Die approbierte Originalversion dieser Dissertation ist an der Hauptbibliothek der Technischen Universität Wien aufgestellt (http://www.ub.tuwien.ac.at).

The approved original version of this thesis is available at the main library of the Vienna University of Technology (http://www.ub.tuwien.ac.at/englweb/).

DISSERTATION

Multi-Mode Power Amplifiers for Mobile Handsets

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

ao. Univ. Prof. Univ. Prof. Dr. Arpad L. Scholtz
E389
Institut für Nachrichtentechnik und Hochfrequenztechnik

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von

Dipl.-Ing. Bernhard Sogl
Matrikelnummer 9826336
Dreyhausenstr. 5
A-1140 Wien

Wien, im März 2010

Abstract

In today's and future mobile communication systems mobile handsets have to support various coexisting standards like GSM, EDGE and WCDMA. To reduce the complexity and cost of separate radios (baseband, transceiver and front end) in the mobile terminal, the development of a single multi-mode radio is required. Up to now multi-mode capability is only implemented for the baseband and transceiver part. The front end still consists of separate paths for each standard. This work deals with the multi-mode capability of the front end power amplifier (PA). The investigations pursue two main goals:

- Reducing the current consumption over a large dynamic output power range, while maintaining high linearity.
- Reducing the sensitivity of PA parameters (e.g. output power, linearity) to antenna impedance variations.

Several concepts are examined to improve a single PA regarding these issues. As outcome of these investigations a novel concept is proposed, which uses a double balanced PA topology. An analytical design procedure for optimization of the proposed multi-mode PA is presented. The excellent results, obtained for the implemented prototype confirm the usefulness of this procedure and the potential of the concept to reduce the number of different PAs in current front end architectures.

The main achievements of this work comprise:

- A scattering parameter based method is proposed, which permits a unified analysis of PA configurations where transistor stages are deactivated only via bias current, not being isolated by RF switches. The efficiency improvement at power back-off and requirements for symmetrical multi-port power combiner networks are derived for the first time.
- A quadrature hybrid as special case of a power combiner is investigated. It offers inherent load insensitivity and the possibility to use its internal isolated port for applying a high efficient low power (LP) path. A detailed study of the mutual influence of high power (HP) path and LP path comes to the result that an independent optimization for highest efficiency is not possible. A compromise has to be accepted. Further considerations show that the best compromise can not be calculated in a closed form. Therefore a novel numerical method is developed to systematically iterate to an optimum.

- A two stage inductively coupled quadrature hybrid, with attenuation < 1.5 dB, amplitude error < 0.75 dB and phase error $< 2.9°$ over a relative bandwidth of 22.2 % is realized in SiGe technology. The design fulfills the requirements for the multi-mode PA and is furthermore suitable for on-chip integration.

- A very time efficient load pull procedure is developed for experimental optimization of PA matching networks. In comparison to conventional load pull methods, here the use of costly and complicated deembedding structures is avoided.

- A prototype of a double balanced switched path PA is fabricated in a low cost 0.35 μm bipolar SiGe technology. At 840 MHz a peak power of 36.1 dBm with 52 % PAE was achieved for GSM operation. In back-off at 27 dBm, 16 dBm and 13 dBm efficiencies of 37 %, 18 % and 15 % for WCDMA signals were measured. For EDGE operation at 30 dBm a peak PAE of 30 % is achieved and held above 10 % over a dynamic output power range of 22 dB, by proper selection of the five amplifier modes. These performance characteristics are comparable to separate single mode PAs, each optimized for GSM/EDGE/WCDMA.

Kurzfassung

In heutigen und zukünftigen mobilen Kommunikationssystemen muss eine Vielzahl von nebeneinander bestehenden Standards, wie GSM, EDGE und WCDMA von den Mobiltelefonen unterstützt werden. Um die Komplexität und Kosten durch getrennte Funksysteme (Basisband, Sendeempfänger und Front-End) in den tragbaren Endgeräten zu reduzieren, ist die Implementierung eines einzelnen "multi-mode" Funksystems notwendig. Bisher wurde die multi-mode Tauglichkeit nur für Basisband und Sendeempfänger umgesetzt. Das Front-End besteht nach wie vor aus getrennten Pfaden, die an jeden Standard individuell angepasst sind. In dieser Arbeit wird die multi-mode Tauglichkeit des Front-End Leistungsverstärkers untersucht. Dabei werden zwei wesentliche Ziele verfolgt:

- Die Verringerung des Stromverbrauchs über einen großen Ausgangsleistungsbereich unter Aufrechterhaltung hoher Linearität.
- Die Verringerung der Empfindlichkeit von Verstärkerparametern (z.B. Ausgangsleistung, Linearität) gegenüber Änderungen der Antennenimpedanz.

In dieser Arbeit werden mehrere Konzepte untersucht, die eine Verbesserung eines einzelnen Verstärkers hinsichtlich dieser Punkte erlauben. Als Ergebnis dieser Untersuchungen wird ein neues Konzept präsentiert, das auf einer doppelt symmetrischen Topologie mit Quadratur-Hybriden basiert. Zusätzlich wird eine rechnerische Entwurfsmethode zur Optimierung des vorgeschlagenen multi-mode Verstärkers vorgestellt. Die ausgezeichneten Messergebnisse, die für den realisierten Prototypverstärker erreicht wurden, bestätigen die Brauchbarkeit der Vorgehensweise und zeigen, dass es mit diesem Konzept möglich ist, die Anzahl unterschiedlicher Leistungsverstärker in momentanen Front-End Architekturen zu reduzieren.

Die wesentlichen Leistungen dieser Arbeit beinhalten:

- Eine auf Streuparameter basierte Berechnungsmethode wird vorgeschlagen, welche eine vereinheitlichte Analyse von Verstärkerkonfigurationen ermöglicht, in denen Transistorstufen nur durch Abschalten des Ruhestroms deaktiviert werden und keine weiteren HF Schalter zur Isolation notwendig sind. Zum ersten Mal kann dadurch die Wirkungsgraderhöhung bei kleinen Leistungen und die Voraussetzungen für symmetrische Leistungskombinationsnetzwerke mit mehreren Toren allgemein abgeleitet werden.

- Der Quadratur-Hybrid als spezieller Fall eines Kombinationsnetzwerks wird weiter untersucht. Dieser bietet neben der Eigenschaft der Lastunempfindlichkeit auch die Möglichkeit, sein internes isoliertes Tor mit einem hoch effizienten Verstärkerpfad für kleine Ausgansleistungen zu verbinden. Eine detaillierte Analyse der gegenseitigen Beeinflussung der Pfade für hohe und niedrige Ausgangsleistung zeigt, dass eine unabhängige Optimierung für höchsten Wirkungsgrad nicht möglich ist. Es muss ein Kompromiss bezüglich des erreichbaren Wirkungsgrades zugelassen werden. Weitere Überlegungen zeigen, dass der best mögliche Kompromiss nicht geschlossen berechnet werden kann. Es wird daher eine neue numerische Methode entwickelt, mit deren Hilfe trotzdem systematisch in wenigen Schritten das Optimum zu erreichen ist.

- Ein zweistufiger indukiv gekoppelter Quadratur-Hybrid wurde in SiGe Technologie realisiert. Dieser besitzt eine Dämpfung von < 1.5 dB, einen Amplitudenfehler von < 0.75 dB und einen Phasenfehler von $< 2.9°$. Die Voraussetzungen für den Einsatz im multi-mode Verstärker werden dadurch erfüllt. Ausserdem kann dieser auf dem Chip zusammen mit den Transistoren integriert werden.

- Für die experiementelle Optimierung des Anpassnetzwerks für Leistungsverstärker wurde ein sehr effizientes und zeitsparendes "load pull" Verfahren entwickelt. Im Vergleich zu üblichen load pull Verfahren wird hier der Einsatz von teuren und aufwändigen Kalibrierstrukturen umgangen.

- Der Prototyp eines Leistungsverstärkers mit geschalteten Pfaden und doppelt symmetrischer Topologie mit Quadratur-Hybriden wurde in einer kostengünstigen 0.35 μm SiGe Bipolar Technologie realisiert. Dieser erreicht bei 840 MHz eine maximale Ausgangleistung von 36.1 dBm mit einem Gesamtwirkungsgrad von 52 % im GSM Betrieb. Bei reduzierten Ausgangsleistungen von 27 dBm, 16 dBm and 13 dBm werden Wirkungsgrade von 37 %, 18 % und 15 % im WCDMA Betrieb erreicht. Im EDGE Betrieb bei 30 dBm Ausgangsleistung wird ein Gesamtwirkungsgrad von 30 % erreicht, der über einen dynamischen Ausgangleistungsbereich von 22 dB über 10 % gehalten werden kann. Dies wird durch geeignetes Schalten zwischen den fünf Betriebsarten des Verstärkers erreicht. Diese Werte sind vergleichbar mit jenen von getrennten Leistungsverstärkern, die einzeln für GSM/EDGE/WCDMA optimiert sind.

Contents

List of Abbreviations

AC	Alternating Current
ACLR	Adjacent Channel Leakage Ratio
ACPR	Adjacent Channel Power Ratio
ADC	Analog Digital Converter
ADS	Advanced Design System: Agilent circuit simulation tool
AMPD	Analog Memoryless Predistortion
B7HFP	Infineon SiGe bipolar technology with $f_T = 33$ GHz
BALUN	*BAL*anced to *UN*balanced
BICMOS	Bipolar Complementary Metal Oxide Semiconductor
BJT	Bipolar Junction Transistor
BO	Back-Off
CDMA	Code Division Multiple Access
CMOS	Complementary Metal Oxide Semiconductor
CW	Constant Wave (= constant frequency)
DAC	Digital Analog Converter
DAT	Distributed Active Transformer
DC	Direct Current
DCS	Digital Cellular System
DMPD	Digital Memoryless Predistortion
DUT	Device Under Test
EDGE	Enhanced Data for GSM Evolution
EER	Envelope Elimination and Restoration
EF	Envelope Following
EM	Electromagnetic
ESD	Electrostatic Sensitive Device
ET	Envelope Tracking
EVM	Error Vector Magnitude
FB	Feedback
FDMA	Frequency Division Multiple Access
FET	Field Effect Transistor
FF	Feedforward
FOM	Figure of Merit
FR4	Flame Resistance 4
GaAs	Gallium Arsenide
GaN	Gallium Nitride
GMSK	Gaussian Minimum Shift Keying
GSM	Global System for Mobile Communication
HB	High Band
HBT	Heterojunction Bipolar Transistor
HICUM	High-Current Model

HP	High Power
IC	Integrated Circuit
IL	Insertion Loss
IMD	Intermodulation Distortion
LB	Low Band
LDMOS	Laterally Diffused MOS
LDNBL	Lightly Doped N Buried Layer
LINC	Linear Amplification using Nonlinear Components
LNA	Low Noise Amplifier
LOCOS	Local Oxidation of Silicon
LP	Low Power
LTE	Long Term Evolution
MB	Mid Band
MEMS	Micro Electro Mechanical System
MEXTRAM	Most Exquisite Transistor Model
MIM / MIS	Metal Isolator Metal / Metal Isolator Semiconductor
MMIC	Monolithic Microwave Integrated Circuits
MOS	Metal Oxide Semiconductor
MP	Medium Power
OFDMA	Orthogonal Frequency Division Multiple Access
PA	Power Amplifier
PAE	Power Added Efficiency
PCB	Printed Circuit Board
PCS	Personal Communication System/Service
pHEMT	Pseudomorphic High Electron Mobility Transistor
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
RF	Radio Frequency
RL	Return Loss
RMS	Root Mean Square
SGP	SPICE Gummel Poon model
SiGe	Silicon Germanium
SMA	SubMiniatur A : Standard RF connector up to 18 GHz
SMD	Surface Mounted Device
SiP	System in Package
SoC	System on Chip
SOI / SOS	Silicon On Insulator / Silicon On Sapphire
SPICE	Simultion Program with Integrated Circuit Emphasis
SSLA	Stage Size- and Load Adaption
TDMA	Time Division Multiple Access
UMTS	Universal Mobile Telecommunication System
VBIC	Vertical Bipolar Inter-Company model
VPNP	Vertical PNP transistor
VSWR	Voltage Standing Wave Ratio
WCDMA	Wideband-CDMA

List of Symbols

A	Area, unit: m^2
a_i	Incident wave to port i
b_i	Reflected wave from port i
C	Capacitance, unit: F
D	Diameter, unit: m
f	Frequency, unit: Hz
$f_{\max}$	Maximum oscillation frequency, unit: Hz
f_{T}	Maximum transit frequency, unit: Hz
f_{S}	Self-resonant frequency, unit: Hz
G	Gain
I	Current, unit: A
j	Imaginary unit
k	Coupling coefficient in a transformer
L	Inductance, unit: H
l	Length, unit: m
n	Number of input ports in a power combiner
N	Total number of ports in a power combiner
P	Power, unit: W
Q	Quality factor
R	Resistance, unit: Ω
r	Radius, unit: m
S_{ij}	Scattering parameters
$\underline{U}$	Unity matrix: $\underline{U} = \mathrm{diag}\,(1, \ldots, 1)$
V	Voltage, unit: V
w	Winding ratio in a transformer
X	Reactance, unit: Ω
Y	Complex admittance, unit: Ω
Z	Complex impedance, unit: Ω

$\mathrm{Re}(\cdot)$	Real part
$\mathrm{Im}(\cdot)$	Imaginary part
$\mathrm{diag}\,(\cdot)$	Diagonal matrix
$\vec{X}$	Vector X
$\hat{X}$	Peak value of X
$\overline{X}$	Average value of X
$\lvert\underline{X}\rvert$ or $\det(\underline{X})$	Determinant of $\underline{X}$
$\underline{X}^{\mathrm{T}}$	Transpose matrix
$\underline{X}^{*}$	Complex conjugate matrix
$\underline{X}^{-1}$	Inverse matrix of $\underline{\mathrm{X}}$
$\underline{X}^{*\mathrm{T}}$	Transpose conjugate matrix

α	Conduction angle
β	Phase error, unit: dB
η	Collector/Drain efficiency, unit: %
ε	Amplitude error, unit: dB
ε_r	Relative permittivity
Γ	Reflection coefficient
λ	Wavelength, unit: m
μ	Permeability, unit: Vs/Am
ω	Angular frequency, unit: Hz
ρ	Resistivity, unit: Ωm
φ	Reflection coefficient angle

Chapter 1

Introduction

1.1 Trends in cellular communication

The recent growth of mobile communication market and increased number of users has pushed demands for modulation standards with higher capacities within the regulated spectrum allocations. Multiple access techniques have developed from simple FDMA (Frequency Division Multiple Access) and TDMA (Time Division Multiple Access) systems to more elaborate CDMA (Code Division Multiple Access) and OFDMA (Orthogonal Frequency Division Multiple Access) systems. These allow a higher number of users in densely populated areas due to higher spectral efficiency and increased data rates. Along with the gained data rate improvements the demand in the wireless market has not been restricted to voice communication, but has also included new services and features such as video and data applications.

