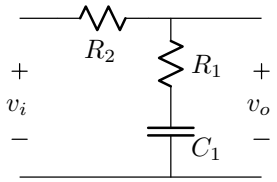

Figure 63: Passive implementation of lag-lead loop filter.

gain is needed in CALLUM 1lin, and an equally reduced loop bandwidth, when processing an EDGE signal. The maximum loop gain is the same for CALLUM 1 and CALLUM 1lin, but surprisingly from simulations CALLUM 1 was found to have a harder time acquiring lock.

The smaller acceptable time delay for W-CDMA, compared to EDGE, is primarily due to the larger bandwidth of the W-CDMA signal, as discussed in section 4.12.2. It can be noted that a time delay as small as 0.1 ns was enough to prevent CALLUM 2 to acquire lock with a W-CDMA signal.

4.13.2 Loop Bandwidth Reduction

From (108) it can be concluded that the acceptable time delay can be increased if the loop bandwidth is reduced. Bandwidth reduction can be achieved in a brute-force approach by decreasing the loop gain, but this will deteriorate the spectral performance in an unacceptable way. A much better solution is to reduce the bandwidth while keeping the low-frequency loop gain unaffected. One such type of frequency compensation, often used in PLL design, is known as lag-lead compensation. It consists of a pole-zero pair in the loop transfer function, which is easy to implement with a simple RC network [107], see Figure 63. As a rule-of-thumb, the zero is placed at approximately one tenth of the loop bandwidth to be achieved; in this way, a robust stability is ensured. A second parameter is the distance between the zero and the pole in the lag-lead filter, here denoted as $E = n_{\mathrm{ll}}/p_{\mathrm{ll}}$ ($E > 1$). The loop bandwidth is reduced by E to a first-order approximation, and an approximately E times larger delay can be accepted, compared to the uncompensated case. It is therefore clear that the factor E should be maximized (in the limit of $E \to \infty$, the lag-lead network can be

Table 4: Acceptable loop delay after compensation still fulfilling the spectrum emission mask.

	Maximum loop delay (ns)		
Standard	CALLUM 1	CALLUM 1 lin.	CALLUM 2
EDGE	32	71	3.4
W-CDMA	3.4	4.5	$<$0.1

based on an integrator), while making sure that the zero keeps the right distance from the bandwidth edge. The phase of the loop gain with the time delay included is given by,

$$\arg G(j\omega) = -\frac{\pi}{2} - \arctan\left(\frac{\omega}{-p_{\mathrm{ll}}}\right) + \arctan\left(\frac{\omega}{-n_{\mathrm{ll}}}\right) - \omega T_d. \tag{120}$$

With the zero set according to the rule-of-thumb, $n_{\mathrm{ll}} = -\omega_0/10$, and $p_{\mathrm{ll}} = n_{\mathrm{ll}}/E$ the phase is now given by,

$$\arg G(j\omega) = -\frac{\pi}{2} - \arctan\left(\frac{\omega}{\frac{\omega_0}{10E}}\right) + \arctan\left(\frac{\omega}{\frac{\omega_0}{10}}\right) - \omega T_d. \tag{121}$$

By placing the zero at least one decade from the bandwidth edge then the phase change of the pole is cancelled by the zero and no phase-margin penalty results.

The inevitable drawback of introducing frequency compensation is that undesired spectral emissions will increase, since the loop gain is reduced at higher frequencies. To compensate for this reduction in resolution for the high frequency contents in the signal, the loop gain has to be increased, counteracting to some extent the loop bandwidth reduction of the lag-lead compensation. A few simulation iterations are usually sufficient to obtain the optimal values for pole, zero, and loop gain.

4.13.3 Acceptable Loop Delay in CALLUM after Compensation

In this section the improvements in time delay insensitivity yielded by the lag-lead compensation are discussed. We have seen that the main goal of frequency compensation is to allow a larger time delay in the loop. Table 4 shows the maximum acceptable time delay for modulated signals still meeting the specifications on spectral emissions, once frequency compensation has been applied. Clearly, the improvements on the uncompensated cases (see again Table 3) are significant for EDGE. As an example, the settings for the CALLUM 2 compensations were the following: the lag-lead zero was placed at $-8\,\mathrm{Mrad/s}$, E was set to 10, and the loop gain was increased by a factor of three, compared to the uncompensated case. These values are readily implementable

in a real-life integrated design. In the open literature implementations of the CALLUM architecture have shown promising spectral performance [81] and other simulations of the system with all baseband blocks implemented at transistor level and other RF blocks given realizable parameters have shown CALLUM 2 suitable for EDGE.

In the case of a first-order system, an upper limit for the maximum acceptable time delay after compensation ($T_{d,\max,\mathrm{comp}}$) can be easily derived from (108) as,

$$T_{d,\max,\mathrm{comp}} \simeq E \cdot \frac{A\beta_{\mathrm{uncomp}}}{A\beta_{\mathrm{comp}}} \cdot T_{d,\max,\mathrm{uncomp}}. \tag{122}$$

As an example, in the case of CALLUM 2, (122) estimates $T_{d,\max,\mathrm{comp}}$ to 3.3 ns, while the simulated value is 3.4 ns. In the case of CALLUM 1lin, however, (122) would overestimate $T_{d,\max,\mathrm{comp}}$ by 80%. Although too optimistic in general, (122) shows the (qualitative) relation linking E, the loop gain before and after compensation, and the acceptable time delays before and after compensation.

To succeed with the compensation strategy there must be room for deterioration of the spectral emissions; if these are already very close to the spectral mask, no significant improvements can be made. Thus, the reason why large improvements were achievable for CALLUM 1 and CALLUM 1lin (and to a lesser extent for CALLUM 2) operating on EDGE was the relatively large distance between mask and actual spectral emissions for large offset frequencies; overly good spectral performances at large offset frequencies were in this case traded for higher insensitivity to time delay (as an example, compensation for CALLUM 1 is shown in Figure 64). For any of the three CALLUM versions operating on W-CDMA, on the contrary, even a small additional time delay was enough to deteriorate the signal spectrum at frequencies where it was very close to the signal mask, thereby failing to comply with it.

It remains to note that CALLUM 1lin performed very well, in terms of maximum acceptable time delay for a given standard. Of course, the implementation cost for the SCG in CALLUM 1lin is higher than that for CALLUM 2, but significantly lower than for CALLUM 1, without any obvious performance drawbacks.

4.14 Comparison of CALLUM Derivatives

As we now approach the end of this chapter it is time to summarize some of the features of the CALLUM derivatives before discussing the design of the circuits. First an architecture must be chosen that is adequate for the task. Quite some work has been done in the field of CALLUM architectures, and the features of each CALLUM derivative is summed up in Table 5. The table needs some additional explanation.

- **Performance**: Reflects the ability of the system to achieve high linearity, large distortion suppression, low error vector magnitude (EVM), spectral purity, etc. (1–5, where 5 is highest rating).
- **Cost**: Sophisticated solutions are expensive in terms of man hours. Other things that influence the cost are: the power consumption, the size of the chip (area),

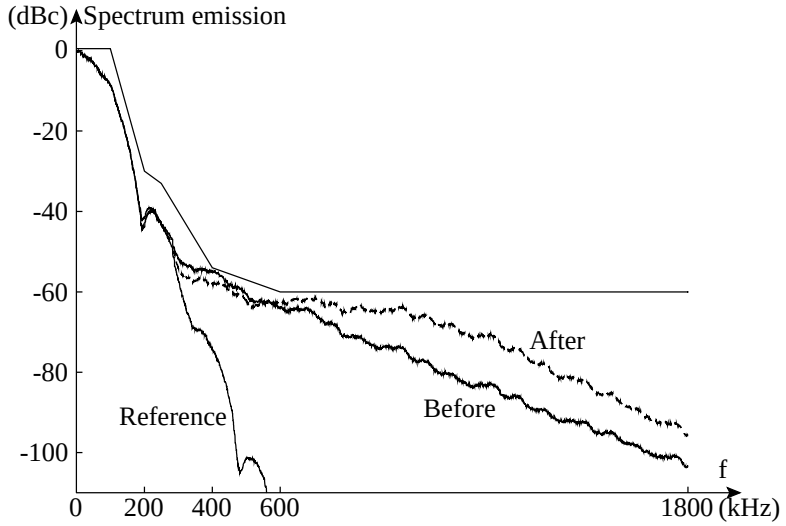

Figure 64: EDGE spectrum for CALLUM 1 before and after lag-lead compensation. $T_d = 10\,\mathrm{ns}$, $n_{\mathrm{ll}} = -15\,\mathrm{Mrad/s}$, $E = 8$, and loop gain doubled compared to the uncompensated case.

Table 5: Overview of various features for each architecture.

Architecture	Performance	Cost	Tracking	Implementation reported
CALLUM 1	5	5	excellent	yes (digital with DSP)
CALLUM 1lin	4	4	very good	no
CALLUM 2	2	2	average	yes (analog)
CALLUM 3	2	2	poor	yes (analog)
VLL	3	3	average	yes

Table 6: Relative cost for each mathematical function, where 1 is lowest and 5 highest cost.

function	cost
summation	1
subtraction	1
multiplication	2
division	4
square-root	5

the complexity, time to market, etc. The initial cost for a digital solution is very high, whereas the computational cost is low. The square-root function of CALLUM 1 has to be stored in a look-up table (LUT) not to introduce any unnecessary time delay in the loop. As a guideline to separate the architectures the costs shown in Table 6 are applied to the control equations. The implementation of the SCG is assumed to be balanced (differential). Therefore, the cost for summation is the same as for subtraction.

- **Tracking**: The tracking capability of the architecture is very dependent on the loop gain. Large loop gain variation in the IQ-plane will result in poor tracking performance.
- **Implementation**: This column tells which architectures that have been implemented so far. Previous implementations have been built up from discrete components or even discrete building blocks. We conclude that all CALLUM architectures could gain significantly by using a high degree of integration.

Chapter 5

Implementation Aspects of CALLUM

This chapter deals with implementation aspects of the CALLUM transmitter architecture. Here a high degree of system integration is targeted for reduced loop delay, higher bandwidth and improved spectral properties compared to a discrete solution. From system simulations at architectural level CALLUM 2 has proven to be suited for the EDGE standard. Therefore, CALLUM 2 was chosen as a test vehicle for this first design. An analog solution was adopted as the SCG of CALLUM 2 is one of the least complex. Only summation and multiplication are needed to realize the control equations. The design process of a CALLUM transmitter rises implementation aspects to the surface, which are not always easy to foresee from a generic signal graph. Experience from circuit design at transistor level is important to pin-point the difficulties of the implementation. The non-ideal behavior of the circuits result in parasitic phenomena that are theoretically studied to gain knowledge about this complex feedback system.

5.1 Circuit Orientated CALLUM Implementation

The generic CALLUM architecture is illustrated in Figure 27. A refined block diagram is shown in Figure 65 at a more hardware based design level. Many of the sub-circuits will be presented down to transistor level together with key simulation results.

A system orientated description of the schematic in Figure 65 follows next. The input to the system is the complex-valued baseband signal connected to the analog signal component generator (SCG). The outputs from the SCG are the control signals to the VCOs, but before feeding the signals to the VCOs they are amplified in two variable-gain amplifiers (VGAs) followed by a frequency compensation filter. The output of the VCO is buffered by a CS-stage to drive the low impedance outside the chip. Both the PA driver and the PA itself are external components, as well as the hybrid combiner and the antenna that forms the load.

The output signal is sensed and attenuated to a level suitable for the quadrature demodulator. The downconversion is accomplished by a mixer pair followed by succeeding low-pass filters. The transmit frequency (channel selection) is set by the local oscillator (LO), from which four phases are formed by the poly-phase filter. To improve the signal swing of the LO signals, the outputs of the poly-phase filter are connected to limiters for near rail-to-rail swing.

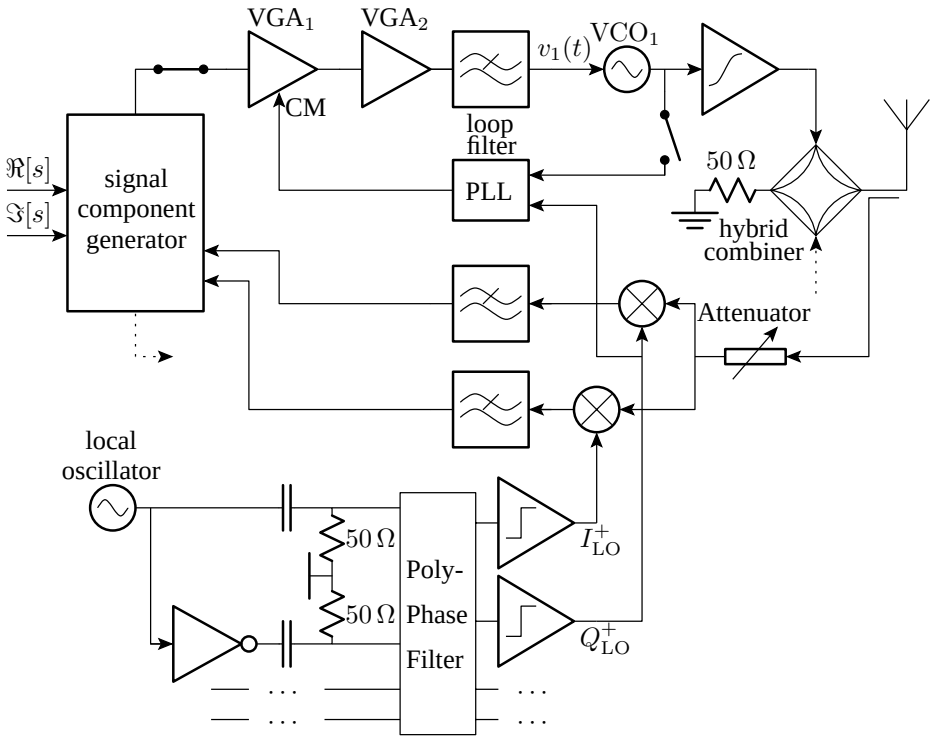

Figure 65: Electrical block level schematic of the CALLUM architecture.

For the sake of clarity all signals are single-ended in Figure 65. The main part of the on-chip blocks are however differential, except for the PLL and some control signals. Consequently, the differential to single-ended conversions (and vice versa) are not shown. In the design baluns, transformers and active components are used for this type of conversion.

In addition to the core of the CALLUM system a frequency synchronization loop is needed to reduce the signal bandwidth seen by the loop [82]. The free-running frequency of the VCO is synchronized to the LO frequency by a phase-locked loop (PLL). During synchronization the normal transmit loop is broken and a smaller loop is formed by changing the state of the switches in the schematic. The time between transmission of bursts of data can be used for synchronization, but for a test system the time can be prolonged. For the VCO to run at the correct frequency a DC correction value is applied to the control port. This value is fed to the common-mode (CM) input of the VGA even when the switches are set for ordinary transmission of data.

A stable and well characterized process was selected for this design. Austrian Micro Systems provides a 0.35 μm CMOS process with the following (selected) features:

- maximum supply voltage, 3.3 V
- poly and metal capacitor
- high resistive poly
- 4 metal layers with thick top metal (40 m$\Omega/\square$)

5.2 Signal Component Generator

The SCG is the core signal processing block of the CALLUM transmitter architecture. This block generates the control voltages, $v_1(t)$ and $v_2(t)$, to VCO_1 and VCO_2, respectively. When it comes down to implementation of the SCG, the input signals are beneficially baseband quantities, since the implementation becomes easier, less costly, more accurate, and less power consuming, compared to an implementation at for instance IF. The input signals to the SCG are the I and Q component of both the baseband information signal and the fed back signal. For the feedback to have a linearizing effect the feedback path must be inherently linear. The outputs of the SCG depend on what set of control equations that are chosen (e.g., CALLUM 1, or CALLUM 2). An early decision has to be taken what set of control equations that should be implemented and if the SCG should be analog or digital. A digital implementation is easily re-programmable to cope with another set of control equations, but demands ADC, DAC, anti-aliasing filter, etc. These blocks do not only consume substantial amount of power; they also introduce latency and quantization error due to sampling [84].

Equation (59) contains three basic functions, i.e., summation, subtraction, and multiplication. All these functions are fairly easy to implement. In radio frequency (RF) design it is very common to use differential (balanced) signals, and the main

reason for many designers is the reduced sensitivity to CM induced noise. However, the use of differential signals have more advantages; since no information is present in the CM component the biasing is easier and fewer coupling capacitors are needed. Also by utilizing the complementary signal of the differential signal pair reduces the need for both summation and subtraction to only one of them. (In this design the subtraction is excluded.)

5.2.1 Summation

We have seen in section 3.3.1 that more efficient combining techniques allow the power amplifiers to affect each other by coupling their outputs. This section treats the summation rather generally, and the results can be applied both on signal summation in CALLUM as well as signal addition in general.

Analog summation is most often carried out in the current domain, as currents ideally are easy to add by connecting the outputs to the same junction, and the accumulated current is measured by a detector. In Figure 66 the sum of the currents $i_1, i_2, \ldots, i_n$ are sensed by the detector. In this ideal case the current sources are ideal, meaning that they have infinite output impedance, and the ideal load for current output is a short circuit formed by the ideal detector. In a real implementation the finite output impedance of the current source as well as the quality of the detector will set a limit of the accuracy of the summation. A very coarse approximation of the current source is a single transistor (e.g., BJT, FET). Yet they share the same general properties; they have finite output impedance, and nonlinear transfer from control signal to output current. The modeling of these devices is quite complex, and in [3] the physical models of the active devices for BJT and FET are reviewed. There are many ways to improve the output impedance, and a lot of inspiration can be found by studying current mirror design. The basic implementation is the simple current mirror and improvements of the output impedance is achieved by using the cascode current mirror or any of the Wilson current mirror implementations [24, 108], at the expense of reduced voltage headroom.

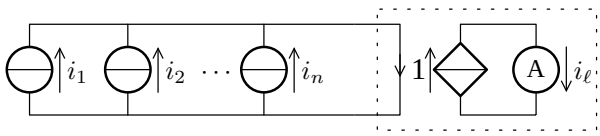

Figure 66: Current addition with ideal source and load impedance.

A more universal solution for improved output impedance and reduced nonlinearity is to apply negative feedback. For a voltage control signal the feedback configuration needed is a series feedback[12]. The nullor circuit and a one-stage FET implemen-

[12] A more correct name is series-series feedback as the current at the output is sensed in series, and the fed back voltage is compared in series with the voltage source.

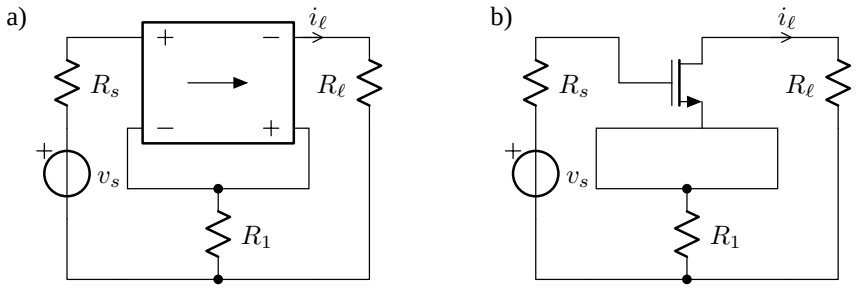

Figure 67: a) A series feedback nullor amplifier. b) A single CS stage implementation of the amplifier.

tation of the series feedback amplifier is shown in Figure 67.

5.2.2 Differential Analog SCG Implementation of CALLUM 2

In this section we will discuss an analog CMOS transistor implementation of the control equation (59) of CALLUM 2. In section 5.2.1 we have seen how the summation (and subtraction) is handled, but the multiplication still remains.

The multiplication is accomplished by a four-quadrant CMOS analog multiplier, based on the folded six-transistor CMOS Gilbert cell in Figure 68. The circuit originates from Gilbert's six-transistor cell, and handles a wide range of input voltages [109]. The valid voltage range of the input signals, v_x and v_y, is further extended by using a folded structure. In many low voltage designs, several levels of stacked transistors are inconsistent with the operating point of the transistors. The tail current, I_T, is part of the bias of the folded multiplier cell. The difference between the two drain currents, i_a^+ and i_a^-, is proportional to the product of v_x and v_y.

The input signals to the SCG should have a DC component such all transistors are kept in the desired region of operation. The nominal value of the DC component is half the supply voltage, $V_{\mathrm{DD}}/2$. The center part of Figure 69 corresponds to v_y in Figure 68, and is the differential implementation of $(I_i - I_o/A_{t\infty})$ in (59). The error signal of the in-phase component is generated from this part of the equation, and since the error signal is very small when the high performance architecture is in transmit mode, any other source of uncertainty has to be low. By using large sized transistors not only the matching error is reduced, also the flicker noise is decreased.

By rewriting the part $(-Q_i - I_i)$ of (59) as $((-Q_i) + (-I_i))$ the corresponding sum is implemented in analogy to v_x in Figure 68. The sign of the signal is just a matter of definition of the ports what is regarded as positive and negative sign of the applied differential input signal.

The second part $(Q_i - Q_o/A_{t\infty})(I_i - Q_i)$ of (59) is implemented by another copy of the circuit in Figure 69. The output signals of this second block are the currents

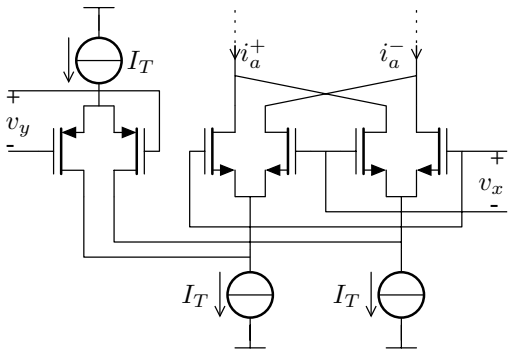

Figure 68: A folded CMOS Gilbert cell used for analog four-quadrant multiplication.

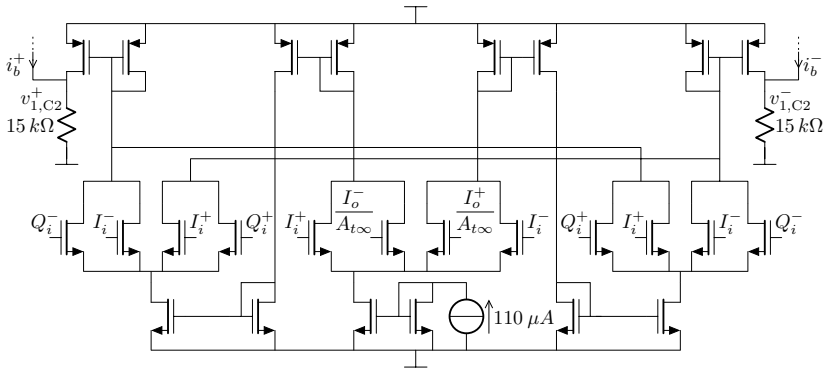

Figure 69: One fourth of the signal component generator of CALLUM 2.

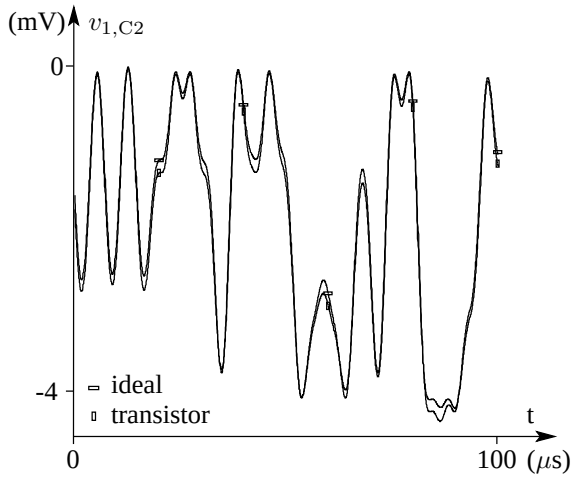

Figure 70: Ideal SCG output versus transistor implementation for an EDGE modulated signal.

indicated by i_b^+ and i_b^- in Figure 69. The output signals of the sub-circuits are summed in a resistor to generate the output signal $v_{1,\mathrm{C2}}$ of the SCG. The control signal $v_{2,\mathrm{C2}}$ is generated in analogy to $v_{1,\mathrm{C2}}$. The total simulated current consumption of the SCG for CALLUM 2 is 2.0 mA.

Simulation of the SCG with EDGE input signal

The transistor implementation of the SCG for CALLUM 2 described above has been simulated and compared with the ideal mathematical realization. The control equation (59) was implemented in Analog Hardware Description Language (AHDL). The input test signal to the SCG was a baseband EDGE modulated signal with 270.833 ksymb/s. For a working transmitter architecture the fed back signal to the SCG closely resembles the input signal. Therefore a slightly attenuated copy of the input signal was used as the feedback signal to the SCG. The differential output from the AHDL block (ideal case) was printed in the same plot as the output, $v_{1,\mathrm{C2}}$, from the CMOS transistor implementation of the SCG, see Figure 70.

The analog implementation of the control equation for CALLUM 2 works well, but a perfect match between the mathematical function and its implementation is hampered by the nonlinear characteristic of the MOS transistor, and also the limited bandwidth of the SCG implementation is in some rare cases visible as slewing.

5.3 The Variable Gain Amplifier

The designer of a large system must be aware of the importance of robustness against parameter variations, parasitics, cross-coupling, mutual coupling between inductors on chip, etc. The experienced designer can be more aggressive since the outcome of this 'multi-variable equation' can be predicted more accurate. The loop gain is one of the most important parameters of the system, since it determines the linearity performance, the stability, and the bandwidth. A VGA should be included to enable adjustment of the loop gain.

5.3.1 Variable Gain Techniques

The VGA can have continuous (analog) or discontinuous (digital) variable gain, and various implementations have been suggested in the literature [110–122]. Common for most variable gain amplifiers is that the conductance of the transistor is changed by varying the bias current. This can be implemented by a variable current source or in a current stealing fashion. Another possibility is to use the MOSFET in its linear region, where the transistor can be used as a variable load or feedback impedance [123]. The square-root dependency of g_m on the bias current for the MOSFET makes it difficult (and power consuming) to obtain a very large range of gain per stage. Often linear-in-dB gain control is desired. This can be accomplished by a linear change of the base-emitter voltage of the bipolar transistor, which results in exponential growth of the current. Only non-feedback stages can be used with this technique, and thereby linearity performance will be limited.

The FET can be used as a variable resistor in its triode region. The $I_\mathrm{D}-V_\mathrm{DS}$ characteristic of a MOSFET in its triode region $V_\mathrm{DS} < V_\mathrm{GS} - V_\mathrm{th}$ is represented by,

$$i_\mathrm{D} = \mu C_\mathrm{ox} \frac{W}{L} \left[(v_\mathrm{GS} - V_\mathrm{th}) v_\mathrm{DS} - \frac{1}{2} v_\mathrm{DS}^2 \right], \tag{123}$$

and the incremental resistance for $V_\mathrm{DS} = 0$ is,

$$r_\mathrm{ds} = \left. \frac{\partial v_\mathrm{DS}}{\partial v_\mathrm{D}} \right|_{V_\mathrm{DS}=0} = \frac{1}{\mu C_\mathrm{ox} \frac{W}{L} (V_\mathrm{GS} - V_\mathrm{th})}. \tag{124}$$

Thus, in order to change r_ds over a given range, $V_\mathrm{GS} - V_\mathrm{th}$ must change by the same ratio. Differentiation of (123) for arbitrary V_DS yields,

$$r_\mathrm{ds} = \frac{\partial v_\mathrm{DS}}{\partial i_\mathrm{D}} = \frac{1}{\mu C_\mathrm{ox} \frac{W}{L} (V_\mathrm{GS} - V_\mathrm{th} - V_\mathrm{DS})}. \tag{125}$$

A rule of thumb, if the resistance variation of r_ds should be in the order of 10:1 then the voltage swing on V_DS must be less than 100 mV (assuming $V_\mathrm{GS} - V_\mathrm{th}$ is limited to a maximum value of about 1 V). The nonlinear behavior of the resistor introduces

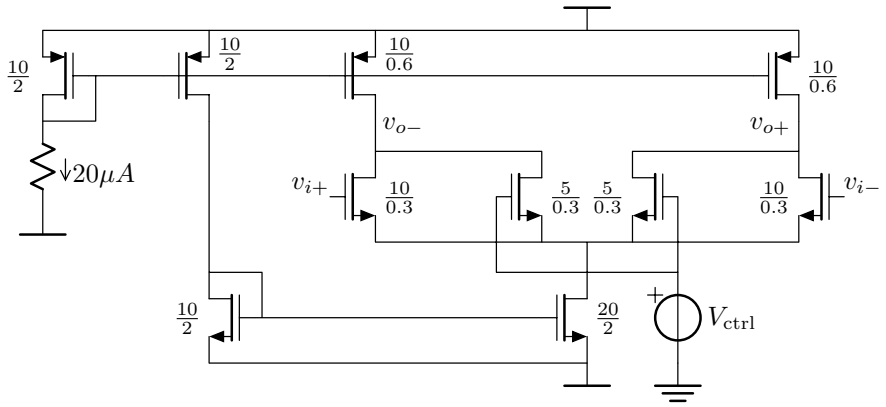

Figure 71: VGA based on current stealing. All ratios in μm.

distortion, which can be diminished by cancelling techniques such as anti-series or anti-parallel connection.

The current stealing VGA technique applied on an anti-series input stage is illustrated in Figure 71. The bias current of the anti-series stage should be chosen in a way that avoids current slewing. This scaling of the current does not change the principle of the circuit, but alters the gain as well as the needed control voltage, V_{ctrl}. A larger control voltage range is needed for larger currents. A variation of the control voltage $\{-0.3\,\mathrm{V} < V_{\mathrm{ctrl}} < 0.2\,\mathrm{V}\}$ results in a variable differential gain of $[10, -20]$ dB.

5.3.2 VGA with Common Mode Control

When implementing equation (59) the gain term, $g_1 k_1 / K_{\mathrm{VCO}}$, must be taken care of by the VGA. This term controls the loop gain and thereby has a strong influence on the stability and spectral performance of the linear transmitter architecture.

As described in [86] the VCOs should be synchronized to the reference frequency to maximize the spectral performance of the system. This will be further motivated in section 5.5.1. The static frequency offset should be corrected, and an efficient way of synchronizing two signals in frequency (and phase) is to use a phase-locked loop (PPL). The static frequency offset corresponds to a DC correction voltage that has to be applied to the VCO. The VGA with CM control is inserted in between the SCG and the VCO. The VGA is described in [106], and shown in Figure 72. The differential input, $(v_i^+,\, v_i^-)$, and the output, $(v_o^+,\, v_o^-)$, are complemented by the CM input, V_{CM}. The circuit will ensure that the CM level of the output signal equals V_{CM}.

The bias current, I_{VGA}, sets the gain and it can easily be varied in the range 3 – 25 V/V with a single-ended parallel RC-load ($R_\ell = 100\,\mathrm{k}\Omega$, $C_\ell = 150\,\mathrm{fF}$). The main problem with low gain is that the -3 dB bandwidth also decreases significantly.

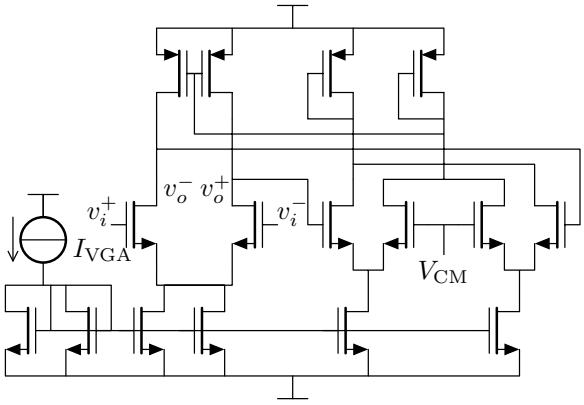

Figure 72: VGA with CM control.

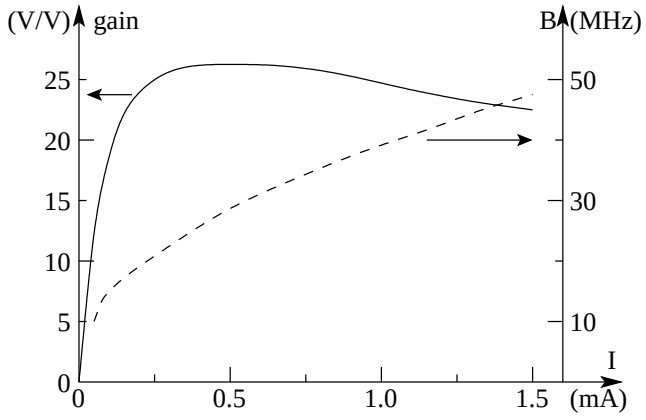

Figure 73: VGA gain transfer and bandwidth versus current.

The gain and bandwidth versus total current consumption for the VGA implemented in the 0.35 μm CMOS process is shown in Figure 73.

5.4 The Voltage Controlled Oscillator

The VCO in the CALLUM system is illustrated in Figure 74. This complementary LC oscillator play an important role in high-frequency circuit design due to its relatively good phase noise, ease of implementation, and differential operation [124]. At the resonance frequency, the current waveform through the tank may be approximated by a sinusoid, due to finite switching time and limited gain of the transistors. The voltage across the tank is then determined by the tail-current times the equivalent resistance of the tank. The voltage always stays within the supply voltage, which is more friendly to the transistors compared to a standard differential LC oscillator having voltage swing above the supply (ideally twice the supply voltage). This increases the risk of transistor break-down caused by too large electrical field in the transistor. In the literature substantial attention has been directed towards the prediction of the phase noise from LC oscillators [124–127]. In summary, to build high performance LC oscillators the designer should use high Q inductors and capacitors/varactors, fast commuting switches, and take measures to reduce the noise contribution from the tail-current source if necessary.

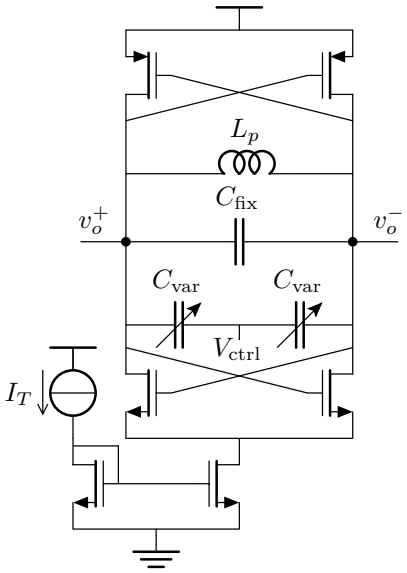

Figure 74: The complementary LC oscillator.

5.4.1 Phase Noise Complications in the Transceiver

The mobile station, MS_1, transmits and the basestation synchronizes its local oscillator (LO) to the channel in which MS_1 is transmitting, see Figure 75a). In the adjacent channel another mobile station, MS_2, is transmitting at the same time. Since the LO signal is not concentrated to a single frequency, also the adjacent channels are down-converted and added to the desired channel, which reduces the SNR of the desired signal. In case of user equipment in the GSM900 band the requirement at 3 MHz offset is -141 dBc/Hz.

Phase noise from the LO in transmitter can block signals in the receiving band (RX) due to far out phase noise, see Figure 75b). This phenomenon sets the far out phase noise requirements.

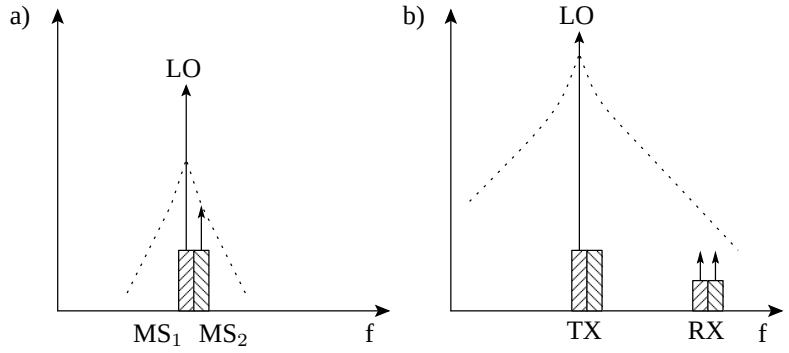

Figure 75: a) LO phase noise causes limited adjacent channel suppression. b) LO phase noise causes spectral leakage into the receiving band.

5.4.2 VCO Inductor

The design of the spiral inductor can be time consuming as it involves optimization of several parameters. The series resistance must be low to get a high Q-value. The Q of the tank is a dominant factor in the phase noise expression. The self-resonance frequency must be higher than the intended frequency of operation. The reference list to work in this area can be made long, but Mohan et al. nicely summarizes the formulas for the inductance value in [128].

A symmetrical inductor with $L = 4.0$ nH is sought. The modified Wheeler formula was implemented in a Maple-script and with $n = 4$ (number of turns), $d_{\text{out}} = 260\,\mu$m (outer diameter), $d_{\text{in}} = 129\,\mu$m (inner diameter), $w = 15\,\mu$m (wire width), $s = 2\,\mu$m (spacing between wires), the correct inductance value was found. The current sheet approximation and data fitted monomial expressions found in [128] gave 4.05 nH and 4.08 nH, respectively. For the AMS 0.35 μm CMOS process the third and fourth layer were connected in parallel with an array of vias at the center of the

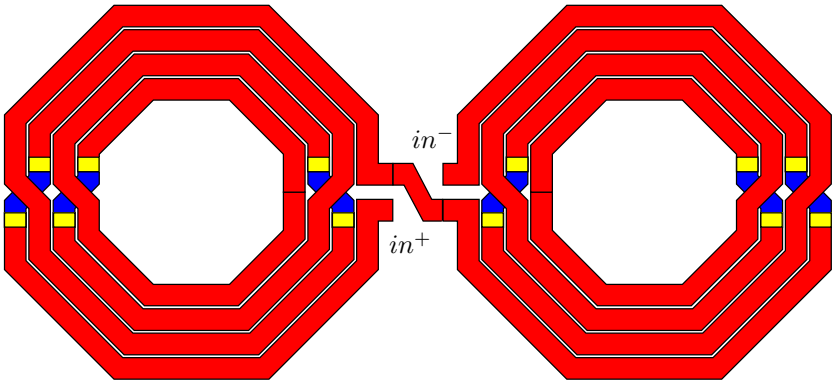

Figure 76: Two inductors connected in anti-series for low magnetic field leakage. Outer diameter $d_{\text{out}} = 260\,\mu$m.

wire, where the current density is low. The DC series resistance was estimated using this formula,

$$R_{\text{ind}} = \frac{R_{\text{sheet}}}{W} \sum_{k=0}^{n-1} \left((d_{\text{out}} - w) - 2k(w+s) \right) \pi. \tag{126}$$

For our inductor $R_{\text{ind}} \simeq 4.35\,\Omega$, which results in $Q \simeq 5.2$. Simulation of the inductor layout using FastHenry[13] gave the desired inductance value and $Q \simeq 5.4$.

The CALLUM architecture uses two VCOs, one in each forward path. These VCOs operate at the same carrier frequency, but on completely different signals. The mutual coupling between the VCOs is undesired, and to minimize the coupling through the magnetic field, the inductor is actually implemented as two symmetrical inductors connected in anti-series, see Figure 76. The magnetic field is concentrated to the device itself and less leakage to the surrounding components. The coupling between the VCOs can also be through the substrate, and to lower that effect the distance separating the VCOs should be as large as possible in combination with guard-rings. Separate supply voltage and ground also help isolating the VCOs from each other.

The variation in resonance frequency of the VCO in the transmit path, found from Monte Carlo simulations, is about $\pm 6\,\%$ from its nominal value, when using the mismatch parameters for the process. When considering only the mismatch between the VCOs, a frequency mismatch of about $0.6\,\%$ is expected from simulations. If there exists an inductive coupling between the VCO inductors the frequency mismatch is diminished. At a low coupling factor of 0.05 the frequency mismatch is $0.4\,\%$, and for a coupling factor of 0.1 the frequency mismatch is only $0.1\,\%$. A wise approach is to reduce the coupling (both capacitive and inductive) as much as possible, even

[13] *www.fastfieldsolvers.com*

though it comes at the expense of larger offset in free-running frequency between the VCOs. This offset frequency error is handled separately and corrected for by a PLL, see section 5.5.

The parasitics of the inductor are modeled using a π-model with inductance and resistance values found above together with parasitic capacitances extracted from the layout. The varactor used is the accumulation-mode MOS capacitor described in [129, 130]. The VCO was simulated with a tail-current of 8 mA, which resulted in a differential amplitude of 2.4 V. Apart from close to the carrier the phase noise characteristic shows a regular second order roll-off with -115 dBc/Hz at 600 kHz offset and -135 dBc/Hz at 3 MHz. The tuning range of the VCO was found to be 956 – 1047 MHz. The sensitivity of the VCO is approximately 60 MHz/V in the middle of the tuning range. In this simulation the capacitive loading effect of the output buffers is not included.

5.5 Design of the PLL for Frequency Synchronization

Jennings et al. reported in [78] that the free-running frequency of the VCOs should be set to the local oscillator (LO) frequency for optimum performance. If there is a difference between the VCO free-running frequency and the LO, the system has to track a signal with an instantaneous frequency increased by the fixed offset. The impact of frequency mismatch is analyzed in more detail in the next section.

5.5.1 Error Caused by VCO Free-Running Frequency Mismatch

In the CALLUM architecture the two VCOs, one in each branch, should be synchronized in frequency to the reference frequency set by the local oscillator (LO) in the feedback path, otherwise this frequency mismatch will cause a voltage (or current) error at the output. In the following the frequency error is assumed to be a static frequency offset. The output signal is an amplified replica of the input plus some additional discrepancy,

$$I_o(t) = A_{t\infty}\{I_i(t) + I_{\rm dis}(t)\} \quad (127)$$

$$Q_o(t) = A_{t\infty}\{Q_i(t) + Q_{\rm dis}(t)\}. \quad (128)$$

The constant frequency offset corresponds to a DC correction voltage that has to be applied to the corresponding VCO according to

$$V_1 = \frac{f_{\rm VCO1} - f_c}{K_{{\rm VCO},f}} = \frac{\Delta f_1}{K_{{\rm VCO},f}}, \quad (129)$$

$$V_2 = \frac{f_{\rm VCO2} - f_c}{K_{{\rm VCO},f}} = \frac{\Delta f_2}{K_{{\rm VCO},f}}, \quad (130)$$

where $f_{\rm VCO1}$ ($f_{\rm VCO2}$) denotes the free-running frequency of $\rm VCO_1$ ($\rm VCO_2$). After straightforward calculation using (46), (127), and (129), the following expression for

$I_{\rm dis}$ was found,

$$\begin{aligned} I_{\rm dis} &= \frac{\pi}{g_1 k_1}\left[\Delta f_1\left(I_i\sqrt{\frac{4r_{\rm max}^2}{r^2}-1}+Q_i\right)\right. \\ &- \left.\Delta f_2\left(I_i\sqrt{\frac{4r_{\rm max}^2}{r^2}-1}-Q_i\right)\right] \qquad (131) \\ Q_{\rm dis} &= \frac{\pi}{g_1 k_1}\left[\Delta f_1\left(Q_i\sqrt{\frac{4r_{\rm max}^2}{r^2}-1}-I_i\right)\right. \\ &- \left.\Delta f_2\left(Q_i\sqrt{\frac{4r_{\rm max}^2}{r^2}-1}+I_i\right)\right]. \qquad (132) \end{aligned}$$

The expression for $Q_{\rm dis}$ was calculated analogously. There exists two obvious ways to decrease the effect of frequency mismatch between the VCOs, namely

1. increase the loop gain,
2. synchronize the VCOs, i.e., remove Δf_1 and Δf_2.

An efficient remedy to minimize the influence of the frequency offset is to apply sufficient loop gain, affected by $g_1 k_1$. However, the amount of loop gain also affects the high frequency behavior, and for large values of the loop gain instability will result for this system. Of course, as much loop gain as possible should be used, but to make the system perform at its best, spectral wise, frequency mismatch has to be minimized by means of frequency synchronization.

5.5.2 VCO Frequency Synchronization

The correction of the frequency mismatch is a true biasing problem, where the biasing and the information should be orthogonal. Often biasing and information are located in non-overlapping frequency bands, but since the baseband signal stretches from DC, the biasing is here more suited for a time division approach. Therefore, the loop is broken during synchronization by changing the state of the sync signal in Figure 77.

The frequency component of the signal is not possible to detect instantaneously. In frequency demodulators an FM-to-PM or FM-to-AM conversion has to be made first, followed by a PM- or AM-demodulator [131]. For the reason given above a direct frequency comparator does not exist, but an efficient way of synchronizing two signals is to use a phase-locked loop (PPL). The PLL includes a phase detector as a sub-block. Often a multiplier is used as a simple phase detector, which can be seen from (133).

$$\begin{aligned} v_d &= v_i(t)v_{\rm osc}(t) \\ &= A_i\sin(\omega_i t+\theta_i)A_{\rm osc}\cos(\omega_i t+\theta_{\rm osc}) \\ &= \frac{A_i A_{\rm osc}}{2}\sin(\theta_i-\theta_{\rm osc})+\frac{A_i A_{\rm osc}}{2}\sin(2\omega_i t+\theta_i+\theta_{\rm osc}) \qquad (133) \end{aligned}$$

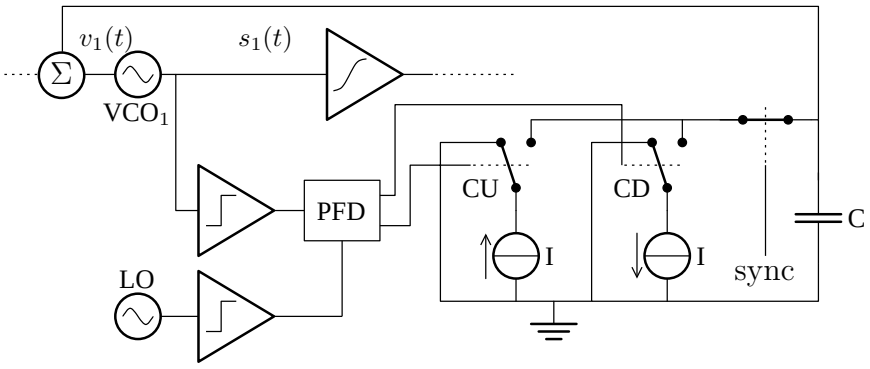

Figure 77: Frequency synchronization of VCO_1 in the transmit path.

If $\sin(\theta_i+\theta_{\mathrm{osc}}) \ll 1$ then $\sin(\theta_i+\theta_{\mathrm{osc}}) \simeq \theta_i+\theta_{\mathrm{osc}}$. Under this condition the multiplier gives an output proportional to the phase difference. This explains why the symbol for a phase detector often is the same as for a multiplier. For further reading about frequency synchronization the following books can be recommended [132–136].

The time to achieve frequency synchronization is less critical for this test system, but to guarantee lock a phase-frequency detector (PFD) is used. During the frequency synchronization phase the correct DC voltage for phase-lock is stored on the capacitor, C, in Figure 77. This node is sensitive to substrate noise, which will be injected during transmission. The capacitor should be large both to withstand a small leakage current without harmful voltage drop and reduced noise sensitivity.

5.5.3 Phase-Locked Loop Design

The target for the design is that the remaining frequency error after frequency synchronization should be less than 10 % of the bandwidth of the input signal. Assume the input signal is an EDGE modulated signal, (200 kHz channel bandwidth), then the error should be less than 20 kHz. As the carrier frequency is around 900 MHz the phase-frequency detector (PFD) must be preceded by a prescaler. A prescaler is a frequency divider with fixed division ratio, and here a prescaler with division ratio of two is used in a cascade of 6 stages to scale down the carrier frequency, $900\,\mathrm{MHz}/2^6 \simeq 14.1\,\mathrm{MHz}$. At this lower frequency the waveform from the PFD is accurate and reliable. The use of a PFD in combination with current charge pump is for instance described in [137].

The pull-in time of the PLL is non-critical for this test circuit, as long as it acquires lock in a few ms. The correct DC voltage applied to the VCO should be stable for at least the time during one transmission burst of data, (approximately. 1 ms). If frequency synchronization should not be done after each transmitted burst of data a longer time-constant is needed. For test purposes a much longer time-constant would

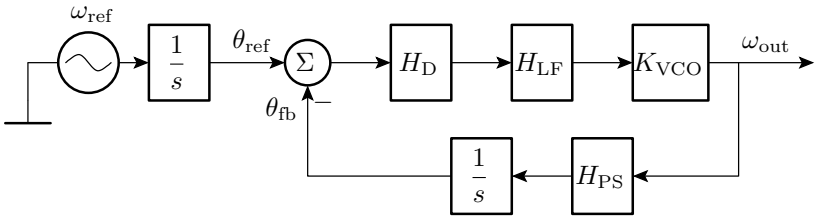

Figure 78: Block schematic of the PLL for baseband calculations.

be convenient, from which the size of the capacitor, C, is set. The transfer function for the VCO is K_{VCO}, and the prescaler is modeled as an attenuator with a transfer function H_{PS}. The PFD and the charge pump (CHP) are treated as one unit. The transfer function for the PFD and the CHP is given by,

$$H_{\mathrm{D}} = \frac{\{I_{\mathrm{chp}}\}_{\max} - \{I_{\mathrm{chp}}\}_{\min}}{4\pi}. \tag{134}$$

For a symmetric current I_{chp} the transfer of the PFD and the CHP can be simplified to $H_{\mathrm{D}} = I_{\mathrm{chp}}/2\pi$. The loop filter is chosen as the simplest practical filter according to [138]. The filter is a series connection of a capacitor and a resistor, i.e, the introduction of a pole-zero pair. The transfer function for the loop filter is given by,

$$H_{\mathrm{LF}} = R_1 + \frac{1}{sC_1} = \frac{1 + sR_1C_1}{sC1}. \tag{135}$$

From Figure 78 the loop gain, $G(s)$, can be formulated as,

$$G(s) = -\frac{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}(1 + sR_1C_1)}{2\pi s^2C_1}. \tag{136}$$

The closed-loop transfer is given by,

$$H(s) = \frac{\theta_{\mathrm{fb}}}{\theta_{\mathrm{ref}}} = \frac{-G(s)}{1 - G(s)}, \tag{137}$$

and the characteristic polynomial for the closed-loop system is,

$$\mathrm{CP}_{\mathrm{closed-loop}} = s^2 + \frac{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}R_1}{2\pi}s + \frac{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}}{2\pi C_1}. \tag{138}$$

Equation (138) should be matched with the equation for the system poles in Butterworth position for maximum flat magnitude (MFM) transfer. The characteristic polynomial for the MFM transfer is given by (use $p_1^{'}p_2^{'} = \omega_0^2$ and $-(p_1^{'} + p_2^{'}) = \sqrt{2}\omega_0$),

$$\mathrm{CP}_{\mathrm{MFM}} = s^2 + \sqrt{2}\omega_0 + \omega_0^2. \tag{139}$$

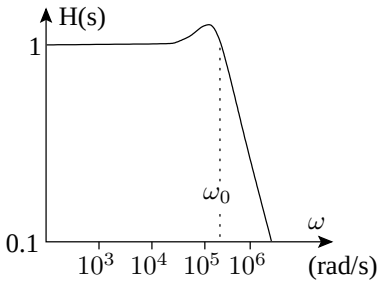

Figure 79: Closed-loop transfer for the PLL with $\omega_0 = 175$ krad/s.

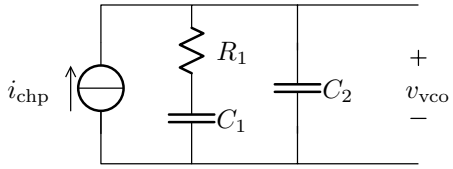

Figure 80: Loop filter for reduced ripple and increased PLL performance.

Identification gives the loop filter values,

$$C_1 = \frac{I_{\text{chp}} K_{\text{VCO}} H_{\text{PS}}}{2\pi\omega_0^2} \tag{140}$$

$$R_1 = \frac{2\pi\sqrt{2}\omega_0}{I_{\text{chp}} K_{\text{VCO}} H_{\text{PS}}}. \tag{141}$$

Most PLLs are designed for maximum flat magnitude, even though in the literature the damping factor $\zeta = 1/\sqrt{2}$ is used instead [107, 132, 134]. The closed-loop transfer is shown in Figure 79 and there is a slight peaking at the upper end of the band, since the loop zero is not a phantom zero.

Second Order Loop Filter

The simplest practical loop filter described in [138] does not filter out the spikes from the switched signal of the charge pump. These high frequency spikes can be filtered out to a certain extent by using an additional capacitor, C_2, in parallel with the R_1C_1 combination. The loop filter is illustrated in Figure 80. The size of C_2 is often chosen about 20 times smaller than C_1 [138]. The PLL is designed for the first order loop filter, and the small capacitor C_2 connected in parallel is assumed to have little influence on the loop dynamics. The extra capacitor makes the filter a second order filter (and

a thereby the PLL a third-order). It is possible to find the exact values of C_1, C_2, R_1, for MFM transfer and real system pole placement, respectively. The transfer function for the second order loop filter is given by,

$$H_{\mathrm{LF2}} = \frac{1 + sR_1C_1}{s\left\{(1 + sR_1C_1)C_2 + C_1\right\}}. \tag{142}$$

In (142) is the zero positioned at $n_{\mathrm{LF2}} = -\frac{1}{R_1C_1}$, and the poles $p_{1,\mathrm{LF2}} = 0$, $p_{2,\mathrm{LF2}} = -\frac{C_1+C_2}{R_1C_1C_2}$. The loop gain is now,

$$G(s) = \frac{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}(1 + sR_1C_1)}{2\pi s^2\left\{(1 + sR_1C_1)C_2 + C_1\right\}}, \tag{143}$$

and the characteristic polynomial for the closed-loop system is,

$$\mathrm{CP}_{\mathrm{closed-loop3}} = s^3 + \frac{C_1 + C_2}{C_1C_2R_1}s^2 + \frac{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}}{2\pi C_2}s + \frac{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}}{2\pi C_1C_2R_1}. \tag{144}$$

For correct system pole placement the characteristic polynomials should be matched, where the characteristic polynomial for Butterworth position of the system poles is given by,

$$\mathrm{CP}_{\mathrm{MFM3}} = s^3 + 2\omega_0 s^2 + 2\omega_0^2 s + \omega_0^3. \tag{145}$$

For increased robustness all system poles are placed on the real axis right on top of each other at ω_0. This pole placement will be referred to as real poles. The characteristic polynomial for real poles is given by,

$$\mathrm{CP}_{\mathrm{real3}} = s^3 + 3\omega_0 s^2 + 3\omega_0^2 s + \omega_0^3. \tag{146}$$

Identification of the CPs gives the loop filter values for MFM,

$$C_1 = \frac{3}{2}\frac{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}}{2\pi\omega_0^2} \tag{147}$$

$$C_2 = \frac{1}{2}\frac{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}}{2\pi\omega_0^2} \tag{148}$$

$$R_1 = \frac{4}{3}\frac{2\pi\omega_0}{I_{\mathrm{chp}}K_{\mathrm{VCO}}H_{\mathrm{PS}}}. \tag{149}$$

It is easy to see that the ratio is 3 between C_1 and C_2, quite a bit from the recommended 20 in [138]. The zero and poles of the loop filter are given by $n_{\mathrm{LF2}} = -\omega_0/2$, and $p_{1,\mathrm{LF2}} = 0$, $p_{2,\mathrm{LF2}} = -2\omega_0$, respectively. The amplitude plot of the closed loop

system shows similar behavior as Figure 79, but with a steeper roll-off, and is not repeated here.

The same identification of the CPs for real poles yields,

$$C_1 = \frac{8}{3}\frac{I_{\rm chp}K_{\rm VCO}H_{\rm PS}}{2\pi\omega_0^2} \tag{150}$$

$$C_2 = \frac{1}{3}\frac{I_{\rm chp}K_{\rm VCO}H_{\rm PS}}{2\pi\omega_0^2} \tag{151}$$

$$R_1 = \frac{9}{8}\frac{2\pi\omega_0}{I_{\rm chp}K_{\rm VCO}H_{\rm PS}}. \tag{152}$$

Here the ratio is 8 between C_1 and C_2, and the zero and poles of the loop filter are given by $n_{\rm LF2} = -\omega_0/3$, and $p_{1,\rm LF2} = 0$, $p_{2,\rm LF2} = -3\omega_0$, respectively.

The system poles are often positioned as real poles due to unknown spread in loop gain, primarily caused by the non-constant gain of the VCO, $K_{\rm VCO}$, and the phase-frequency detector and charge pump combination.

Short Description of the Implementation

The signals from the VCO and the LO are prescaled by a six-stage divide-by-two cascade. The prescaler was presented by Chen in [139], and it can operate accurately on a 900 MHz signal when using the 0.35 μm CMOS process. The cascade of prescalers are followed by a limiter implemented as a two-stage non-feedback amplifier in an ASCS-CPCS configuration. The signals from the limiters closely resembles square-waves, which are fed to the phase-frequency detector (PFD). This is the conventional PFD found in [140] with the extra inverters for correct timing of the signals. The outputs from the PFD are two binary signals, charge up (CU) and charge down (CD), that are the inputs to the charge pump illustrated in Figure 81. The output from the CHP is a current that is transformed to a voltage via the loop filter shown in Figure 80. The voltage over the loop filter is fed to the CM pin of the VGA depicted in Figure 72, which provides this control signal to the VCO, and thereby the loop is closed.

5.6 Downconversion

The downconversion part of the CALLUM transmitter shifts the information from RF down to baseband. In this case a direct downconversion topology (also known as homodyne) was chosen. In a homodyne receiver the frequency translation is performed in one stage. The advantages with this structure are simplicity, fewer components, and no need for image frequency rejection. Among the drawbacks the sensitivity to 1/f-noise and DC offset can be mentioned [141].

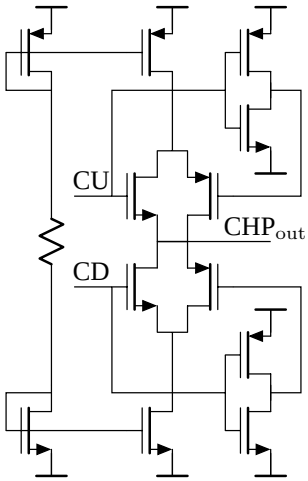

Figure 81: Symmetrical CMOS charge pump.

The conversion gain is defined as the ratio between the amplitude of the baseband and the RF signal. The conversion gain depends on the actual implementation and the waveform of the LO signal. An ideal situation regarding noise and linearity is to have ideal switches, instantly commuting from the off-state to the on-state, and vice versa. Here the voltage switches are approximated by field effect transistors. The waveform of the LO signal, controlling the state of each transistor, should be such that the time spent in the region between the off- and on-state is minimized. Normally a rather large square-wave or sinusoid will do.

The linearizing effect of the feedback system relies on a linear feedback factor. This means that the downconversion must be highly accurate if the transmitter should be highly linear. The design space is limited by a few fundamental boundaries, i.e., the noise floor, the distortion, and the clipping, as seen in Figure 82. To get a certain signal-to-noise ratio (SNR) the input level has to be in a certian interval not to be degrade by neither noise nor distortion.

5.6.1 Block Level Schematic of the Downconversion

In Figure 83 the block schematic of the downconversion is illustrated. The mixing core is synthesized by two passive mixers, and since the RF signal source is not a perfect voltage source, the RF signal is sensitive to loading effects. This loading effect comes from the mixers and in particular from the varying load over time. This type of loading is harmful for the linearity of the downconverter, and to reduce the cross-coupling between the mixers, each input is isolated by a voltage follower. Other ways to isloate the mixer inputs from each other exist, and connection of series resistors

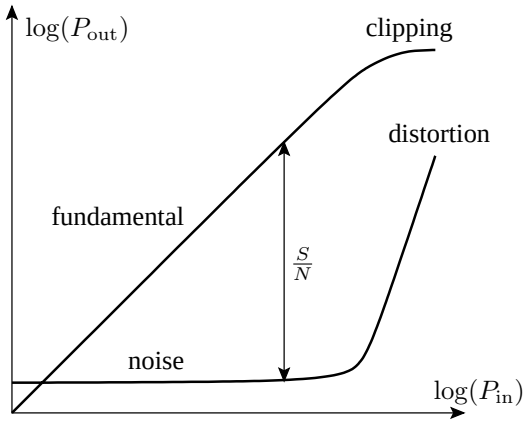

Figure 82: Fundamental limitation of the dynamic range.

can be mentioned as one example of a simple solution. However, it is both noisy and has limited isolation. The output RF signal of the CALLUM transmitter is large, and there is no need for gain in the feedback path, which motivates the chosen passive mixer structure.

Isolating Voltage Follower

The input of the downconverter should be matched to $50\,\Omega$ input resistance for maximum power transfer. Often a common-gate stage is used for simple and accurate input matching, but here a more brute force solution is adopted. After a coupling capacitor the signal is shunted with a $50\,\Omega$ resistor. The main penalty with this solution is the added noise, but on the other hand area consuming inductors are not needed as when using a tuned CG-stage. The voltage follower is illustrated in Figure 84, and it can be seen that the DC level drops by one threshold voltage from $V_{\mathrm{DD}}/2$ at the output of the voltage follower. This voltage drop has to be compensated for before feeding the signal to the SCG. A level shift at the end of the downconverter adds the necessary DC voltage.

Mixer

The main task of the downconversion is to perform a shift in frequency from RF to a lower frequency, e.g., IF or baseband (also known as zero IF). In a multiplicative manner of two sinusoids this is accomplished by,

$$A\cos(\omega_1 t)B\cos(\omega_2 t) = \frac{AB}{2}\{\cos((\omega_1-\omega_2)t) + \cos((\omega_1+\omega_2)t)\}. \qquad (153)$$

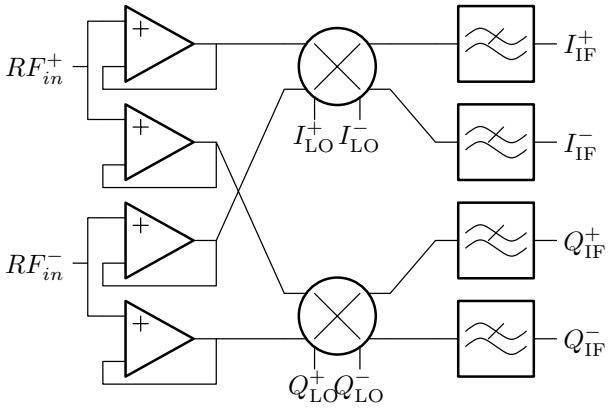

Figure 83: Block schematic of the downconversion.

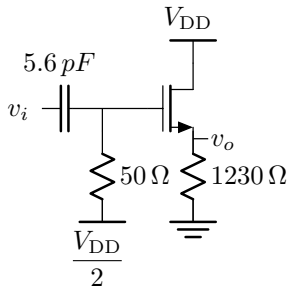

Figure 84: CS implementation of the voltage follower for isolation of the RF inputs.

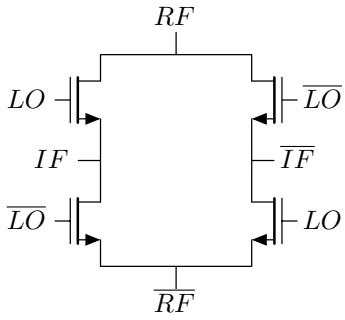

Figure 85: The passive double-balanced mixer implemented with FETs.

The sum (difference) frequency is filtered out when performing down (up) conversion. A bit simplified, only half of the energy is used and the other half is filtered out. Typically one of the signals is the RF ($A\cos(\omega_1 t)$) and the other one is the LO ($B\cos(\omega_2 t)$). In case of a passive mixer the LO signal can be approximated by a square-wave with a fundamental amplitude of $4B/\pi$. By using the result in (153) the maximum conversion gain for the passive mixer is $G_c = 2AB/\pi$.

The passive mixer has some attractive properties, such as the potential for extremely low-voltage and low-power operation, as well as low 1/f-noise. The CMOS technology offers excellent voltage switches, and the RF signal is switched directly in the voltage domain. Thereby the square-wave dependency of the transistor is avoided, which is beneficial for the linearity of the mixer [7]. The passive double-balanced mixer is shown in Figure 85.

Low-Pass Filter

The low-pass filter after the mixer filters out the sum frequency, and the -3 dB frequency is chosen such the low-pass filter does not contribute significantly to the phase degradation. A first order low-pass RC-filter with a cut-off frequency of about 100 MHz was found as a good compromise from simulations. If larger attenuation of the sum frequency is needed, then increased order low-pass filter is demanded.

Simulation of the Downconversion

Three different mixers have been tested, i.e., the active double-balanced Gilbert mixer, the passive mixer, and the passive mixer with a bootstrapping technique to improve the linearity [142]. In summary, the simulations showed that the Gilbert mixer gave about 5 dB of gain and worse linearity performance compared to the passive mixer. Since the gain is not needed, the Gilbert mixer was abandoned. The linearity improvements of the bootstrapped passive mixer is due to a more stable duty cycle of the switches in the passive mixer. This is accomplished by adding the IF signal to the LO signal [142].