The evolution of standards, developed for cellular multiple access, can basically be divided into three generations. The first generation (1G) was based on analog frequency modulation schemes, providing voice communication for a limited number of users. The second generation (2G) comprises the first standards with digital modulation formats and TDMA/FDMA access, which provide higher network capability, higher security and better communication quality. These include GSM (Global System for Mobile Communications) [3GPP 09,a], as first transnational cellular communication standard for voice communication. Due to the constant envelope GMSK (Gaussian Minimum Shift Keying) modulation format very robust and efficient systems could be implemented. By adaption with the packet-switched 2.5G standard GPRS (General Packet Radio Service), network utilization rate could be further improved, while sharing the same network infrastructure with GSM. Higher data rates were achieved by the 2.75G extension EDGE (Enhanced Data for GSM Evolution), operating in the same frequency bands as GSM. The used modulation format 8-PSK (Phase Shift Keying) causes a varying transmit signal envelope, requiring linear systems with respect to amplitude and phase. The third generation (3G) comprises CDMA and OFDMA standards, namely UMTS (Universal Mobile Telecommunication System) / WCDMA (Wideband-CDMA) [3GPP 09,b, 3GPP 07,b] and

LTE (Long Term Evolution) [3GPP 07,a]. WCDMA was introduced as first globally standardized multimedia systems. Users occupy the same frequency spectrum, being separated by code orthogonality. The modulation format HPSK (Hybrid Phase Shift Keying) leads to a non-constant signal envelope with even higher requirements on system linearity than EDGE. The 3.5G extension HSPA (High Speed Packet Access) allows higher data rates (28 . . . 84 Mbit/s) but at the expense of higher peak to average ratios and more stringent linearity requirements. The latest 3.9G evolution LTE was introduced as an extension to existing 3G systems. It represents a cost-effective radio-access technology, which uses the same frequency allocations as WCDMA (Fig. 1.1), but supplies even time critical applications like online gaming. The spectral efficiency is improved by flexible transmission bandwidths (1.4 . . . 20 MHz), which can be adapted to the user's data rate requirements. This is gained by the OFDMA or similar SC-FDMA (Single Carrier-FDMA) technique, where a scalable number of orthogonal frequency subcarriers is used to adapt the occupied bandwidth. The use of many modulated carriers also reduces inter-symbol interference in multipath environments, as the corresponding symbol duration can be made larger than the delay spread of the transmission channel. With a temporal guard interval (cyclic extension) to each symbol the different propagation delays of the subcarriers and thus the inter symbol interference can be completely eliminated. This robustness together with diversity transceivers allows data rates of up to 75 Mbit/s in uplink and 300 Mbit/s in downlink. The used modulation formats QPSK / 16 QAM require high amplitude and phase linearity of the radio equipment as for WCDMA or HSPA. The system specifications of LTE, regarding power range, linearity, and frequency are very similar to WCDMA, why the same RF radio front end will be used.

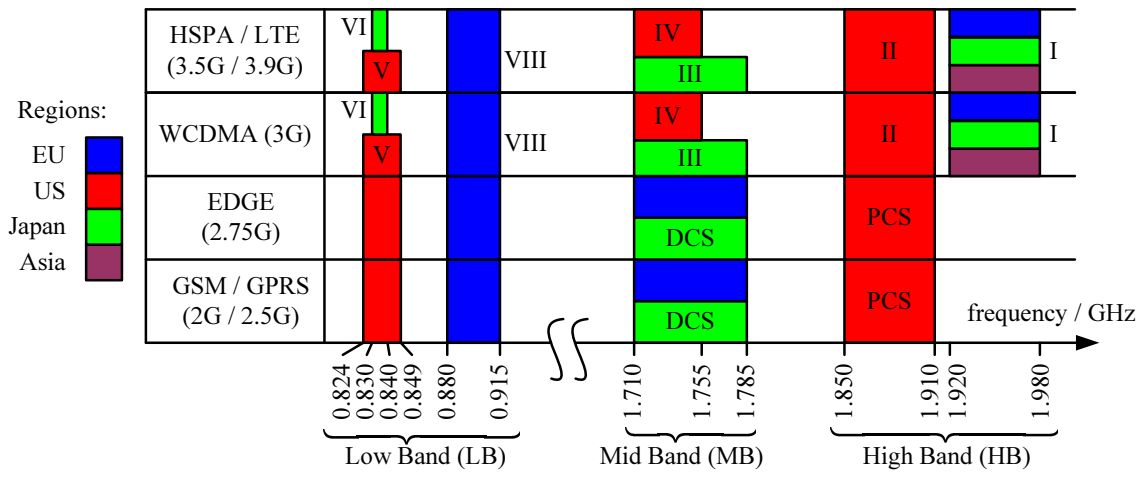


Figure 1.1: Frequency allocation of mobile communication standards. (PCS . . . Personal Communication System/Service; DCS . . . Digital Cellular System)

Each of the introduced standards offers optimized network utilization for its designated field of application (voice, data or multimedia). The coexisting infrastructure (base stations, telephone switches,...) of 2G and 3G networks currently allows par-

allel operation in areas where available. This requires handsets, with the ability to connect to multiple systems. Having a look at the geographical expansion of 2G and 3G networks it becomes apparent that currently worldwide similar frequency bands are occupied, which can be subsumed as low band (LB), mid band (MB) and high band (HB) (Fig. 1.1). Multi-band radio frequency (RF) systems, which operate in these three bands are required by the market to allow world-wide distribution of handsets. The challenge, however remains on designing multi-mode systems, which fulfill the air interface specifications (Table 1.1) with a minimum number of components for a competitive price.

Parameters	GSM		EDGE		WCDMA	LTE
	LB	MB/HB	LB	MB/HB	LB/MB/HB	LB/MB/HB
Modulation	GMSK		8-PSK		HPSK	QPSK/16QAM
Duplex mode	TDD/FDD		TDD/FDD		TDD/FDD	TDD/FDD
Multiple access	TDMA/FDMA		TDMA/FDMA		CDMA	SC-/OFDMA
Bandwidth Δf (MHz/Channel)	0.2		0.2		5	1.4...20
Max. datarate (kbit/s)	22.8		384		2000	300000
$P_{\text{out,max}}$ (dBm)	33	30	27	26	24	23
P_{out} control range (dB)	28	30	22	21	68	63
Peak-average ratio (dB)	0		3.2		3.5...7	7.5/8.5
Peak EVM (%)	–		20		17.5	17.5
Peak phase Error (degree)	20		–		–	–
ACLR (dBc)						
@±5 MHz	–		–		< -33	–
@±10 MHz	–		–		< -43	–
ACLR (dBc)						
@$\pm\frac{(\Delta f+3.84\text{ MHz})}{2}$	–		–		–	< -33
@$\pm\frac{(\Delta f+11.52\text{ MHz})}{2}$	–		–		–	< -36
Mod. Spec. (dBc)						
@±400 kHz	< -60		< -54		–	–
@±600 kHz	< -60		< -60		–	–
Noise ($\frac{\text{dBm}}{100\text{ kHz}}$)						
@20 MHz	< -79		–		–	–

Table 1.1: Transmitter air interface specifications of GSM, EDGE, WCDMA and LTE.

1.2 PAs in mobile handsets

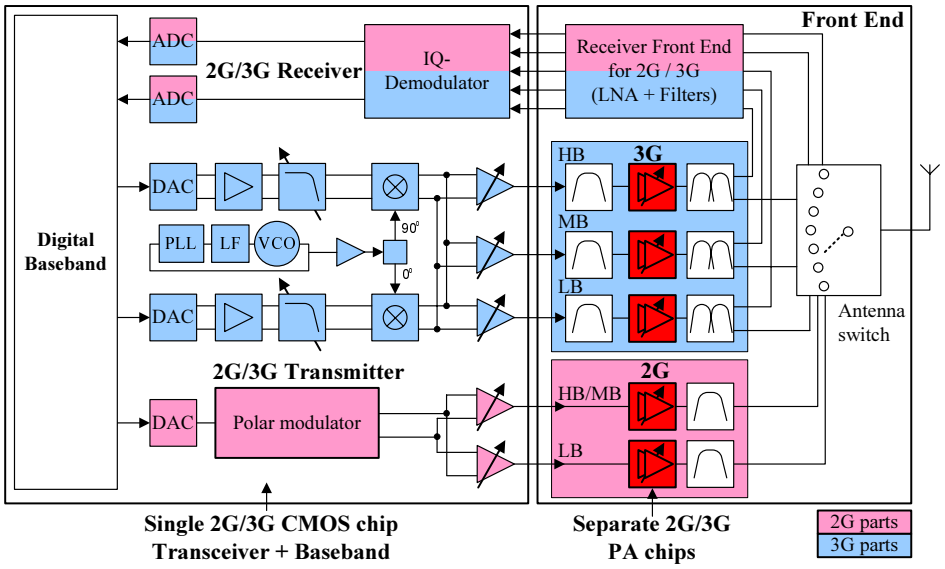


Figure 1.2: Block diagram of a typical GSM/EDGE/WCDMA cellular radio [Infineon Tec 08, Freescale Se 08].

The ability of a mobile handset to connect to multiple systems demands higher integration of functionality, currently obtained by implementation of multiple parallel transmit paths. The block-diagram of a typical multi-mode multi-band cellular radio is shown in Fig. 1.2. This includes the baseband processing, transmitter and receiver (transceiver) and the front end (PA, LNA, filters, duplexers, antenna switch). Due to innovations in design techniques the number of discrete active components has reduced over the years. The focus was laid on full integration of transceiver and baseband using low-cost Si based CMOS technology. The form factors have reduced, as illustrated in Fig. 1.3, but the PA units still consume a remarkable area. An own PA module is necessary for each standard, requiring separate control logic, transmit filters and multi-pole antenna switches. Major attempts are now the higher integration of PA circuits using single chain topologies, to enable transmission of 2G/3G signals by a single PA for each band. In addition trends emerge to replace currently used expensive GaAs technologies for PAs by Si-based technologies to further reduce cost and size by integrating the PA together with the transceiver. This requires not only to be technically feasible but also commercially reasonable to really happen. This will depend on the balance between increased development costs and savings of production costs.

The reason why a separate PA is still used for each standard is its impact on the overall performance of the handset, in particular its talk time. The PA has to amplify the signal, generated by the transmitter, to the necessary output power level with smallest possible current consumption, while fulfilling linearity, stability, robustness and noise specifications. The varying environment of the PA like supply

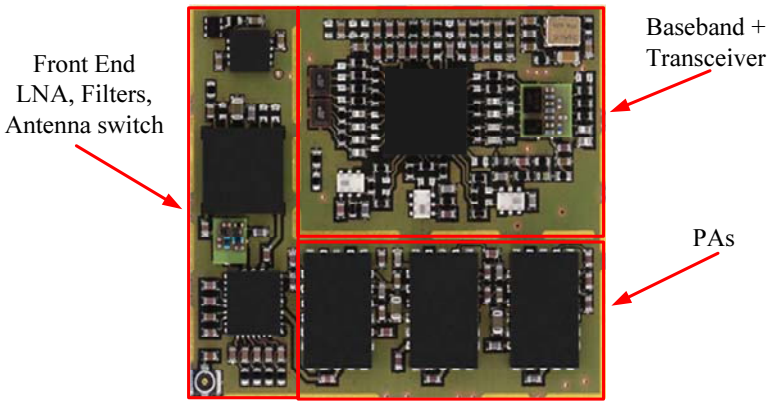


Figure 1.3: Photograph of a GSM/EDGE/WCDMA cellular phone's RF parts.

voltage changes in the range of 2.5 . . . 5.5 V and load impedance variations up to a VSWR=1:10, require a design with high margins for maintaining efficient and stable operation without reducing integrity of the transmit signal. Including standards with different specifications increases the necessary margin, which is then mainly reflected in increased power consumption and reduced talk-time.

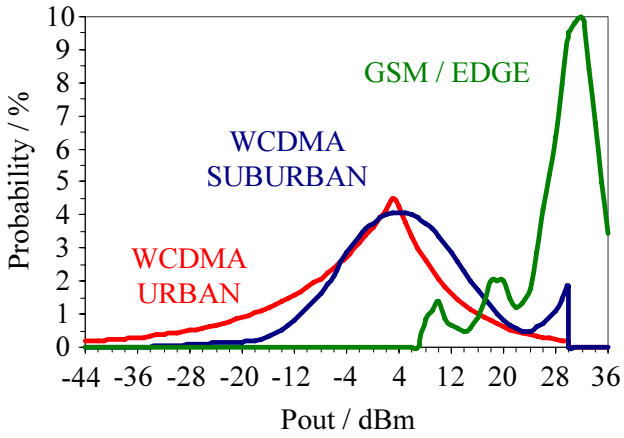


Figure 1.4: Output power probability density function (PDF) for GSM/EDGE/WCDMA [Egolf 07].

Another reason for the poor efficiency of multi-mode handset PAs is found in system dependent regulation of transmit power. The power transmitted by the mobile unit is adjusted to compensate environmental effects, such as distance to the base station and fading. Statistical analysis provided by service providers [Egolf 07] have shown that 2G systems operate near maximum output power most of the time, in 3G systems the operation at maximum levels is unlikely. The most likely transmit power level is about 30 dB smaller than for 2G systems, as shown in Fig. 1.4. Due to the inherent drop of PA efficiency with reduced output power it is apparent that separate 2G and 3G PAs, optimized for the dedicated power range, can more effectively reduce

the system's current consumption. Future multi-mode PAs therefore have to cover a very large dynamic range with high DC to RF conversion efficiency.

Performance degradation and ruggedness under mismatch conditions is another concern for single-chain PA implementation. Separate PA modules, as currently implemented in GaAs technologies, have remarkable RF performance under nominal load conditions and withstand high mismatch conditions without damage. Considering multi-mode PAs in Si technologies, a compromise has to be made between nominal RF performance and operation under mismatch, mostly at cost of efficiency. Multi-mode PAs therefore also have to adapt concepts for improved insensitivity to load mismatch. For cost reasons (size, number of components) these should be capable of being integrated without additional external active control chips.

The main challenges of achieving the goal of high power, highly efficient, linear single chain PA modules can not only be addressed by technological developments. New concepts and topologies have to be established to improve classical design approaches and increase efficiency and load insensitivity of the PA over a large linear output power range.

The aim of this work is to find new solutions, beyond existing concepts, to achieve better linearity versus efficiency compromises over a large dynamic range with load insensitive behavior. The theoretical findings are of general nature, being applicable to any combination of 2G/3G standards. Due to the strong industrial-related focus of the project, the developed prototype and the experimental evaluation concentrates on GSM, EDGE and WCDMA.

1.3 Thesis outline

Chapter 2 will give a brief review of PA characteristics and operation modes.

Chapter 3 introduces several efficiency improvement techniques and discusses impacts of load variations and possibilities for their reduction. The balanced PA is further examined and a novel concept for a multi-mode PA is analyzed in detail.

Chapter 4 concentrates on the implementation of the multi-mode PA in a low cost SiGe technology. The technology is presented and relevant circuit design issues (on-chip quadrature hybrid implementation, simulations and experimental optimization) for the prototype realization are considered.

Chapter 5 summarizes the main findings and gives a proposal for further work.

Chapter 2

Power Amplifier Characteristics

To evaluate the performance of a power amplifier, system dependent parameters have to be defined. While specifications only regulate upper limits, which have to be fulfilled, the parameters are also used for comparing quality of different PAs for the same application. This chapter provides definitions of PA parameters, important for GSM/EDGE/WCDMA systems. Very briefly a review of basic PA operation classes will be given at the end to show the inherent efficiency linearity trade-off in classical PA designs.

2.1 Basic Parameters

2.1.1 Output Power and Power Distribution Function

The most important parameter for a PA is the maximum output power $P_{\mathrm{out,max}}$, which can be delivered to a specified load R_{L}. This can be calculated by $P_{\mathrm{out,max}} = 1/2\ \mathrm{Re}(V_{\mathrm{max}} I^*_{\mathrm{max}})$, which reduces for a sinusoidal signal to $P_{\mathrm{out,max}} = \hat{V}^2_{\mathrm{max}}/(2R_{\mathrm{L}})$. When the drive level for the PA is reduced, the voltage peak $\hat{V}$ reduces and less output power is delivered to the load. This operating condition below $P_{\mathrm{out,max}}$ is called power back-off BO and the amount of BO is usually given in dB, related to the maximum output power ($\mathrm{BO} = 10\lg\left(P_{\mathrm{out}}/P_{\mathrm{out,max}}\right)$ dB). A statistical analysis of the rate with which a PA is operated at BO leads to the so called power distribution function (PDF). This gives information on the probability for occurrence of specific output power levels, as shown in Fig. 1.4.

2.1.2 Gain and 1 dB compression point

The gain is defined as the relation between output power and input power $G = P_{\mathrm{out}}/P_{\mathrm{in}}$. Depending on the kind of measured power there exist four different small signal gain definitions: transducer gain, operating power gain, available power gain and maximum available gain (see Appendix C). The usual definition for PAs is the

transducer gain, which accounts for mismatch loss at the input and the output. At large BO power levels where the gain characteristic versus output power is almost constant the PA is operated in its linear range and the gain value is called linear gain G_{lin}. When a PA is driven towards its maximum output power the gain starts to drop. The power level at which the gain is reduced by -1 dB is named the one dB compression point. This can be referred to the input $P_{\mathrm{in},-1\,\mathrm{dB}}$ or the output $P_{\mathrm{out},-1\,\mathrm{dB}}$. At maximum output power the minimum gain value is obtained, which is then called the saturated power gain G_{sat}.

2.1.3 Efficiency

The PA is a device, which basically converts the DC power, drawn from a supply (in handsets this is the battery), to an RF signal power. The efficiency, also known as DC to RF efficiency or collector/drain efficiency is defined by:

$$\eta = \frac{P_{\mathrm{out}}}{P_{\mathrm{DC}}}, \tag{2.1}$$

with P_{DC} as the DC power supplied to the collector/drain contact of the PA. As the contribution of the input drive power can considerably worsen the overall power budget, an alternative definition is given by the power added efficiency:

$$\mathrm{PAE} = \frac{P_{\mathrm{out}} - P_{\mathrm{in}}}{P'_{\mathrm{DC}}} = \frac{P_{\mathrm{out}}}{P'_{\mathrm{DC}}}\left(1 - \frac{1}{G}\right), \tag{2.2}$$

where P'_{DC} subsumes the total supply power, including driver stages. The PAE converges to η for very large gain values (> 10 dB) and small pre-driver power consumption. Efficiency in a PA reduces significantly with increased BO power levels. To consider the variations of output power over time, the signal statistics in terms of the PDF can be used to define an average efficiency [Sevic 97]:

$$\overline{\eta} = \frac{\int P_{\mathrm{out}} \cdot \mathrm{PDF}(P_{\mathrm{out}})\,\mathrm{d}P_{\mathrm{out}}}{\int P_{\mathrm{dc}}(P_{\mathrm{out}}) \cdot \mathrm{PDF}(P_{\mathrm{out}})\,\mathrm{d}P_{\mathrm{out}}} \tag{2.3}$$

This measure is more useful for comparison of PAs, when considering systems with large BO power probability, such as WCDMA.

2.1.4 Linearity

The transmitted information in wireless communication systems is embedded in the amplitude and the phase variation of the RF signal. The tolerated limits of distortion for recovering the information and not interfering with other users can be expressed with different measures, depending on the system specifications. In general the nonlinearities in a PA can be traced back to AM-AM and AM-PM distortion mechanisms [Kennington 00, Vuolevi 03]. The AM-AM characteristic basically describes the compression of output signal amplitude with increasing input

signal level. The AM-PM characteristic reflects the undesired nonlinear impact of input amplitude variation on output phase. Common measures deduced from single tone and two tone excitation, such as $P_{-1\mathrm{dB}}$ and intermodulation distortion (IMD), are often insufficient to describe the PA's linearity, when driven with complex modulation signals. Additional electrically and thermally induced memory effects lead to waveform dependent non-linearities [Vuolevi 01]. Therefore other system relevant measures have to be introduced like EVM, modulation spectrum and ACLR/ACPR, which are described in the following.

EVM

For digital modulation schemes the information data is mapped to so called symbols in a constellation diagram (IQ-diagram). Each symbol is identified by a vector of certain amplitude and phase. Due to various impairments in the system the vector is distorted in amplitude and phase, leading to potential symbol detection errors. The error vector $\vec{E}(k)$ is a measure of how far the actual detected constellation points $\vec{M}(k)$ deviate from the ideal locations $\vec{S}(k)$: $\vec{E}(k) = \vec{M}(k) - \vec{S}(k)$, where k denotes the consecutive number of samples. A graphical illustration is shown in Fig. 2.1.

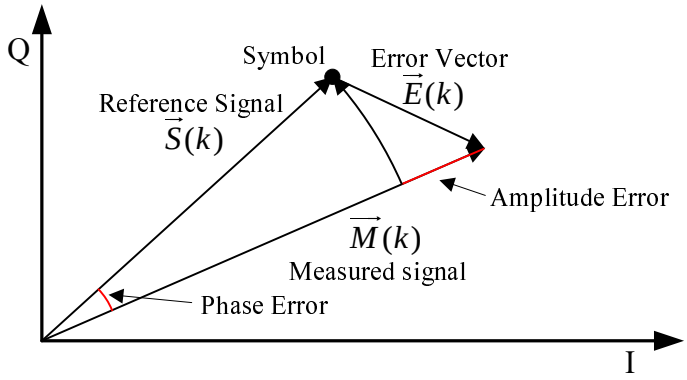


Figure 2.1: EVM measurement illustration in the IQ-diagram.

Since the deviation is of statistical nature, the root mean square (RMS) EVM is defined as ratio of the mean powers between the error vector and the ideal signal, measured for many samples K:

$$\mathrm{EVM}_{\mathrm{RMS}} = \sqrt{\sum_{k\in K} |\vec{E}(k)|^2 \Big/ \sum_{k\in K} |\vec{S}(k)|^2} \tag{2.4}$$

The maximum EVM is also specified for 2G and 3G signals:

$$\mathrm{EVM}_{\mathrm{peak}} = \max_{k\in K} \sqrt{|\vec{E}(k)| \Big/ \frac{1}{K}\sum_{k\in K} |\vec{S}(k)|^2} \tag{2.5}$$

Details on measurement procedures can be found in [Agilent 02].

Phase Error

For signals, where all information is contained in the phase, as in GSM, the signal quality is also specified by the maximum phase error between the transmitted and the expected ideal signal. This is similar to tracing the AM-PM characteristic of the PA, but the complex valued transmit signal is used instead of a single tone.

Spectral Mask and Modulation Spectrum

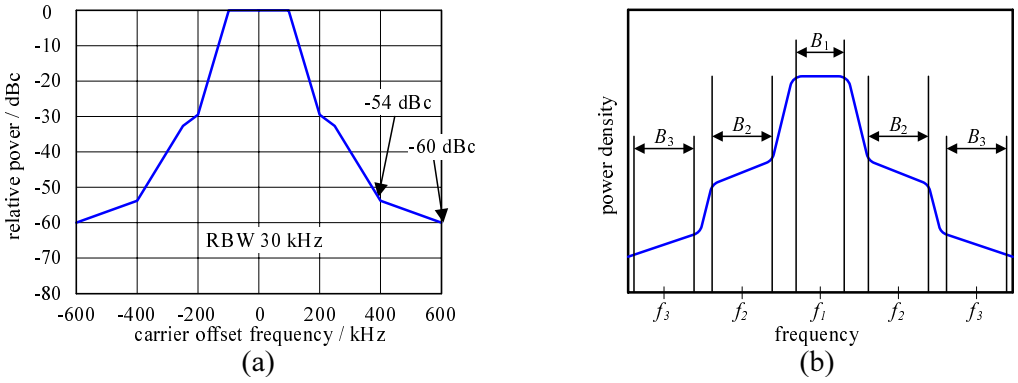


Figure 2.2: Spectral emission measures: (a) Spectral mask, (b) ACLR

Excessive radiation at frequencies beyond the channel bandwidth lead to interference with neighbor channels. To describe this out of band emission, the spectral mask or modulation spectrum (Mod. Spec.) is specified in 2G systems. This is defined as the maximum allowed power density outside the signal band, normalized to the power density within the signal band, as a function of the frequency offset from the carrier. The power density is measured within a certain resolution bandwidth RBW. Fig. 2.2 (a) shows the specification for GSM/EDGE.

ACLR/ACPR

A similar measure to the spectral mask is the adjacent channel leakage/power ratio (ACLR/ACPR). This is specified for 3G systems and replaces the definition of spurious out of band emissions for 2G systems. The ACLR is defined as the ratio of the total power in the transmission channel with bandwidth B_1 to the total power within a certain bandwidth (B_2) in the channel adjacent to the transmission channel at a certain offset frequency f_2 to the carrier frequency f_1. An illustration is provided in Fig. 2.2 (b). For WCDMA the bandwidths for power integration are 3.84 MHz.

2.1.5 VSWR

In a mobile handset the performance and reliability of a PA is subject to load impedance changes. The load variations occur due to antenna impedance changes

caused by use in real life scenarios (e.g. grip of phone). The specifications for output power reduction and device degradation or destruction are usually given for a maximum voltage standing wave ratio (VSWR) of the load. As the load reflection coefficient value Γ_L is readily accessible with a network analyzer, the VSWR can be expressed by:

$$\mathrm{VSWR} = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|} \tag{2.6}$$

As Γ_L is related to a nominal load impedance Z_0 (usually 50 Ω), the VSWR can be seen as a relative measure of changes in the impedance magnitude.

2.2 Parameter table for GSM/EDGE/WCDMA

The air interface system specifications for GSM/EDGE/WCDMA can be found in Table 1.1. These requirements are defined at the antenna of the whole transmitter system. The relation between the system requirements and the PA requirements is determined by the circuit environment of the PA. This typically consists of:

- Output interface circuits such as harmonic and duplex filters, matching network and switch for separation of transmit and receive signals. These components mainly add additional losses and some mismatch, which reduce output power and efficiency of the PA module, which has to be considered in the specification.
- Input transceiver, which generates the transmit signal at a low power level ($0 \ldots 10$ dBm). The generated signal has a certain amount of ACLR/EVM and noise, thus decreasing the PA's linearity and noise figure budget compared to system specifications.
- Power supply, which is often coming directly from the battery or stabilized by a DC/DC converter. Allowed supply variations in the charge and discharge cycles of the battery have to be specified.

The signal degradation at the input, the output and through the varying supply has to be considered and the PA is designed with margin to the system specifications. As the actual margin values depend on the specific components and architectures, used for the system, only typical values can be given as indicator for a cellular PA design. Table 2.1 sums up the typical requirements for cellular PAs.

Parameters	GSM		EDGE		WCDMA	
	LB	MB/HB	LB	MB/HB	LB	MB/HB
Supply Voltage (V)	2.5 . . . 5.5		2.5 . . . 5.5		2.5 . . . 5.5	
$P_{\text{out,max}}$ (dBm)	36	33	30	29	27	27
PAE @ $P_{\text{out,max}}$ (%)	> 50		> 20		> 40	
typ. G_{lin} (dB)	40		30		30	
ACLR (dBc) @±5 MHz	–		–		-35	
ACLR (dBc) @±10 MHz	–		–		-45	
Mod. Spec. (dBc) @±400 kHz	-64		-58		–	
Mod. Spec. (dBc) @±600 kHz	-64		-64		–	
Average EVM (%)	–		$< 5...9$		–	
VSWR_{max}, no damage	10:1		10:1		10:1	

Table 2.1: PA module specifications of mobile communication standards (GSM, EDGE and WCDMA).

2.3 Operation Classes

As the operation modes of power amplifiers are well explained in literature and broadly covered in textbooks such as [Cripps 99, Lee 98, Krauss 80, Shirvani 03], the following section will only give a brief overview of the classification to form an understanding of the inherent tradeoff between linearity and efficiency.

A typical PA circuit consists of an active device and a load or matching network. The classification of PAs can be done on basis of the active device's operation mode: current source or switching mode. A further distinguishing characteristic is the kind of load represented to the PA: filter with load adjustment only at the fundamental frequency or tuning of fundamental and harmonic frequencies.

Current source PAs include classes A, AB, B and C, while classes D, E and F are switching PAs - depending on how hard the active device is driven. For class A, AB, B, C and D only fundamental frequencies are matched, while class E and F operation requires additional tuning of harmonics.