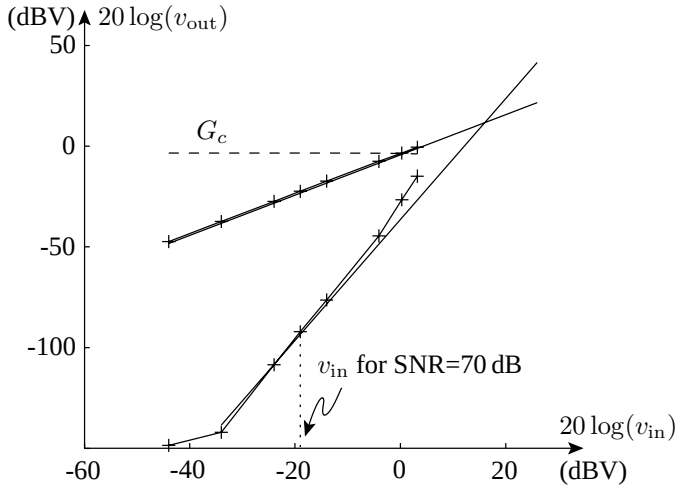

Figure 86: Simulation results for the complete downconversion with passive mixers.

However, simulations show that the IIP_3 is increased by a few dB, but to achieve the desired SNR$\geq$70 dB a smaller signal is actually required than for the standard passive mixer. The simulation results for the complete downconverter with the passive mixer is plotted in Figure 86. The downconverter is broadband and will give the same result for any IF up to several MHz.

The dynamic range of the downconversion has to be superior to the desired dynamic range of the transmitter. The intermodulation distortion causes spectral emissions to rise, and in particular the requirements set by the spectrum mask at the adjacent channel are the most difficult to fulfill. For EDGE the spectrum mask demands -54 dBc at 400 kHz offset and -60 dBc in the range 600 – 1800 kHz. The input level, $v_{\rm in}$, of the downconverter is attenuated such that SNR$\geq$70 dB, see Figure 86.

5.7 Quadrature Signal Generation

The receiver needs quadrature signals to the detect the information concealed in the in-phase and quadrature-phase of the RF signal. There are numerous ways to create quadrature LO signals, and in those instances where square-wave signals are acceptable, a solution is to use a VCO at double frequency followed by a digital divider [7]. An interesting feature with this technique is that the coupling between LO signal and RF signal is significantly reduced. Another way to generate quadrature LO signals is to use a single signal source followed by a poly-phase filter. The poly-phase filter is often a passive RC-CR filter with one or more stages. More stages mean more broadband, better phase balance, but also more noise and signal attenuation [143, 144].

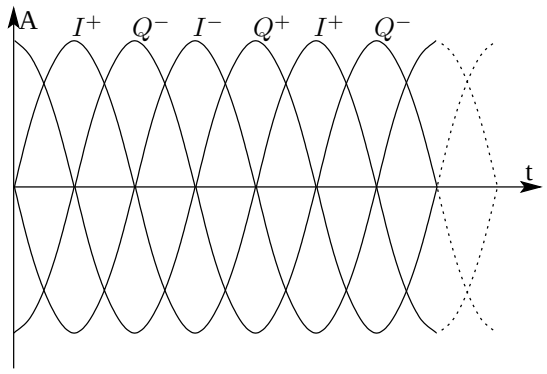

Figure 87: Order of the quadrature balanced signals.

The most direct way of quadrature signal generation is to use a quadrature VCO [145, 146]. To date, this exotic solution is not by far as widely accepted as the other methods mentioned. The phase relations of the quadrature output signals are shown in Figure 87.

5.7.1 The Poly-Phase Filter

A practical poly-phase network is presented in [143], and a n^{th} order passive RC poly-phase filter is shown in Figure 88. The filter values are determined by the center frequency of operation according to,

$$\omega_c = \frac{1}{R_1 C_1}. \tag{154}$$

The value of the capacitor is often determined by the available chip area (or acceptable cost) and the resistor by the amount of acceptable added noise and/or the maximum allowed attenuation. In the normal case, the LO drive signals to the poly-phase filter are large, which reduces the importance of noise optimization.

No finite network can provide both a constant phase shift and a constant gain over an infinite frequency range, so one must settle for approximations that work well only over some limited range [7]. The number of stages in the poly-phase filter is chosen with respect to the quality in phase and amplitude of the drive signal, as well as the demands on the quality after the poly-phase filter over the entire frequency range.

In this design an on-chip second order poly-phase filter was implemented to create the necessary LO phases. In addition to this practical passive poly-phase network, a few other blocks were implemented to produce the quadrature LO signals, see Figure 89. For test purpose the channel select frequency is taken from a signal generator. Of course, for more product orientated design a higher degree of integration is necessary, and the channel frequency will be generated from a piezo electric crystal

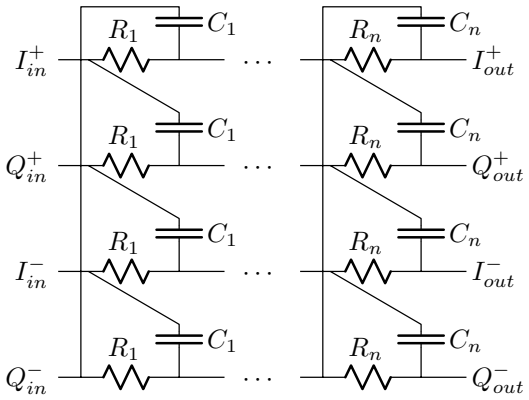

Figure 88: n:th order poly-phase filter.

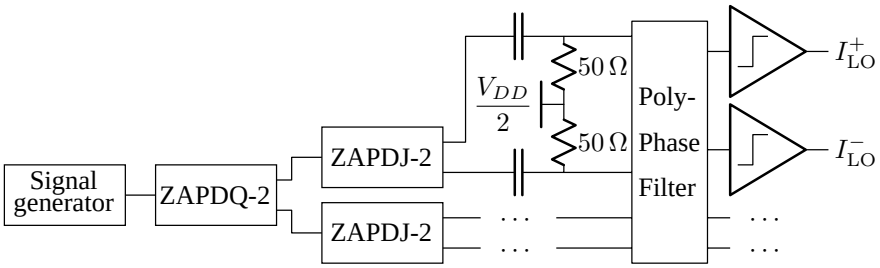

Figure 89: Block schematic of the quadrature signal generation.

reference and a sophisticated PLL. Here the LO signal from the generator is split into quadrature signals via a ZAPDQ-2[14]. Each of these signals are transformed to differential (balanced) signals by a ZAPDJ-2. Before the balanced quadrature signals enter the chip the bias component is removed by coupling capacitors. The input matching is accomplished by shunting the signal path with a 50 Ω resistor. The poly-phase filter is followed by limiters, which produce square-wave output signals with an amplitude limited by the supply voltage. The limiter is further described in 5.7.2.

5.7.2 Post-Poly-Phase Filter Limiter

The output of each branch of the poly-phase filter drives a limiter. The limiter discriminates the amplitude information from the signal to a binary signal. The phase information is however preserved and the output is ideally a square-wave signal with rail-to-rail swing. The limiter itself is a two-stage open-loop amplifier. The input

[14]Mini-Circuits, *www.minicircuits.com*

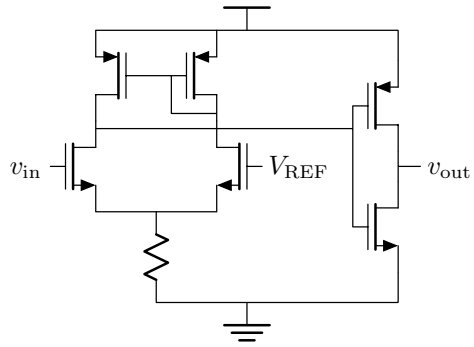

Figure 90: The limiter succeeding the poly-phase filter.

stage is an anti-series common-source (ASCS) stage followed by a complementary-parallel common-source (CPCS) stage, also known as a digital inverter. The biasing of the input stage is facilitated by use of an active load as it automatically balances the quiescent currents and doubles the gain of the stage [24]. The implementation of the limiter is shown in Figure 90. The input signal, v_{in}, is compared to a reference level, V_{REF}, here chosen to half the supply voltage, and the output is v_{out}.

5.7.3 Simulation of the Quadrature Signal Generation

The on-chip part of the quadrature signal generation was simulated to find an appropriate amplitude of the input signals to the poly-phase network. The amplitude of each input to the device under test (DUT) is given by v_{in} in Figure 91, and the peak-to-peak value together with the average of the output signal were measured. The goal is to have rail-to-rail switching ($v_{\mathrm{peak-to-peak}} = 3.3\,\mathrm{V}$) symmetrically located between the supply rails ($v_{\mathrm{avg}} = 1.65\,\mathrm{V}$). From Figure 91 a reasonable input amplitude for each phasor should be 0.4 V or above for full swing and an average voltage near the middle of the supply.

The maximum output voltage level from the Rohde & Schwarz SMIQ 06 B (available in the laboratory) is 1 V, and this signal is then split into four phasors according to Figure 89. The power is ideally conserved by the power splitters and since the impedance level is not changed ($50\,\Omega$ in this case) the resulting output level is 0.5 V per input. However, the power splitters have some attenuation and the measured output level is approximately 0.35 V over the frequency range 800 – 1000 MHz. This amplitude level is on the low side of what can be considered as a suitable input amplitude. The output level of the signal generator is therefore boosted by an external amplifier, e.g., Mini-Circuits ZHL-2-8, before entering the poly-phase network.

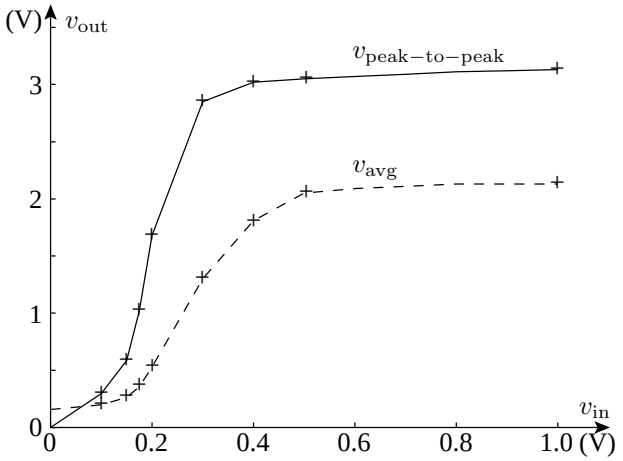

Figure 91: Simulation results for the quadrature generation block.

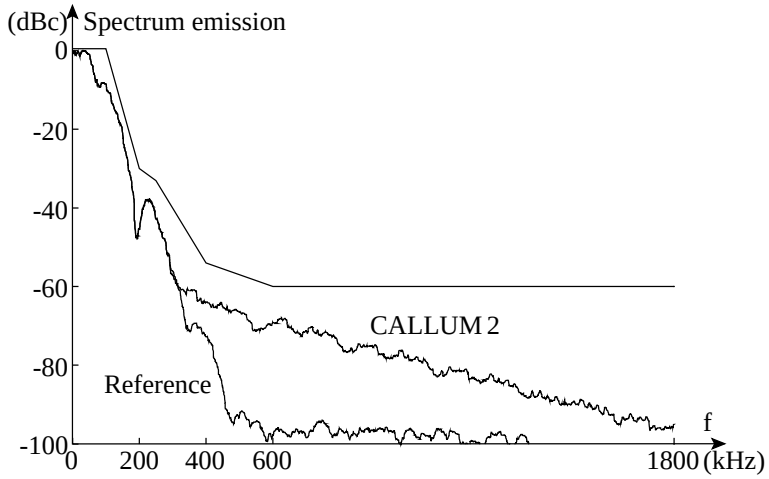

Figure 92: The simulated spectrum from CALLUM 2 together with the reference (input) signal and the spectrum emission mask for EDGE.

5.8 Simulation of CALLUM 2

The blocks in the CALLUM 2 transmitter architecture are now described, and to study the effect of a non-ideal SCG (see section 5.2.2) and VGA (see section 5.3.2) in terms of transmit spectrum emissions. The schematic view was built at block level using AHDL, apart from the SCG and VGA, which are at transistor level. The loop is assumed to be delayless such the influence of the implemented blocks could be isolated. Normally, the effect of the loop delay on spectrum emissions is minor as long as the stability of the loop can be guaranteed. As the loop delay is increased and the phase margin becomes small, the far out spectrum rises as a last warning before instability.

The simulated spectra are based on a random data sequence containing 33 ksymbs, using a $3\pi/8$-shifted 8PSK modulation and filtering according to the EDGE standard. The loop gain was increased until the system could not achieve lock, but an extensive search for parameter optimization has not been performed. The EDGE spectrum emission mask for mobile stations in the GSM 900 band is plotted in Figure 92 together with the simulated spectrum of the transmitted signal for CALLUM 2 and the reference. From the reference it is possible to detect any spectrum degradation, and gives a quick estimate of the SNR. The gap to the spectrum mask is $\simeq$8 dB at the adjacent channel 400 kHz off center. This is a desirable design margin. From the simulated spectral performance the SCG and VGA implementation proved to be appropriate for an EDGE modulated signal.

Chapter 6

Summary and Contributions

In this chapter, we first give a summary of the general results of this thesis. We continue with a presentation of the principal results of each paper.

The designers' strive to fulfill the consumers' demand of new data hungry services, e.g., Internet browsing, video and image messages, has conducted the development of new standards which use linear digital modulation schemes. The utilization of a linear modulation in a mobile environment requires a linear transmitter and a linear power amplifier in particular, which prevent the spectral properties of the modulated signal to be deteriorated by intermodulation distortion. A nonlinear transmitter causes the spectrum to rise in the adjacent channels and will degrade the signal-to-noise ratio (SNR). Lower SNR means reduced data rate for users of these neighboring channels.

A linear power amplifier is a necessity that cannot be compromised and the limited battery capacity in mobile equipment calls for efficient transmitters. The power amplifier is one of the most power consuming parts. For high power amplifiers in basestations the efficiency parameter is equally important, but for another reason: the heat dissipation. It is the combination of sufficient linearity and high power efficiency that makes the transmitter problem so delicate. We have seen several techniques to build a linear transmitter, where the most common solution is to apply linearization to a power efficient but nonlinear power amplifier. Feedforward, feedback, and predistortion are examples of established linearization techniques.

In an amplifier with a constant-envelope input signal the nonlinearities of the power amplifier characteristic is not exercised. By using a signal decomposition of the amplitude- and phase-modulated input signal such the power amplifier is fed with a constant-envelope phase-modulated signal, a linear transfer is possible even when using nonlinear amplifiers. LINC and CALLUM are two viable examples of this technique.

6.1 General Results

LINC and CALLUM share the same fundamental properties since they both can have 100 % efficiency for all signal levels, and they rely on the same idea of signal decomposition and recombination. The spectral regrowth appearing inside the linear transmitter architecture as the linear baseband signal is converted into two constant-envelope signals, s_1 and s_2, has been studied for $3\pi/8$-shifted 8PSK and $\pi/4$-shifted

QPSK modulation, respectively. The bandwidth expansion is significant, which puts even tighter demands on the circuits as they have to operate on relatively more wideband signals. The control signals to the VCOs, v_1 and v_2, for the CALLUM architecture have proven to be even more wideband than s_1 and s_2. From a coarse approximation, the instantaneous frequency of s_1 (and s_2) is shown to be proportional to the modulation depth. Therefore, suitable modulations are those with $m_d < 10$. Any modulation allowing zero crossings causes severe problems for some derivatives of CALLUM as a zero appears in the denominator of the control equations. This should be avoided as it forces the system out of lock.

The ideal efficiency can only be approached by using a signal combination technique with power amplifier interaction. Compared to perfect isolation of the PAs, interaction increases the distortion level. This combination technique is not intended for LINC, but for CALLUM as it has a global feedback loop that can correct these added errors. To keep the efficiency at decent level when the output power is reduced, the power amplifier should be supplied with a variable bias, primarily a variable supply voltage.

The control equations for several versions of the CALLUM architecture based on a Cartesian representation have been derived, and an investigation of the loop properties based on linearized control equations has been performed. The analytical loop gain equations are displayed in 3D contour plots, and already at this point CALLUM 1 and CALLUM 1lin are predicted superior to CALLUM 2 in terms of distortion suppression. This has been shown by baseband modeled architectural system simulations on both the EDGE and W-CDMA standard. The Achilles' heel of CALLUM 2 is its lack of loop gain for low envelope signals, which limits its application to the EDGE standard. The simplicity of the control equations for CALLUM 2 is attractive for an analog implementation of the CALLUM transmitter architecture.

The loop time delay originating from wave propagation and tuned filters will effectively limit the achievable loop bandwidth and the versatility of the CALLUM system. The time delay together with the integrator effect of the VCO are the main contributors to the phase curve. The effect of time delay enforced the use of a band limiting lag-lead compensation to increase the robustness of the system. The lag-lead compensation network keeps the low-frequency loop gain. This type of frequency compensation is very common in PLL designs, with which CALLUM has many similarities.

The building blocks needed for implementation of a CALLUM 2 system have been presented, and the main focus was directed to the differential analog implementation of the SCG. The mutual coupling between the VCOs has negative effect on the system performance. To reduce this coupling a special inductor layout was used, i.e., two symmetrical inductors connected in anti-series. A PLL with a phase-frequency detector was used for frequency synchronization of the VCOs to the reference frequency. The downconversion was based on a passive mixer core that translates the RF signal down to baseband. The CALLUM 2 system has been simulated with the SCG and the VGA blocks at transistor level and proven to be adequate for a modulated baseband signal according to the EDGE standard.

6.2 Results of Research Papers

This section provides a brief presentation of each paper included in the thesis. Furthermore, the contributions by each paper in the field of analog radio electronics are discussed. The papers are written by the author, who also conducted the modeling, development, programming, and analysis that led to the material presented in the papers. The coauthors assisted in various ways, e.g., as the persons with the idea of the investigation, by proof reading the drafts of manuscript.

6.2.1 Paper I

The first paper, *Analysis and Implementation of a Semi-Integrated Buck Converter with Static Feedback Control*, has a natural application in a handset as several voltage levels are required for optimum performance of the analog and digital part of the system. The power from the transmitter is controlled over a large range of different power levels, and the supply voltage to the power amplifier is adjusted to improve the power efficiency for lower output levels.

The most common feedback control is PID-regulation of the output voltage by adjusting the duty cycle of the signal to the switches of the converter. The PID-regulator is designed around the equilibrium, i.e., the steady-state solution, which leaves most of the state-space without taken into consideration. Since stability cannot be guaranteed more than in a small region around the steady-state operating point, several additional circuits, (e.g., up-ramp, duty-cycle limitation, overvoltage crowbar circuit), must be added to support the controller.

Instead a large-signal control approach is taken. The Lyapunov function represents the energy coupled to the state-space representation $\dot{x} = Ax + Bu$, and by assigning the Lyapunov function $V(x) = x^T Qx$, the time derivative of the energy function decays as fast as possible if $u = -\text{sign}(B^T Qx)$, where the input to the system is limited by $u \in [-1, 1]$. This motivates switching control action and the system dynamics get fast. The controller implemented in this paper acts very much in the same way, which explains its excellent dynamic performance. The switching action divides the state-space into two halves. In each half the system has specific dynamics connected to a certain switch action. The proposed control for the Buck converter is proven to be globally asymptotically stable using numerical tools in combination with successor and Lyapunov theory. This property is one of the most favorable features of a control system. The converter is robust and will work regardless of load condition or other disturbances.

The DC-DC converter is autonomous (meaning that no clock signal is needed). The steady-state ripple and the switching frequency are predicted by using describing function analysis. The frequency planning becomes more involved as the switching frequency depends on the load, the loop delay, and the hysteresis of the comparator. The switches and the control circuit of a Buck converter with static feedback control have been implemented in a standard digital 0.6 μm CMOS process. The design supports low voltage operation and a bootstrapping technique is used for the floating

switch of the converter. The simulated ripple of the implemented converter is in the order of $\pm 3\,\%$, and the efficiency of the complete converter is approximately $75\,\%$ at $1.5\,\mathrm{V}$ output on a $10\,\Omega$ resistive load. The inductor and capacitor are external components, but the rest of the converter is integrated and occupies $1.9\,\mathrm{mm}^2$. The start-up is very quick ($\simeq 10\,\mu\mathrm{s}$) and so is the recovery from any disturbance.

6.2.2 Paper II

The second paper, *Bandwidth Considerations for a CALLUM Transmitter Architecture*, presents the CALLUM architecture and a derivation of the control equations for the system, together with a study of the relation between the spectral widening and the modulation depth of the baseband signal. The spectum of the information signal, $s(t)$, is shown together with the spectrum of the constant-envelope signal $s_1(t)$, and $v_1(t)$, for $3\pi/8$-shifted 8PSK modulation used in EDGE, and $\pi/4$-shifted QPSK modulation, respectively. The bandwidth of the constant-envelope signal in the loop is apparently dependent on the modulation depth of the signal, and the control signal to the VCO, $v_1(t)$, was found to be spectrally wider than the constant-envelope signal, $s_1(t)$. A simple approximate formula is derived, linking the instantaneous frequency of the constant-envelope signals to the modulation depth of the information signal. Most importantly, the instantaneous frequency is (to first order approximation) proportional to the modulation depth of the information signal. Therefore, some modulation schemes, e.g., $3\pi/8$-shifted 8PSK, and $\pi/4$-shifted QPSK, are more suitable due to their limited modulation depth.

Further, the effect of the frequency mismatch between the VCOs and the channel frequency present in the CALLUM architecture is examined as the effective bandwidth of the signal is increased by the amount of frequency mismatch [82]. From a closed expression the error of the baseband signal was found to be proportional to the frequency mismatch. Since the loop gain is dropping at higher frequencies and the loop time delay degrades the phase-margin, the regrowth of the information signal should be effectively minimized for optimal performance. It is suggested that the resulting bandwidth expansion is removed with frequency synchronization by means of a phase-locked loop (PLL) operating in between transmitted bursts of data.

6.2.3 Paper III

It has not been possible to analyze a CALLUM transmitter as a linear feedback system, due to the nonlinear nature of the control equations governing it. The main purpose of the third paper, *Spectrum Emission Considerations for Baseband-Modeled CALLUM Architectures*, is the derivation of linearized equations for the nonlinear control equations, which enables the use of linear network theory in the study of CALLUM. In particular, it can be used to analyze the stability and maximize the bandwidth of the system. The paper only covers continuous versions of CALLUM suitable for implementation based on a Cartesian representation as they promise good signal matching

properties between the I and Q path in the feedback loop.

The dramatic reduction in simulation time is the major drive behind the realization of a baseband model for CALLUM. The proposed baseband model can be handled by time domain simulators, without any complex notation. Instead the phase between the I and Q path must be $\pi/2$ radians to bring about the complex implementation. To guarantee this relative phase shift between the VCOs, they are described as mathematical functions in AHDL.

CALLUM is governed by a set of nonlinear control equations, from which basic features like loop gain, bandwidth, and stability are difficult to obtain. The loop gain calculations are based on linearized equations valid around a static bias point. A reduced baseband model is used for loop gain calculations for the three CALLUM versions under test, that is, CALLUM 1, CALLUM 1lin, and CALLUM 2. An unifying approach in the loop gain calculations was possible since the three CALLUM versions differ only in the control equations. The generic structure of the linearized baseband CALLUM model is the same in all cases. The loop gain equations are given and graphically visualized in the IQ-plane as 3D contour plots, and the information available in these plots supports the investigation of system stability and the related design of frequency compensation.

Simulation examples are presented on how three different CALLUM architectures behave for EDGE and W-CDMA signals. In addition, the effects of loop time delay are also considered. The loop delay is the ultimate limiting factor in terms of bandwidth for all feedback-based linear transmitter architectures. The spectrum of the transmitted signal for each CALLUM version is compared to the EDGE spectrum emission mask. The magnitude of the loop gain was adjusted such CALLUM 2 barely passed the spectrum emission mask. The gap to CALLUM 1 and CALLUM 1lin was significant, and it can be concluded that dropping the denominator terms in the control equations for CALLUM 2 have severe effects on the spectral properties. The same simulation setting, apart from a loop gain scaling, was used also on a W-CDMA signal. It has been found from transient simulations that CALLUM 2 loses lock now and then, always in conjunction with small amplitudes of the input signal. This unfavorable behavior of CALLUM 2 makes it very unsuitable for operations on a W-CDMA signal. The performance for both CALLUM 1 and its linearized sibling is strikingly good, as it tracks the reference perfectly.

The effects of time loop delay on a CALLUM system are investigated by isolating them from other sources. A table is given with maximum acceptable loop delay to fulfill the spectrum emission mask for each CALLUM version and for the two standards already targeted – EDGE and W-CDMA. To improve the robustness against time delay the bandwidth is reduced while keeping the low-frequency loop gain unaffected by a lag-lead compensation network. It was shown that to succeed with the compensation strategy there must be room for deterioration of the spectral emissions; if these are already very close to the spectral mask, no significant improvements can be made. Thus, the reason why large improvements were achievable for CALLUM 1 and CALLUM 1lin (and to a lesser extent for CALLUM 2) operating on EDGE was

the relatively large distance between mask and actual spectral emissions for large offset frequencies; overly good spectral performances at large offset frequencies were in this case traded for higher insensitivity to time delay. Simulations showed a large performance difference in favor of CALLUM 1 and CALLUM 1lin, compared to CALLUM 2, whose strong point is the significantly reduced design complexity. The EDGE standard can be handled by all three implementations, but the larger signal bandwidth of W-CDMA prevents the use of CALLUM 2, when realistic time delays in the feedback loop are accounted for. The applied compensation showed to be truly efficient on each of the CALLUM versions working on an EDGE modulated signal.

6.2.4 Paper IV

The fourth paper, *Implementation of the Signal Component Generator of a CALLUM 2 Transmitter Architecture in CMOS Technology*, shows a fully integrated solution of the SCG in a 0.35 μm CMOS process. One of the most important advantages with integration is less loop time delay, which allows larger bandwidth signals and better spectral performance of the feedback system. The SCG generates the control voltages $v_1(t)$ and $v_2(t)$ for VCO_1 and VCO_2, respectively, from the input and the feedback signals.

The implementation of the control equation for CALLUM 2 is based on a fully differential analog structure. The use of differential signals have advantages; since no information is in the common-mode component the biasing is easier and fewer coupling capacitors are needed. Also by utilizing the complementary signal of the differential signal pair the need for both summation and subtraction are reduced to only one of them. The multiplication is achieved by a four-quadrant CMOS analog multiplier based on the folded six-transistor CMOS Gilbert cell. The transistor implementation of the SCG for CALLUM 2 has been simulated and compared to the ideal mathematical realization, and it works well, but a perfect match between the mathematical function and its implementation is hampered by the nonlinear gain of the MOS transistor. The total current consumption of the SCG is only 2.0 mA from a 3.3 V supply when implemented in a 0.35 μm CMOS process.

To maximize the system performance, control of the loop gain via a VGA with common-mode control is important. A VGA is presented and it is inserted in between the SCG and the VCO to adjust the loop gain. As described in Paper II the VCOs should be synchronized to the reference frequency to maximize the spectral performance of the system. The static frequency offset corresponds to a DC correction voltage that has to be applied to the VCO. Therefore the need for common-mode control of the VGA output. The presented SCG and VGA were simulated at transistor level in a baseband modeled CALLUM 2 architecture. The model contains a mixture of blocks at transistor level and AHDL. The modelling is further described in Paper III. The input signal is the $3\pi/8$-shifted 8PSK modulation used in EDGE, and from the simulated spectrum of the transmitted signal for CALLUM 2, both the SCG and VGA implementation proved to be suited for EDGE.

6.2.5 Paper V

The fifth paper, *Analytical Expression of the Efficiency of Phantom Zero Compensation Applied on Negative-Feedback Amplifiers*, reviews the phantom zero compensation technique, applied on negative-feedback (NFB) amplifiers. The paper also includes an analysis of the important efficiency parameter, δ, of the implemented phantom zero. Phantom zero compensation is easy and straightforward to implement and considered as the most favorable compensation technique available. The effect of the efficiency on the root locus is presented, and it has been found that $\delta \geq 7$ will give near ideal behavior of the applied phantom zero. For efficiency values as low as 5 acceptable compensation is achieved, but only with a phantom zero with infinite efficiency the estimated MFM bandwidth, ω_0, can be achieved.

For phantom zero efficiency we assume the zero to be well separated in frequency from the loop poles. Any capacitance giving rise to a loop pole can be modeled as a short circuit. The ordinary feedback factor, β, contains information about the loop poles and the efficiency, but the extraction of the efficiency tends to be messy if no approximations are made. Therefore we define β_{ph} as the current-to-current feedback factor, which only contains the information about the phantom zero. It is desirable to be able to calculate the efficiency of the phantom zero (with simple hand-calculations), since all possible implementations will not have the same efficiency and thereby not the same impact on the root locus. From an example of a two-stage BJT NFB amplifier the simplicity and practicable were presented. It can be concluded from the expression of the efficiency that the degree of success of the phantom zero compensation depends on the impedance of the load and the source, relative to their ideal values.

Appendix A

Chip Implementation

During the implementation of the circuits of the CALLUM 2 system on silicon, the strengths and the weaknesses of the CALLUM architecture became more apparent. Valuable information is gained from simulations of the circuits at transistor level, and for instance the baseband model can be refined to include parasitics that are difficult to foresee for the unexperienced designer. The CALLUM system differs from other types of linearization techniques, but has many similarities with PLLs and Cartesian modulation feedback. Therefore all knowledge of the CALLUM transmitter helps the designer to succeed with the project. In this appendix the outcome of the chip implementation is discussed.

A.1 The Chip

A chip containing the CALLUM 2 core has been fabricated in AMS 0.35 μm CMOS process. The circuits implemented are described in chapter 5. The design of the CALLUM 2 transmitter architecture is based on the schematic in Figure 65. The design is mainly differential to solve the bias problem and to increase the robustness against common-mode induced noise. During the floor planning of the chip, see Figure 93, the VCOs were given the highest priority. The VCOs were placed far away from each other to reduce the mutual coupling. Important for good matching, the VCO were duplicated in a copy and paste fashion (not mirrored in the symmetry axis). The RF downconversion was placed in the upper left corner with the quadrature signal generation slightly below. To reduce the coupling between the RF and LO, the wires should enter the chip orthogonally, which also would make the practical aspects of the printed circuit board (PCB) and the placement of the connectors easier.

The chip photo is shown in Figure 94. The area of the design is set by the pad frame and the rather large distance used between the VCOs. The total chip area (inluding the pad frame) is $2.1 \times 2.1\,\mu\mathrm{m}^2$. To a large extent the layout is dominated by the two VCOs, which consume substantial area due to on chip inductors. A significant number of pads are used to control bias voltages and currents, and to probe certian nodes in the design. The total current consumption for the chip is approximately 110 mA, where each CS-driver of the VCO output consumes 20 mA to generate 1 $\mathrm{V_{peak}}$ as the impedance level outside the chip is 50 Ω. A significant contribution to the current consumption comes from the VCOs with another 20 mA.

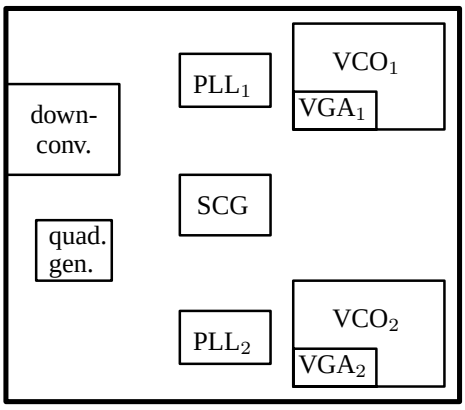

Figure 93: Floor planning of the chip.

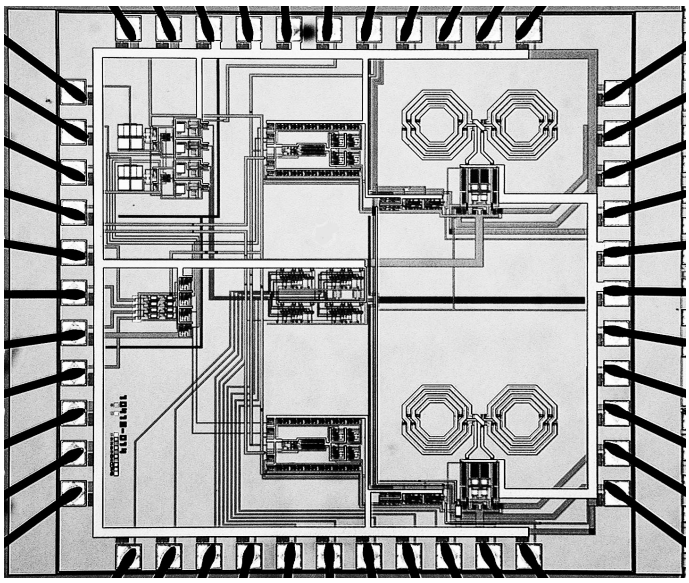

Figure 94: Chip photo of the CALLUM 2 system.