2.3.1 Class A, AB, B and C

This family of PAs has similar circuit configuration, and distinction is based upon the biasing conditions. As shown in Fig. 2.4 the voltage waveform at the collector V_{CE} is purely sinusoidal, while the current through the transistor I_{C} is truncated below zero. How long the device is turned on in one RF cycle is expressed by the so called conduction angle α. It is obvious that dissipation of power at the transistor only occurs, if V_{CE} and I_{C} are > 0 at the same time. The main target of reducing α is the reduction of dissipation and thereby the increase of efficiency. To see the effect on linearity the α dependant current waveform can be decomposed into its

harmonic contents by Fourier analysis [Cripps 99]:

$$I_{\mathrm{DC}} = \frac{I_{\max}}{2\pi} \int_{-\alpha/2}^{\alpha/2} \frac{\cos\theta - \cos\frac{\alpha}{2}}{1-\cos\frac{\alpha}{2}}\,\mathrm{d}\theta = \frac{I_{\max}}{2\pi}\frac{2\sin\frac{\alpha}{2} - \alpha\,\cos\frac{\alpha}{2}}{1-\cos\frac{\alpha}{2}} \tag{2.7}$$

$$\hat{I}_{\mathrm{n}} = \frac{I_{\max}}{\pi} \int_{-\alpha/2}^{\alpha/2} \frac{\cos\theta - \cos\frac{\alpha}{2}}{1-\cos\frac{\alpha}{2}}\,\cos n\theta\,\mathrm{d}\theta = \lim_{n'\to n} \frac{2I_{\max}}{\pi}\,\frac{\sin(n'\frac{\alpha}{2})\cos(\frac{\alpha}{2}) - n'\cos(n'\frac{\alpha}{2})\sin(\frac{\alpha}{2})}{(n'^3-n)(1-\cos(\frac{\alpha}{2}))} \tag{2.8}$$

where I_{DC} is the DC component and $\hat{I}_{\mathrm{C,n}}$ the amplitude of the n^{th} harmonic. The graphical representation in Fig. 2.3 (a) shows that the DC component is highest at class A bias and decreases monotonically with the conduction angle. The fundamental component increases slightly initially and then drops with the conduction angle. The fundamental at class B bias reaches the same value as for class A bias. Thus, class B is capable of delivering the same amount of output power with a substantially lower DC drain current, which results in improved efficiency. The other frequency components start to grow as the conduction angle decreases and converge to the same level. Whereas out of band harmonics can be filtered, in band distortion due to intermodulation will degrade linearity in terms of signal quality. Hence, special attention has to be paid to odd harmonics, which rise substantially in class C operation and have moderate or even zero value in class A, AB and B, compared to the fundamental. For that reason highest signal amplitude linearity can only be obtained in class A and B, moderate linearity in class AB and almost no linearity in class C operation. Efficiency and output power, achievable by this range of amplifiers is readily computed by: $P_{\mathrm{out}}(\alpha) = V_{\mathrm{DC}}\hat{I}_1(\alpha)$ and $\eta(\alpha) = \hat{I}_1(\alpha)/(2I_{\mathrm{DC}}(\alpha))$, using (2.7) and (2.8). The results are illustrated in Fig. 2.3 (b). The efficiency of the amplifier can be increased up to theoretically 100 % by reducing the conduction angle. This is only obtained at the expense of output power, which diminishes rapidly in deep class C operation. In order to obtain higher output power values and achieve additional efficiency it is necessary to allow the voltage waveform to take on non-sinusoidal shapes. This is the basic idea behind the following operation classes. Amplitude modulation of the input signal is then no longer reflected in the output signals, making these classes nonlinear.

2.3.2 Class F

The basic idea is to operate the amplifier in class B bias ($\alpha = 180°$) and to add odd harmonics to the voltage waveform so that it begins to increasingly resemble a square wave [Raab 77,a, Raab 02, Kee 03]. This requires a load network that has resonance at one or more harmonic frequencies as well as at the fundamental frequency. With a theoretical load network, which tunes an infinite number of harmonics, the collector voltage approaches a square wave out of phase with the drain current. Therefore efficiency can be increased to 100 %, which is usually associated with poor linearity.

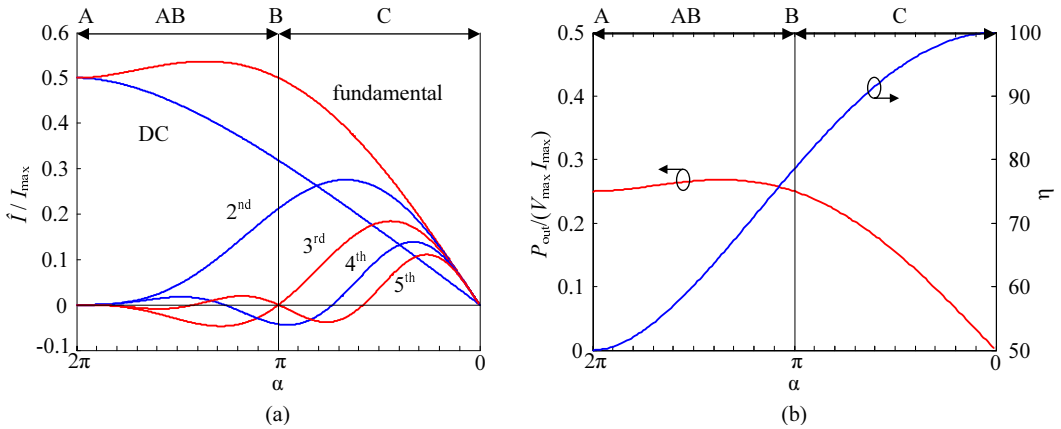


Figure 2.3: Reduced conduction angle classes. (a) Fourier decomposition of current. (b) Output power and collector efficiency.

However the linearity can be improved to meet the design standards by tuning less harmonics and thus compromising on efficiency. It should be further noticed that the current amplitude at the fundamental is the same as in class B operation, but the fundamental voltage amplitude is $\frac{4}{\pi}$ times higher[1]. Thus, output power will be also 1.05 dB higher. Main drawback of this amplifier class is the complexity of the waveform shaping network and its implementation at microwave frequencies.

2.3.3 Class E

In this class the transistor is driven to act as a switch. The load network is designed to resonate with the collector-emitter capacitance of the transistor, to reduce voltage and current waveform overlap during switching transitions. This is achieved by discharging the capacitance at the collector node at the start of every cycle, consequently generating a voltage across the transistor of zero value and zero slope at the time of turn-on [Sokal 75, Raab 77,b]. Although in theory 100 % efficiency can be expected, this concept has serious disadvantages. The peak collector voltage can be more than 3 times higher than the supply voltage [Krauss 80], causing a potential threat to the amplifier's reliability. The current waveform includes odd harmonics, as indicated in Fig. 2.4, effecting insufficient linearity for most applications.

2.3.4 Class D

This, at microwave frequencies hypothetical class of operation, assumes the device to behave as an ideal switch. While the voltage and current waveforms are similar to ideal class F operation, the output network is only tuned to the fundamental. The collector capacitance being present is not considered. This reduces switching speed

[1]Fourier decomposition of the square wave voltage waveform gives: $1-\frac{4}{\pi}\sin(\omega_0 t)+\frac{4}{9\pi}\sin(3\omega_0 t)-\frac{4}{25\pi}\sin(5\omega_0 t)\ldots$

and efficiency due to an increased voltage-current overlap. On top of that strong amplitude nonlinearity is caused by the hard driven transistor. These drawbacks limit a practical implementation to low frequency applications with phase modulated signals.

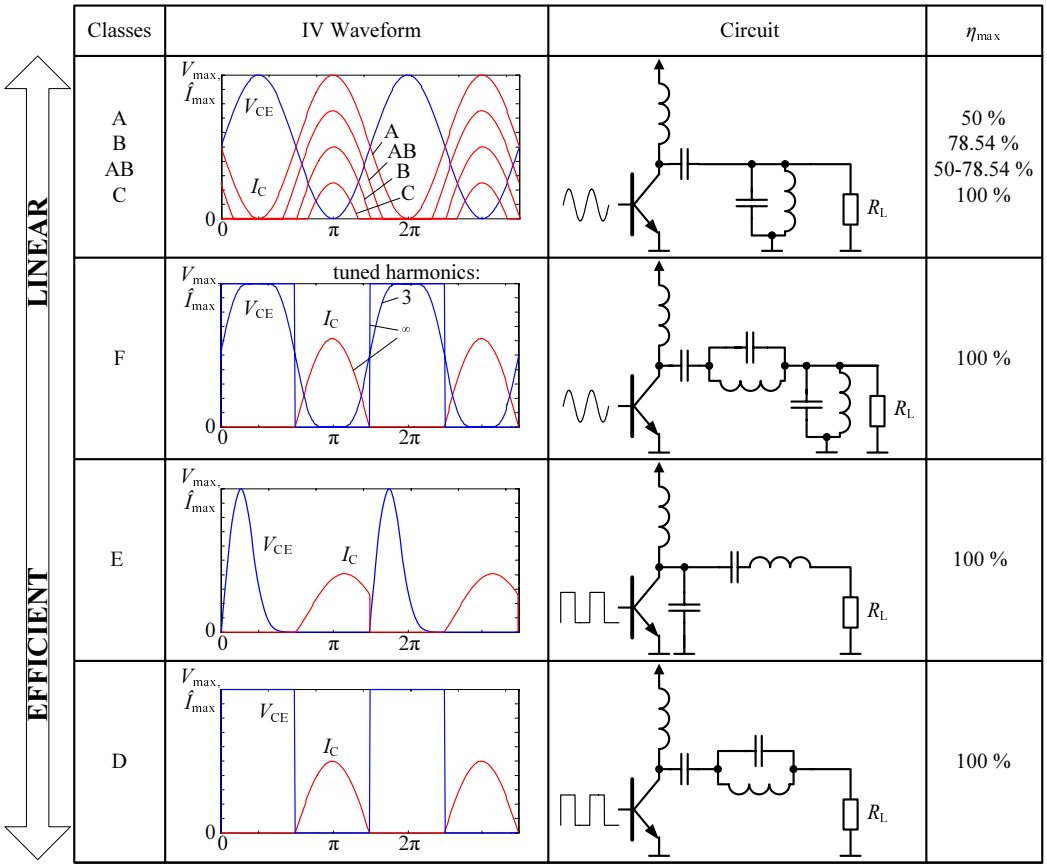


Figure 2.4: Amplifier classification: Voltage and current waveforms, simplified circuits and maximum achievable efficiency.

Among the classes of operation it has to be taken into account that a real transistor exhibits a finite saturation voltage (> 0) and intrinsic parasitics, which lead to reduction of efficiency, due to imperfect waveform shaping. Generally spoken the design of highly efficient PAs in reality demands the smallest possible power dissipation in the device. The exact waveforms will differ from the ideal presented, but a small I-V product at the transistor is always desired. Tuning to high efficiency operation can reduce performance significantly with regard to measures as output power or gain (as in class C) and linearity (as in class C, D, E, F). Hence, the level of improved efficiency always has to be weighted against other degradation mechanisms.

Chapter 3

Multi-Mode PA Concepts

The demand for handset power amplifiers fulfilling more than one cellular standard rises due to expected saving of costs in mobile phone production. The increasing technical requirements for these multi-mode PAs by far exceed the performance limits of conventional amplifier classes, presented in the previous chapter. Besides of output power and bandwidth constraints the limits are primarily given by the fundamental tradeoff between linearity and efficiency. In a stand-alone amplifier only one of these two objectives can be increased at the expense of the other. A variety of enhancement techniques have been developed over the years to overcome this tradeoff. Each of them has use-oriented advantages and disadvantages but a general solution has not been found yet. In this chapter enhancement techniques are classified, shortly described and finally rated for their potential usage in multi-mode handset applications. Further improvements of PA performance can be achieved by using more than one of these techniques simultaneously. The resulting circuit complexity is contrasted with the benefits in this chapter as well. Another important characteristic of the PA is its response to load impedance changes at the antenna. This can have three consequences: reversible performance degradation, irreversible performance degradation due to aging and irreversible destruction of the transistor. Methods for reducing these effects will be discussed.

Finally the concept for the implemented multi-mode power amplifier is presented. It is based on novel combinations of some of the presented techniques. The stage size- and load adaption and the power combining technique turn out to be most promising for optimizing the various parameters in the multi-mode designs. This way efficiency and load independence are improved without degrading linearity. The theoretical performance limits on the efficiency due to different power loss mechanisms caused by dissipation or bad isolation of parallel amplifier paths are derived and discussed for both concepts.

3.1 Power amplifier architectures

The arrangement of building blocks defines the architecture of a power amplifier. A variety of architectures can be found in literature [Cripps 99, Cripps 02, Kennington 00]. Most of them are efficiency enhancement concepts which were invented to extend the high efficient operation range, where the amplifier is in or close to saturation, to lower power levels. The basic building blocks for all these concepts are amplifiers in two different configurations: single ended and push-pull (also called differential) topologies. They are designed following the principles in Chapter 2.3. The performance of these blocks defines the theoretical peak power efficiency of the overall amplifier. The peak efficiency will not be improved by these architectures, it is defined by the technology in use and the load matching design. This has to be kept in mind when thinking about efficiency enhancement architectures. Fig. 3.1 illustrates the classification into basic PA topologies and deduced from this some of the most frequently used performance enhancement methods. These are further subdivided into methods which result in continuous PA characteristics (i.e. PAE vs. P_{out} curves) and methods where the curves are discontinuous due to e.g. on and off switching of stages. The performance benefit of the latter one is usually less sensitive to variations in production at a relatively low complexity level. The stepwise change of PA properties enforce on the other hand a more sophisticated system power regulation. Enhancement techniques belonging to both groups will be discussed in detail in Chapter 3.2. Following a closer look is taken onto the core topologies (single ended and push-pull) and their differences which are mainly influenced by the technology in use.

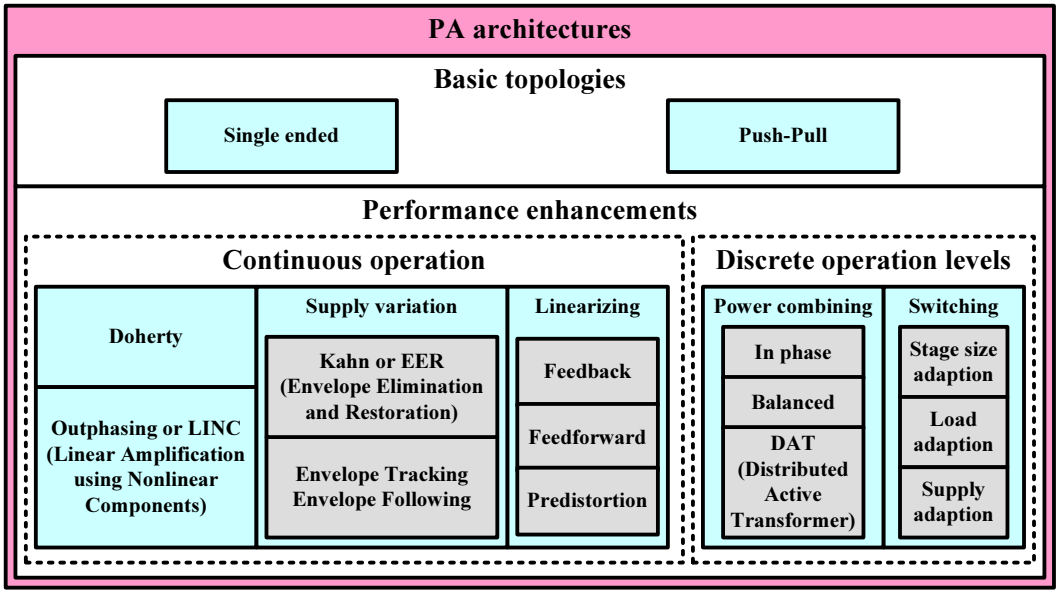


Figure 3.1: Classification of power amplifier architectures.

3.1.1 Single ended amplifiers

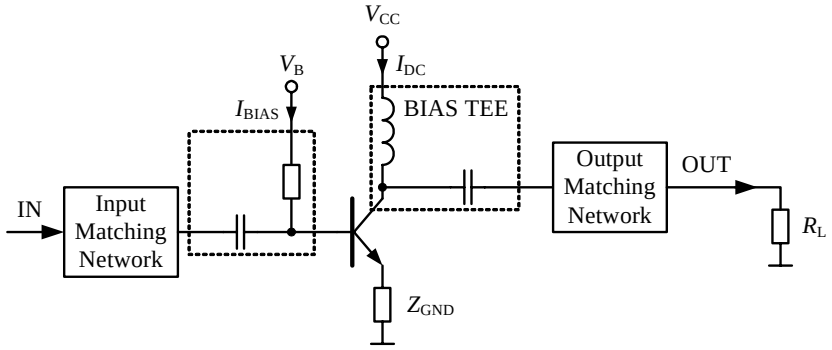


Figure 3.2: Single ended one stage power amplifier block diagram.

The single ended power amplifier with the transistor in a common-emitter configuration is shown in Fig. 3.2. The part count for this configuration is low, it consists of input and output matching networks with proper biasing and RF decoupling at the input and the output (bias tee). Its simplicity is one of the reasons why this topology is frequently used in commercial products. The design of matching networks is straight forward and there is no need for additional power combining components which may add losses and decrease efficiency. Hence, this topology has the potential to realize highest peak efficiencies. Another advantage is the short duration of circuit simulations. Simulation of large signal characteristics with higher level transistor models (e.g. HICUM model incorporating more than 90 parameters, with harmonic balance simulator, see Chapter 4.1.1) show easier and faster convergence, if the number of devices is kept low. Therefore a complete circuit simulation containing all matching networks and parasitics will save development time compared to differential or power combining topologies, using at least twice the number of active devices. The potential for a pure single ended design for improving PAE at back-off is very low. For transmission of linear signals the typical operation class will be somewhere between A and B. Therefore this topology has to be combined with performance enhancement techniques from Fig. 3.1 when used in mobile terminals.

In RF applications with amplification of constant envelope signals like GSM where high output powers in the range of $\sim$ 36 dBm are needed the single ended topology suffers from several drawbacks:

Impedance sensitive ground connection

The ground connection of large size transistor stages require very low impedance. Every parasitic resistance or inductance, subsumed in Z_{GND} in Fig. 3.2 causes a negative feedback which reduces gain, output power and therefore efficiency. Ground connections made with bond wires can become problematic at higher frequencies. Assuming GSM at 2 GHz the reactance of a bond wire with typical length of 1 mm will be about $2\pi f 1$ nH = 12.5 Ω. The voltage drop across this reactance reduces the RF voltage swing between base and emitter ($\hat{V}_{\mathrm{BE}}$) and consequently the gain of

the device. As this will also reduce the current through the transistor the output power will be reduced as well. For low frequencies in the MHz range the inductance from bond wires has no severe impact, Z_{GND} is dominated by the bond wire's series resistance of typically several $\mathrm{m}\Omega s$.

A method to reduce inductance and series resistance is using several bond wires in parallel. The overall inductance will be reduced more effectively when the wires have small mutual inductance, which means large distance to each other. Therefore this solution requires a lot of chip and printed circuit board (PCB) space.

Using a flip-chip package where the active area of the chip is flipped over, facing downwards, reduces inductance more effectively. Instead of wire bonds metal bumps of solder (diameters of $\sim$ 80...300 μm) are directly connected from the emitter top metal connection to the PCB ground. The inductance will be about 1/10 compared to wires.

The most effective method to reduce the ground connection impedance is a through-connection from the active area to the backside of the chip. This can be realized as a metal via through the silicon or a highly doped area in a highly conductive substrate - so called sinker. For a low impedance a short distance through the substrate is necessary, which is achieved by grinding down the wafers to $\sim$ 80...160 μm thickness. An additional backside metallization further reduces contact resistance to the PCB. In comparison to a flip-chip package the inductance can be reduced by more than 1/10 because of the larger area of a sinker compared to bumps. Also the heat dissipation to the PCB will be more effective. More information on this sinker technology will be given in Chapter 4.1.

Large single device transistor area

In order to obtain the desired RF power of the final transistor several transistor base cells are parallel circuited. With increasing number of base cells the input and output impedances of the transistor decrease, making the design of low loss and broadband matching networks more difficult. At the input an interstage matching between driver stage and output stage on chip can be challenging as an impedance of about $0.5 \ldots 2\ \Omega$[1] has to be transformed to about 20 Ω[2]. Losses in the on-chip matching network which are mainly caused by spiral coils with low quality factors of $5 \ldots 10$ will reduce gain and therefore PAE. At the output the necessary load resistance R_{opt}, presented to a transistor which is capable of delivering 36 dBm output power is only 1.5 Ω ($= \frac{V_{\mathrm{DC}}^2}{2P_{\mathrm{out,max}}}$, supply voltage $V_{\mathrm{DC}} = 3.5$ V). This has to be transformed from a 50 Ω load with a network of very high quality factor to obtain maximum power transfer and maximum PAE. The large transformation ratio requires a multi-stage matching network to achieve a large bandwidth. Nevertheless there exists a theoretical limit for the gain-bandwidth product of the matching network [Bode 45, Fano 50], why the bandwidth can only be increased by additional losses which compromise output power and lead to a reduced efficiency. An increase of the necessary load resistance would reduce the transformation ratio and therefore

[1] Typical bias dependent values for the used SiGe transistor delivering 36 dBm output power.
[2] Typical value for $\sim$ 20 dBm driver power.

improve bandwidth and efficiency. For the same output power this increase can be achieved by a higher supply voltage level: $R_{\text{opt}} \propto V_{\text{DC}}^2/P_{\text{out,max}}$. Due to the squared relation a two times higher supply voltage causes a four times higher load resistance and four times lower transformation ratio. PAs in handsets are usually supplied directly from the battery (nominal voltage 3.5 V). The bandwidth and efficiency of single ended designs could therefore be increased by a supply voltage up-conversion, using a DC-DC converter. The benefits will be limited by the efficiency of the DC-DC converter and the breakdown voltage of the devices. For the bipolar technology used in this thesis the breakdown voltage is 5.5 V. Hence a 2.5 times smaller transformation ratio can be obtained in best case, combined with smaller reliability of the device and smaller overall efficiency.

Another disadvantage of the large transistor area can be found in the layout. At frequencies in the GHz range the distributed connection lines to each base cell become a severe problem due to inhomogeneous signal distribution among the unit cells of the whole transistor block. Consequently the output power is reduced as the RF currents do not add in-phase at the output. An analysis of this effect and the proposal of a compensation network for balancing the BJTs base connections to reduce this effect can be found in [Bakalski 04].

Thermal heat concentration

With increasing transistor area the thermal stability of BJTs becomes a problem due to the inherent positive thermal feedback properties of the bipolar device. With increasing output power the base cells heat up mutually and hot-spot formation in the center of the transistor block takes place [Rudolph 04]. This changes the DC and RF properties of the amplifier and in worst case causes break-down of the device. To counteract this effect a high thermal conductive substrate as well as low resistive thermal connection of the chip to the housing or to the carrier is necessary. In this regard the use of Si (148 $\text{W/m} \cdot \text{K}$) gives an advantage compared to GaAs (46 $\text{W/m} \cdot \text{K}$)[3]. Another possibility to improve thermal stability is to introduce a resistor at the base or the emitter for thermal feedback. Hence the internal base emitter voltage is reduced with increasing current, partially compensating the positive thermal feedback effect. Increasing the transistor area by moving the transistor base cells apart counteracts the thermal interaction. Besides the higher area required, the RF parasitics for the whole block increase which has to be reviewed carefully in the specific implementation.

3.1.2 Push-Pull amplifiers

The push-pull amplifier topology makes use of power combination in a special manner. The input signal is divided into two equal signals 180° out of phase being amplified by two identical devices and combined with 180° phase shift at the output. The combining structure is called BALUN (BALanced to UNbalanced). Fig. 3.3 (b) shows a simplified circuit with complementary transistors (NPN and PNP) and

[3]Values taken from [Palankovski 99]

Fig. 3.3 (a) shows a transformer based PA with the same type of transistors (NPN). Today's technologies do not allow the development of PNP transistors with the same performance as NPN. PNPs have smaller gain at GHz frequencies which can only be increased by accepting a smaller breakdown voltage. Complementary circuits for high power are amongst others mainly used in audio amplifiers.

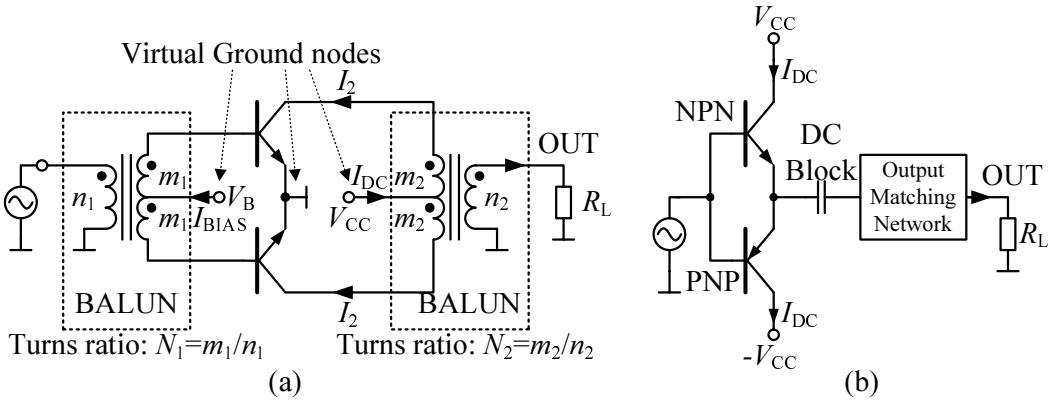


Figure 3.3: Push-Pull amplifier topologies: (a) Transformer based PA for RF applications, (b) Complementary configuration for low frequencies.

Transformer based push-pull amplifiers have several advantages compared to single ended amplifiers [Shumaker 02]:

- Virtual ground: At the common connection of the emitters the RF currents cancel with respect to ground. The resulting virtual AC ground at this node allows a separation of RF signals from the PCB ground. As this virtual ground node is on-chip the connection from the emitter to the node is short with small inductance. Therefore the negative feedback effect, as described for single ended amplifiers, is substantially reduced. As the RF signals cancel at this node, the DC biasing can be applied there, without need of additional blocking capacitors.

- Load line impedance: The voltage swing at the load is double the collector voltage swing at one device, while the current swing is halved. Therefore the load impedance is four times higher than for a single ended amplifier delivering the same output power.

- Even order harmonic cancellation: The harmonics generated by the differential stage at even multiples of the fundamental frequency are in phase at the collectors of the transistors. Hence, they do not appear at the load since the BALUN suppresses every common mode signal from its input ports.

- Matching: The load impedance matching can be accomplished by the turns ratio of the transformer.

- DC biasing: Center tapped transformers (see Fig. 3.3 (a)) can be used to apply the DC bias without additional needs for bias tees.

This topology has also some disadvantages:

- Symmetry: The degree to which the benefits of a push-pull design can be used depends on the device matching and input signal matching between the two sides of the circuit. Hence, a monolithic integration (with both transistors on one die) is favorable for push-pull amplifier implementation.
- BALUN implementation: Along with the symmetry the realization of BALUNs with exact 180° phase relation is difficult, especially at the output where low losses ($\ll 1$ dB) are required for a high efficient design.
- Isolation: With conventional baluns the isolation between the two sides is only 6 dB which may cause instability.

The implementation of push-pull power amplifiers for a SiGe based HBT technology was shown in [Bakalski 03]. The described advantages allow broadband and high efficient designs even with low Q on-chip matching networks. Nevertheless more effort has to be done to improve the efficiency at power back-off levels larger than 20 dB.

3.2 Efficiency enhancement architectures

Optimization of the two key parameters efficiency and linearity underlie the conventional design of a linear power amplifier, regardless if a single ended or push-pull topology is used. The matching networks for optimum linearity or optimum efficiency require different tuning and optimum performance for both cannot be achieved simultaneously. Performance enhancement architectures have two basic approaches to defeat this tradeoff barrier. The first one is improving the efficiency of a linear but rather inefficient PA, being termed 'efficiency enhancement techniques'. The second one is the attempt to improve linearity of a nonlinear but efficient PA, being termed 'linearization techniques'. If the linearity is improved to an amount where the specification is fulfilled, then both techniques will lead to an improved efficiency. For PA applications in mobile terminals the improvement of the overall mean efficiency is of main importance to extend the talk-time. This can be accomplished by improving the instantaneous back-off or peak efficiency or both. While linearization techniques mainly affect the peak efficiency, efficiency enhancement techniques improve the PA efficiency at back-off. In modern communication systems, most of the time, the PA output power needs to be much lower than the peak, so its overall efficiency is much less than the peak efficiency. Therefore, improving the PA efficiency at back-off can potentially improve the overall efficiency significantly. Nevertheless, linearization techniques, especially the predistortion technique can be combined with others to further enhance performance.

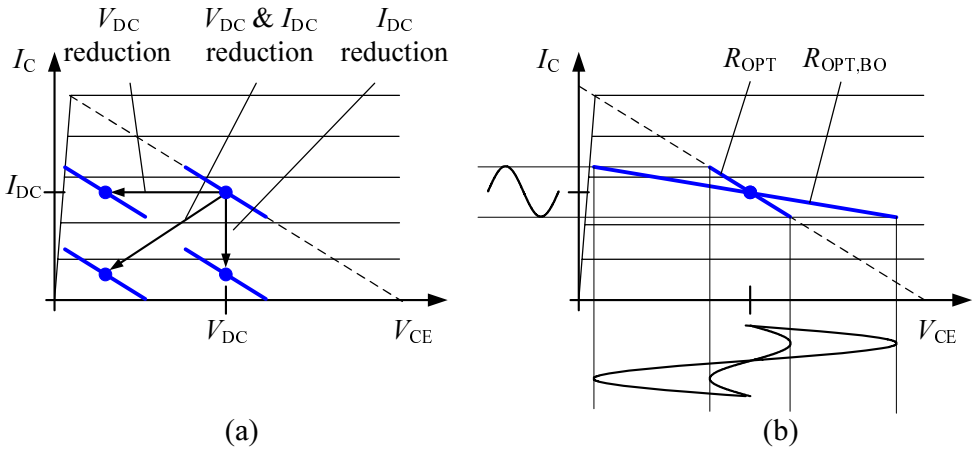


Figure 3.4: Efficiency enhancement methods: (a) Supply voltage and bias current reduction; (b) load impedance adaption.