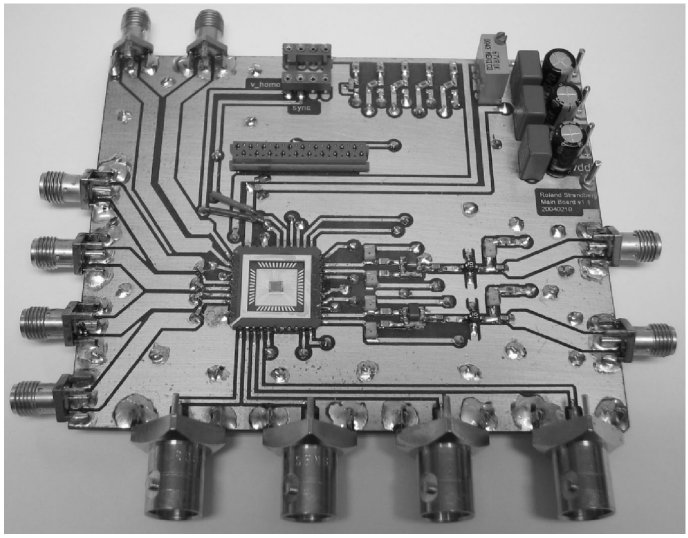

Figure 95: The main board.

A.2 Printed Circuit Board

The test chip was mounted on a two layer PCB measuring $10 \times 8\,\mathrm{cm}$. For large scale integration the external bias to the chip should be effectively minimized. For an intermediate step in the development of the system it was placed outside. The main reason for having all these currents adjustable is flexibility in terms of control of loop gain, voltage swing in the VCOs, drive capability of the output buffers, etc. The chip with 44 pins was mounted on the PCB according to Figure 95. The number of inputs to the chip is substantial as most inputs are differential. The differential I-Q baseband signals are applied to the BNC connectors, the differential quadrature LO signals are fed to the SMAs on the left hand side, whereas the RF feedback signals are connected to the SMAs at the top. The signal outputs from the chip (the SMAs to the right) are the constant-envelope vectors that are the inputs to the PA drivers.

A.2.1 Auxiliary Board

By having a module based test setup the main board can be swiftly exchanged for another. Therefore an auxiliary PCB was designed, see Figure 96, to contain the loop filter for the frequency synchronization accomplished by the PLL together with most bias signals for the SCG, VGAs and VCOs. Much soldering can be saved by having all bias on separate board(s), done once and then reused for any new test setup.

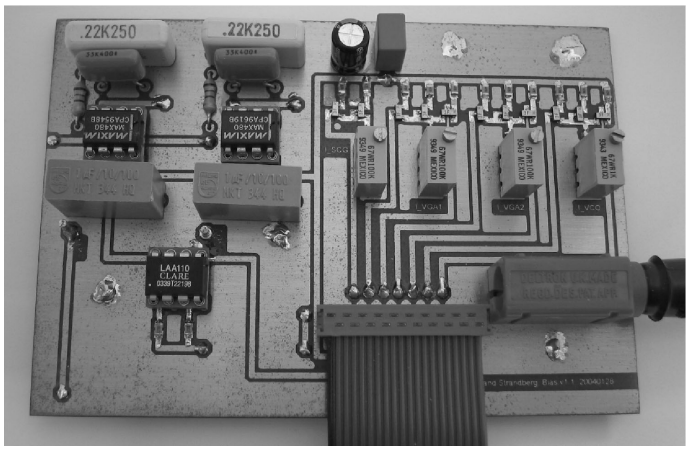

Figure 96: The auxiliary board.

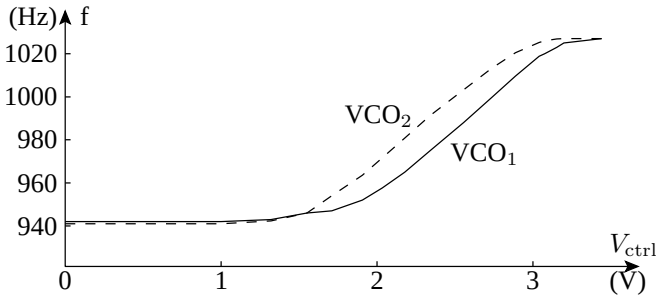

Figure 97: Tuning range at $I_T = 10\,\mathrm{mA}$.

A.3 Measuring the VCO Characteristic

The characteristics of the VCOs were measured for a tail-current, $I_T = 10\,\mathrm{mA}$, and the plot is shown in Figure 97. Both VCOs have almost the same minimum and maximum frequency, meaning that C_{ox} is similar for the two VCOs. It is likely that the MOS varactors have different threshold voltage, V_{th}. Even though the VCOs differ in their tuning characteristic they have the same sensitivity in the middle of the band, $K_{\mathrm{VCO},f} = 60\,\mathrm{MHz/V}$. The effect of the mutual coupling, discussed in section 5.4.2, must be low otherwise the measurement results would not differ that much. The effort with careful design of the inductors and the spacious layout paid off, and the static frequency error is handled separately by the PLL.

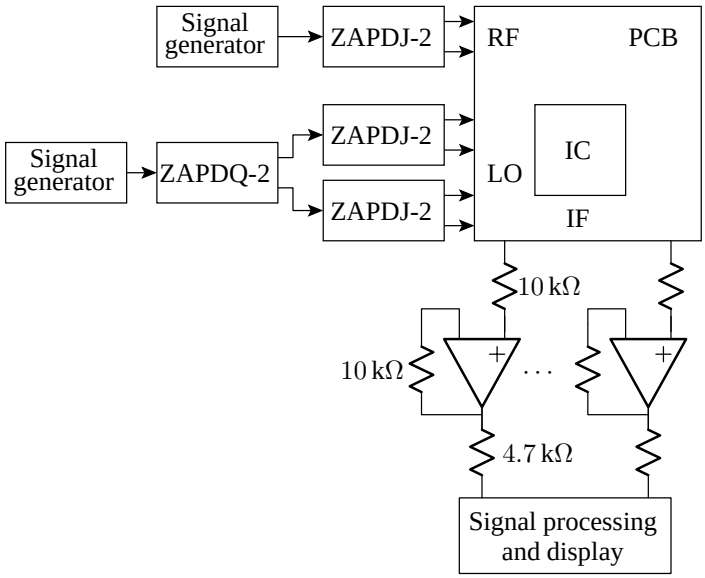

Figure 98: Block schematic of the downconversion.

A.4 Measuring the Downconverter

The downconversion part of the design needs a rather strong LO signal. The Rohde & Schwarz SMIQ 06 B signal generator has a maximum output amplitude of 1 V. Since the signal is divided into four phasors the signal amplitude can be only 0.5 V, if no loss is accounted for in the power slitters. The actual maximal measured amplitude at the input of the poly-phase filter is 0.37 V, due to some loss in cables and splitters. This value is on the low side accoring to Figure 91, and therefore the amplitude of the LO signal is boosted by an external amplifier (here a Mini-Circuits ZHL-2-8 power amplifier was used). The measurement setup is shown in Figure 98, where the downconverted baseband signals are buffered by voltage followers having high input impedance. Unfortunately the downconverter did not work, and the reason has not been identified. Measurements of the reflection coefficient, s_{11}, show about -10 dB for all RF ports, and the chip consumes the expected amount of current.

A.5 Additional Simulations Based on Measurement Observations

An encountered problem is the control voltage to the VCO, which is not tunable over the complete voltage range as shown in Figure 99. The control voltage to the VCO should be a stable voltage unaffected by the pulsating current due to charging and discharging of the varactor. For proper operation the drive signal to the VCO must

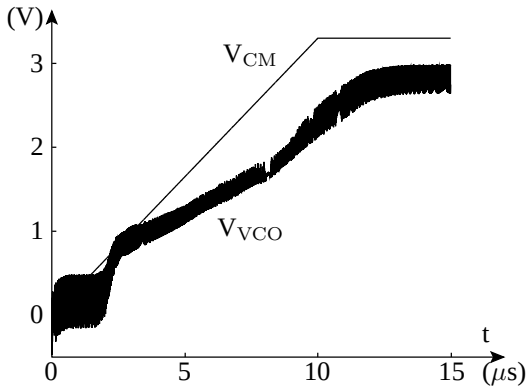

Figure 99: VGA control voltage.

be from a device with voltage output characteristic, whereas the output of the VGA is not an ideal voltage output. The unfortunate placement of the loop filter further degrades the output impedance. That is why the RF signal of the VCO makes the control signal look very noisy. To solve the problem the lag-lead compensation filter should be buffered before connected to the control port of the VCO.

To significantly reduce the need for current drive capability, the varactor should be made as small as possible. To keep the tuning range a coarse tuning can be made by a switched bank of capacitors and the varactor only used for fine tuning of the frequency.

References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell Systems Technical Journal*, vol. 27, pp. 379–432, 623–656, July, Oct. 1948.

[2] J. Davidse and E. H. Nordholt, “Basic considerations concerning the application of semicustom IC techniques for the processing of analogue electronic signals,” *Journal on Semi-Custom ICs*, vol. 5, no. 2, pp. 5–11, 1987.

[3] M. Lantz, *Systematic Design of Linear Feedback Amplifiers*. PhD thesis, University of Lund, Department of Electroscience, 2002.

[4] K. Kiasaleh and T. He, “On the performance of DQPSK communication systems impaired by timing error, mixer imbalance, and frequency nonselective slow rayleigh fading,” *IEEE Transactions on Vehicular Technology*, vol. 46, pp. 642–652, Aug. 1997.

[5] L. Sundström, *Digital RF Power Amplifier Linearisers*. PhD thesis, Lund University, Lund, Sweden, Aug. 1995.

[6] B. Shi and L. Sundström, “Investigation of a highly efficient LINC amplifier topology,” in *IEEE Vehicular Technology Conference*, vol. 2, pp. 1215–1219, Oct. 2001.

[7] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2003.

[8] E. Westesson and L. Sundström, “A complex polynomial predistorter chip in CMOS for baseband or if linearization of RF power amplifiers,” *IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 1, pp. 206–209, 1999.

[9] T. Rahkonen and T. Kankaala, “An analog predistortion integrated circuit for linearizing power amplifiers,” *Midwest Symposium on Circuits and Systems*, pp. 480–483, 9–12 Aug. 1998.

[10] Y. Nagata, “Linear amplification technique for digital mobile communications,” in *Proceedings of the 39th IEEE Vehicular Technology Conference*, pp. 159–164, May 1989.

[11] A. Mansell and A. Bateman, “Practical implementation issues for adaptive preditortion transmitter linearisation,” *IEE Colloquium on ’Linear RF Amplifers and Transmitters’ (Digest No:1994/089)*, pp. 5/1–5, Apr. 1994.

[12] S. P. Stapleton and J. K. Carvers, “A new technique for adaptation of linearization predistorters,” in *Proceedings of the 41st IEEE Vehicular Technology Conference*, pp. 753–758, May 1991.

[13] K.-J. Cho, D.-H. Jang, S.-H. Kim, J.-H. Kim, B. Lee, N.-Y. Kim, J.-C. Lee, and S. P. Stapleton, "Multi-order predistortion of power amplifiers using a second harmonic based technique," *IEEE Microwave and Wireless Components Letters*, vol. 13, pp. 452–454, Oct. 2003.

[14] H. S. Black, "Translating system," *filed 3 February 1925, U.S. Patent No. 1,686,792*, Oct. 1928.

[15] H. S. Black, "Stabilized feedback amplifiers," *Bell Systems Technical Journal*, vol. 13, pp. 1–18, Jan. 1934.

[16] P. B. Kenington, P. A. Warr, and R. J. Wilkinson, "Analysis of instability in feedforward loop," *Electronics Letters*, vol. 33, no. 20, pp. 1669–1671, 1997.

[17] R. J. Wilkinson, P. B. Kenington, and J. D. Marvill, "Power amplification techniques for linear TDMA base stations," in *Proceedings from IEEE Global Telecommunications Conference*, pp. 74–78, Dec. 1992.

[18] R. K. Manas, "Distortion cancellation performance of miniature delay filters for feed-forward linear power amplifiers," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 49, pp. 1592–1595, Nov. 2002.

[19] R. J. Wilkinson and P. B. Kenington, "Specification of error amplifiers for use in feedforward transmitters," *IEE proceedings part G*, vol. 139, pp. 477–480, Aug. 1992.

[20] R. D. Stewart and F. F. Tusubira, "Feedforward linearisation of 950 amplifiers," *IEE Proceedings*, vol. 135, pp. 347–349, Oct. 1988.

[21] K. J. Parsons and P. B. Kenington, "Feedforward linearisation of 950 amplifiers," *IEE Proceedings*, vol. 141, pp. 140–144, Apr. 1994.

[22] E. H. Nordholt, *Design of High-Performance Negative-Feedback Amplifiers*. Delftse Uitgevers Maatschappij, 1993.

[23] E. M. Cherry and D. E. Hooper, *Amplifying Devices and Low-Pass Amplifier Design*. New York: John Wiley and Sons, 1968.

[24] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. New York: Wiley, third ed., 1993.

[25] C. J. M. Verhoeven, A. van Staveren, G. L. E. Monna, M. H. L. Kouwenhoven, and E. Yildiz, *Structured Electronic Design: Negative-Feedback Amplifiers*. Kluwer Academic Publishers, new ed., 2003.

[26] M. Johansson and T. Mattsson, "Linearised high-efficiency power amplifier for PCN," *Electronics Letters*, vol. 27, pp. 762–764, Apr. 1991.

[27] M. Faulkner and M. A. Briffa, "Amplifier linearisation using RF feedback and feedforward techniques," in *Proceedings of the 45th IEEE Vehicular Technology Conference*, pp. 525–529, July 1995.

[28] V. Petrovic, "Reduction of spurious emission from radio transmitters by means of modulation feedback," in *IEE Conference on Radio Spectrum Conservation Techniques*, pp. 44–49, Sep. 1983.

[29] V. Petrovic and C. N. Smith, "Reduction of intermodulation distortion by means of modulation feedback," in *IEE Collocquium on intermodulation-causes, effects and mitigation, London*, pp. 8/1–8/8, Apr. 1984.

[30] V. Petrovic and A. N. Brown, "Application of Cartesian feedback to HF SSB transmitters," in *Proc. from IEE Conference on HF communication systems and techniques*, pp. 81–85, Feb. 1985.

[31] M. Johansson, "Linearization of RF power amplifiers using Cartesian feedback," Teknisk Litentiat Thesis LUTEDX/(TETE-7051)/1-107(1991), Lund University, Department of Applied Electronics, Nov. 1991.

[32] M. Johansson and L. Sundström, "Linearisation of RF multicarrier amplifiers using Cartesian feedback," *Electronics Letters*, vol. 30, pp. 1110–1112, July 1994.

[33] M. Boloorian and J. P. McGeehan, "The frequency-hopped cartesian feedback linear transmitter," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 688–706, Nov. 1996.

[34] N. Sornin, M. Massei, L. Perraud, and C. Pinatel, "A robust cartesian feedback loop for a 802.11 a/b/g CMOS transmitter," in *IEEE Radio Frequency Integrated Circuits Symposium*, pp. 145–148, June 2004.

[35] F. Carrara, A. Scuderi, and G. Palmisano, "Wide-bandwidth fully integrated cartesian feedback transmitter," in *IEEE Custom Integrated Circuits Conference*, vol. 1, pp. 451–454, Sep. 2003.

[36] M. Faulkner, D. Contos, and M. A. Briffa, "Performance of automatic phase adjustment using supply current minimisation in a RF feedback lineariser," in *Proceedings of the eighth IEEE International Symposium on Personal, Indoor and Mobile Radio Comunications*, vol. 3, pp. 858–862, Sep. 1997.

[37] M. A. Briffa and M. Faulkner, "Stability analysis of cartesian feedback linearisation for amplifiers with small non-linearities," *IEE Proceedings - Communications*, vol. 143, no. 4, pp. 212–218, 1996.

[38] Y. Ohishi, M. Minowa, E. Fukuda, and T. Takano, "Cartesian feedback amplifier with soft landing," in *Proceedings of the third IEEE International Symposium on Personal, Indoor and Mobile Radio Comunications*, pp. 402–406, Oct. 1992.

[39] M. Boloorian and J. P. McGeehan, "Linearisation of frequency-hopped transmitters using cartesian feedback," in *Proceedings of the 45th IEEE Vehicular Technology Conference*, pp. 520–524, July 1995.

[40] J. L. Dawson and T. H. Lee, "Automatic phase alignment for a fully integrated cartesian feedback power amplifier system," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 2269–2279, Dec. 2003.

[41] M. Boloorian and J. P. McGeehan, "Phase-lag compensated cartesian feedback transmitter," *Electronics Letters*, vol. 32, pp. 1547–1548, Aug. 1996.

[42] M. A. Briffa and M. Faulkner, "Dynamically biased cartesian feedback linearization," in *Proceedings of the 43rd IEEE Vehicular Technology Conference*, pp. 672–675, May 1993.

[43] M. Johansson, T. Mattsson, L. Sundström, and M. Faulkner, "Linearization of multi-carrier power amplifiers," in *Proceedings of the 43rd IEEE Vehicular Technology Conference*, pp. 684–687, May 1993.

[44] V. Petrovic and W. Gosling, "Polar-loop transmitter," *Electronics Letters*, vol. 15, pp. 286–288, May 1979.

[45] E. McCune and W. Sander, "EDGE transmitter alternative using nonlinear polar modulation," in *Proceedings of the 2003 International Symposium on Circuits and Systems, ISCAS*, vol. 3, pp. III–594–III–597, May 2003.

[46] L. R. Kahn, "Single-sideband transmission by envelope elimination and restoration," *Proceedings of the IRE*, vol. 40, pp. 803–806, July 1952.

[47] L. R. Kahn, "Comparison of linear single-sideband transmitters with envelope elimination and restoration single-sideband transmitters," *Proceedings of the IRE*, pp. 1706–1712, Dec. 1956.

[48] D. K. Su and W. J. McFarland, "An IC for linearizing RF power amplifiers using envelope elimination and restoration," *IEEE Journal of Solid-State Circuits*, vol. 13, pp. 2252–2258, Dec. 1998.

[49] M. D. Weiss, F. H. Raab, and Z. B. Popović, "Linearity of X-band class-F power amplifiers in high-efficiency transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, pp. 1174–1179, June 2001.

[50] D. Rudolph, "Out-of-band emissions of digital transmissions using Kahn EER technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 1979–1983, Aug. 2002.

[51] F. H. Raab, B. E. Sigmon, R. G. Myers, and R. M. Jackson, "L-band transmitter using Kahn EER technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, pp. 2220–2225, Dec. 1998.

[52] F. H. Raab, "Drive modulation in Kahn-technique transmitters," *IEEE MTT-S International Microwave Symposium Digest*, pp. 811–814, 1999.

[53] D. Rudolph, "Kahn EER technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 548–552, Feb. 2003.

[54] F. H. Raab and D. J. Rupp, "High-efficiency single-sideband HF/VHF transmitter based upon envelope elimination and restoration," in *International Conference on HF Radio Systems and Techniques*, pp. 21–25, July 1994.

[55] F. H. Raab, "Intermodulation distortion in Kahn-technique transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, pp. 2273–2278, Dec. 1996.

[56] G. Baudoin, C. Berland, M. Villegas, and A. Diet, "Influence of time and processing mismatches between phase and envelope signals in linearization systems using envelope elimination and restoration, application to Hiperlan2," *IEEE MTT-S International Microwave Symposium Digest*, vol. 3, pp. 2149–2152, 2003.

[57] D. C. Cox, "Linear amplification with nonlinear components," *IEEE Transactions on Communications*, vol. 22, pp. 1942–1945, Dec. 1974.

[58] H. Chireix, "High power outphasing modulation," *Proceedings of the IRE*, vol. 23, pp. 1370–1392, Nov. 1935.

[59] L. F. Gaudernack, "A phase-opposition system of amplitude modulation," *Proceedings of the IRE*, vol. 26, pp. 983–1008, Aug. 1938.

[60] B. Shi and L. Sundström, "A LINC transmitter using a new signal component separator architecture," in *IEEE Vehicular Technology Conference*, vol. 3, pp. 1909–1913, May 2000.

[61] B. Shi and L. Sundström, "A translinear-based chip for linear LINC transmitters," *Symposium on VLSI Circuits. Digest of Technical Papers*, pp. 58–61, June 2000.

[62] B. Shi and L. Sundström, "A 200-MHz IF BiCMOS signal component separator for linear LINC transmitters," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 987–993, July 2000.

[63] B. Shi and L. Sundström, "A novel design using translinear circuit for linear LINC transmitters," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 64–67, May 2000.

[64] B. Shi and L. Sundström, "An IF CMOS signal component separator chip for LINC transmitters," in *IEEE Custom Integrated Circuits Conference*, pp. 49–52, May 2001.

[65] A. J. Rustako and Y. S. Yeh, "A wide-band phase-feedback inverse-sine phase modulator with application toward a LINC amplifier," *IEEE Transactions on Communications*, vol. 24, pp. 1139–1143, Oct. 1976.

[66] S. A. Hetzel, A. Bateman, and J. P. McGeehan, "LINC transmitter," *Electronics Letters*, vol. 27, pp. 844–845, May 1991.

[67] L. Sundström, "Effects of reconstruction filters and sampling rate for a digital signal component separator on LINC transmitter performance," *Electronics Letters*, vol. 31, pp. 1124–1125, July 1995.

[68] L. Sundström, "Spectral sensitivity of LINC transmitters to quadrature modulator misalignment," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 1474–1487, July 2000.

[69] X. Zhang, L. E. Larson, P. M. Asbeck, and P. Nanawa, "Gain/phase imbalance-minimization techniques for LINC transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, pp. 2507–2516, Dec. 2001.

[70] F. Casadevall and J. J. Olmos, "On the behavior of the LINC transmitter," in *Proceedings of the 40th IEEE Vehicular Technology Conference*, pp. 29–34, May 1990.

[71] F. J. Casadevall and A. Valdovinos, "Performance analysis of QAM modulations applied to the LINC transmitter," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 399–406, Nov. 1993.

[72] L. Sundström, "Automatic adjustment of gain and phase imbalances in LINC transmitters," *Electronics Letters*, vol. 31, pp. 155–156, Feb. 1995.

[73] S. Ampem-Darko and H. S. Al-Raweshidy, "Gain/phase imbalance cancellation technique in LINC transmitters," *Electronics Letters*, vol. 34, pp. 2093–2094, Oct. 1998.

[74] X. Zhang and L. E. Larson, "Gain and phase error-free LINC transmitter," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 1986–1994, Sep. 2000.

[75] B. Shi and L. Sundström, "A time-continuous optimization method for automatic adjustment of gain and phase imbalances in feedforward and LINC transmitters," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 1, pp. I–45–I–48, May 2003.

[76] A. Bateman, "The combined analogue locked loop universal modulator (CALLUM)," in *Proceedings of the 42nd IEEE Vehicular Technology Conference*, pp. 759–763, May 1992.

[77] K. Y. Chan and A. Bateman, "Linear modulators based on RF synthesis: Realization and analysis," *IEEE Transactions on Circuits and Systems*, vol. 42, pp. 321–333, June 1995.

[78] D. J. Jennings, A. Bateman, and J. P. McGeehan, "Adjacent channel power and error-vector magnitude performance of reduced complexity CALLUM systems," *IEE Proceedings on Communications*, vol. 146, pp. 297–302, Oct. 1999.

[79] K. Y. Chan and A. Bateman, "Analytical and measured performance of the combined analogue locked loop universal modulator (CALLUM)," in *IEE Proceedings on Communications*, pp. 297–306, Oct. 1995.

[80] D. J. Jennings and J. P. McGeehan, "Hardware implementation of optimal CALLUM transmitter," *Electronics Letters*, vol. 34, pp. 1816–1817, Sep. 1998.

[81] D. J. Jennings and J. P. McGeehan, "A high-efficiency RF transmitter using VCO-derived synthesis: CALLUM," in *Proceedings of the Radio and Wireless Conference*, pp. 137–140, 1998.

[82] D. J. Jennings and J. P. McGeehan, "A high-efficiency RF transmitter using VCO-derived synthesis: CALLUM," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, pp. 715–721, June 1999.

[83] D. C. Cox and R. P. Leck, "A VHF implementation of a LINC amplifier," *IEEE Transactions on Communications*, vol. 24, pp. 1018–1022, Sep. 1976.

[84] L. Sundström, "The effect of quantization in a digital signal component separator for LINC transmitters," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 346–352, May 1996.

[85] S. A. Hetzel, A. Bateman, and J. P. McGeehan, "A LINC transmitter," in *Proceedings of the 41st IEEE Vehicular Technology Conference*, pp. 133–137, May 1991.

[86] R. Strandberg, P. Andreani, and L. Sundström, "Bandwidth considerations for a CALLUM transmitter architecture," *IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 4, pp. 25–28, 26–29 May 2002.

[87] C. P. Conradi, R. H. Johnston, and J. G. McRory, "Evaluation of a lossless combiner in a LINC transmitter," in *IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 1, pp. 105–110, May 1999.

[88] L. Sundström and M. Johansson, "Effect of modulation scheme on LINC transmitter power efficiency," *Electronics Letters*, vol. 30, pp. 1643–1645, Sep. 1994.

[89] R. A. Langridge, T. Thornton, P. M. Asbeck, and L. E. Larson, "A power re-use technique for improved efficiency of outphasing microwave power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, pp. 1467–1470, Aug. 1999.

[90] G. Gonzalez, *Microwave Transistor Amplifiers*. Englewood Cliffs: Prentice Hall, 1984.

[91] F. H. Raab, "Efficiency of outphasing RF power-amplifier systems," *IEEE Transactions on Communications*, vol. 33, pp. 1094–1099, Oct. 1985.

[92] B. Stengel and W. R. Eisenstadt, "LINC power amplifier combiner method efficiency optimization," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 229–234, Jan. 2000.

[93] P. Dent, "Hybrid chireix/doherty amplifiers and methods," *U.S. Patent Application No. 6,133,788*.

[94] R. Strandberg and J. Yuan, "Analysis and implementation of a semi-integrated buck converter with static feedback control," *Midwest Symposium on Circuits and Systems*, vol. 2, pp. 934–937, 8–11 Aug. 2000.

[95] A. Bateman, M. Li, and K. Chan, "Implementation of the LINC transmitter using the combined analogue locked loop universal modulator (CALLUM)," in *Proceedings of the 7th IEE European Conference on Mobile and Personal Communications*, pp. 31–37, Dec. 1993.

[96] K. Y. Chan, A. Bateman, and M. Li, "Analysis and realisation of the LINC transmitter using the combined analogue locked loop universal modulator (CALLUM)," in *Proceedings of the 44th IEEE Vehicular Technology Conference*, pp. 484–488, June 1994.

[97] D. Senderowicz, D. A. Hodges, and P. R. Gray, "An nmos integrated vector-locked loop," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 1164–1167, 1982.

[98] A. Bateman and K. Y. Chan, "Methods and apparatus for modulating, demodulating and amplifying," *U.S. Patent No. 5,939,951*, Aug. 17 1999.

[99] R. Kaunisto, "A vector-locked loop for power amplifier linearization," *IEEE MTT-S International Microwave Symposium Digest*, pp. 673–676, 2004.

[100] W. C. Lindsey and C. M. Chie, *Phase-Locked Loops*. New York: IEEE Press, 1986.

[101] C. J. M. Verhoeven, A. van Staveren, G. L. E. Monna, M. H. L. Kouwenhoven, and E. Yildiz, *Structured Electronic Design*. Delft, july 2002.

[102] R. Strandberg and J. Piper, "Analytical expression of the efficiency of phantom zero compensation applied on negative-feedback amplifiers," *Norchip Conference*, pp. 87–90, 8–9 Nov. 2004.

[103] K. J. Åström, *Reglerteori*. Stockholm: Almqvist & Wiksell Förlag AB, second ed., 1976.

[104] L. Weinberg, *Network Analysis and Synthesis*. New York: McGraw-Hill, first ed., 1962.

[105] H. Meyr, *Regelungstechnik und Systemtheorie*. Druckerei Mainz GmbH, Aachen, 1998.

[106] D. A. Johns and K. Martin, *Analog Integrated Circuit Design*. Wiley, 1997.

[107] R. E. Best, *Phase-Lock Loops*. McGraw-Hill, second ed., 1993.

[108] B.-A. Molin, *Analog konstruktion i CMOS*. Lund University, 1997.

[109] J. N. Babanezhad and G. C. Temes, "A 20-v four-quadrant CMOS analog multiplier," *IEEE Journal of Solid-State Circuits*, vol. 20, pp. 1158–1168, Dec. 1985.

[110] D. Coffing, E. Main, M. Randol, and G. Szklarz, "A variable gain amplifier with 50-dB control range for 900-MHz applications," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 1169–1175, Sep. 2002.

[111] P. Orsatti, F. Piazza, and Q. Huang, "A 20-mA-receive, 55-mA-transmit, single-chip GSM transceiver in 25-μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 1869–1880, Dec. 1999.

[112] A. Deiss, D. Pfaff, and Q. Huang, "A 200-MHz sub-mA RF front end for wireless hearing aid applications," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 977–986, July 2000.

[113] W. C. Song, C. J. Oh, G. H. Cho, and H. B. Jung, "High frequency/high dynamic range CMOS VGA," *Electronics Letters*, vol. 36, pp. 1096–1098, June 2000.

[114] C. W. Mangelsdorf, "A variable gain CMOS amplifier with exponential gain control," *Symposium on VLSI Circuits. Digest of Technical Papers*, pp. 146–149, 2000.

[115] R. G. Meyer and W. D. Mack, "A DC to 1-GHz differential monolithic variable-gain amplifier," *IEEE Journal of Solid-State Circuits*, vol. 26, pp. 1673–1680, Nov. 1991.

[116] G. S. Sahota and C. J. Persico, "High dynamic range variable-gain amplifier for CDMA wireless applications," *Solid-State Circuits Conference, Digest of Technical Papers*, pp. 374–375, 488, 1997.

[117] Tzu-Wang Pan and Asad A. Abidi, "A 50-dB variable gain amplifier using parasitic bipolar transistors in CMOS," *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 951–961, Aug. 1989.

[118] R. G. Meyer and W. D. Mack, "A wideband low-noise variable-gain BiC-MOS transimpedance amplifier," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 701–706, June 1994.

[119] Z. Wang and W. Guggenbuhl, "A voltage-controllable linear MOS transconductor using bias offset technique," *IEEE Journal of Solid-State Circuits*, vol. 25, pp. 315–317, Feb. 1990.

[120] R. Harjani, "A low-power CMOS VGA for 50 Mb/s disk drive read channels," *IEEE Transactions on Circuits and Systems–II: Analog and Digital Signal Processing*, vol. 42, pp. 370–376, June 1995.

[121] T. Yamaji, N. Kanou, and T. Itakura, "A temperature-stable CMOS variable-gain amplifier with 80-dB linearly controlled gain range," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 553–558, May 2002.

[122] W. C. Song, C. J. Oh, G. H. Cho, and H. B. Jung, "A 200MHz/90dB gain range CMOS VGA," *Proceedings of the Second IEEE Asia Pacific Conference on ASICs, 2000*, pp. 1–4, 2000.

[123] A. A. Abidi, "Gigahertz transresistance amplifiers in fine line NMOS," *IEEE Journal of Solid-State Circuits*, vol. 19, pp. 986–994, Dec. 1984.

[124] A. Hajimiri and T. Lee, "Design issues in CMOS differential LC oscillators," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 717–724, May 1999.

[125] A. Hajimiri and T. Lee, "A general theory of phase noise in electrical oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 179–194, Feb. 1998.

[126] E. Hegazi, H. Sjöland, and A. A. Abidi, "A filtering technique to lower lc oscillator phase noise," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 1921–1930, Dec. 2001.

[127] P. Andreani, "Phase noise analysis of the LC-tank CMOS oscillator," *Norchip Conference*, pp. 147–150, 8–9 Nov. 2004.

[128] S. S. Mohan, M. del Mar Hershenson, S. P. Boyd, and T. H. Lee, "Simple accurate expressions for planar spiral inductances," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 1419–1424, Oct. 1999.

[129] P. Andreani and S. Mattisson, "On the use of MOS varactors in RF VCOs," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 905–910, June 2000.

[130] P. Andreani and S. Mattisson, "A 1.8-GHz CMOS VCO tuned by an accumulation-mode MOS varactor," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. I–315–I–318, 2000.

[131] M. H. L. Kouwenhoven, *High-Performance Frequency-Demodulation Systems*. PhD thesis, Delft University of Technology, Delft, The Netherlands, Mar. 1998.

[132] F. M. Gardner, *Phaselock Techniques*. New York: Wiley, second ed., 1979.

[133] U. L. Rohde, *Digital PLL Frequency Synthesizers*. Englewood Cliffs: Prentice Hall, new ed., 1983.

[134] W. F. Egan, *Phase-Lock Basics*. New York: John Wiley and Sons, new ed., 1998.

[135] W. F. Egan, *Frequency Synthesis by Phase Lock*. New York: John Wiley and Sons, second ed., 2000.

[136] A. Blanchard, *Phase-Locked Loops*. New York: John Wiley and Sons, new ed., 1976.

[137] F. M. Gardner, "Charge-pump phase-lock loops," *IEEE Transactions on Communications*, vol. COM-28, pp. 1849–1858, Nov. 1980.