The methods for efficiency enhancement can best be understood by assuming an amplifier being that much in back-off that it operates at high linearity. The conditions are similar to a class A amplifier where the bias current and voltage are larger than the current and voltage amplitudes, to avoid clipping. These relations are shown in Fig. 3.4(a). Recalling the definition of the collector efficiency

$$\eta = \frac{P_{\text{out}}}{P_{\text{DC}}} \tag{3.1}$$

one can see that there are two principal possibilities for an improvement:

1. A reduction of P_{DC} at equal or increased P_{out}.
2. An increase of P_{out} at equal or reduced P_{DC}.

In both cases the linearity must not be decreased, or at least the specification still has to be fulfilled. These actions are clear from a mathematical view point but have different consequences on the enhancement method. For case one the bias current (I_{DC}) or supply voltage (V_{DC}) can be reduced in such a way that the RF signal's current and voltage levels do not cross zero. The signal shifts in the output characteristics are indicated in Fig. 3.4 (a). Concerning the implementation the current reduction approach offers the advantage of simplicity but is less effective. The reason for this is that the DC current at back-off is composed of two contributors. First the bias current which is adjusted by a current mirror. Second the bias current due to self biasing of the transistor. While the first one can be controlled, the second one strongly depends on the class of operation. In a class B or class C amplifier the bias current already varies with the signal amplitude, hence this technique cannot be applied to influence efficiency. Enhancement methods from Fig. 3.1 which fall into the category of bias point adjustment are supply adaption, envelope elimination and restoration (EER), envelope tracking (ET), envelope following (EF), aiming at

a reduced V_{DC}. Those which target a reduced I_{DC} are stage size adaption and power combining.

The second approach to improve efficiency at back-off by increasing the RF power - without increasing P_{DC} - can be realized by adapting the load impedance. In Fig. 3.4 (b) this possibility is illustrated. At the same current amplitude level, a two times larger load resistance doubles the voltage amplitude, hence P_{out} increases by a factor of 2. As long as no waveform clipping occurs the DC power remains constant and consequently the efficiency is improved. This approach can unexceptionally be applied to all operation classes, but an implementation for linear amplification may be difficult due to the level dependent altering output power. Enhancement methods which are based on active load variation are Doherty and outphasing amplifiers, those which use passive load variation are load adaption and power combining PAs (Fig. 3.1).

3.2.1 Doherty amplifier

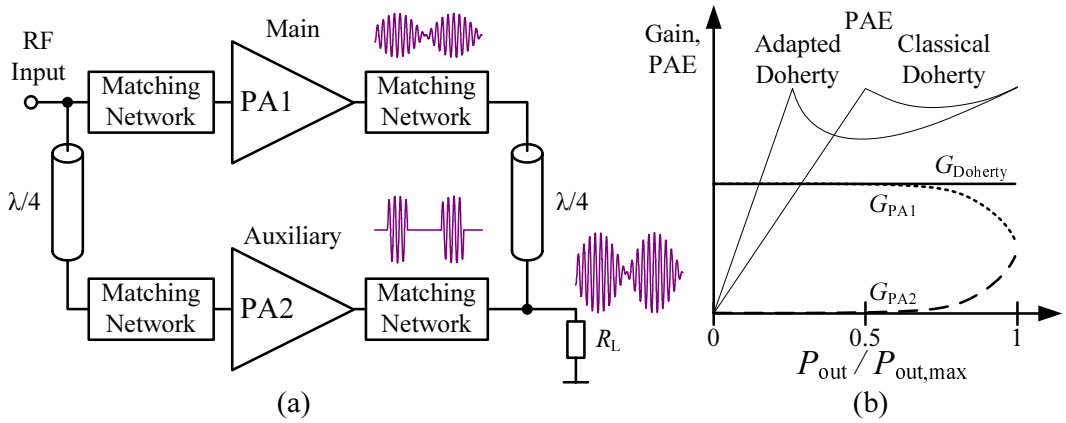


Figure 3.5: (a) Doherty PA block diagram; (b) Gain and PAE characteristics versus normalized output power.

A classical Doherty power amplifier [Doherty 36] consists of two amplifier stages of same transistor size in a configuration as shown in Fig. 3.5. At low input drive levels the main amplifier PA1 operates in its linear range (class A or AB) and the auxiliary amplifier PA2 is turned off via bias (class B or C). When PA1 saturates at higher drive levels then PA2 contributes to the total output power until it saturates as well. Consequently there exist two peaks for the power added efficiency, one at maximum output power and the other one at 3...6 dB back-off, depending on the saturation of PA1. The second peak can be shifted by adapting the power division ratio (e.g. by using different load impedances and supply voltages in the two PAs [Iwamoto 01]).

There exist several design challenges for the realization of a Doherty amplifier. First, the gain characteristics of the two amplifiers have to compensate for a flat response to obtain a linear amplification (see Fig. 3.5 (b)). Not only the AM-AM but also

the AM-PM characteristics have to match which can cause problems of yield in high volume production for commercial products. Second, the small bandwidth and size of the quarter-wave lines. As the 90° phase shift of the lines is only obtained in a small frequency band, the whole amplifier has a narrow frequency characteristic. The length of a quarter-wave line at 2 GHz is about 20 mm, supposing an FR4 substrate with $\varepsilon_{r,eff} \sim 3$. Lumped element equivalents for the lines can be used instead to reduce size, although they will add more losses and decrease efficiency. Due to the dimensions of the lines MMIC implementations of the Doherty amplifier have been limited to very high frequency (above a few tens of GHz) [McCarroll 00, Kobayashi 00, Campbell 99]. Third, the sensitivity to load variations. Analyzes in [Hammi 06, Hammi 07, Messaoudi 07] have shown that the linearity of a Doherty amplifier at constant VSWR over phase is lower compared to a single ended class AB amplifier. Yet, the Doherty architecture is an interesting topic for research but multi-mode implementations for handset MMICs are difficult unless combined with other performance enhancement techniques such as bias tracking [Lee 08], predistortion [Sim 07] and stage switching [Apel 07].

3.2.2 Outphasing amplifier

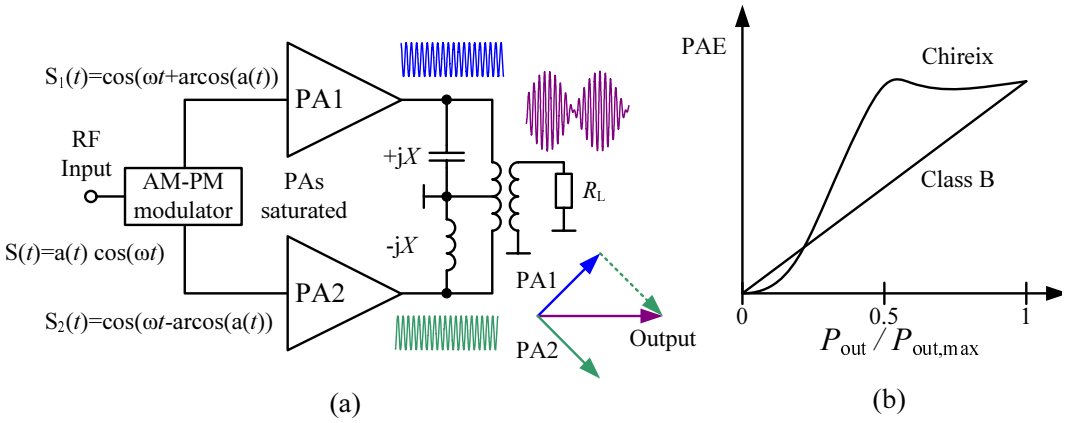


Figure 3.6: (a) Outphasing PA block diagram; (b) PAE characteristic versus output power of outphasing amplifier compared to a single ended class B amplifier.

An outphasing amplifier (also known as Chireix [Chireix 35] or LINC - LInear amplification using Nonlinear Components) is based on the idea of decomposing an amplitude-modulated signal into two constant envelope signals with different time-varying phases. These signals can be amplified by two nonlinear but highly efficient PAs driven in saturation. Combination of the output signals again produces an amplitude-modulated but amplified signal. Fig. 3.6 (a) shows a block diagram of this approach. The amount of efficiency improvement over the whole power range relies on the kind of power combination at the output. A simple in-phase combiner is very inefficient in this regard. At maximum output power the signals produced

in the individual branches add in phase. Hence, the overall efficiency equals the efficiency of the individual amplifiers, only reduced by the combiner loss. At back-off the output power is reduced by partly compensation of the output powers from the individual branches. As a result efficiency decreases rapidly, as more power is used to keep the amplifiers in saturation than power is delivered to the load. At 3 dB back-off the efficiency would be reduced by 50%, therefore not gaining any improvement compared to a single stage amplifier. For this reason the combiner structure additionally has to compensate for the signal phase difference which produces a virtual reactive load component to each amplifier at back-off (a derivation can be found in [Cripps 06]). As proposed by Chireix [Chireix 35], two shunt reactances ($\pm jX$) at the inputs of the combiner can effectively produce a purely resistive load at a particular amplitude level, maximizing efficiency at that level. Fig. 3.6(b) shows the theoretical improvement with this combining structure versus a class B amplifier. By adjustment of X the power level of the second peak can be shifted. Therefore, similar to a Doherty amplifier the efficiency can be held above a certain level in the upper power region. In contrast, at low power levels the efficiency drops even below a class B amplifier level, making this concept unattractive for applications with power distribution functions having their peak at these levels, e.g. WCDMA.

Difficulties for the implementation of an outphasing amplifier are accurate generation of the phase modulated input signals and bandwidth limitation due to the tuning elements at the output. At microwave frequencies the Chireix compensation circuit to the power combiner can hardly be realized due to non-ideal effects of the transistors in saturation, consequently the second peak in the PAE characteristics is not distinct. Additionally there is no inherent immunity to load variations. Moreover, in [Birafane 04] a derivation can be found which states that load variations decrease the linearity in a way that it cannot be improved with simple predistortion. Although this concept seems to be promising at first sight the sensitivity and complexity in implementation are major drawbacks.

3.2.3 Supply variation

The most frequently referred supply variation schemes are envelope tracking, envelope following and envelope elimination and restoration (EER or Kahn technique). The block schematics of these techniques are shown in Fig. 3.7 (a), (c). The basic idea of these techniques is to reduce the DC supply voltage according to the amplitude of the RF signal. Hence, maximum efficiency is maintained for PA2 due to the full rail-to-rail voltage swing even at power back-off.

For the implementation of an *EER amplifier* the input signal is separated into an RF constant envelope signal (containing the phase information) and a low frequency (LF) envelope signal (containing the amplitude information). The signal from amplifier PA2 is then modulated with the envelope signal from PA1. The related design challenges of this technique are:

- Time alignment between the supply and RF path: Any misalignment of the

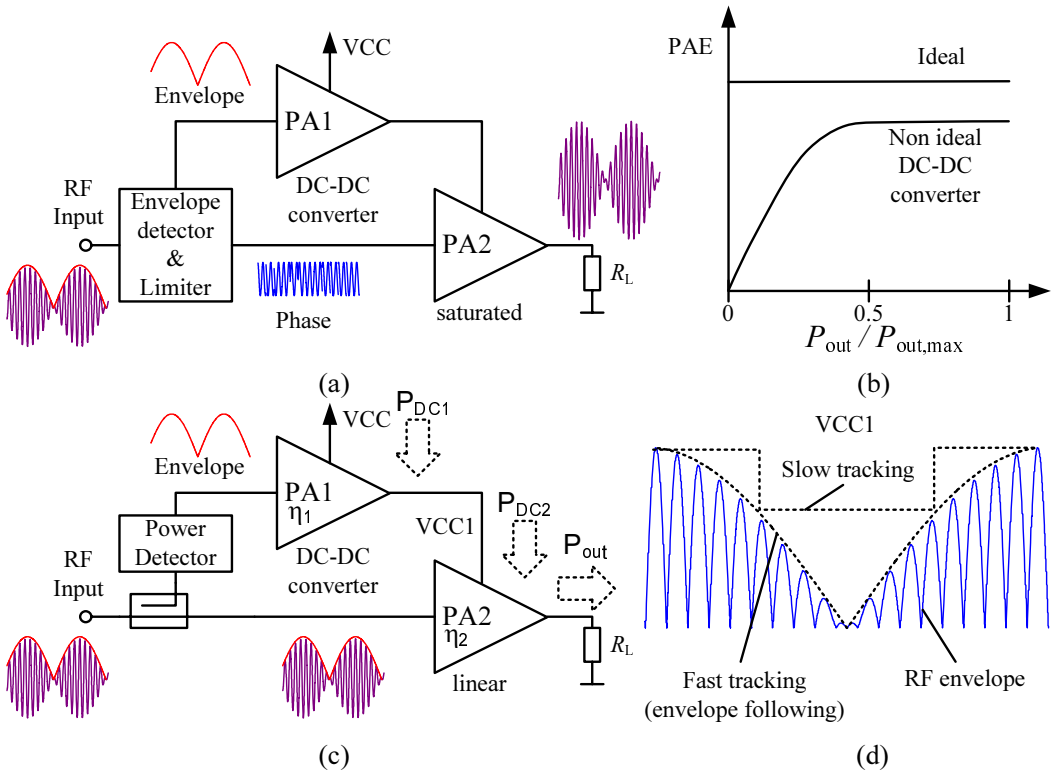


Figure 3.7: (a) EER technique; (b) PAE characteristic versus output power of EER amplifier with ideal and real DC-DC converter; (c) Envelope tracking block schematic; (d) Tracking options: fast and slow.

amplitude and phase information when modulating PA2 leads to severe distortion of the output signal.

- Phase distortion due to supply variations: The RF amplifier must exhibit limited AM-PM distortion as a function of supply voltage if acceptable small spectral regrowth and EVM is to be achieved.

- Low loss and wide-band supply modulator required: Even if PA2 operates with high efficiency the amplifier PA1 has to be a high efficient audio amplifier or switching regulator (DC-DC converter) with several MHz of bandwidth. It has to be noted that the envelope spectrum of an EDGE signal has about five times larger bandwidth than its composite baseband signal (containing amplitude and phase information) [Cripps 06]. A wider bandwidth necessitates an increase in switching speed which results in higher dynamic losses in the switching regulator which then lowers the efficiency.

The high accuracy with which the envelope has to be replicated in an EER amplifier can be overcome using *envelope tracking*. Fig. 3.7 (d) shows the supply voltage to PA2 at slow and fast tracking options compared to the envelope of the signal. Main difference to EER is that PA2 is a linear amplifier (class A, AB, B) and not

fully saturated. The average efficiency is enhanced due to reduced power consumption at lower power levels. The problem of a high efficient supply regulator is still not overcome with this technique. Different bias supplies can be used instead to switch between discrete supply levels (e.g. with FET switches providing low series resistance). However, different supply levels in mobile handsets have to be generated again with fixed DC-DC converters. In general the overall collector efficiency of a stacked configuration PA with voltage regulator can be calculated as (variables referred to Fig. 3.7 (c)):

$$\eta = \frac{P_{\mathrm{OUT}}}{P_{\mathrm{DC1}}} = \frac{P_{\mathrm{OUT}}}{P_{\mathrm{DC2}}} \cdot \frac{P_{\mathrm{DC2}}}{P_{\mathrm{DC1}}} = \eta_2 \cdot \eta_1, \tag{3.2}$$

This shows that the overall efficiency, especially at high output power will always be reduced in direct proportion to the DC-DC converter efficiency.