[138] S. Mirabbasi and K. Martin, "Design of loop filter in phase-locked loops," *Electronics Letters*, vol. 35, pp. 1801–1802, Oct. 1999.

[139] R. Y. Chen, "High-speed CMOS frequency divider," *Electronics Letters*, vol. 33, pp. 1864–1865, Oct. 1997.

[140] H. O. Johansson, "A simple precharged CMOS phase frequency detector," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 295–299, Feb. 1998.

[141] B. Razavi, "Challenges in portable RF transceiver design," *IEEE Circuits and Devices Magazine*, vol. 12, pp. 12–25, Sep. 1996.

[142] F. Tillman and H. Sjöland, "A bootstrapping technique to improve the linearity of CMOS passive mixers," *Symposium on VLSI Circuits. Digest of Technical Papers*, pp. 221–222, June 2003.

[143] M. J. Gingell, "Single sideband modulation using sequence asymmetric polyphase networks," *Electrical Communication*, vol. 48, no. 1 and 2, pp. 21–25, 1973.

[144] F. Behbahani, Y. Kishigami, J. Leete, and A. A. Abidi, "CMOS mixers and polyphase filters for large image rejection," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 873–887, June 2001.

[145] P. van de Ven, J. van der Tang, D. Kasperkovitz, and A. van Roermund, "An optimally coupled 5 GHz quadrature LC oscillator," *Symposium on VLSI Circuits. Digest of Technical Papers*, pp. 115–118, 2001.

[146] P. Andreani, A. Bonfanti, L. Romanó, and C. Samori, "Analysis and design of a 1.8-GHz CMOS LC quadrature VCO," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 1737–1747, Dec. 2002.

Paper I

Paper I

Analysis and Implementation of a Semi-Integrated Buck Converter with Static Feedback Control

Abstract

The switches and the control circuit of a Buck converter with static feedback control have been implemented in a standard digital 0.6 μm. CMOS process. The control used is a direct large-signal approach, and the converter has been proven to be globally asymptotically stable using the theory of successors in combination with Lyapunov's direct method. One very important design parameter is the output ripple, and prediction has shown good agreement between the describing function method and simulations. The simulated ripple of the implemented converter is in the order of $\pm 3\,\%$, and the efficiency of the complete converter is approximately 75 % at 1.5 V output on a 10 Ω resistive load. The inductor and capacitor are external components, but the rest of the converter is integrated and occupies 1.9 mm^2. The start-up is very quick ($\simeq$10 μs) and so is the recovery from any disturbance.

Based on: Roland Strandberg and Jiren Yuan, "Analysis and Implementation of a Semi-Integrated Buck Converter with Static Feedback Control," *Midwest Symposium on Circuits and Systems*, vol. 2, pp. 934–937, 2000.

1 Introduction

Frequently in mobile applications, e.g., mobile phones, portable computers, there is need for different voltage levels internally. The necessity for several voltages arises when to maximize performance of a system. The voltage to the analog part is advantageously chosen high to achieve good signal to noise ratio (SNR), but on the other hand problems with hot electrons (oxide breakdown caused by too high electrical fields in the semiconductor) might arise. As far as digital is concerned low voltage is worth to strive for, since the power consumption decreases quadratically with decreasing voltage supply. Although the speed of the logic also decreases is not a problem in numerous applications.

A straightforward solution to this problem is to use several batteries. However, it is definitely more convenient to have only one single battery supplying the whole circuit, in terms of complexity, cost, and size. Single battery supply obviously involves voltage conversion to supply each block with an appropriate voltage.

It is fair to assume that the digital part of the system is more robust and noise insensitive, than the analog part. Therefore it is best to down-convert the voltage, since the ripple appearing at the output of the converter is substantially stronger than at the battery supply line. Here the battery is assumed to act as a voltage source with rather low output resistance. In very uncritical applications a pure resistive divider can be utilized for down conversion, which in fact is the most common way to bias a series feedback transistor stage. The resistive divider suffers from two fatal drawbacks. First, the efficiency can not be higher than $V_{\mathrm{out}}/V_{\mathrm{supply}}$, and second, it is not controllable. Switched converters do not suffer from these drawbacks, as the efficiency can be 100 % (theoretically) and sophisticated control action can be taken.

In the following section the standard Buck converter (step-down converter) topology is introduced. In section 3 the mathematics are formulated and a large-signal control approach is taken. The start-up dynamics is displayed in a phase-plane portrait and global asymptotic stability is shown followed by steady-state ripple and switching frequency prediction. In section 4 some implementation aspects are put forward, such as complete schematic and remedy for low voltage operation followed by the conclusions in section 5.

2 Buck Converter Architecture

There are a few simple schemes of converters that work very well. That is why other schemes solving the same task are not inquired. However, a Buck converter transforms one voltage to another lower voltage, ideally with no energy loss. It consists of two switches, two reactive components, and a controller forming a closed-loop system, see Fig. 1. The two reactive components, an inductor and a capacitor, work for intermediate storage of energy. The transfer function is low-pass, and performs all the necessary filtering of the input signal.

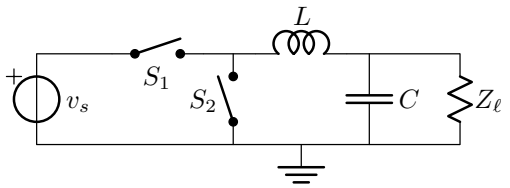

Figure 1: A Buck converter, where the controller is omitted.

In order to get a low cut-off frequency, and thereby good filtering, both the inductance and capacitance must be large, which contradicts both the size and cost issues. A common solution to overcome bulky filtering components is to increase the switching frequency. Unfortunately will also the switching losses increase.

3 Buck Converter with Static Feedback

The most common feedback control is PID-regulation (proportional, integrating, differentiating) to control the output voltage by adjusting the duty cycle of the converter. The averaging method for controller design is frequently used, see [1–3] and references therein. This is a simple and useful method, but it has some very distinct drawbacks like that only the bulk behavior is represented (ripple excluded), and the control is designed around the equilibrium (steady-state solution), which leaves most of the state-space without taken into consideration. Since stability cannot be guaranteed more than in a small region around the steady-state operating point, several additional circuits, (e.g., up-ramp, duty-cycle limitation, etc.), must be added to support the controller.

3.1 Motivation for Switch Action

Consider a Lyapunov function $V(x) = x^T Q x$, where the matrix Q is positive definite. The Lyapunov function represents the energy coupled to the state-space representation $\dot{x} = Ax + Bu$. Assume that the input to the system is limited, which is the normal case, $u \in [-1, 1]$. The time derivative of the energy function is given by, $\dot{V}(x(t)) = x^T(A^T Q + AQ)x + 2B^T Q x u$, and the energy decays as fast as possible if $u = -\text{sign}(B^T Q x)$. This motivates switching action and the system dynamics get fast. The controller implemented in this work acts very much in the same way, which explains its excellent dynamic performance.

The state equation for the on state of the circuit in Fig. 1 with resistive load, R, can be formulated as follows,

$$\dot{x} = Ax + Bu \tag{1}$$

$$A = \begin{pmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{pmatrix} \tag{2}$$

$$B = \begin{pmatrix} \frac{1}{L} \\ 0 \end{pmatrix}. \tag{3}$$

The control signal u determines the switch action and is chosen as,

$$u = \frac{1 - \text{sign}(x_2 - V_r)}{2} V_s, \tag{4}$$

where V_r is the output reference voltage and V_s the supply voltage to the circuit. $x_1(t)$, $x_2(t)$ are the inductor current and capacitor voltage, respectively. This control ($u(t)$) is a direct large-signal control approach and the DC-DC converter is self-oscillating. Explicitly, this means that there is no need for any additional clock signal on-chip.

3.2 Buck Converter Phase-Plane Portrait

The trajectory of a second order system is suitable for phase-plane portrait representation. A trajectory is a combination of the states $x(t)$ forming a path in time plotted in the state space. In the phase-plane portrait a boundary is introduced to separate the system into two halves. In each half the system has specific dynamics connected to certain switch action. This kind of systems are often called hybrid systems, which are systems that involve interaction between discrete and continuous dynamics. Simulation of the converter during start-up is shown in Fig. 2. Before studying stability of the converter the behavior around the switching surface must be determined in order to use the correct tools. The applied control does not make it to a sliding mode control, which can be seen in Fig. 2. This can also be shown in a more formal way by the following relations,

$$\dot{x} = \begin{cases} f^+(x) & \sigma(x) > 0 \\ ? & \sigma(x) = 0 \\ f^-(x) & \sigma(x) < 0 \end{cases} \tag{5}$$

where $\sigma(x) = x_2 - V_r$ and defines the boundary. Here, $\{X_i\}_{i \in I} \subseteq \mathbf{R}^n$ is a partition of the state space into a number of closed (possibly) unbounded polyhedral cells, and I is the index set of the cells. For $x(t) \in X_2$ where $\sigma(x) > 0$ the control is $u = 0$. $x(t) \in X_1$ implies $u = V_s$. The sliding set is given by $x(t)$ such that these relations hold:

$$\begin{array}{lcl} \sigma(x) & = & 0 \\ (\nabla\sigma(x))^T f^+ & < & 0 \\ (\nabla\sigma(x))^T f^- & > & 0 \end{array} \tag{6}$$

The constraints contradict, saying $x_1 < V_r/R$ and at the same time $x_1 > V_r/R$. Therefore no sliding mode operation appears in the converter with this control law.

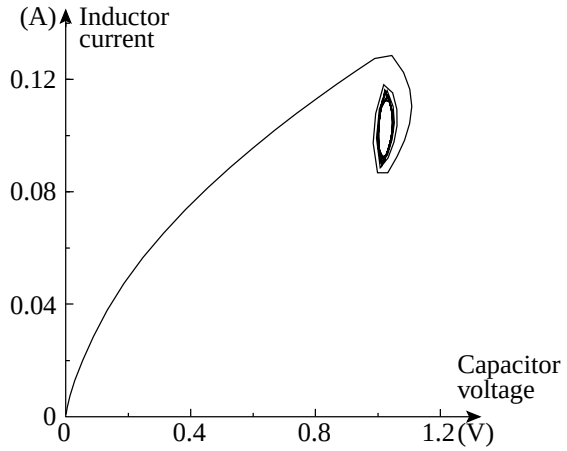

Figure 2: The converter very rapidly approaches the switching surface, and then moves to a stable limit cycle. The maximum width of the limit cycle in the horizontal direction is equivalent to the steady state ripple.

3.3 Stability of the Buck Converter

The construction of a Lyapunov function is one of the most important problem in system theory. Lyapunov functions are most often used to prove stability. Among the wide range of applications of the Lyapunov function controller synthesis can be mentioned. Hybrid systems are systems that involve interaction between discrete and continuous dynamics, which describes switched power converters well. In some cases it is possible to find a quadratic or elliptic Lyapunov function to a hybrid system.

To prove stability the theory of successors in combination with Lyapunov's direct method is used. If the switching frequency $f_s \rightarrow \infty$ the ripple tends to zero, which implies that the equilibrium is given by $x^0 = (V_r/R, V_r)^T$. The active switch must be controlled in such a manner that in each partition of the state-space the trajectory has the equilibrium point in the opposite half plane [4]. Let the n:th intersection with the switching boundary be denoted $x_{1,n}$, (Note that $x_2 = V_r$ on the switching surface for $n = 1, 2, \dots$). Assume the Lyapunov function candidate $V(x) = (x_{1,n} - x_1^0)^2$. The Lyapunov function is positive semi-definite and radially unbounded. If $(x_{1,n+2} - x_1^0)^2 \leq (x_{1,n} - x_1^0)^2$ then the system is globally stable and will approach the equilibrium or a limit cycle. A Buck converter with the proposed static feedback control has this property independent of the choice of component values.

Global stability is one of the most favorable parameters of a control system. This fits well in the trend in designing robust systems that work regardless of load condition, disturbances, etc.

3.4 Prediction of Ripple and Switching Frequency

Now we know that this converter is globally stable, and it tends to a stable limit cycle. In the design flow, the ripple frequently appears as part of the specification. Therefore, it is interesting to predict, preferably analytically, the amplitude of the ripple as well as its frequency. There is a powerful tool for this, i.e., describing function analysis, (discussed in detail in [5]). Describing functions are easily applicable when the feedback system can be transformed into a system according to Fig. 3. The system consists of a linear part, $G(s)$, and a nonlinear part, $f(\cdot)$, connected in a closed loop. The nonlinearity can be exchanged for an amplitude and frequency dependent gain, $N(A,\omega)$. $f(\cdot)$ is thereby transformed to a quasi-linear block. For the approximations to give accurate results the transfer function $G(s)$ should be low-pass, i.e., $|G(ni\omega)| \ll |G(i\omega)|$ for $n = 2, 3, \ldots$ The relation from which both the amplitude and the frequency of the stable oscillation can be determined is given by,

$$G(i\omega) = -\frac{1}{N(A,\omega)}. \tag{7}$$

This formula can be intuitively found by breaking the feedback-loop in Fig. 3 and then go one revolution in the system. For stable oscillation the same signal as the one started with should be achieved when returning to the starting point, meaning $G(i\omega)N(A,\omega) = -1$. This turns out to be exactly the well known Barkhausen's criterion for sustained stable oscillation. The nonlinearity in the converter is modeled as a relay with hysteresis. The hysteresis appears due to delay in the feedback loop, but also when implementing the comparator. The hysteresis is necessary to prevent from very high switching frequency, normally referred to as scattering, that might damage the circuit. The nonlinearity in Fig. 3 has the describing function,

$$N(A) = \frac{4H}{\pi A}\left(\sqrt{1-\left(\frac{D}{A}\right)^2} - i\frac{D}{A}\right), \tag{8}$$

where $H = V_{\mathrm{DD}}/2$. Since the switches are not ideal, the on-resistance $R_{\mathrm{ds,on}}$ of the switch transistors has to be included. This implies a transfer function

$$G(i\omega) = \frac{R}{i\omega(RR_{\mathrm{ds,on}}C + L) + R + R_{\mathrm{ds,on}} - \omega^2 LRC}. \tag{9}$$

The frequency of oscillation can be found by solving $\Im\{G(i\omega)\} = \Im\{-1/N(A)\}$. Unfortunately, this equation is of order four and most suitable for numerical solution. However, once the oscillation frequency is found the amplitude of the oscillation is given by this closed expression,

$$A = D\frac{R^2CL\omega^2 - R^2 - RR_{\mathrm{ds,on}}}{R^2R_{\mathrm{ds,on}}C\omega + RL\omega}. \tag{10}$$

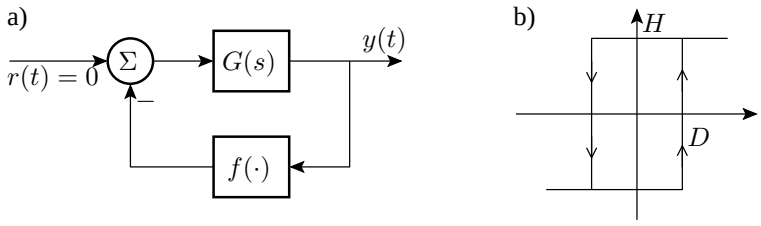

Figure 3: The system for describing function analysis is shown in a). In b) is the characteristics of the nonlinearity shown; a relay with hysteresis.

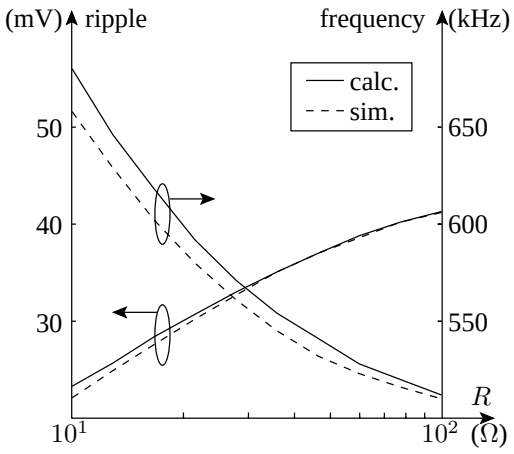

Figure 4: Comparison between calculated ripple and frequency versus simulated ditto.

The ripple and the switching frequency in steady state were calculated with the describing function method and compared to simulations performed in Cadence. The parameters are chosen as $D = 1.7\,\text{mV}$ and $R_{\text{ds,on}} = 1.1\,\Omega$. The agreement is good for both amplitude and frequency estimation, as the deviation is less than 5 % over the whole load range, Fig. 4. The size of the ripple and the switching frequency are determined by the hysteresis in the comparator and the parameters R, L, and C. Large inductance and capacitance will make the system less damped, and thereby increase the ripple while the switching frequency decreases.

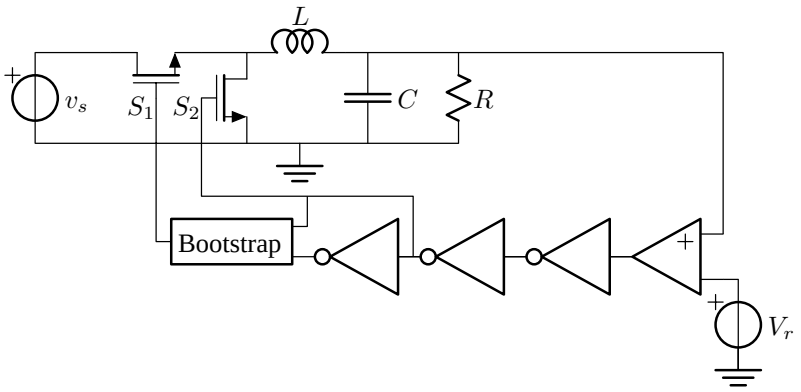

Figure 5: A schematic of the converter including the controller.

4 Implementation Aspects

So far power converters are forced to use external components due to value and quality. Integrated inductors have shown moderate or bad quality factor and will degrade the efficiency. The long-term goal is to implement every circuitry of a system on a single chip. Therefore a standard digital 0.6 μm CMOS process was used. Indeed, CMOS is not well suited to implement power switches in, but is considered as a low cost process. Both the voltage capability and the on-resistance are obstacles that limit the design. The switches have to be very wide, $W \in [1000,\ 10000]\mu$m, to get a reasonably low $R_{\mathrm{ds,on}}$. A schematic of the converter including the controller is shown in Fig. 5. The output voltage is compared with a reference voltage and the output from this reference block is a binary signal, i.e., 0 or V_{DD}. This comparator is implemented by cascading two inherently nonlinear amplifying stages, i.e., Anti-Series Common-Source (ASCS) followed by a Common-Source (CS) stage. The anti-series stage is also known as a differential pair. As discussed earlier the hysteresis of the switches is mainly due to the delay in the feedback loop, and is not explicitly implemented in the comparator. The cascaded inverters are successively scaled in order to drive large capacitive load.

4.1 Low Voltage Operation

In low voltage applications the threshold voltage occupies a significant part of the voltage range. The floating switch in the converter was chosen n-type, implying that the gate voltage must be higher than V_{DD} otherwise a threshold voltage will remain across the transistor [6, 7]. The voltage up-conversion is made with a coarse bootstrapped digital switch illustrated in Fig. 6. The excess voltage applied to the switch transistor can cause reliability problem. One conclusion from [7] is that transient stress is less harmful than DC stress. The bootstrap capacitor was chosen five times

greater than the gate capacitance, which gives an upper voltage limit,

$$V_{\mathrm{boot}} = 2V_{\mathrm{DD}} \frac{C_{\mathrm{boot}}}{C_{\mathrm{boot}} + C_{\mathrm{gate}}} - V_{\mathrm{th}}. \tag{11}$$

An equally area efficient alternative is to use a three times wider pMOS instead, which makes the bootstrapping superfluous.

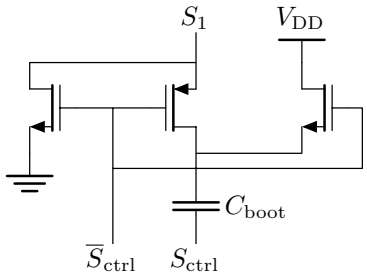

Figure 6: The bootstrap circuit for increased control voltage.

5 Conclusions

Flexibility and optimization demand several voltage supplies, and instead of several batteries switched power converters are used. The Buck converter has worked as a test vehicle in the investigation of implementing converters for small power levels on-chip. The semi-integrated Buck converter with static feedback control has been implemented in a standard digital 0.6 μm CMOS process. The chosen feedback is a direct large-signal approach, and the converter has been proven to be globally asymptotically stable using numerical tools in combination with successor and Lyapunov theory. The ripple has been given an analytical expression by use of describing function analysis. The method proves good amplitude and frequency estimation.

References

[1] P. T. Krein, *Elements of Power Electronics*. Oxford University Press, 1998.

[2] A. J. Forsyth and S. V. Mollov, "Modeling and control of DC-DC converters," *Power Engineering Journal*, vol. 12, pp. 229–236, Oct. 1999.

[3] R. D. Middlebrook, "Small-signal modeling of pulse-width modulated switched-mode power converters," *Proceedings of the IEEE*, vol. 76, pp. 343–354, Apr. 1998.

[4] R. Munzert and P. T. Krein, "Issues in boundary control of power convertors," in *27th Annual IEEE Power Electronics Specialists Conference*, vol. 1, pp. 810–816, 23-27 June 1996.

[5] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs: Prentice Hall, new ed., 1991.

[6] J. Steensgaard, "Bootstrapped low-voltage analog switches," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 29–32, 30 May-2 June 1999.

[7] A. M. Abo and P. R. Gray, "A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 599–606, May 1999.

Paper II

Paper II

Bandwidth Considerations for a CALLUM Transmitter Architecture

Abstract

This article presents an investigation of the combined analog locked loop universal modulator (CALLUM) linear transmitter architecture. A simple approximate formula is derived, linking the instantaneous frequency of the CALLUM signals to the modulation depth of the information signal. Further, the effect of the frequency mismatch between the VCOs present in the CALLUM architecture is examined, and the need for frequency synchronization via a phase-locked loop is discussed.

Based on: Roland Strandberg, Pietro Andreani and Lars Sundström, "Bandwidth Considerations for a CALLUM Transmitter Architecture," *IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 4, pp. 25–28, 2002.

1 Introduction

It is well known that for a certain channel capacity and a given signal-to-noise ratio (SNR) there is a corresponding demand on the radio channel bandwidth. The available frequency band is a very scarce resource, and a wise utilization is of the utmost importance. For a frequency band used for communication the system specification declares the width of each channel, wherein the modulated signal must fit. The transmitted signal must also fit into the spectrum mask, not to disturb any of its neighbors. From any imperfection in the transmitter chain the spectral properties change; the signal is subject to spectral regrowth. This spectral widening of the signal not only hampers full use of the signal itself, but also contaminates the adjacent channels (lowers the SNR), and thus reduce the channels' capacity.

Since the available bandwidth is limited, an extensive search for spectrum efficient modulation schemes has been carried out. A general trend is the move from phase or frequency modulated signals, that is, constant envelope signals, towards more spectrally efficient modulation schemes, which all incorporate non-constant envelope signals. These modulation schemes have information in both phase and amplitude, and require linear processing, otherwise the information is deteriorated. This demand for linear processing is often hampered by the power amplifier (PA) in the radio transmitter chain. Both linear and efficient power amplification can be accomplished with a linear architecture called LINC (LInear amplification using Nonlinear Components), first described in [1]. The LINC architecture is a "divide and conquer" technique, which transforms an amplitude and phase modulated signal into two constant-envelope phase modulated signals. Because of the constant envelope the information is not affected even when a grossly nonlinear PA is used [2]. This opens up the possibility to use amplifiers driven into deep saturation or even highly efficient switching PAs. The approach described so far is also the foundation for the CALLUM architecture, which offers linear amplification and theoretically 100 % efficiency for all output levels [3]. The fact that CALLUM is a feedback system makes this architecture less dependent on matching between the two transmitter paths.

This article presents the CALLUM architecture (section 2) and a derivation of the control equations for the system, together with a study of the relation between spectral widening and modulation depth of the baseband signal (section 3). The influence of frequency mismatch is also investigated, and it is suggested that the resulting bandwidth expansion is removed via frequency synchronization (section 4).

2 CALLUM Architecture

CALLUM is a linear transmitter architecture, and can produce a linear output even with grossly nonlinear PAs. The basic principle is like in LINC, but a few major differences separate CALLUM and LINC. The CALLUM topology includes feedback of the output signal, advantageously taken near the antenna, see Fig. 1. The output

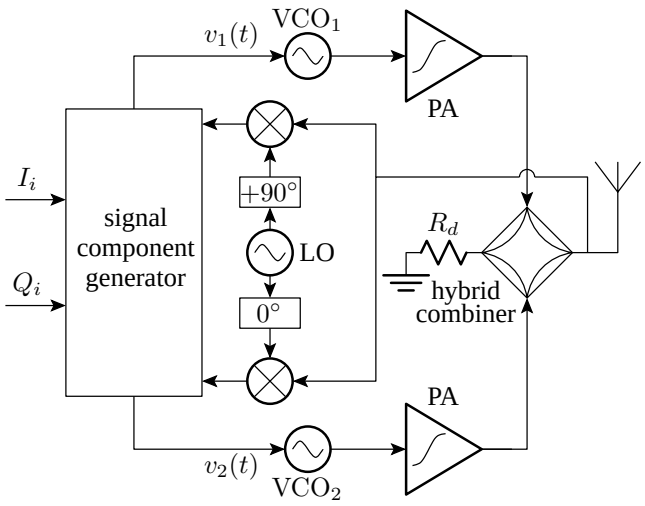

Figure 1: The generic CALLUM transmitter architecture.

signal is measured and downconverted in a complete receiver with some attenuation, equal to the wanted asymptotic gain, $A_{t\infty}$.

In general distortion can be reduced in different ways depending on the nature of the nonlinearity: 1) if the nonlinearity is known, it can easily be canceled, 2) matched nonlinearities are removed by adopting balanced circuits, 3) unknown nonlinearities are reduced by applying negative feedback. The most universal remedy against distortion is to apply negative feedback, and anything in the CALLUM system that produces distortion, (e.g., mismatch between branches, and nonlinearities), is suppressed to a certain extent, depending on the loop gain. As a drawback, a feedback system has the disadvantages of an increased complexity and of possible stability hazards.

The signal component generator (SCG), creating the control voltages, $v_1(t)$ and $v_2(t)$, to the VCOs from the input and feedback signals, works at baseband or at intermediate frequency (IF). In comparison with LINC, where the upconversion is already made and the SCG works directly on the RF signal, the SCG in CALLUM has the benefit to perform the signal processing at low frequency, which means lower power consumption and more easily met signal constraints on the SCG.

3 Signal Component Generator

In the SCG the baseband signal $s(t)$ is mathematically manipulated and two separate control signals are fed to the VCOs, to form two phase modulated and frequency translated signals, $s_1(t)$ and $s_2(t)$. The representation of the baseband signal is given

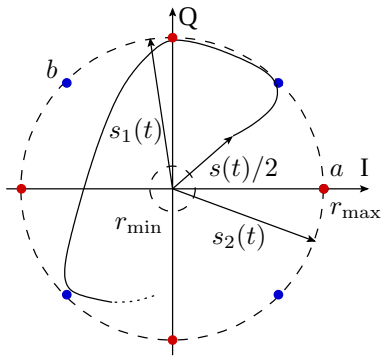

Figure 2: A $\pi/4$-shifted QPSK modulation scheme showing a stylized trajectory of the modulated signal, $s(t)$, and the constant envelope signals, $s_1(t)$ and $s_2(t)$.

by

$$s(t) = r(t)\cos(2\pi f_c t + \phi(t)), \tag{1}$$

where $r(t)$ is the amplitude, $\phi(t)$ is the phase modulation, and f_c is the center frequency. The goal is to find the appropriate control signals to the VCOs, $v_1(t)$ and $v_2(t)$, described either in a polar or a Cartesian form. However, if the equations imply a polar feedback implementation, the amplitude and phase information will be detected in two completely different paths, with different and unknown delays. By instead using a Cartesian representation, and integration on silicon, the nominally identical I and Q paths will have much better matched delays.

The amplitude of the input signal is

$$r(t) = \sqrt{I_i(t)^2 + Q_i(t)^2}, \tag{2}$$

where the index i refers to the input signal. Vector summation of the two constant envelope signals, shown in Fig. 2, is the principle for generating the output signal in the CALLUM system,

$$s(t) = s_1(t) + s_2(t). \tag{3}$$

In our search for a closed expression for the control voltages, $v_1(t)$ and $v_2(t)$, for the closed loop system, we start by approximating the time derivative of the input signal with a difference ratio. The system is assumed to only suffer from a small delay τ around the loop, and by introducing a parameter, $k_1 = 1/\tau$, the time derivative is

$$\dot{I}_i(t) \simeq k_1 \left(I_i(t) - \frac{I_o(t-\tau)}{A_{t\infty}} \right). \tag{4}$$

The delay around the loop is expected to be frequency independent for narrow band signals.

The CALLUM system is first studied in an open loop configuration. By using (3) and the fact that the frequency of VCO_1 can be controlled with a voltage, $v_1(t)$, the approximate closed loop control voltage is given by (5). An analogous equation can be found for $v_2(t)$. In (5) the time derivative of the input signal is already replaced by difference ratios; $K_{\mathrm{VCO},f}$ is the sensitivity of the VCO, $r_{\max}$ is the amplitude of $s_1(t)$ (or $s_2(t)$), and g_1 is a gain factor. Instead of letting the loop gain amplify the error signal directly, equation (5) contains amplitude scaling for improving performance for all positions in the complex plane.

$$\begin{aligned} v_1(t) &= \frac{g_1 k_1}{2\pi K_{\mathrm{VCO},f}} \left\{ \left(I_i(t) - \frac{I_o(t-\tau)}{A_{t\infty}} \right) \left(-\frac{Q_i(t)}{r(t)^2} - \frac{I_i(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}} \right) \right. \\ &\quad \left. + \left(Q_i(t) - \frac{Q_o(t-\tau)}{A_{t\infty}} \right) \left(\frac{I_i(t)}{r(t)^2} - \frac{Q_i(t)}{r(t)\sqrt{4r_{\max}^2 - r(t)^2}} \right) \right\} \end{aligned} \tag{5}$$

3.1 Spectrum Widening

The information signal, $s(t)$, is shown in Fig. 3 and 4, together with the signals $s_1(t)$ and $v_1(t)$ for Enhanced Data rates for GSM Evolution (EDGE) and $\pi/4$-shifted Quadrature Phase Shift Keying (QPSK) modulations, respectively. The bandwidth of the constant-envelope signal in the loop is apparently dependent on the modulation depth of the signal, and the control signal to the VCO is spectrally wider than the constant-envelope signal, $s_1(t)$. Since the loop gain is decreasing at higher frequencies, large spectral regrowth limits the performance of the CALLUM architecture. The true implementation of the loop will limit the bandwidth of the signal, (e.g., the signal fed to the VCO will be a lowpass filtered version of the exact signal), and the degradation of the information signal has so far not been reported.

In order to find an approximate formula describing the instantaneous frequency of the constant envelope signals in CALLUM systems, we consider the constellation diagram for the modulation, see Fig. 2. The trajectory has to move from one point to another in the modulation diagram in exactly the time between two consecutive symbols, T_s. The distance between the points a and b is r_{ab}. For signals near $r_{\min}$ the angular velocity of $s_1(t)$ (and $s_2(t)$) is almost the same as for $s(t)$. If the signal moves at a constant speed from a to b, then the maximum instantaneous frequency of the vector $s_1(t)$ (or $s_2(t)$) is,

$$f_{\max} = \frac{r_{ab}}{2\pi T_s r_{\min}} \leq \frac{m_d}{\pi T_s}. \tag{6}$$

$r_{\min}$ equals the smallest output amplitude for the transmitted sequence, and the modulation depth, m_d, of a signal is defined as the ratio between the maximum and minimum amplitude. Equation (6) predicts $f_{\max} \simeq 2.9/T_s$, and simulations give

$f_{\max} \simeq 3/T_s$ for $\pi/4$-shifted QPSK modulation ($m_d \simeq 10$). Similarly, for an EDGE modulated signal ($m_d \simeq 7$), the predicted $f_{\max}$ is $2.2/T_s$, and the simulated $f_{\max}$ is $1.4/T_s$. Most importantly, (6) shows that the instantaneous frequency is proportional to the modulation depth of the information signal.

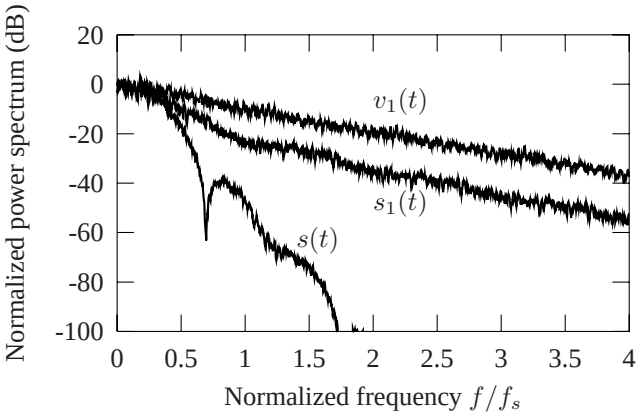

Figure 3: EDGE modulation with modulation depth $m_{\mathrm{d}} \simeq 7$.