Reference	Technology	η_{peak}	P_{OUT}	WCDMA	GSM/EDGE
[Lee 06]	0.5 μm CMOS	94%	30 dBm	–	✓
[Blanken 08]	0.25 μm CMOS	85%	35 dBm	✓	✓
[Tombak 08]	0.5 μm LDMOS	90%	35 dBm	✓	✓

Table 3.1: DC-DC converter performances for GSM/EDGE/WCDMA supply variation from a 3.5 V supply.

Current reports on high efficient DC-DC converters for multi-mode handset power amplifiers are listed in Table 3.1. High peak efficiencies up to 90% with bandwidths covering even WCDMA signals are feasible. Nevertheless the converters are implemented on separate dies with off-chip coils of high quality factor. Therefore an integration to handset applications increases part count and complexity of the PA.

Summing up, supply variation architectures for integrated handset power amplifiers are feasible but the integration of DC-DC converters and the reduced linearity due to the changing supply voltage are drawbacks which prevent simple integration into a transceiver without additional changes in the concept. A polar modulation transmitter with digital predistortion is necessary to benefit an increased average PAE from supply variation techniques.

3.2.4 Linearizing

Linearization can be divided into feedback, feedforward and predistortion techniques [Cripps 99]. They are well documented in literature [Kennington 00, Cripps 02]. The concern of linearization is mostly not to increase efficiency at large power back-off but to meet linearity constraints especially in basestation transmitters where specifications of spectral distortion levels are much higher than the specified requirements in handset systems. The effort for implementation of feedback and feedforward linearization techniques is very high and a lot of additional components to the PA are necessary which increase size and current consumption of the overall system. This

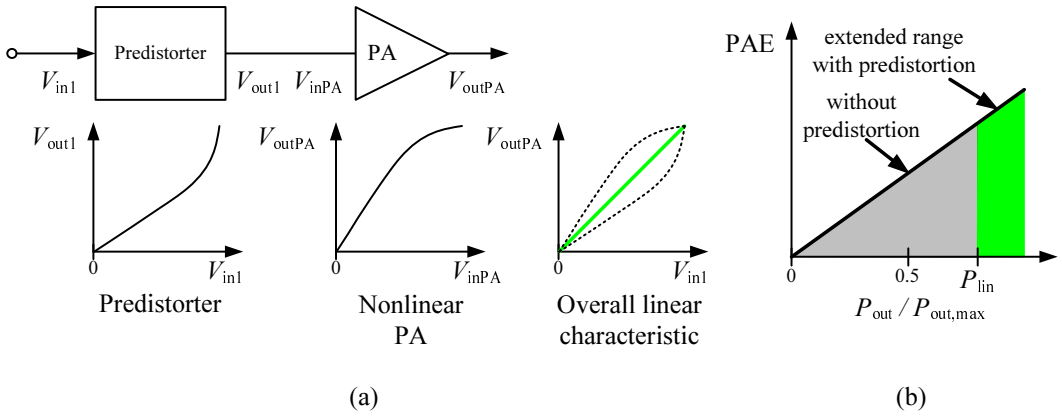


Figure 3.8: (a) Block schematic of linearization with predistortion. (b) Efficiency enhancement with Linearization

approach is therefore not suitable for handset applications. A much easier approach is simple predistortion. Fig. 3.8 (a) shows the principle block schematic. The nonlinear characteristic of the PA is compensated by a nonlinear block in front of the PA which generates a level dependent distortion opposite to the distortion of the PA. The correction has to be done for the amplitude (AM-AM) as well as for the phase (AM-PM). Fig. 3.8 (b) shows the benefit of the linearization. Without predistortion an operation only up to P_{lin} with sufficient linearity is possible. With predistortion this range can in theory be extended to the maximum output power of the saturated PA, still maintaining linearity.

The implementation of this nonlinear predistorter can be analog or digital. While analog predistortion is easy to implement (e.g. using a diode [Yamauchi 96]) the resulting linearity improvement is only in the range of 3...5 dB of ACLR [Hau 01]. Digital predistortion has higher flexibility. It can be implemented in base band using the signal processing infrastructure of modern transceivers. The problem of predistortion for usage within mobile handsets is mainly the determination of the amplifier's AM-AM and AM-PM characteristics under real operating conditions and process dependent device variations and VSWR. An elementary approach is to use prior laboratory measurements on one or several PA samples and write the characteristics into a static look-up table or calculate the characteristics in time, using complex polynomial expressions. Any changes in the characteristics will immediately degrade the correction process, if the complexity of an adaptive approach is omitted. These changes can be caused by memory effects, temperature drifts, device aging and process variations. A feed-back loop could compensate for these influences but then stability issues have to be considered.

Over all linearizing techniques (for efficiency improvement) in multi-mode power amplifiers are feasible but the benefit in efficiency enhancement will only be at peak power when the amplifier is in compression. It is not possible to improve PAE at large back-off which is necessary for WCDMA signals. Hence, for the application in multi-mode systems predistortion has to be combined with other techniques to

reduce the quiescent current of the amplifier such as stage size- and load adaption.

3.2.5 Stage size- and load adaption (SSLA)

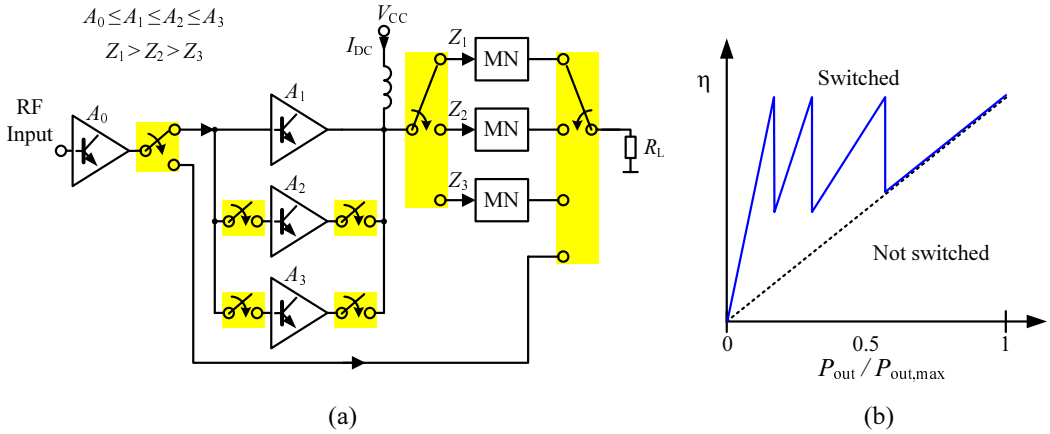


Figure 3.9: (a) Block schematic of PA with switched matching network, adapted transistor size and stage bypassing. (b) Theoretical efficiency enhancement at back-off.

This straight method to enhance efficiency at arbitrary back-off levels is based on direct adaption of transistor size and load impedance of the final amplifier stage at discrete switching levels. Fig. 3.9 (a) shows exemplary one of many possible implementations. Two improvements are offered from this concept.

First, the quiescent current consumption is reduced stepwise by turning off stages which operate far below their maximum output power capability. These amplifier stages are bypassed with other stages of smaller size or with a through connection directly from the first stage to the load matching network.

Second, the operating class of the amplifier is adapted at back-off by changing over different matching networks. This load impedance tuning can improve efficiency, even if the size of the final stage is not reduced. A relation between the efficiency at back-off and the load impedance will be derived in the following. The starting point for the analysis are the first two fourier components of the truncated current waveform through the transistor (see (2.7) and (2.8)). They can be written as:

$$I_{DC} = I_{max} \cdot f_1(\alpha) \tag{3.3}$$

$$\hat{I}_{RF} = I_{max} \cdot f_2(\alpha) \tag{3.4}$$

where I_{DC} and $\hat{I}_{RF}$ are the DC component of the current and the amplitude of the fundamental, respectively. The remaining fourier components are assumed to be shorted by the matching network. I_{max} is the maximum value of the current and $f_1(\alpha)$ and $f_2(\alpha)$ are functions of the conduction angle α. If the drive level reduces, the maximum current swing will reduce as well, which can be described by substitution of $I_{max} \rightarrow \frac{I_{max}}{A_{BO}}$, where A_{BO} specifies the amount of reduction. An optimum load

impedance is characterized by the condition that the voltage swing at the output for a specific amplitude of the fundamental current has an amplitude of V_{DC} (voltage range $0 \ldots 2V_{\mathrm{DC}}$): $R_{\mathrm{opt}} = V_{\mathrm{DC}}/\hat{I}_{\mathrm{RF}}$. With these assumptions, the output power P_{out} and the DC power consumption P_{DC} can be calculated as functions of the optimum load impedance and the amplitude reduction.

$$P_{\mathrm{out}} = \frac{1}{2}\, V_{\mathrm{DC}} \cdot \hat{I}_{\mathrm{RF}} = \frac{1}{2}\, R_{\mathrm{opt}}\, \frac{I_{\mathrm{max}}^2}{A_{\mathrm{BO}}^2}\, f_2^2(\alpha) \tag{3.5}$$

$$P_{\mathrm{DC}} = V_{\mathrm{DC}} \cdot I_{\mathrm{DC}} = V_{\mathrm{DC}}\, \frac{I_{\mathrm{max}}}{A_{\mathrm{BO}}}\, f_1(\alpha) \tag{3.6}$$

Hence, the efficiency at back-off follows from the above as:

$$\eta(A_{\mathrm{BO}}) = \frac{P_{\mathrm{out}}}{P_{\mathrm{DC}}} = \frac{1}{2}\, \frac{R_{\mathrm{opt}}}{A_{\mathrm{BO}}} \cdot \frac{I_{\mathrm{max}}}{V_{\mathrm{DC}}} \cdot \frac{f_2^2(\alpha)}{f_1(\alpha)} \tag{3.7}$$

This shows that with reduction of the current swing (increasing A_{BO}), efficiency reduces linearly. If the optimum load impedance R_{opt} is increased, in order to keep the value of the fraction $\frac{R_{\mathrm{opt}}}{A_{\mathrm{BO}}}$ constant, efficiency is kept constant with the drive level. As a consequence the continuous tuning of R_{opt} causes a nonlinear behavior of the amplifier. This results from (3.5), where one can observe a linear instead of a quadratic dependence between the output power and the current level. To avoid severe linearity degradation, it is better to switch the load at only few power levels to its optimum value and kept it constant over the new power range. As an abrupt load change during transmission may disrupt the connection, switching has to be performed during transmission breaks, e.g. between "frames" during the power regulation process with the base station.

The definition of power levels for switching and the subsequently adjusted load value require previous knowledge of the following signal and its maximum output power. At system level the complexity of the transceiver base band, especially the memory will increase when these requirements are implemented. In the case of multi-mode applications an adjustment of the amplifier can be done more conveniently when changing between transmission standards, as a precise knowledge of the instantaneous signal power is not necessary. The PA is then reconfigured to the peak power of the current standard (see Table 2.1). Considering GSM, EDGE and WCDMA standards the dynamic range of the amplifier from -40...36 dBm can effectively be divided into three ranges with two switching levels, defined by the peaks of the corresponding PDFs (Fig. 1.4). The first one will be 33...36 dBm for pure GSM operation, the second one 16...33 dBm for GSM/EDGE and WCDMA and the last one -40...16 dBm mainly for WCDMA in urban environments. In this regard the optimum impedances can be calculated, assuming a supply voltage of 3.5 V. For GSM and ideal class B operation $R_{\mathrm{opt}} = V_{\mathrm{DC}}^2/(2P_{\mathrm{out,max}})$ will be 1.53 Ω. At 3 dB power back-off ($A_{\mathrm{BO}} = 2$) it will be $R_{\mathrm{opt,max}} \cdot A_{\mathrm{BO}} = 3\ \Omega$, hence twice as large. In order to get similar efficiency at 16 dBm ($A_{\mathrm{BO}} = 100$) for WCDMA R_{opt} will be 153 Ω. These values show the difficulty in matching network design for high efficient multi-mode amplifiers. The 50 Ω output impedance has to be transformed to lower and higher values, requiring not only different element values but also different topologies.

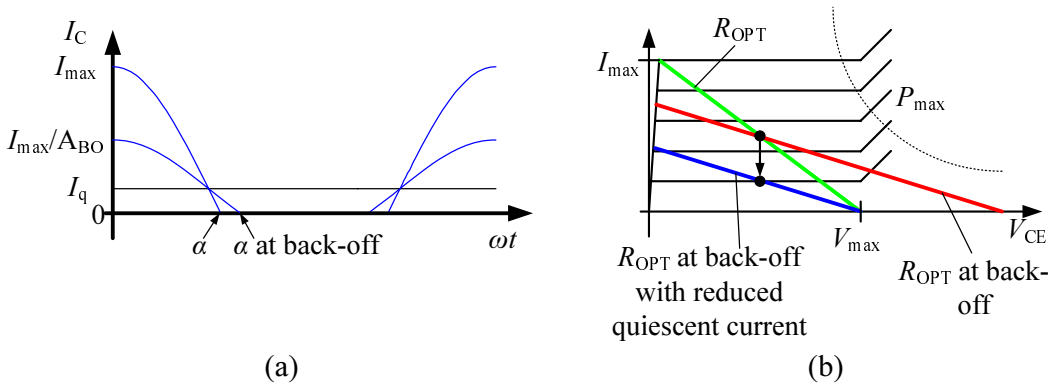


Figure 3.10: (a) Current waveform and conduction angle α at full swing and back-off. (b) Load line adaption for class A operation at full power and in back-off.

It has to be noted that the above analysis is only valid, if the current conduction angle does not change with the amplitude reduction in back-off and, if the DC power consumption reduces with the drive level. This is true for class B or AB operation with a small quiescent current I_q. Fig. 3.10 (a) illustrates this relation. If the conduction angle is very large as in the case of a class A amplifier the back-off efficiency cannot be increased only with load impedance changes. Here the power consumption is dominated by the quiescent current. Fig. 3.10 (b) shows that an increase of R_{opt} would result in voltage swings larger than V_{max} (breakdown voltage of the device). Hence, the quiescent current has to be reduced as well to shift the load line to a save operation area. This has also the advantage that for lower power levels the saturation voltage becomes smaller, hence when the peak current amplitude is maximum the instantaneous collector emitter voltage is smaller, and therefore less power is dissipated in the transistor.

Besides the optimum load impedance, also the transistor size has influence on the efficiency. The optimum operation point for the final amplifier's transistor cell depends not only on the overall current but also on the current density in each unit cell (see current versus f_T dependence in Chapter 4.1.1). In order to keep the current density constant, at low power levels a smaller transistor cell, which is a fraction of the large transistor, has to be used instead. Thereby parasitic capacitances are also reduced which do not have to be compensated by the matching network, wherefore the bandwidth of the matching network can be increased. The isolation between these transistor cells of different size has to be good not to degrade performance in each of the operation states.