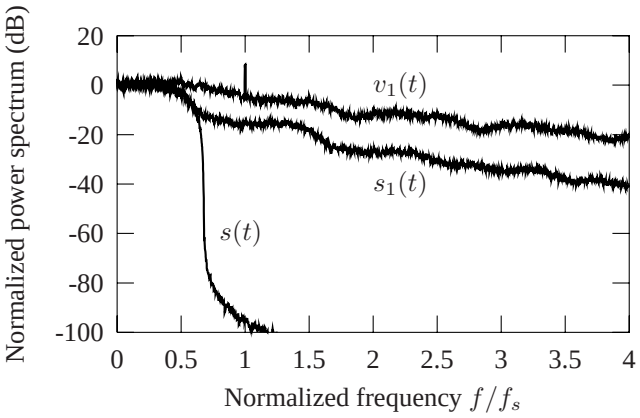

Figure 4: $\pi/4$-shifted QPSK modulation with modulation depth $m_{\mathrm{d}} \simeq 10$.

4 VCO Free-Running Frequency Mismatch

In the CALLUM architecture two VCOs are involved, and the LO in the feedback path sets the reference frequency. The two VCOs, one in each branch, should synchronize in frequency to the reference LO, otherwise this frequency mismatch will cause a voltage (or current) error at the output. In the following the frequency error is assumed to be a static frequency offset. The output signal is an amplified replica of the input plus some additional discrepancy,

$$I_o(t) = A_{t\infty}(I_i(t) + I_{\mathrm{dis}}(t)). \tag{7}$$

The constant frequency offset corresponds to a DC correction voltage that has to be applied to the corresponding VCO according to

$$\begin{aligned} V_1 &= \frac{f_{\mathrm{VCO1}} - f_c}{K_{\mathrm{VCO},f}} = \frac{\Delta f_1}{K_{\mathrm{VCO},f}}, \\ V_2 &= \frac{f_{\mathrm{VCO2}} - f_c}{K_{\mathrm{VCO},f}} = \frac{\Delta f_2}{K_{\mathrm{VCO},f}}, \end{aligned} \tag{8}$$

where f_{VCO1} (f_{VCO2}) denotes the free-running frequency of VCO_1 (VCO_2). After straightforward calculation using (5), (7), and (8), the following expression is found:

$$\begin{aligned} I_{\mathrm{dis}} &= \frac{\pi}{g_1 k_1} \left\{ \Delta f_1 \left(I_i \sqrt{\frac{4r_{\max}^2}{r^2} - 1} + Q_i \right) \right. \\ &- \left. \Delta f_2 \left(I_i \sqrt{\frac{4r_{\max}^2}{r^2} - 1} - Q_i \right) \right\}. \end{aligned} \tag{9}$$

There exists two obvious ways in order to decrease the effect of frequency mismatch between the VCOs, namely

1. increase the loop gain,
2. synchronize the VCOs, i.e., remove Δf_1 and Δf_2.

An efficient remedy to minimize the influence of the frequency offset is to apply sufficiently much loop gain, affected by $g_1 k_1$. However, the amount of loop gain also affects the high frequency behavior, and for large values of the loop gain instability is expected for this system. Of course, as much loop gain as possible should be used, but in order to make the system perform at its best, frequency mismatch has to be minimized by means of frequency synchronization.

4.1 VCO Frequency Synchronization

The effective bandwidth of the signal is increased by the amount of frequency mismatch [4]. Since the loop gain is falling off at higher frequencies, the regrowth of the

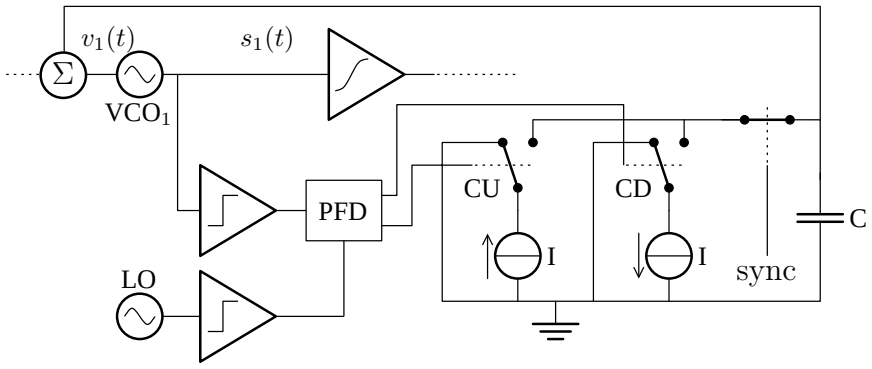

Figure 5: Frequency synchronization of VCO$_1$ in the transmit path.

information signal should be minimized for optimal performance. An efficient way of synchronizing two signals is by using a phase-locked loop (PPL). In Fig. 5 the frequency synchronization of the VCO in the transmit path is shown. The PLL includes VCO, limiters, phase-frequency detector (PFD), charge pump, and filter as the main blocks.

The bandwidth of the frequency synchronization loop must be low, in order not to compensate for the information signal as well. Since the baseband signal stretches from DC, we enforce frequency lock by closing the PLL between bursts of transmitted data. The loop is subsequently opened during transmission by changing the state of the *sync* signal, see Fig. 5. The time to achieve frequency synchronization is not critical for this test system, and to guarantee the PLL functionality a PFD is used. The offset voltage needed to synchronize the VCO is stored in a sufficiently large capacitor.

5 Conclusions

In this article we have investigated bandwidth related issues in CALLUM linear transmitter architecture, which offers linear amplification and theoretically 100 % efficiency for all output levels. Based on some simple parameters, an approximate equation predicts that the instantaneous frequency of the constant envelope signal is proportional to the modulation depth of the information signal. Therefore, some modulation schemes, (e.g., EDGE, $\pi/4$-shifted QPSK) are more suitable due to their limited modulation depth. In the implementation of the transmitter architecture the free-running frequency of the VCO in the transmit path will differ from the channel frequency set by the frequency reference. It has been shown that this frequency error cause an error at the output, and further expand the signal bandwidth of $s_1(t)$ and $v_1(t)$ within the loop. To reduce this error on the output signal an PLL is the necessary add on circuitry to perform the frequency synchronization.

References

[1] D. C. Cox, "Linear amplification with nonlinear components," *IEEE Transactions on Communications*, vol. 22, pp. 1942–1945, Dec. 1974.

[2] F. J. Casadevall and A. Valdovinos, "Performance analysis of QAM modulations applied to the LINC transmitter," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 399–406, Nov. 1993.

[3] K. Y. Chan and A. Bateman, "Linear modulators based on RF synthesis: Realization and analysis," *IEEE Transactions on Circuits and Systems*, vol. 42, pp. 321–333, June 1995.

[4] D. J. Jennings and J. P. McGeehan, "A high-efficiency RF transmitter using VCO-derived synthesis: CALLUM," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, pp. 715–721, June 1999.

Paper III

Paper III

Spectrum Emission Considerations for Baseband-Modeled CALLUM Architectures

Abstract

Linear transmitters based on the CALLUM architecture are attractive, as they promise both high efficiency and high linearity.

To date, it has not been possible to analyze a CALLUM transmitter as a linear feedback network, due to the non-linear nature of the control equations governing it. The main purpose of the present work has been the derivation of a linearized model for the control equations, which enables the use of linear network theory in the study of CALLUM. In particular, it can be used to analyze the stability and maximize the bandwidth of the system.

Simulation examples are presented on how three different CALLUM architectures behave for EDGE and W-CDMA signals. In addition, we have considered the effects of loop time delay, which is the ultimate limiting factor for all feedback-based linear transmitter architectures, in particular for large bandwidths. Finally, it is shown how frequency compensation of the feedback loop improves insensitivity to the loop time delay.

Based on: Roland Strandberg, Pietro Andreani and Lars Sundström, "Spectrum Emission Considerations for Baseband-Modeled CALLUM Architectures," *IEEE Transactions on Microwave Theory and Techniques*, Accepted for publication.

1 Introduction

The need of utilizing the available frequency spectrum as efficiently as possible has resulted in wireless communication standards abandoning the traditional constant-amplitude modulation schemes (such as GMSK in GSM), adopting instead non-constant envelope modulations, as in the Enhanced Data rates for GSM Evolution (EDGE) standard. Common to all non-constant envelope modulation schemes is that the information is not only impressed on the phase, but also on the amplitude of the carrier. This imposes stringent demands on the linearity of the power amplifier (PA) used in the radio transmitter, and makes the design of linear PAs highly non-trivial, especially when a major challenge in wireless applications is to realize low power circuits. In fact, it is well known that linearity and efficiency trade with each other.

Highly efficient linear power amplification can be achieved by linearizing a non-linear but very efficient PA. Established linearization techniques are feedback (either polar or cartesian [1, 2]), feed-forward [3], and predistortion [4]. Alternatively, an intrinsically linear transmitter architecture can be adopted. The Combined Analog Locked Loop Universal Modulator (CALLUM), the subject of the present paper, is one such linear architecture, having a theoretical efficiency of 100% for all output levels [5]. CALLUM is an expansion of the LInear amplification using Nonlinear Components (LINC, first described in [6]), since it is based on a LINC core placed in a feedback loop. LINC and CALLUM are based on a kind of divide-and-conquer approach, which first transforms an amplitude- and phase-modulated signal into two constant-envelope phase-modulated signals, and then recombines them after they have been power amplified. Since a constant-envelope signal can be processed by a grossly non-linear PA without information loss [7], the two constant-envelope signals can now be amplified by highly efficient but highly non-linear class-C PAs, or even by switching PAs (such as class-E PAs, having a 100% theoretical efficiency [8]). Although linear and efficient power amplification is particularly attractive for battery-powered wireless handsets, it could be fruitfully employed in base stations as well, in order to decrease the overall power dissipation. The diminished cooling requirements would result in cheaper and more compact base stations. However, despite the promises offered by the different techniques just described, it should be admitted that major breakthroughs in the art of linear and efficient radio-frequency (RF) PA design have remained elusive to date.

Many variations around the basic CALLUM architecture, usually referred to as CALLUM 1, have appeared in the open literature. Chan and Bateman [5] reviewed no less than six: CALLUM 1, 2, 3, 4, VLL, and higher-order. The various CALLUM implementations were investigated in terms of stability in the IQ-plane and speed estimations performed by step response analysis, which were supplemented with measurements on physical implementation using discrete building blocks. Several CALLUM designs were simulated in [9], where their RF output spectra for a TETRA-like baseband modulation were presented, together with actual measurements for CALLUM 2. This work was an important contribution to the understanding of the properties of the

different CALLUM versions, and the goal of the present paper is to further investigate these properties, mainly through the study of the feedback loop present in all CALLUM designs.

We will confine our analysis to those implementations that are both continuous (i.e., with loop coefficients independent of the input signal value) and based on cartesian feedback, as opposed to polar feedback. This second condition is important in guaranteeing matched time delays in the two paths of the feedback signal, which is crucial for achieving high linearity. As a consequence, we will focus on CALLUM 1 and CALLUM 2, together with an intermediate approach [9], here referred to as CALLUM 1 with linearized denominator (CALLUM 1lin).

The rest of this paper is organized as follows: the CALLUM architecture is presented at block level in section 2, and an equivalent baseband model used for the loop gain calculations is presented in section 3. The loop gain equations are graphically visualized in the IQ-plane, and are shown to contain important information for the subsequent stability analysis. In section 4 spectral emissions for three different versions of CALLUM are compared when operating on an EDGE and on a W-CDMA signal, respectively. Finally, the effect of time delay in the feedback loop is presented in section 5, together with suitable frequency compensation techniques improving the stability properties of the system.

2 System Description

We have seen that in a CALLUM architecture the input signal, which is assumed to be both amplitude- and phase-modulated, is divided into two constant-envelope phase-modulated signals, which are recombined after amplification. The feedback signal is advantageously taken as close to the antenna as possible (constrained by stability issues, see section 5), in order to minimize the number of error sources that can not be corrected by the feedback loop. It is well known that the nonlinearities included in the loop are reduced by the action of negative feedback, which is a very powerful and robust remedy against distortion regardless of what causes it (e.g., mismatch between ideally identical branches, nonlinear transfers, etc). The degree of distortion suppression is proportional to the magnitude of the signal gain in the loop. A major drawback of this approach, on the other hand, is that any feedback system has an increased complexity and is prone to instability in general. The stability properties are fundamental and will be studied in detail in section 5.

The generic CALLUM architecture is shown in Fig. 1 at the functional block level. The signal component generator (SCG) is the heart of the architecture, and the algorithm implemented in the SCG will differ depending on the kind of CALLUM realized (or, equivalently, depending on the chosen set of *control equations*).

In this paper only continuous versions of CALLUM suitable for implementation based on a cartesian representation are considered. Cartesian representation promises good signal matching properties between the I and Q path in the feedback, especially

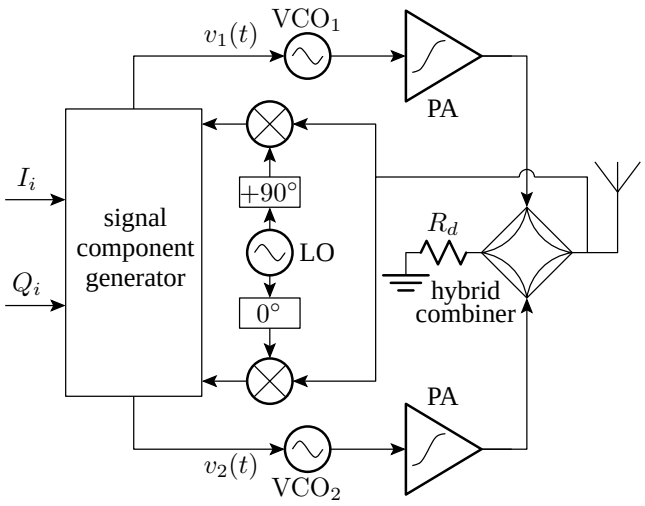

Figure 1: The generic CALLUM transmitter architecture.

in terms of delay [10]. Starting from the input signals and the feedback signals, the SCG generates the control voltages $v_1(t)$ and $v_2(t)$ for two voltage-controlled oscillators (VCO$_1$ and VCO$_2$). The SCG operates on signals either at baseband, or at an intermediate frequency. A derivation of the control equations for the original version of CALLUM, here to be referred to as CALLUM 1, was presented in [10]. The control equation for $v_1(t)$ is repeated here in (1) for convenience, since it is the basis for two more CALLUM versions to be presented in the following (an analogous equation describes $v_2(t)$).

$$
\begin{aligned}
v_{1,\mathrm{C1}} &= \frac{g_1 k_1}{K_{\mathrm{VCO}}} \left\{ \left(I_i(t) - \frac{I_o(t-\tau)}{A_{t\infty}} \right) \left(-\frac{Q_i(t)}{r_i(t)^2} - \frac{I_i(t)}{r_i(t)\sqrt{4r_{\max}^2 - r_i(t)^2}} \right) \right. \\
&\quad \left. + \left(Q_i(t) - \frac{Q_o(t-\tau)}{A_{t\infty}} \right) \left(\frac{I_i(t)}{r_i(t)^2} - \frac{Q_i(t)}{r_i(t)\sqrt{4r_{\max}^2 - r_i(t)^2}} \right) \right\} \qquad (1) \\
v_{1,\mathrm{C1lin}} &= \frac{g_1 k_1}{K_{\mathrm{VCO}}} \left\{ \left(I_i(t) - \frac{I_o(t-\tau)}{A_{t\infty}} \right) \left(-\frac{Q_i(t)}{r_i(t)^2} - \frac{I_i(t)}{2r_{\max} r_i(t)} \right) \right. \\
&\quad \left. + \left(Q_i(t) - \frac{Q_o(t-\tau)}{A_{t\infty}} \right) \left(\frac{I_i(t)}{r_i(t)^2} - \frac{Q_i(t)}{2r_{\max} r_i(t)} \right) \right\} \qquad (2) \\
v_{1,\mathrm{C2}} &= \frac{g_1 k_1}{K_{\mathrm{VCO}}} \left\{ \left(I_i(t) - \frac{I_o(t-\tau)}{A_{t\infty}} \right) \Big(-Q_i(t) - I_i(t) \Big) \right. \\
&\quad \left. + \left(Q_i(t) - \frac{Q_o(t-\tau)}{A_{t\infty}} \right) \Big(I_i(t) - Q_i(t) \Big) \right\} \qquad (3)
\end{aligned}
$$

Index C1 in (1) denotes the control equation for the CALLUM 1 version of the CALLUM design, while subscripts i and o refer to the input and output signals, respectively; K_{VCO} is the gain factor of the VCO (in $\mathrm{rps/V}$), $A_{t\infty}$ is the asymptotic closed-loop transfer from input to output [11], g_1 and k_1 are gain factors, and r_i is the magnitude of the complex input vector. r_i should be chosen such that the maximum input signal amplitude generates the maximum output signal for best power efficiency. The value for $r_{\max}$ is given by $r_{\max} = (1/2)\max(r_i)$.

It should now be noted that the square root term in (1) poses severe challenges for the actual circuit implementation, and avoiding it might be advantageous from an overall point of view. A straightforward simplification can be obtained by replacing the square root with its Taylor series expansion, truncated at the first order [9]. This results in CALLUM 1 with linearized denominator (to be referred to as CALLUM 1lin), whose control equation is given by (2). It is possible to make a further very significant simplification of (2), in order to improve designability even more. Since a division is a costly operation, denominator terms in the control equation for CALLUM 1lin can be disregarded altogether; the resulting CALLUM version is known as CALLUM 2 [5, 12, 13], with control equation given by (3). Of course, possibly severely degraded signal performances are expected in CALLUM 2.

2.1 Baseband Modeling of CALLUM System

The dramatic reduction in simulation time is the major drive behind the realization of a baseband model for CALLUM. In a baseband simulation the carrier frequency is not present, and as a result simulations based on a baseband model are far more efficient in terms of computational workload than simulations performed on the system with RF signals.

The carrier frequency f_c can be removed because its only effect is to rotate the coordinate system (the complete IQ-diagram), without affecting the information contained at baseband. In all CALLUM architectures both an up- and a down-conversion in frequency are performed, which are bypassed in the baseband model. It is of course crucial that the baseband model is able to capture all relevant features and possible problems of the original RF system.

Since the carrier frequency is 0 Hz in a baseband simulation, and a reference phase is therefore lacking, the signal representation has to be made in the complex domain. A normal circuit simulator based on time domain analysis (e.g., spectreRF in Cadence) cannot handle complex signals directly, which calls for a workaround. The definition of the baseband output signal, $s_o(t)$, and its relation to the constant-envelope vectors, $s_{o1}(t)$ and $s_{o2}(t)$, is shown in Fig. 2 and given by,

$$s_{o1} = \Re\{s_{o1}\} + j\Im\{s_{o1}\} \quad (4)$$

$$s_{o2} = \Re\{s_{o2}\} + j\Im\{s_{o2}\} \quad (5)$$

$$s_o = \Re\{s_o\} + j\Im\{s_o\} = s_{o1} + s_{o2}, \quad (6)$$

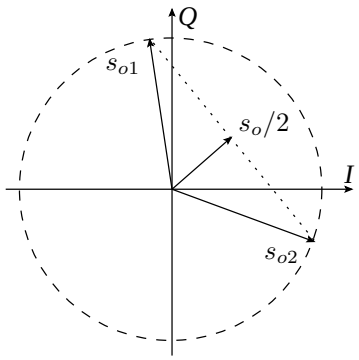

Figure 2: The vector addition of the constant envelope signals to reconstruct an amplified replica of the input signal.

where $\Re\{s_o\}$ ($\Im\{s_o\}$) denotes the real (imaginary) part of the output signal s_o. It is now easy to identify the output I and Q baseband components as,

$$I_o = \Re\{s_{o1}\} + \Re\{s_{o2}\} \quad (7)$$

$$Q_o = \Im\{s_{o1}\} + \Im\{s_{o2}\}. \quad (8)$$

The baseband circuit of Fig. 3 results directly from the above equations, while the factor K_{FB} models the gain (or attenuation) in the frequency down-conversion. The phase between $\mathrm{VCO}_{\cos}$ and $\mathrm{VCO}_{\sin}$ is $\pi/2$ radians to accommodate the complex implementation. To guarantee this relative phase shift between the VCOs, they are described as mathematical functions in AHDL[1]. To repeat, the baseband representation is of great importance, since it makes possible the simulation of complete bursts of transmitted data in a reasonable amount of time. This, in turn, enables the assessment of the qualifying features for the different CALLUM versions.

3 Loop Gain Calculations Based on Linearized Equations

As is clear from (1)-(3), CALLUM is governed by a set of nonlinear control equations, from which basic features like loop gain, bandwidth, and stability are difficult to obtain. For these reasons, these equations will be linearized around a static bias point.

Since the three CALLUM versions differ only in the control equations, the generic structure of the linearized baseband CALLUM model, displayed in Fig. 4, is the same

[1] Analog Hardware Description Language is handled by many circuit simulators (spectre in our case).

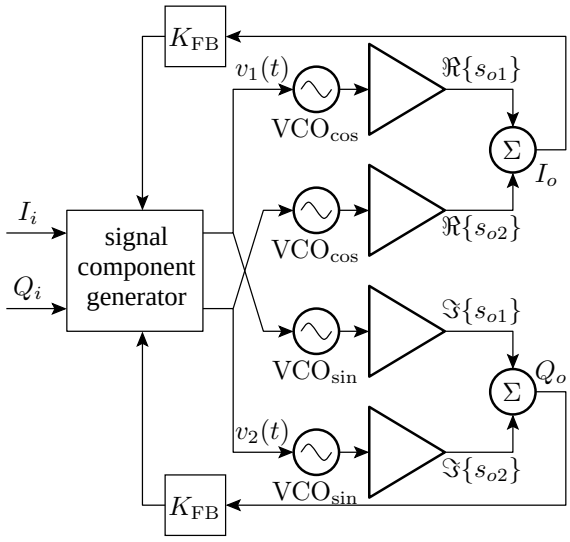

Figure 3: Baseband model for a generic CALLUM system with real-valued signal paths.

in all cases. As with phase-locked loops (PLLs), the relevant input and output signals are phases, while inside the loop different phase-to-voltage and voltage-to-phase conversions take place. The model in Fig. 4 is roughly equivalent to half of the circuit in Fig. 3, in that the cross-coupling between the I and Q part of the baseband equivalent has been neglected. This rather drastic simplification can be justified as follows: since the goal is to have a linear signal transfer, the required amount of loop gain is substantial, and therefore the I (Q) component of the signal propagating through the Q (I) part of the system will be significantly attenuated[2]. This simplified model is only used for loop gain calculations. Closed-loop simulations are of course performed on the complete model of Fig. 3.

Throughout the rest of the paper, the following notation is adopted, which hopefully will make the linearization procedure easy to follow: we define x_a as the total signal, X_A its bias component, and $\tilde{x}_a$ its small-signal variation (i.e., $x_a = X_A + \tilde{x}_a$). We begin the linearization procedure by noting that the error signal for a constant input signal (I_I, Q_I) goes to zero as time goes to infinity, due to the presence of an integrator in the loop (it is well known from PLL theory that an oscillator behaves as an ideal integrator in the baseband PLL model [14]). At steady state, the output signal will be equal to the input signal scaled by the asymptotic gain: $I_O = I_I A_{t\infty},\ Q_O = Q_I A_{t\infty}$. Now, in order to perform a loop gain analysis, all variations for the input signals are

[2]The I (Q) component is seen as a disturbance by the Q (I) part of the system, and as such is largely suppressed by it.

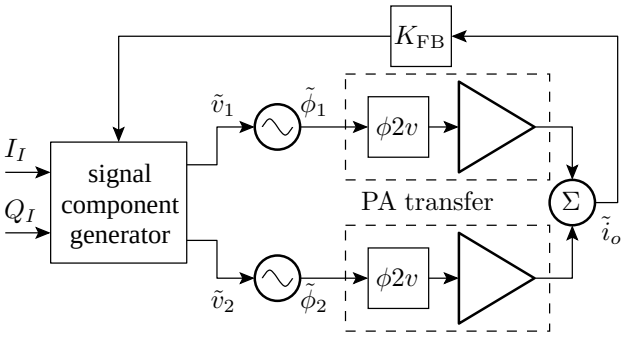

Figure 4: Reduced baseband model used for loop gain calculations.

set to zero, whereas their bias components are fed to the system. The non-linear system is linearized around this quiescent point, and the loop is opened at a suitable node (for instance, at the output of the signal adder) for open-loop calculations.

From the outputs of the VCOs to the output of the CALLUM architecture a phase-to-amplitude conversion takes place, indicated with $\phi 2v$ in Fig. 4. The trigonometric function performed by the PA and the combiner is a projection of the output signal from the PAs onto the I-axis, as shown in Fig. 5. The in-phase output signal is given

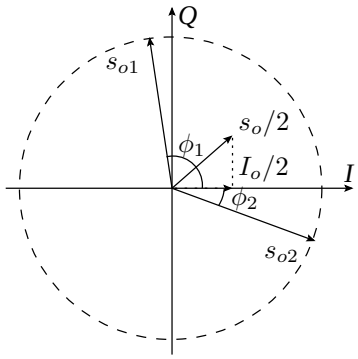

Figure 5: Projection of the output signal on the I-axis.

by

$$I_o = \Re\{s_{o1}\} + \Re\{s_{o2}\} = K_{\mathrm{PA}} V_{\mathrm{VCO}} (\cos\phi_1 + \cos\phi_2), \tag{9}$$

where K_{PA} is the transfer of the power amplifier (here considered as a constant), and V_{VCO} is the amplitude of the VCO signal. The baseband input signal level is determined by $r_{\max}$, which depends on the feedback gain, the PA gain, and the VCO

gain ($r_{\max} = V_{\rm VCO} K_{\rm PA} K_{\rm FB}$). The transfer function for the VCO is given by $\tilde{\phi}_1 = K_{\rm VCO}\tilde{v}_1/s$, where s is the complex angular frequency as usual.

Straightforward linearization of (9) yields the output signal as

$$\tilde{i}_o = -K_{\rm PA} V_{\rm VCO} \left\{ \sin(\Phi_1)\tilde{\phi}_1 + \sin(\Phi_2)\tilde{\phi}_2 \right\}, \tag{10}$$

where $\Phi_1 = \Phi + \alpha$, $\Phi_2 = \Phi - \alpha$, and Φ and α are given by

$$\Phi = \arctan\left(\frac{Q_I}{I_I}\right) + n\pi \tag{11}$$

$$\alpha = \arccos\left(\frac{\sqrt{I_I^2 + Q_I^2}}{2r_{\max}}\right), \tag{12}$$

where $n = 0$ if $I_I \geq 0$, and $n = 1$ if $I_I < 0$. From the model in Fig. 4 we can write the loop gain as

$$A\beta = -\frac{K_{\rm FB} K_{\rm PA} K_{\rm VCO} V_{\rm VCO}}{s} \cdot \left(\sin(\Phi_1) K_{{\rm SCG},{\rm v}_1} + \sin(\Phi_2) K_{{\rm SCG},{\rm v}_2} \right), \tag{13}$$

where $K_{{\rm SCG},{\rm v}_1}$ ($K_{{\rm SCG},{\rm v}_2}$) is the SCG transfer relative to signal v_1 (v_2).

3.1 Linearization of the Control Signals

In the expression for the loop gain (13) the transfers for the SCG, $K_{{\rm SCG},{\rm v}_1}$ and $K_{{\rm SCG},{\rm v}_2}$, have not been derived yet. The $K_{{\rm SCG},{\rm v}_1}$ transfer is found by linearizing the control equations (1), (2), and (3), respectively. Note that direct linearization of these equations include the effect of $K_{\rm FB} = 1/A_{t\infty}$, which is not part the SCG transfer. Therefore the result should be divided by $K_{\rm FB}$ in order to include the effect of the gain in the feedback path only once in the loop gain expression (13). As indicated in Fig. 4, only a small variation in the output signal, $\tilde{i}_o$, is considered. Thus, the linearized transfers of the SCG are calculated in presence of the following signal: $\tilde{i}_i = 0$, $\tilde{q}_i = 0$, $\tilde{q}_0 = 0$, $Q_O = Q_I A_{t\infty}$, $I_O = I_I A_{t\infty}$. These transfers for the three CALLUM derivatives are given by,

$$K_{{\rm SCG},{\rm v}_1,{\rm C1}} = \frac{g_1 k_1}{K_{\rm VCO}} \left(\frac{Q_I}{r_I^2} + \frac{I_I}{r_I \sqrt{4r_{\max}^2 - r_I^2}} \right) \tag{14}$$

$$K_{{\rm SCG},{\rm v}_1,{\rm C1lin}} = \frac{g_1 k_1}{K_{\rm VCO}} \left(\frac{Q_I}{r_I^2} + \frac{I_I}{2r_{\max} r_I} \right) \tag{15}$$

$$K_{\mathrm{SCG,v_1,C2}} = \frac{g_1 k_1}{K_{\mathrm{VCO}}}(Q_I + I_I). \tag{16}$$

Here r_I is the amplitude of the input signal taken in the operating point:

$$r_I = \sqrt{I_I^2 + Q_I^2}. \tag{17}$$

The SCG transfer for v_2 only differs from $K_{\mathrm{SCG,v_1}}$ by having a minus sign in front of the I_I term.

3.2 Loop Gain Characteristics

All information to express the loop gain for any of the CALLUM versions is now available. The loop gain is found by inserting the transfers of the SCG in the generic expression for the loop gain (13). The fairly complicated loop gain expression prevents a direct interpretation; hence, the absolute value of the loop gain is plotted in the IQ-diagram, which gives a very quick indication of the expected performance of the chosen implementation. Fig. 6 shows the magnitude of the loop gain for CALLUM 1 for the in-phase component, plotted at a fix frequency. The maximum value of the loop gain is set to 100 to enable an easy comparison between the CALLUM versions. Due to the loop gain normalization, the chosen frequency is arbitrary.

The loop gain for CALLUM 1 is constant over the whole valid input range $0 < r_i \leq 2r_{\max}$ in the IQ-plane, which is not obvious from the loop gain expression. This is unique for the original CALLUM 1 control law, as it has a continuous amplitude normalization in the loop gain expression. For the quadrature-phase component variation, the loop gain will show the same pattern, rotated by $\pm 90^\circ$. Since the same information is contained in the in-phase and quadrature plots, in the following we will not make this distinction.

For CALLUM 1lin the first-order approximation $r_i\sqrt{4r_{\max}^2 - r_i^2} = 2r_{\max}r_i$ is used in the SCG control equation, which changes the loop gain compared to CALLUM 1. It can be shown that the first-order Taylor series expansion is accurate for $I_I \ll r_{max}$, when similar results as for CALLUM 1 are expected. The 3D loop gain plot for CALLUM 1lin is shown in Fig. 7, where the loop gain clearly drops for large values of I_I. A more accurate approximation (i.e., a higher order Taylor expansion) would reduce the drop in loop gain, at the expense of a much increased circuit complexity.

Both CALLUM 1 and CALLUM 1lin have high implementation costs, due to the relatively complicated mathematical operations (square root and/or division) that have to be performed. The strive for reduced complexity resulted in CALLUM 2, where the price to be payed is small loop gain (or lack thereof) at low signal levels, and a strongly varying loop gain within the amplitude interval of interest $0 < r_i \leq 2r_{\max}$, as shown in Fig. 8. The valid input range is limited at the lower end by the bandwidth of the closed loop system [10], and at the upper end by signal handling capability (i.e., by the onset of clipping).

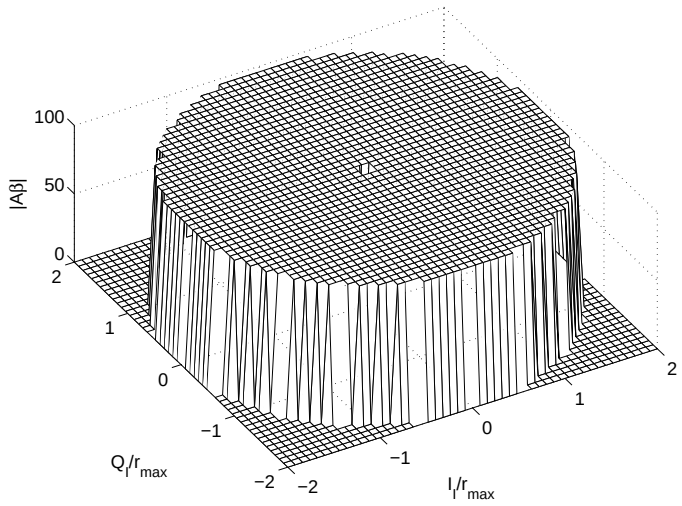

Figure 6: The loop gain is shown for a CALLUM 1 system implementation. The loop gain refers to the in-phase component variation plotted at a fix frequency. The loop gain is set to zero outside the valid input range.

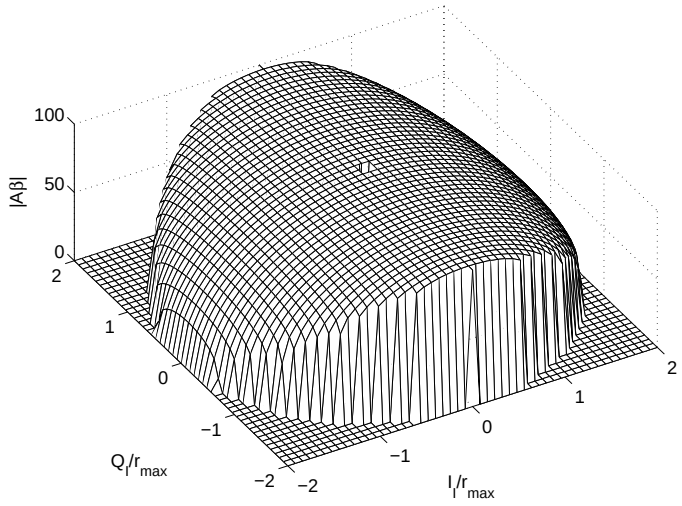

Figure 7: Loop gain for CALLUM 1 with linearized denominator.

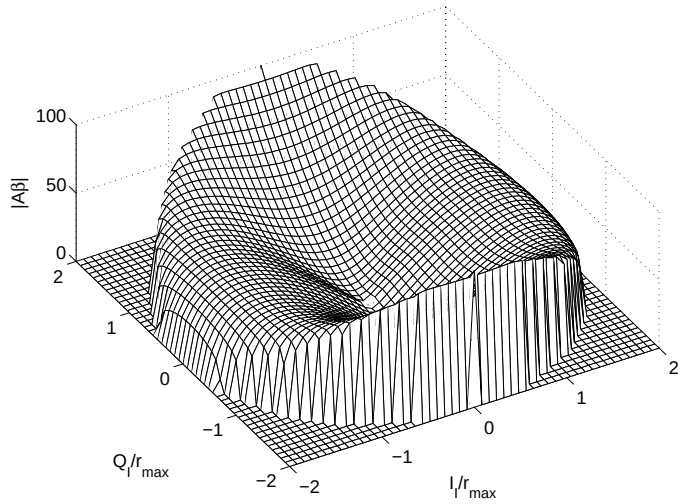

Figure 8: Loop gain 3D plot for CALLUM 2. The region with low loop gain for low envelope signal is one of the main drawbacks for CALLUM 2 compared to CALLUM 1.

The information available in the 3D contour plots of the loop gain allows the investigation of feedback stability and the related design of frequency compensation. A prediction where the loop gain has its largest value in the IQ-plane will serve as the point selected for compensation purposes, since this is the worst-case scenario in feedback systems. Clearly, in all CALLUM versions maximum loop gain is achieved when $I_I = 2r_{\max}$. Thus, frequency compensation must be performed in presence of maximum envelope for the input signal.

4 Comparisons between Baseband-Simulated Spectral Performances

The three CALLUM versions under test, that is, CALLUM 1, CALLUM 1lin, and CALLUM 2, have been modeled in a circuit simulator (spectre/spectreRF) at baseband. The schematic views were built at the block level using AHDL as description language. From the resulting simulations, the relative ranking of the different implementations, in terms of the amount of loop gain needed to correctly process a signal with a given modulation, and of the maximum acceptable loop delay, is presented in the next sections.

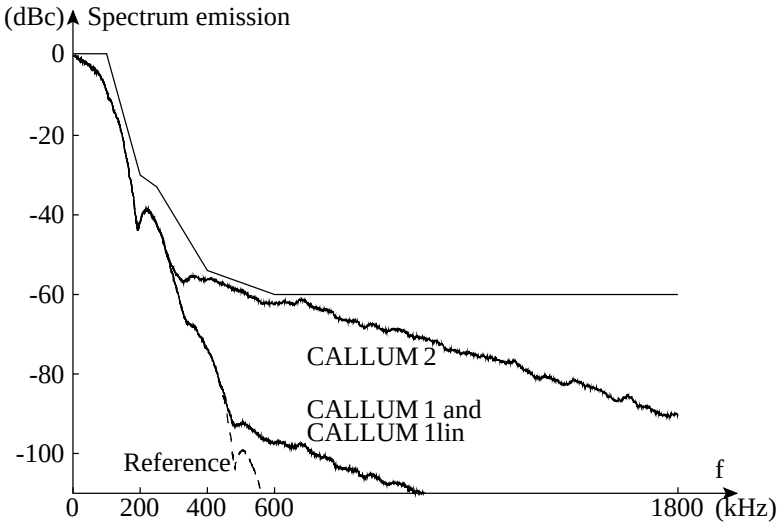

Figure 9: The spectrum emission mask for mobile handset equipment for EDGE transmit spectrum is shown together with the normalized input and output signals from different implementations of CALLUM.

4.1 CALLUM and Spectrum Emission Mask for EDGE

The first issue investigated is how the implementation of the SCG affects the performance of the system when operating on an EDGE-modulated signal. The loop gain plots in the IQ-plane (Figs. 6-8) provide the information to adjust the peak magnitude of the loop gain for each CALLUM version to a constant value. This allows easy comparison of the spectral properties for each implementation. The simulated spectra are based on a random data sequence containing 33 ksymbols, using a $3\pi/8$-shifted 8PSK modulation and filtering according to the EDGE standard. The maximum peak amplitude of the input signal was chosen to be 98% of its valid input range ($2r_{\max}$), in order to avoid the problem of having a zero at the denominator of the control equations in for example CALLUM 1. The simulated spectrum is processed through a 30 kHz filter bandwidth and normalized to the power in 30 kHz of the carrier. The technical specifications covering GSM/EDGE radio transmission can be found in TS 05.05 at *3GPP* and *ETSI*[3].

The EDGE spectrum emission mask for mobile stations in the GSM 900 band is plotted in Fig. 9, together with the spectrum of transmitted signal for each CALLUM version. The input signal serves as a reference, from which it is possible to detect any spectrum degradation. The magnitude of the loop gain was adjusted in such a way that CALLUM 2 barely passed the spectrum emission mask. The gap to CALLUM 1 and

[3]See *www.3gpp.org* and *www.etsi.org* for the complete standard specifications.

CALLUM 1lin is significant, and it can be concluded that dropping the denominator term in the control equations for CALLUM 2 has severe effects on the spectral properties. The low loop gain region for low and medium signal amplitudes relative to $r_{\max}$ (Fig. 8) has really a strong impact on the performance of CALLUM 2. The spectral performances of CALLUM 1 and CALLUM 1lin, on the other hand, are much better as a result of a more leveled loop gain in the IQ-plane for these CALLUM version. It can be noted that CALLUM 1lin approximates the original CALLUM 1 very closely, and for the tested modulation they perform equally well.

4.2 CALLUM and Spectrum Emission Mask for W-CDMA

In section 4.1 the peak loop gain was adjusted such that CALLUM 2 just passed the spectrum emission mask test for EDGE. The same simulation setting, apart from a loop gain scaling, was used also on a W-CDMA signal. Since the chip rate of a W-CDMA signal is 3.84 Mchips/s, compared to only 270.833 ksymb/s for EDGE, a bandwidth normalization was done. This normalization removes the effect of the larger bandwidth of the W-CDMA signal, and was accomplished by increasing the loop gain by 14 times, which results in a 14 times larger loop bandwidth for this first order system transfer.

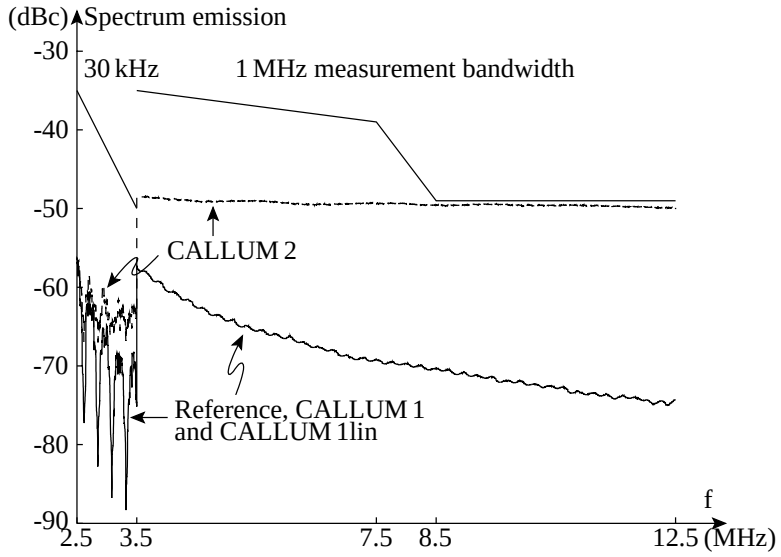

Figure 10: The spectrum emission mask for user equipment (UE) for W-CDMA transmit spectrum is shown together with the reference and outputs for different implementations of CALLUM. Only the part of the spectrum covered by the spectrum emission mask requirement is plotted.

Fig. 10 shows the simulated output spectra for the CALLUM versions together with the spectrum of the input signal. The output power spectra obtained from the simulator were filtered before they were compared to the mask according to the standard TS 25.101, in which the details are found. The out-of-channel emission is specified relative the root raised cosine (RRC) filtered mean power of the modulated carrier. Close to the carrier (2.5 – 3.5 MHz) a 30 kHz measurement bandwidth is used, while at higher offsets (3.5 – 12.5 MHz) the bandwidth becomes 1 MHz. Again, CALLUM 2 barely passes the spectrum emission mask test. A second test was to calculate the adjacent channel leakage power ratio (ACLR) for CALLUM 2. The ACLR is the ratio of the RRC filtered mean power centered on the assigned channel frequency, to the RRC filtered mean power centered on an adjacent channel frequency. In the closest neighbor channel the space to the spectrum emission mask is fairly large, and the ACLR is more than 10 dB better than the specification demands. In the second neighbor channel, however, located ± 10 MHz relative the carrier, the spectrum almost touches the mask and the ACLR is calculated to 43.7 dB, which is more or less at the lower limit of the specification (≥ 43 dB).

It can be suspected from the shape of the spectrum of CALLUM 2, due to its white-like character, that some sort of narrow spikes are superimposed on the signal in the time domain. In fact, it has been found from transient simulations that CALLUM 2 loses lock now and then, always in conjunction with small amplitudes of the input signal. Turning again to Fig. 8, it is clear that the loop gain is low for small to medium signal amplitudes, and it is well known that the bandwidth of a system having only a single loop pole is proportional to the loop gain. Thus, the loop bandwidth decreases for small signal amplitudes, with the result that the loop cannot follow the signal in some (rare) cases. After the loop is un-locked, the process of lock acquisition starts almost immediately. Lock acquisition is a nonlinear and fast process (compared to the modulated signal), which produces glitch-like disturbances in the output signal, hence the white-noise-like floor in Fig. 10. This unfavorable behavior of CALLUM 2 makes it very unsuitable for operations on a W-CDMA signal. In principle, it is true that it is always possible to increase the loop gain to get improved performance; however, the unavoidable presence of parasitic phenomena such as time delays and high frequency poles sets strong limitations to this brute-force approach, as it increases the risk of instability, due to the large loop bandwidth. Hence, it is the limited bandwidth of the CALLUM 2 system that makes its application on a W-CDMA signal difficult. Further, the modulation used in W-CDMA, the so-called hybrid phase shift keying (HPSK), also known as orthogonal-complex quadrature phase shift keying (OCQPSK), allows zero crossings, and in any CALLUM architecture zero crossings give rise to a spectrally very wide signal within the loop. We can conclude that building a linear transmitter for W-CDMA based on CALLUM 2 is a real challenge in terms of the bandwidth needed.

The performance for both CALLUM 1 and its linearized sibling is strikingly good, as it tracks the reference perfectly. Actually, it would be possible to degrade the spectral performance and gain in design robustness. In section 5.1 robustness will be inves-

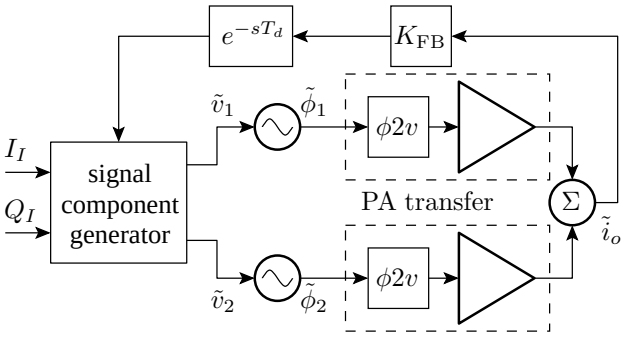

Figure 11: Model used for loop gain calculations including delay.

tigated in terms of the maximum acceptable loop delay, and in section 5.2 a bandwidth reduction technique will be applied to further enhance insensitivity to time delay.

5 Accounting for Time Delay in the Loop

A pure time delay is often used to compensate for a complexity reduction in the model of a given system. In this way, a high-order system can be represented by a low-order model, plus an appropriate time delay. A time delay always reduces the stability of a system, and limits its achievable response time. In the CALLUM architecture this delay originates from the fact that the PA output is filtered by resonant circuits (introducing a delay roughly proportional to their quality factor), and that the signal has to travel a physical distance both on chip and on the printed circuit board.

The effects of time delay on a CALLUM system are here investigated by isolating them from other sources; therefore, we will assume that the only pole present in the loop is the one due to the integrating action of the VCO. The introduction of the delay T_d in the model used for loop gain calculations is shown in Fig. 11. The loop gain results found in section 3.2 can handle this time delay simply by multiplying (13) with the term e^{-sT_d}, and will not be repeated here. Noting that the phase shift associated to the time delay is $-\omega T_d$, the maximum loop bandwidth $\omega_{0,\max}$ is given by

$$\omega_{0,\max} = \frac{90^\circ - \phi_m}{180^\circ} \cdot \frac{\pi}{T_d}, \tag{18}$$

where ϕ_m is the phase margin of the feedback signal, and the integrator accounts for a -90° phase shift. Equation (18) can be interpreted in two ways: either the maximum loop bandwidth can be determined, once phase margin and delay are known; or, for a given standard (e.g., EDGE, W-CDMA, TETRA), the maximum allowed time delay can be determined for the desired phase margin. This second interpretation, however, is unfortunately much less straightforward than might be thought, since the loop must

Table 1: Maximum acceptable loop delay to fulfill the spectrum emission mask.

	Maximum loop delay (ns)		
Standard	CALLUM 1	CALLUM 1 lin.	CALLUM 2
EDGE	11	48	1.0
W-CDMA	3.4	4.5	$<$0.1

be able to handle bandwidths much larger than the bandwidth of the original input signal. This is an undesired consequence of the non-linear conversion of an amplitude- and phase-modulated signal into two constant-amplitude signals.

5.1 Acceptable Loop Delay in CALLUM

Since the time delay in the loop is expected to be the limiting factor for the loop bandwidth in most real-life implementations, we will next examine its impact on CALLUM performances.

The maximum acceptable time delay has been simulated for each CALLUM version and for the two standards already targeted – EDGE and W-CDMA. To create the necessary conditions for the simulations, the magnitude of the loop gain was adjusted, for each CALLUM version, to a level resulting in a close fit of the spectral emission, compared to the mask for the standard under test[4]. Thereafter, a time-delay was added, and the largest acceptable time delay for which the spectrum emission requirements were still fulfilled was noted in Table 1. For small time delays the spectral emission is not noticeably degraded, but as the delay increases, the far-out power spectrum starts increasing (similar observations were reported in [15]). Further increments in time delay prevent the circuit from acquiring lock altogether.

The large difference between the maximum allowable time delay for CALLUM 1lin and CALLUM 2, respectively, is due to the fact that some fifty times lower peak loop gain is needed in CALLUM 1lin, and an equally reduced loop bandwidth, when processing an EDGE signal. The maximum loop gain is the same for CALLUM 1 and CALLUM 1lin, but surprisingly CALLUM 1 was found to have a harder time acquiring lock.

The smaller acceptable time delay for W-CDMA, compared to EDGE, is primarily due to the larger bandwidth of the W-CDMA signal, as discussed in section 4.2. It can be noted that a time delay as small as 0.1 ns was enough to prevent CALLUM 2 to acquire lock with a W-CDMA signal.

[4]It should be noted that in these simulations the loop gain varied between different CALLUM versions, while it was the same for all CALLUM versions in the simulations of section IV.

5.2 Loop Bandwidth Reduction

From (18) it can be concluded that the acceptable time delay can be increased if the loop bandwidth is reduced. Bandwidth reduction can be achieved in a brute-force approach by decreasing the loop gain, but this will deteriorate the spectral performances in an unacceptable way. A much better solution is to reduce the bandwidth while keeping the low-frequency loop gain unaffected. One such type of frequency compensation, often used in PLL design, is known as lag-lead compensation. It consists in the introduction of a pole-zero pair in the loop transfer function, which is easy to implement with a simple RC network [16], see Fig. 12. As a rule-of-thumb, the zero

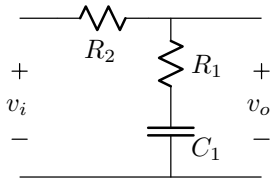

Figure 12: Passive implementation of lag-lead loop filter.

is placed at approximately one tenth of the loop bandwidth to be achieved; in this way, a robust stability is ensured. A second parameter is the distance between the zero and the pole in the lag-lead filter, here denoted as $E = n_{\mathrm{ll}}/p_{\mathrm{ll}}$ ($E > 1$). The loop bandwidth is reduced by E to a first-order approximation, and an approximately E times larger delay can be accepted, compared to the uncompensated case. It is therefore clear that the factor E should be maximized (in the limit of $E \to \infty$, the lag-lead network can be based on an integrator), while making sure that the zero keeps the right distance from the bandwidth edge. The inevitable drawback of introducing frequency compensation is that undesired spectral emissions will increase, since the loop gain is reduced at higher frequencies and more energy in the adjacent channels will appear as a consequence. To compensate for this reduction in resolution for the high frequency contents in the signal, the loop gain has to be increased, counteracting to some extent the loop bandwidth reduction of the lag-lead compensation. A few simulation iterations are usually sufficient to obtain the optimal values for pole, zero, and loop gain.

The improvements in time delay insensitivity yielded by the lag-lead compensation are discussed in the next section.

5.3 Acceptable Loop Delay in CALLUM after Compensation

We have seen that the goal of frequency compensation is to allow a larger time delay in the loop. Table 2 shows the maximum acceptable time delay for modulated signals still meeting the specifications on spectral emissions, once frequency compensation has been applied. Clearly, the improvements on the uncompensated cases (see again

Table 2: Acceptable loop delay after compensation still fulfilling the spectrum emission mask.

	Maximum loop delay (ns)		
Standard	CALLUM 1	CALLUM 1 lin.	CALLUM 2
EDGE	32	71	3.4
W-CDMA	3.4	4.5	$<$0.1

Table 1) are significant for EDGE. As an example, the settings for the CALLUM 2 compensations were the following: the lag-lead zero was placed at $-8\mathrm{Mrad/s}$, E was ten, and the loop gain was increased by a factor three, compared to the uncompensated case. These values are readily implementable in a real-life integrated design.

In the case of a first-order system, an upper limit for the maximum acceptable time delay after compensation ($T_{d,\max,\mathrm{comp}}$) can be easily derived from (18) as

$$T_{d,\max,\mathrm{comp}} \simeq E \cdot \frac{A\beta_{\mathrm{uncomp}}}{A\beta_{\mathrm{comp}}} \cdot T_{d,\max,\mathrm{uncomp}}. \tag{19}$$

As an example, in the case of CALLUM 2, (19) estimates $T_{d,\max,\mathrm{comp}}$ to 3.3 ns, while the simulated value is 3.4 ns. In the case of CALLUM 1lin, however, (19) would overestimate $T_{d,\max,\mathrm{comp}}$ by 80%. Although too optimistic in general, (19) shows the (qualitative) relation linking E, the loop gains before and after compensation, and the acceptable time delays before and after compensation.

To succeed with the compensation strategy there must be room for deterioration of the spectral emissions; if these are already very close to the spectral mask, no significant improvements can be made. Thus, the reason why large improvements were achievable for CALLUM 1 and CALLUM 1lin (and to a lesser extent for CALLUM 2) operating on EDGE was the relatively large distance between mask and actual spectral emissions for large offset frequencies; overly good spectral performances at large offset frequencies were in this case traded for higher insensitivity to time delay (as an example, compensation for CALLUM 1 is shown in Fig 13). For any of the three CALLUM versions operating on W-CDMA, on the contrary, even a small additional time delay was enough to deteriorate the signal spectrum at frequencies where it lied very close to the signal mask, thereby failing to comply with it.

It remains to note that CALLUM 1lin performed very well, in terms of maximum acceptable time delay for a certain standard. Of course, the implementation cost for the SCG in CALLUM 1lin is higher than that for CALLUM 2, but significantly lower than for CALLUM 1, without any obvious performance drawbacks.

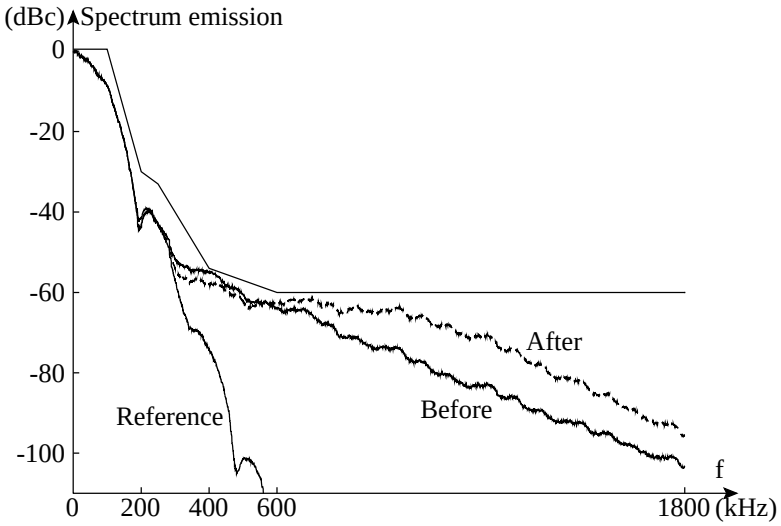

Figure 13: EDGE spectrum for CALLUM 1 before and after lag-lead compensation. $T_d = 10$ ns, $n_{\mathrm{ll}} = -15$ Mrad/s, $E = 8$, and loop gain doubled compared to the uncompensated case.

6 Conclusions

This paper has investigated the behavior of three different versions of the CALLUM transmitter architecture, which have been referred to as CALLUM 1, CALLUM 1lin, and CALLUM 2.

To perform accurate and time-efficient simulations on a large number of transmitted symbols (a prerequisite to calculate the spectrum of the modulated output signal), a baseband model for the generic CALLUM implementation was used. Further, the (non-linear) baseband model was linearized around a bias point, in order to enable the study of feedback stability through a small-signal analysis. The loop gain equations for the three CALLUM versions were explicitly derived and plotted over the IQ-plane; these data contain the relevant information needed for stability analysis and subsequent frequency compensation.

The three CALLUM versions were compared in terms of spectral emission performance, for an equal peak magnitude of the loop gain and for operation on two different standards, EDGE and W-CDMA, representative for signals modulated both in phase and in amplitude. As the loop bandwidth is strongly coupled to both the loop gain and the time delay in the loop, the peak loop gain was held constant to simplify comparison between the different CALLUM versions. Simulations showed a large performance difference in favor of CALLUM 1 and CALLUM 1lin, compared to CALLUM 2, whose strong point is however the significantly reduced design complex-

ity. The EDGE standard can be handled by all three implementations, but the larger signal bandwidth of W-CDMA prevents the use of CALLUM 2, when realistic time delays in the feedback loop are accounted for.

Loop time delay is in fact a key parameter in the design of the CALLUM feedback network, since in all practical implementations it is this delay that limits the achievable loop bandwidth. The maximum time delay compatible with spectral emission requirements was simulated for the three CALLUM versions both before and after frequency compensation. The applied compensation showed to be truly efficient on each of the CALLUM versions working on an EDGE signal, while the high data rate and the modulation nature for W-CDMA demand that either CALLUM 1 or CALLUM 1lin be employed. In particular, CALLUM 1lin appears to be the most attractive general-purpose CALLUM design, offering a very robust transmitter for a simplified CALLUM implementation.

References

[1] M. Briffa, M. Faulkner, and J. Macleod, "RF amplifier linearization using Cartesian feedback," in *First International Workshop on Mobile and Personal Communications Systems*, (Adelaide, Australia), pp. 343–348, Nov. 1992.

[2] M. Johansson and T. Mattsson, "Transmitter linearization using Cartesian feedback for linear TDMA modulation," in *Proceedings of the 41st IEEE Vehicular Technology Conference*, pp. 439–444, May 1991.

[3] P. B. Kenington, R. J. Wilkinson, and J. D. Marvill, "The design of highly linear broadband power amplifiers," *IEE Colloquium on Solid-State Power Amplifiers*, pp. 5/1–5/4, Dec. 1991.

[4] E. Westesson and L. Sundström, "A complex polynomial predistorter chip in CMOS for baseband or if linearization of RF power amplifiers," *IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 1, pp. 206–209, 1999.

[5] K. Y. Chan and A. Bateman, "Linear modulators based on RF synthesis: Realization and analysis," *IEEE Transactions on Circuits and Systems*, vol. 42, pp. 321–333, June 1995.

[6] D. C. Cox, "Linear amplification with nonlinear components," *IEEE Transactions on Communications*, vol. 22, pp. 1942–1945, Dec. 1974.

[7] F. J. Casadevall and A. Valdovinos, "Performance analysis of QAM modulations applied to the LINC transmitter," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 399–406, Nov. 1993.

[8] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popović, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 814–826, Mar. 2002.

[9] D. J. Jennings, A. Bateman, and J. P. McGeehan, "Adjacent channel power and error-vector magnitude performance of reduced complexity CALLUM systems," *IEE Proceedings on Communications*, vol. 146, pp. 297–302, Oct. 1999.

[10] R. Strandberg, P. Andreani, and L. Sundström, "Bandwidth considerations for a CALLUM transmitter architecture," *IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 4, pp. 25–28, 26–29 May 2002.

[11] E. H. Nordholt, *Design of High-Performance Negative-Feedback Amplifiers*. Delftse Uitgevers Maatschappij, 1993.

[12] D. J. Jennings and J. P. McGeehan, "A high-efficiency RF transmitter using VCO-derived synthesis: CALLUM," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, pp. 715–721, June 1999.

[13] D. J. Jennings and J. P. McGeehan, "A high-efficiency RF transmitter using VCO-derived synthesis: CALLUM," in *Proceedings of the Radio and Wireless Conference*, pp. 137–140, 1998.

[14] W. C. Lindsey and C. M. Chie, *Phase-Locked Loops*. New York: IEEE Press, 1986.

[15] M. A. Briffa and M. Faulkner, "Stability analysis of cartesian feedback linearisation for amplifiers with small non-linearities," *IEE Proceedings - Communications*, vol. 143, no. 4, pp. 212–218, 1996.

[16] R. E. Best, *Phase-Lock Loops*. McGraw-Hill, second ed., 1993.

Paper IV

Paper IV

Implementation of the Signal Component Generator of a CALLUM 2 Transmitter Architecture in CMOS Technology

Abstract

This article presents an analog implementation of the signal component generator (SCG) of the CALLUM 2 linear transmitter architecture. The proposed SCG is suited for integration in a standard 0.35 μm CMOS process, and has from simulations proven to be adequate when operating on an EDGE modulated baseband signal with a data rate of 270.833 ksymb/s. The total current consumption of the SCG is 2.0 mA from a 3.3 V supply. A variable-gain amplifier (VGA) with common-mode (CM) control is presented, and the VGA is inserted in between the SCG and the voltage-controlled oscillator (VCO) to adjust the loop gain, which has strong influence on the stability and spectral performance of the linear transmitter architecture.

Based on: Roland Strandberg, Pietro Andreani, and Lars Sundström, "Implementation of the Signal Component Generator of a CALLUM 2 Transmitter Architecture in CMOS Technology," *Norchip Conference*, Nov. 2004.

1 Introduction

Methods of linear amplification, which allow the use of nonlinear but power efficient amplifiers, divide the information into a phase and an amplitude component (e.g., envelope elimination and restoration (EER)), or split the input signal into two constant-envelope signals such the vector addition is equal to an amplified replica of the input signal (e.g., LINC, CALLUM). The generation of these constant-envelope vectors in the LINC case requires some rather complex operations on the input signal, and the functions are implemented in a block known as the signal component separator. The SCG in the CALLUM architecture generates the control voltages to the VCOs, which in turn generate these constant-envelope vectors that are amplified by the power amplifiers (PAs).

A discrete implementation of a CALLUM 2 system has been reported in [1]. In this article we will show a fully integrated solution of the SCG in a 0.35 μm CMOS process. One of the most important advantages with integration is less loop time delay, which allows larger bandwidth signals and better spectral performance of the feedback system. To maximize the system performance, control of the loop gain via a VGA with CM control is important and its implementation is also presented.

2 CALLUM System Description

The Combined Analog Locked Loop Universal Modulator (CALLUM) is an intrinsically linear architecture, having a theoretical efficiency of 100 % for all output levels. CALLUM is based on a LINC-like core placed in a feedback loop. The input signal of a CALLUM transmitter architecture is assumed to be both amplitude- and phase-modulated, and this signal together with the feedback signal produce the control voltages to the VCOs, which in turn generate these two constant-envelope phase-modulated signals that are amplified by the PAs. The amplified signals are recombined to generate the output signal, which ultimately is an amplified replica of the input signal [2]. The generic CALLUM architecture is shown in Fig. 1 at the functional block level. The output signal is sensed and fed back to the SCG, and any nonlinearity within the loop is suppressed by the amount of loop gain under the assumption of a linear feedback factor.

In this article only continuous versions of CALLUM suitable for implementation based on a Cartesian representation are considered. The matching properties of a Cartesian representation are promising since the I and Q path are identical, e.g., opposed to a polar representation [2]. All input and output signals are treated by the SCG according to the chosen set of control equations to generate the control signals sent to the VCOs. Depending on what set of control equations that is implemented in the SCG the CALLUM architecture gets different names. A derivation of the control equations for the original version of CALLUM, here to be referred to as CALLUM 1, was presented in [2]. The SCG generates the control voltages $v_1(t)$ and $v_2(t)$ for VCO_1 and

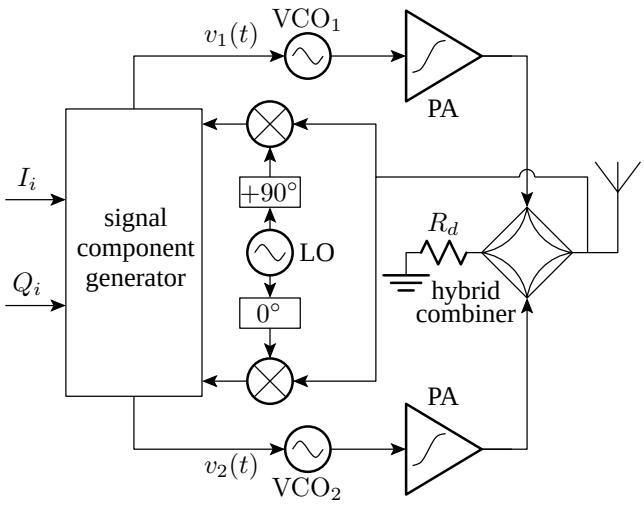

Figure 1: The generic CALLUM transmitter architecture.

VCO$_2$, respectively, from the input and the feedback signals. The simplified version of CALLUM 1 resulted in CALLUM 2, and the control equation is shown in (1).

$$v_{1,\mathrm{C2}}(t) = \frac{g_1 k_1}{K_{\mathrm{VCO}}} \left\{ \left(I_i(t) - \frac{I_o(t-\tau)}{A_{t\infty}} \right) \left(-Q_i(t) - I_i(t) \right) + \left(Q_i(t) - \frac{Q_o(t-\tau)}{A_{t\infty}} \right) \left(I_i(t) - Q_i(t) \right) \right\} \quad (1)$$

The subscripts i and o refer to the input and output signals, respectively; K_{VCO} is the gain factor of the VCO ($\mathrm{rps/V}$), $A_{t\infty}$ is the asymptotic closed-loop transfer from input to output, g_1 and k_1 are gain factors, and r_i is the magnitude of the complex-valued input vector. Depending on what is easiest to comply with when implementing the system, either r_i should be chosen such that the maximum input signal amplitude generates the maximum output signal for best power efficiency, or the value for $r_{\max}$ chosen as $r_{\max} = \max(r_i)/2$. Since the supply voltage of the PA determines $r_{\max}$ it is more likely that r_i is the parameter to adjust.

3 Differential Analog SCG Implementation of CALLUM 2

The input signals to the SCG are the I and Q components of both the information signal and the fed back signal. The outputs of the SCG are the control voltages to

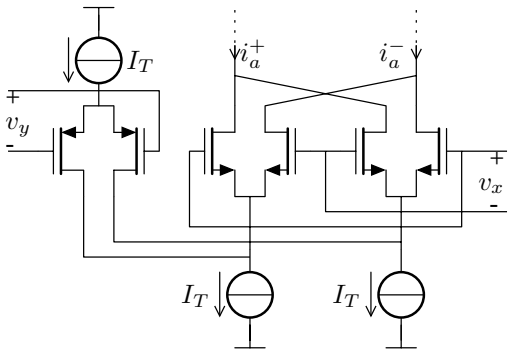

Figure 2: A folded CMOS Gilbert cell.

the VCOs, and depending on what set of control equations that are chosen, the SCG implementation will differ; or in case of a digital implementation, the correct set of equations has to be programmed. The control equations for CALLUM 2 is a compromise between circuit design complexity and linearity performance, and some aspects of the performance differences of the various CALLUM derivatives are reported in [3]. However, in this section we will show an analog CMOS transistor implementation of the control equation (1) of CALLUM 2.

The equation (1) is built from three basic functions, i.e., summation, subtraction, and multiplication. All these functions are fairly easy to implement. In radio frequency (RF) design it is very common to use differential (balanced) signals, and the main reason for many designers is the reduced sensitivity to CM induced noise. However, the use of differential signals has more advantages; since no information is in the CM component the biasing becomes easier and fewer coupling capacitors are needed. Also by utilizing the complementary signal of the differential signal pair reduces the need for both summation and subtraction to only one of them. In this design the subtraction is excluded.

The multiplication is accomplished by a four-quadrant CMOS analog multiplier, based on the folded six-transistor CMOS Gilbert cell in Fig. 2. The circuit originates from Gilbert's six-transistor cell, and handles a wide range of input voltages [4]. The valid voltage range of the input signals, v_x and v_y, is further extended by using a folded structure. In many low-voltage designs several levels of stacked transistors are inconsistent with the operating point of the transistors. The tail current, I_T, is part of the bias of the folded multiplier cell. The difference between the two drain currents, i_a^+ and i_a^-, is proportional to the product of v_x and v_y.

The input signals to the SCG should have a DC component such all transistors are kept in the right region of operation. The nominal value of the DC component of the input signals is half the supply voltage, $V_{\mathrm{DD}}/2$. The center part of Fig. 3 corresponds to v_y in Fig. 2, and is the differential implementation of $(I_i - I_o/A_{t\infty})$ in (1). The

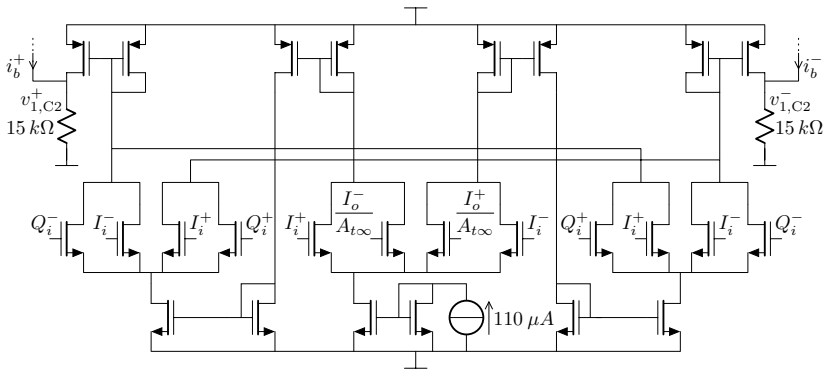

Figure 3: One fourth of the signal component generator of CALLUM 2.

error signal of the in-phase component is generated from this part of the equation, and since the error signal is very small when the high performance architecture is in transmit mode, any other source of uncertainty has to be low. By using large sized transistors not only the matching error is reduced, also the flicker noise is decreased. By rewriting the part $(-Q_i - I_i)$ of (1) as $((-Q_i) + (-I_i))$ the corresponding sum is implemented in analogy to v_x in Fig. 2. The sign of the signal is just a matter of definition of the ports what is regarded as positive and negative sign of the applied differential input signal.

The second part $(Q_i - Q_o/A_{t\infty})(I_i - Q_i)$ of (1) is implemented by another copy of the circuit in Fig. 3, of course with other input signals. The output signals of this second block are the currents indicated by i_b^+ and i_b^- in Fig. 3. The output signals of the sub-circuits are summed in a resistor to generate the output signal $v_{1,\mathrm{C2}}$ of the SCG. The control signal $v_{2,\mathrm{C2}}$ is generated in analogy to $v_{1,\mathrm{C2}}$.

3.1 Simulation of the SCG with EDGE input signal

The transistor implementation of the SCG for CALLUM 2 described above has been simulated and compared with the ideal mathematical realization. The control equation (1) was implemented in Analog Hardware Description Language (AHDL), which is handled by many circuit simulators (Spectre in our case). The input test signal to the SCG was a baseband EDGE (Enhanced Data rates for GSM Evolution) modulated signal with 270.833 ksymb/s. For a working transmitter architecture the fed back signal to the SCG is substantially and slightly attenuated copy of the input signal. The differential output from the AHDL block (ideal case) was printed in the same plot as the output, $v_{1,\mathrm{C2}}$, from the CMOS transistor implementation of the SCG, see Fig. 4.

The analog implementation of the control equation for CALLUM 2 works well,

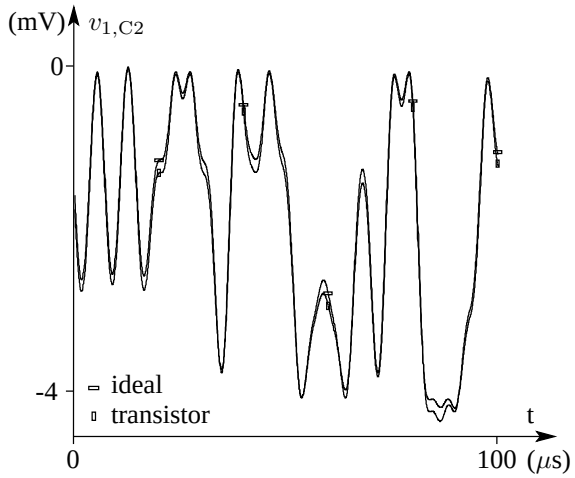

Figure 4: Ideal SCG output versus transistor implementation for an EDGE modulated signal.

but a perfect match between the mathematical function and its implementation is hampered by the nonlinear gain of the MOS transistor, and also the limited bandwidth of the SCG implementation is in some rare cases visible as slewing.

4 VGA with CM Control

The equation (1) is now realized apart from the gain term, $g_1 k_1 / K_{\mathrm{VCO}}$. This term controls the loop gain and thereby has strong influence on the stability and spectral performance of the linear transmitter architecture. CALLUM is governed by a set of nonlinear control equations, from which basic features like loop gain, bandwidth, and stability are difficult to relate to, even though they are meaningful in each static bias point.

As described in [2] the VCOs should be synchronized to the reference frequency to maximize the spectral performance of the system. The static frequency offset can be corrected and an efficient way of synchronizing two signals is by using a phase-locked loop (PPL). The static frequency offset corresponds to a DC correction voltage that has to be applied to the VCO. The VGA with CM control is inserted in between the SCG and the VCO. The VGA is described in [5], and shown in Fig. 5. The differential input, (v_i^+, v_i^-), and output, (v_o^+, v_o^-), are complemented by the CM input, V_{CM}. The circuit will make sure that the CM level of the output signal equals V_{CM}.

The bias current, I_{VGA}, sets the gain, and it can easily be varied in the range 3 – 25 V/V with a single-ended parallel RC-load ($R_\ell = 100\,\mathrm{k\Omega}$, $C_\ell = 150\,\mathrm{fF}$). The main problem with low gain is that the -3 dB bandwidth also decreases significantly. The

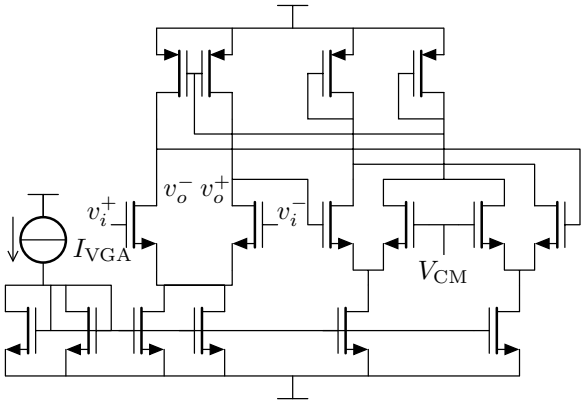

Figure 5: VGA with CM control.

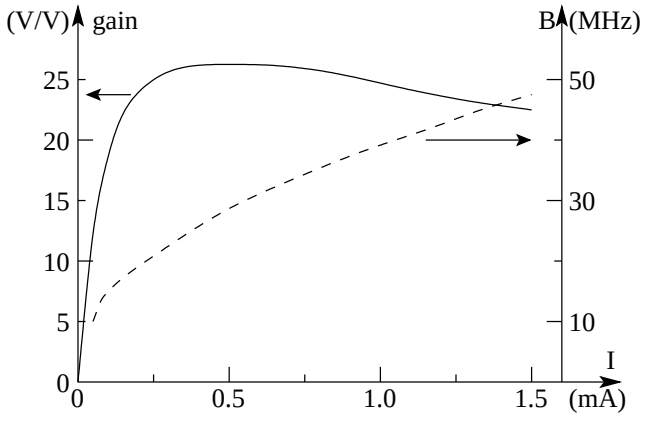

Figure 6: VGA gain transfer and bandwidth versus current.

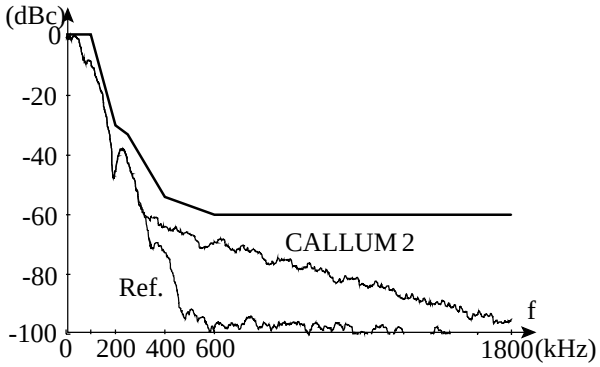

Figure 7: The simulated spectrum from CALLUM 2 together with the reference (input) signal and the spectrum emission mask for EDGE.

gain and bandwidth versus total current consumption of the VGA implemented in a 0.35 μm CMOS process is shown in Fig. 6.

5 Simulation of CALLUM 2

The SCG and VGA presented in this article were simulated at transistor level in a baseband modeled CALLUM 2 architecture. The model contains a mixture of blocks at transistor level and AHDL, and is described further in [6]. The simulated spectra are based on a random data sequence containing 33 ksymbols, using a $3\pi/8$-shifted 8PSK modulation and filtering according to the EDGE standard. The simulated spectrum was processed according to the technical specifications covering GSM/EDGE radio transmission, which can be found in TS 05.05 at *3GPP*. The EDGE spectrum emission mask for mobile stations in the GSM 900 band is plotted in Fig. 7 together with the simulated spectrum of the transmitted signal for CALLUM 2 and a reference, from which it is possible to detect any spectrum degradation. Both the SCG and VGA implementation have proven to be adequate for an EDGE modulated signal.

6 Conclusions

In this article we have proposed an implementation of a fully differential analog SCG for the CALLUM 2 linear transmitter architecture. The proposed SCG is suited for integration in a standard CMOS process, and has proven to be proper when operating on an EDGE modulated baseband signal. The total current consumption of the SCG is only 2.0 mA from a 3.3 V supply when implemented in a 0.35 μm CMOS process.

A VGA has been presented, and the need for a VGA with CM control is twofold.

First, the variable-gain is to adjust the loop gain, which has strong influence on the stability and spectral performance of the architecture. Second, the CM feature is used to correct the static frequency offset between the VCO and the reference frequency. This is accomplished by a PLL operating in between transmission of the bursts of data.

References

[1] D. J. Jennings and J. P. McGeehan, "A high-efficiency RF transmitter using VCO-derived synthesis: CALLUM," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, pp. 715–721, June 1999.

[2] R. Strandberg, P. Andreani, and L. Sundström, "Bandwidth considerations for a CALLUM transmitter architecture," *IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 4, pp. 25–28, 26–29 May 2002.

[3] D. J. Jennings, A. Bateman, and J. P. McGeehan, "Adjacent channel power and error-vector magnitude performance of reduced complexity CALLUM systems," *IEE Proceedings on Communications*, vol. 146, pp. 297–302, Oct. 1999.

[4] J. N. Babanezhad and G. C. Temes, "A 20-v four-quadrant CMOS analog multiplier," *IEEE Journal of Solid-State Circuits*, vol. 20, pp. 1158–1168, Dec. 1985.

[5] D. A. Johns and K. Martin, *Analog Integrated Circuit Design*. Wiley, 1997.

[6] R. Strandberg, P. Andreani, and L. Sundström, "Spectrum emission considerations for baseband-modeled CALLUM architectures," *IEEE Transactions on Microwave Theory and Techniques*, Accepted for publication.

Paper V

Paper V

Analytical Expression of the Efficiency of Phantom Zero Compensation Applied on Negative-Feedback Amplifiers

Abstract

This article reviews the phantom zero compensation technique applied on negative-feedback (NFB) amplifiers, followed by an analysis of the important efficiency parameter, δ, of the implemented phantom zero. The effect of the efficiency on the root locus is presented, and it has been found that $\delta \geq 7$ will give near ideal behavior of the applied phantom zero. A reduced small signal model of the amplifier is presented along with a modified feedback factor, β_{ph}, which yields a simple analytical expression of the efficiency. The theory is independent of technology (BJT, FET, etc.), and exemplified on a two-stage BJT NFB amplifier.

Based on: Roland Strandberg and Johan Piper, "Analytical Expression of the Efficiency of Phantom Zero Compensation Applied on Negative-Feedback Amplifiers," *Norchip Conference*, Nov. 2004.

1 Introduction

One of the most basic and important function in any radio communication system (or processing of information in general) is amplification. Associated with the implementation of the amplifier numerous parasitics appear, e.g., power consumption, noise, and distortion. Back in the nineteen twenties Black struggled to improve the Bell System's new open-wire telephone system. For the system to operate properly the major obstacle was the distortion, and after persistent search the idea of NFB amplifier came to him in a flash [1]. He realized that the nonlinearity of such amplifiers would be suppressed by the amount of loop gain applied. Ever since has the NFB technique been applied to build high-performance amplifiers, which is designed from proper selection of active elements (e.g., BJTs and/or FETs), frequency compensation, and biasing [2].

In this article we focus solely on the frequency compensation known as the phantom zero technique. The next section is a short review of the most important characteristics of phantom zero compensation. It shows, in an unified way, how and where to implement the phantom zero in a NFB amplifier. Section 3 is an investigation of the efficiency of the phantom zero. The effect of the efficiency is shown in the root locus, whereas the modeling and calculations are in section 4. After an example the article is closed by the conclusions in section 5.

2 Phantom Zero Compensation

Frequency compensation is one of the most important design step when building any feedback system, e.g., phase-locked loop, $\Sigma\Delta$-converter, amplifier. At the same time, frequency compensation is one of the hardest tasks to solve when implementing a NFB amplifier. The scope of this article is limited to the most attractive compensation technique, namely the phantom zero. Some basic knowledge about phantom zero compensation is assumed, and an introduction to the subject and terminology is found in [3].

A phantom zero is visible in the loop but not in the system transfer. That is why it is called a phantom zero and not only a loop zero. However, the phantom zero alters the shape of the root locus, and makes the system pole sum smaller [4]. For a zero should become a phantom zero it must be outside the nullor implementation,

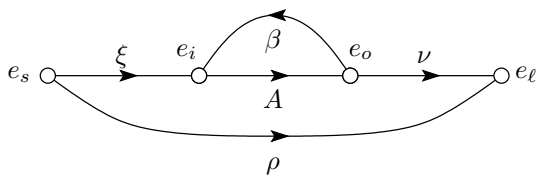

Figure 1: Signal-flow graph describing the asymptotic gain model.

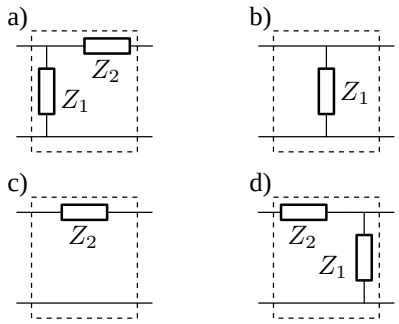

Figure 2: Feedback network for a) voltage, b) transadmittance, c) transimpedance, d) current amplifier.

actually in the feedback factor, β, where the amplifier is modeled using the asymptotic gain model [3]. The model is shown in Fig. 1, where the branches represent the four network parameters ξ, ν, β, and ρ. The parameters correspond to the effects of input and output loading, feedback, and direct feed-through, respectively. The branch A is called the reference variable and reflects the nullor implementation. e_s (e_ℓ) denotes the electrical input (output) signal, represented either in the current or voltage domain.

The zero must be implemented in the feedback factor, β, which depends on the feedback network, the input, and the output. A zero appears when the feedback factor increases at high frequencies, which simply means that more of the signal at the output of the nullor implementation should be fed back to its input at high frequencies. In Fig. 2 are all feedback networks for single loop amplifiers shown, and a phantom zero appears when $|Z_1|$ increases or when $|Z_2|$ decreases at high frequencies. In the case of resistive voltage feedback network an inductor in series with Z_1 or a capacitor in parallel with Z_2 will give rise to a first order phantom zero.

If the implementation of a phantom zero in the feedback network failed, another possibility is either at the input or at the output of the amplifier. A phantom zero will appear when the source or the load impedance is made more ideal at high frequencies. By connecting a component in such way that the source impedance for a shunt (series) input can be modeled by an open (short) circuit will give a phantom zero, see Fig. 3a) and 3b). The ideal series (shunt) output is a short (open) circuit and this can be accomplished with a capacitor in parallel (an inductor in series) with the load, see Fig. 3c) and 3d). The available signals indicated in Fig. 3 are i_f, v_f, i_o, and v_o, respectively, and to increase the feedback factor, β, the denoted parameters i_i, v_i, i_ℓ, and v_ℓ should be maximized at high frequencies to introduce a phantom zero. More examples and further details are found in [5].

To succeed with the phantom zero compensation $m-1$ zeros are needed, where m is the number of loop poles. It is clear that a first order system does not need any compensation and a third order system needs two zeros (possibly complex). For a sec-

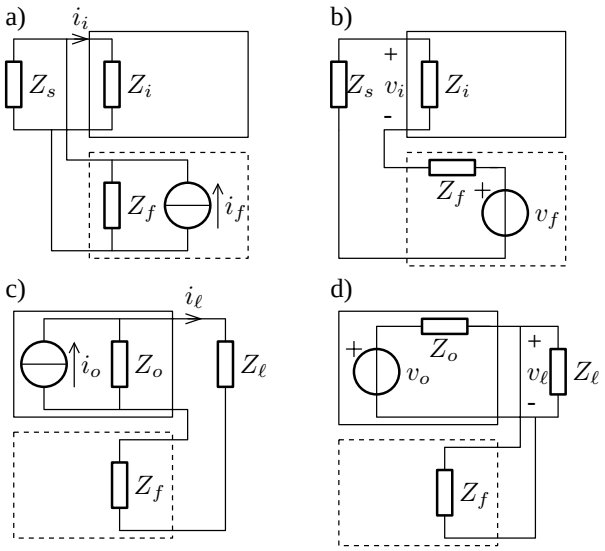

Figure 3: Top: a) shunt input, b) series input. Bottom: c) series output, d) shunt output.

ond order system there is a simple expression describing the position of the phantom zero, $n_{\rm ph}$, according to [2],

$$n_{\rm ph} = \frac{-\omega_0^2}{-\left(p_1^{'} + p_2^{'}\right) + (p_1 + p_2)}, \tag{1}$$

where ω_0 is the closed loop bandwidth, which can be estimated from the loop gain poles (LP) product given by, LP $= |(1 - A\beta)\prod_{k=1}^{m} p_k|) = \omega_0^m$. Here m denotes the number of dominant loop poles. The LP product is actually a generalization of the gain bandwidth (GB) product [3]. p_1 and p_2 are the loop poles, and $p_1^{'}$ and $p_2^{'}$ are the system poles. The phantom zero compensation makes the system pole sum smaller, and the relation between system pole sum and the desired system frequency characteristics is given in Table 1. Butterworth (or maximum flat magnitude (MFM) transfer) gives the largest bandwidth without amplitude peaking, Bessel is known for its linear phase characteristic, and for robust design a real double pole is often chosen.

3 Efficiency of the Phantom Zero

Phantom zero compensation is easy and straight forward to implement, and considered as the most favorable compensation technique available. A very important aspect is

Table 1: Relation between system pole sum and the desired system frequency characteristics.

Characteristics	$p_1^{'} + p_2^{'}$
Butterworth, MFM	$-\sqrt{2}\omega_0$
Bessel	$-\sqrt{3}\omega_0$
Real double pole	$-2\omega_0$

the efficiency of the phantom zero, but this is very seldom seen or discussed in the open literature. Let us first define the efficiency, δ, of the phantom zero as,

$$\delta \equiv \frac{p_{\rm ph}}{n_{\rm ph}}, \tag{2}$$

where $p_{\rm ph}$ is the parasitic pole that appears when a phantom zero, $n_{\rm ph}$, is introduced. There is a saying *"no zero without a pole"*, which seems to make sense, since every component known to date has limited bandwidth. This unwanted pole limits the efficiency of the applied phantom zero compensation. The effect of the efficiency is captured by any circuit simulator, but only gives vague guidelines to the designer, e.g., how to improve the efficiency. A simple analytical expression describing the efficiency will be presented in section 4, and the result might be of interest for the designer.

From the pole-zero map it is straight forward to calculate the efficiency of the phantom zero (if the associated parasitic pole can be identified) using (2). Unfortunately can most circuit simulators neither produce the pole-zero map nor the root locus. In Fig. 4 is a second order system compensated with a zero. The loop poles are located at $p_1 = -1\,\mathrm{krad/s}$, $p_2 = -2\,\mathrm{krad/s}$, and the zero, $n_{\rm ph} = -15\,\mathrm{krad/s}$. For the closed loop system to have MFM transfer one of the complex system poles should be located at the intersection between the half-circle indicating the MFM bandwidth, ω_0, and the 45° line, meaning the system poles should be at $p_{1,2}^{'} = -\omega_0(1 \pm j)/\sqrt{2}$. The zero is shown in Fig. 4 for the sake of clarity even though the zero must be a phantom zero for MFM characteristics, and the position of the parasitic pole, $p_{\rm ph}$, depends on the efficiency and is only indicated in the figure. The numbers in Fig. 4 indicate the efficiency of the zero, and for $\delta = 1$ is the pole $p_{\rm ph}$ right on top of $n_{\rm ph}$ and the root locus is basically a straight line parallel with the $j\omega$-axis that intersects the real axis at $\sigma = -1.5\,\mathrm{krad/s}$. When the efficiency increases the root locus bends more and more into the left-half-plane, and for $\delta = \infty$ the familiar circle is achieved and the MFM bandwidth, $\omega_0 = 19\,\mathrm{krad/s}$, can be fulfilled.

As mentioned before, frequency compensation is about controlling the location of the system poles to their desired position, and from Fig. 4 it is clear that only with a phantom zero with infinite efficiency can the estimated MFM bandwidth, ω_0, be achieved. For non-ideal efficiency the system will have a damping factor $\zeta < 1/\sqrt{2}$, and as a consequence will peaking appear in the amplitude plot of the closed loop system, see Fig. 5.

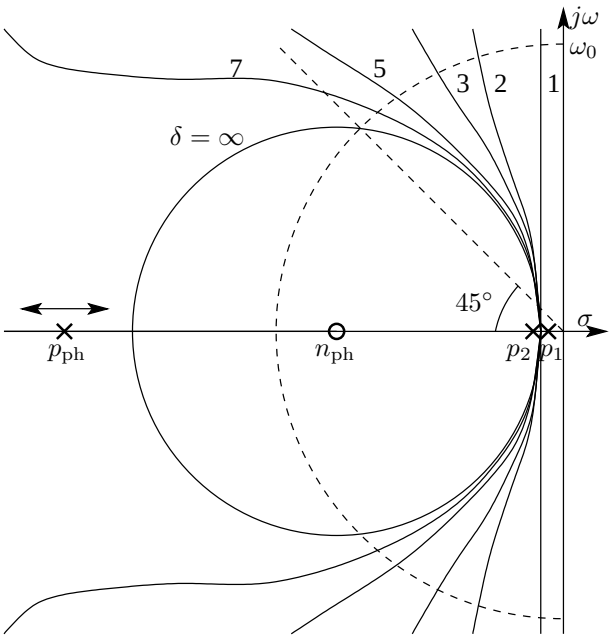

Figure 4: Root locus for the second order system compensated with a phantom zero versus efficiency.

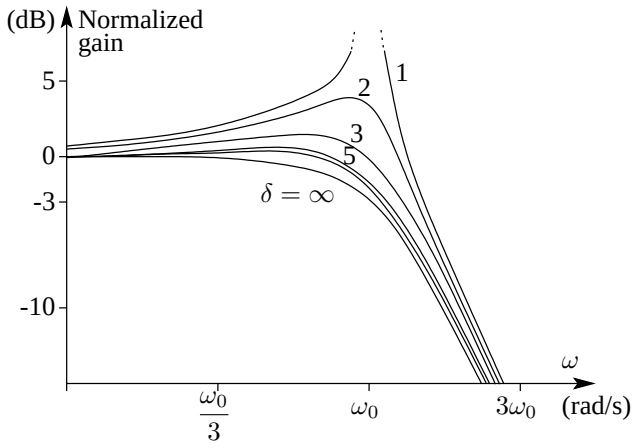

Figure 5: The amplitude plot of the second order system compensated with a phantom zero versus efficiency.

Table 2: The efficiency of the phantom zero and its impact on the circuit, where *high* is most desired.

Efficiency				Impact on root locus
1	$\leq$	δ	< 3	low
3	$\leq$	δ	< 7	moderate
7	$\leq$	δ		high

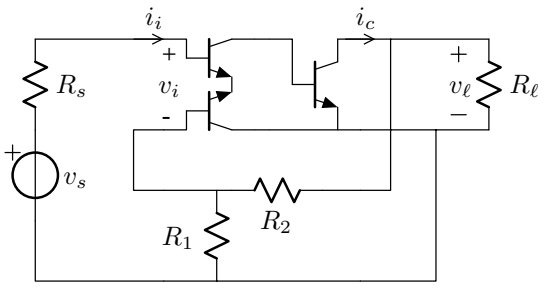

Figure 6: A two-stage voltage amplifier implemented with the ASCE-CE combination together with resistive feedback network.

We see that a good approximation of the ideal efficiency case is achieved for $\delta \geq 7$ and acceptable compensation is achieved for efficiency values as low as 5. The Fig. 4 is an example when the phantom zero is well separated from the loop poles (a factor of ten or more). As long as the phantom zero is well separated from the loop poles the variation of the root locus is very small. The guideline for the impact on the root locus versus efficiency is based upon the observation in the root locus, and given in Table 2.

4 Approximation of the Efficiency

The feedback factor, β, for the amplifier in Fig. 6 will only contain one loop pole, originating from the input capacitance, c_i, of the anti-series common-emitter (ASCE) stage. When the phantom zero is implemented the zero and the parasitic pole will also appear in β. Let us write down the feedback factor as,

$$\beta(s) = \left.\frac{v_i}{i_c}\right|_{Q_s=0} = \left.\frac{z_i i_i}{i_c}\right|_{Q_s=0} = \left.\frac{r_i}{1+sr_ic_i}\frac{i_i}{i_c}\right|_{Q_s=0}, \tag{3}$$

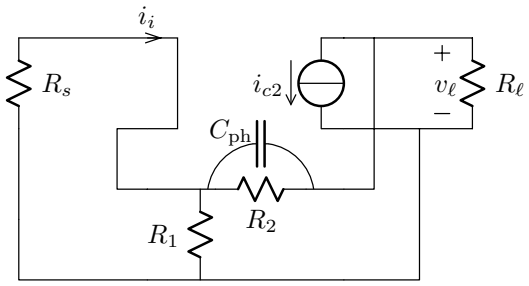

Figure 7: Small signal schematic for calculation of the modified feedback factor, $\beta_{\rm ph}$.

where $Q_s = 0$ corresponds to the source signal set to zero, v_i is the input signal to the reference variable A, and i_c the output signal of controlled source[1]. z_i is the input impedance of the first amplifying stage, and is a parallel coupling of r_i and c_i. It can be mentioned that all information about the phantom zero is in the feedback factor, $\beta(s)$, but the expression tends to be messy if no approximations are made.

Since we assume the phantom zero to be well separated in frequency from the loop poles, any capacitance giving rise to a loop pole can be modeled as a short circuit, i.e., equivalent to make $c_i \to \infty$ in (3). The remaining part of β now contains only the information about the phantom zero, and we define $\beta_{\rm ph}$ as the current-to-current feedback factor according to,

$$\beta_{\rm ph} = \left.\frac{i_i}{i_c}\right|_{Q_s=0,\, c_i\to\infty} = K(0)\frac{1-\dfrac{s}{n_{\rm ph}}}{1-\dfrac{s}{p_{\rm ph}}}, \tag{4}$$

where $K(0)$ is a constant gain.

It is desirable to be able to calculate the efficiency of the phantom zero, preferably by simple hand-calculations, since all possible implementations will not have the same efficiency, and thereby not the same impact on the root locus. In Fig. 7 is a phantom zero implemented in the feedback network, the input signal is set to zero, and the input capacitor is modeled as a short circuit. Calculation of the modified feedback factor for this example yields,

$$\beta_{\rm ph} = \frac{R_\ell R_1(1+sR_2C_{\rm ph})}{(R_\ell+R_2)(R_1+R_s)+R_1R_s+sR_2C_{\rm ph}\left(R_\ell(R_1+R_s)+R_1R_s\right)}. \tag{5}$$

From $\beta_{\rm ph}$ it is simple to calculate the efficiency, which is given by,

$$\delta = \frac{(R_\ell+R_2)(R_1+R_s)+R_1R_s}{R_\ell(R_1+R_s)+R_1R_s}. \tag{6}$$

[1]The controlled source is here modeled as an independent source as we calculate the feedback factor.

In the case of ideal signal source, $R_s = 0$, the expression for the efficiency reduces significantly, $\delta = (R_\ell + R_2)/R_\ell$, and for the compensation to be successful should $R_2 \gg R_\ell$, according to the guidelines in Table 2. This demand can sometimes be met as the designer is free to select the impedance level of the dimensionless feedback network. However, both the noise and distortion performance of the amplifier depend on the impedance level of the feedback network. Another important observation is that the phantom zero compensation will fail if both the source and the load take their ideal values ($R_s \to 0,\ R_\ell \to \infty$).

5 Conclusions

This article has focused on the most attractive compensation technique, i.e., the phantom zero technique, when operating on NFB amplifiers that should achieve MFM transfer. A strength with phantom zero compensation is its straight forward implementation, which explains part of its wide usage. We have defined an important design parameter, namely the efficiency of the phantom zero and shown its effect on the root locus. With support of the root locus has a simple guideline been presented, predicting the level of success of the applied compensation. A simple analytical expression for the efficiency is the best source of information to help the designer where to implement the phantom zero, i.e., in the feedback network, or at the input or the output of the amplifier. A simplified small signal model along with a modified feedback factor, $\beta_{\rm ph}$, has also been presented and the application of the theory has been exemplified on a two-stage BJT NFB amplifier. From the expression of the efficiency it can be concluded that the degree of success of the phantom zero compensation depends on the impedance of the load and the source, relative to their ideal values.

References

[1] H. S. Black, "Stabilized feedback amplifiers," *Bell Systems Technical Journal*, vol. 13, pp. 1–18, Jan. 1934.

[2] E. H. Nordholt, *Design of High-Performance Negative-Feedback Amplifiers*. Delftse Uitgevers Maatschappij, 1993.

[3] C. J. M. Verhoeven, A. van Staveren, G. L. E. Monna, M. H. L. Kouwenhoven, and E. Yildiz, *Structured Electronic Design: Negative-Feedback Amplifiers*. Kluwer Academic Publishers, new ed., 2003.

[4] M. Lantz, *Systematic Design of Linear Feedback Amplifiers*. PhD thesis, University of Lund, Department of Electroscience, 2002.

[5] J. Davidse, *Analog Electronic Circuit Design*. Englewood Cliffs: Prentice Hall, new ed., 1991.