The above described effects and benefits from the stage switching method demand series switches with good isolation in off state, small insertion loss in on state and linear behavior with small current consumption in both states. For integration into mobile terminals semiconductor switches (FET or diode) are used due to their small size and low cost, compared to mechanical switches. They have to pass large RF power, hence, being designed with a large power handling capability. When applied

directly at the output of large transistor stages they have to bear high currents (0.5...1 A). Moreover, a very small series on-resistance ($\ll 1\ \Omega$) is necessary for not reducing the voltage swing at the load ($\sim 1.5\ \Omega$ from previous calculation). An improvement could be to insert the switch not directly at the amplifier but closer to the 50 Ω output after a pre-matching network, where the overall impedance level is higher, current levels are smaller and the series resistance is more negligible. This circumstance is depicted in Fig. 3.11. The matching networks are assumed to be

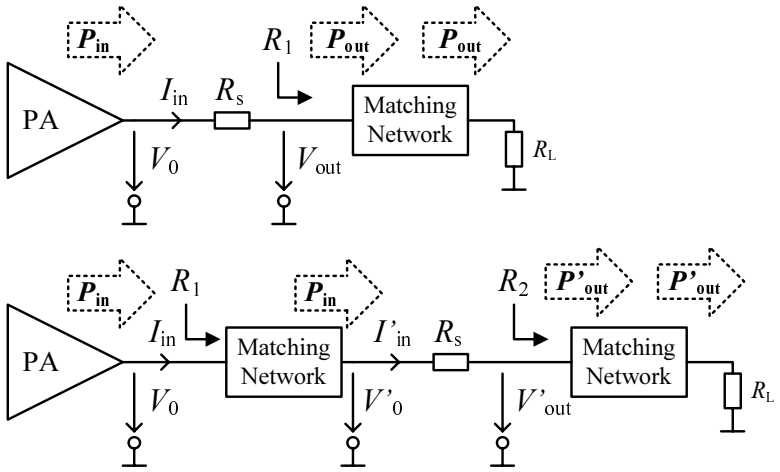


Figure 3.11: Block schematic for calculation of the effect of a series switch resistance R_s on the overall output power.

ideally lossless for the following derivation. In the first case the switch is applied before the matching network and the output power is given by:

$$P_\mathrm{out} = \bar{I}_\mathrm{in} \cdot \bar{V}_\mathrm{out} = \bar{I}_\mathrm{in}\ \bar{V}_0 \cdot \frac{R_1}{R_1 + R_\mathrm{s}} = P_\mathrm{in} \cdot \frac{R_1}{R_1 + R_\mathrm{s}} \tag{3.8}$$

A similar equation is obtained for the second case where the switch is placed between a pre-matching and an output matching network:

$$P'_\mathrm{out} = \bar{I}'_\mathrm{in} \cdot \bar{V}'_\mathrm{out} = \bar{I}'_\mathrm{in}\ \bar{V}'_0 \cdot \frac{R_2}{R_2 + R_\mathrm{s}} = P_\mathrm{in} \cdot \frac{R_2}{R_2 + R_\mathrm{s}} \tag{3.9}$$

The ratio is written as

$$\frac{P_\mathrm{out}}{P'_\mathrm{out}} = m \cdot \frac{R_1/m + R_\mathrm{s}}{R_1 + R_\mathrm{s}}, \tag{3.10}$$

with the impedance transformation ratio $m = \frac{R_1}{R_2}$. In Fig. 3.12 the power ratio is illustrated for $R_1 = 1.5\ \Omega$. Assuming a switch with a series resistance of only 0.4 Ω more than 0.9 dB output power can be saved, if it is placed after a pre-matching network with transformation ratio of $m = 0.1$ ($R_2 = 15$).

Instead of a series resistance, especially at higher frequencies, the losses of a switch are described by its insertion loss (IL in dB). Referring to Fig. 3.13 this is calculated

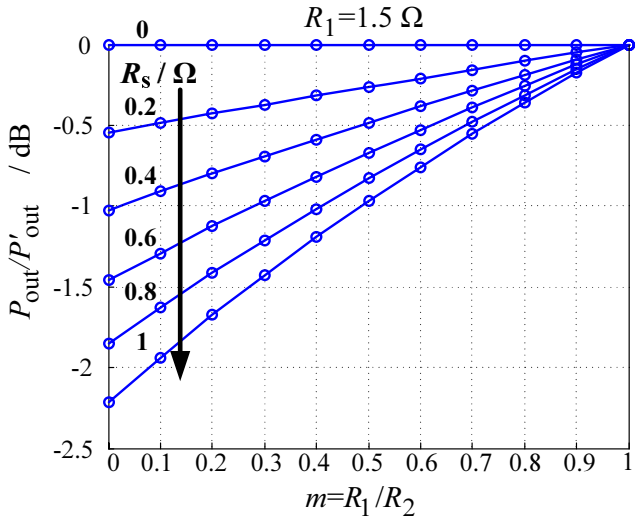


Figure 3.12: Output power reduction due to a series resistance R_s in the matching network positioned after an impedance transformation ratio of m.

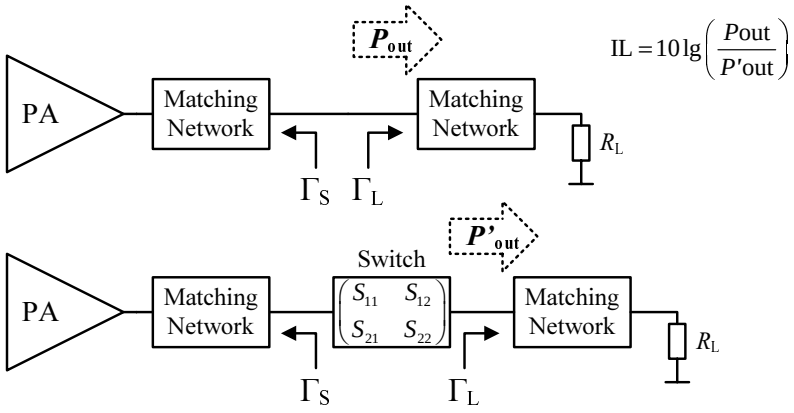


Figure 3.13: Block schematic for definition of the insertion loss (IL).

by:

$$\mathrm{IL} = 10\,\lg\left|\frac{(1-\Gamma_S S_{11})(1-\Gamma_L S_{22}) - \Gamma_S\Gamma_L S_{21}S_{12}}{S_{21}(1-\Gamma_S\Gamma_L)}\right|^2 \tag{3.11}$$

Hence, it also accounts for losses from reflections due to parasitic capacitances and inductances within the switch. For comparability, in data sheets, the IL is usually specified with $\Gamma_S = \Gamma_L = 0$, i.e. $\mathrm{IL} = 10\,\lg|S_{21}|^{-2}$. Typical values of 0.1...0.8 dB, measured within a 50 Ω related system, can be obtained.

With knowledge of the additional IL at the output of the amplifier, the reduction of PAE can easily be calculated with:

$$\frac{\mathrm{PAE}'}{\mathrm{PAE}} = \frac{G\;10^{-(\mathrm{IL}/10)} - 1}{G-1}\,, \tag{3.12}$$

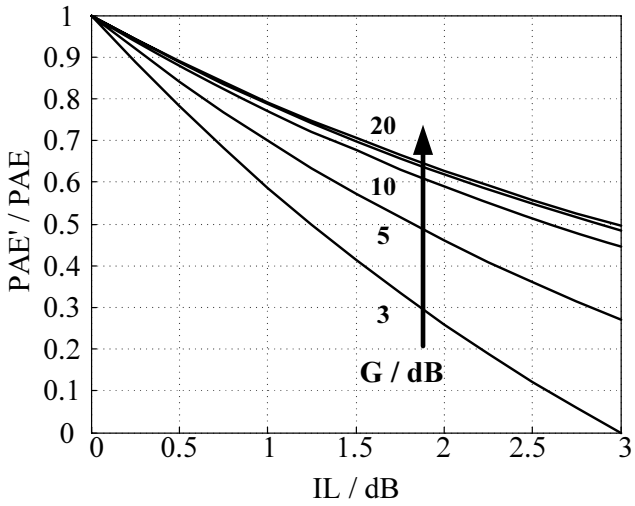


Figure 3.14: Reduction of PAE due to insertion loss (IL) at the output of the PA at different gain values.

Reference	Technology	IL @0.9 GHz	IL @1.9 GHz	Isolation dB	$P_{\max}$ dBm
[Tosaka 03]	GaAs-pHEMT	0.21 dB	0.35 dB	> 30	36
[Masuda 98]	GaAs-HJFET	0.50 dB	0.60 dB	> 20	38
[Tinella 06]	SOI CMOS	0.66 dB	0.79 dB	> 30	36
[Peregrine 08]	SOS CMOS	0.75 dB	0.80 dB	> 55	33
[Lee 05]	MEMS	0.23 dB	0.27 dB	> 48	–
[Infineon 08]	Si-PIN Diode	0.07 dB	0.08 dB	> 7	36

Table 3.2: Semiconductor switch technologies and performance comparison.

where PAE is the power added efficiency of the PA, PAE' is the power added efficiency reduced by the IL and G is the gain of the PA. For large gain values of the amplifier this becomes a relation of the collector efficiencies:

$$\lim_{G\to\infty} \frac{\mathrm{PAE}'}{\mathrm{PAE}} = \frac{\eta}{\eta'} = 10^{-(\mathrm{IL}/10)} \tag{3.13}$$

Fig. 3.14 illustrates equation (3.12) at several gain values. For $G > 10$ dB the curves are not significantly diverging from each other. If $G < 10$ dB, which might be the case if a single amplifier stage is operated in saturation, PAE decreases rapidly even for small IL. Depending on the gain of the amplifier an insertion loss of 0.5 dB reduces the overall PAE by 11...20%, which shows the significant effect of even small losses.

Table 3.2 lists actual bench marks of switches in different technologies for multi-mode handset applications. It has to be noted that all of the presented switches, except the PIN diodes are multi-pole switches, i.e. providing more than one input

or output. A fundamental problem in switch design is to maintain an acceptable insertion loss and isolation at the same time. The transistor's ON-resistance (R_{ON}) affects insertion loss and its OFF-capacitance (C_{OFF}) reduces isolation. A figure of merit (FOM) used to compare different technologies is the product $R_{\mathrm{ON}} \cdot C_{\mathrm{OFF}}$. Methods like stacking or parallel connection of transistors can reduce one of the two parameters, while the other is increased, hence the FOM remains constant. A trade-off between insertion loss and isolation has to be found for each application. For a small R_{ON} GaAs based pHEMT switches are favorably used due to their larger electron mobility compared to MOS devices. Good isolation up to 30 dB can be obtained when several of these transistors are connected in series. This reduces C_{OFF} and at the same time increases the breakdown voltage. Bandwidths covering all GSM and UMTS bands from 0.7...2 GHz are achievable without major degradation in performance. The current consumption for pHEMT switches is very low, typically < 100 μA. This is one reason why pHEMT switches are more frequently used in PA front-ends than PIN diode switches, which have by far smaller insertion loss but worse isolation and consume bias currents in the range of 5...20 mA. Compared to FET switches where the control voltage is essentially isolated from the main signal path [Gotch 07], the biasing of PIN diodes must be applied to the RF path which can result in reduced bandwidth or large size due to additional chokes. Additionally, electrical characteristics of PIN diodes vary with the drive level at large power, thus producing nonlinear distortion. In literature also MEMS based switches have been found which have about 20 dB better isolation than pHEMT with comparable insertion loss but are more costly in production.

The concept of stage size- and load adaption for application to multi-mode amplifiers is most attractive due to the direct choice of levels for reduction of power consumption. The transistor stages for each power range together with their matching networks can be optimized independently for the best trade-off between linearity and efficiency. On the other hand the integration of the PA together with several matching networks and switches into modules may become a costly issue. Besides the increasing size and complexity of the circuit also the use of separate technologies for the amplifier, the switches and the matching networks increase manufacturing costs. Solutions which manage the PA adaption with less or even without switches are preferable in this concern, even if the performance improvement is smaller. This is addressed in the next chapter, where possible benefits from topologies using power combiners are discussed.

3.2.6 Power Combining

To overcome shortcomings of semiconductor PA technologies, power combining is an efficient method to increase the amplifier's output power. The power of identical amplifier stages is coherently summed with a multi port power combiner network, as shown in Fig. 3.15 (a). The efficiency enhancement at power back-off is achieved by successively switching off M out of N amplifiers ($M < N$). The reduced number of active amplifiers mainly has two advantages. First the overall bias current is reduced

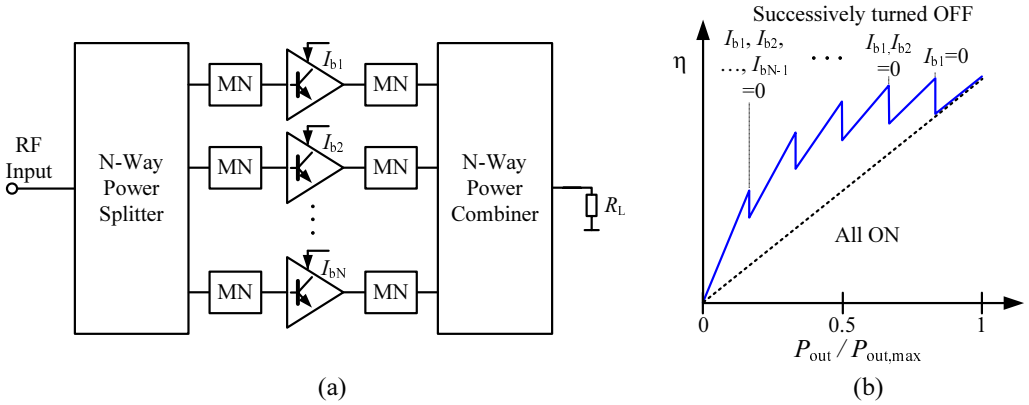


Figure 3.15: (a) Block Schematic of PA with power combining network (b) Theoretical efficiency enhancement at back-off.

and second each of the remaining amplifiers delivers higher output power at back-off and hence operates in a more efficient region. The achievable peak efficiency of the reconfigured PA is always smaller than at maximum output power and reduces with the number of active amplifiers. This characteristic is shown in Fig. 3.15 (b) and is one of the main differences to the stage size adaption concept with switches. The reason for this can be found in the reduced combining efficiency of the power combiner network. The combining efficiency is defined as the ratio of combiner output power to the total combiner input power. With all amplifiers being active this ratio subsumes the *resistive combiner loss* and losses introduced by *amplitude* and *phase errors* due to not identical amplifiers and imperfect symmetry of the network. A quantitative analysis of these errors can be found in Appendix A.1. With only a few of all amplifiers being active, additional losses occur due to *mismatch* caused by impedance changes (load pulling) at the ports of active amplifiers and due to *dissipated power* at the deactivated amplifiers. This reduction of combiner efficiency is known as "graceful degradation" and was earlier described in [Ernst 77, Saleh 80, Eom 94] for the case of device failure in a combined linear amplifier structure. Since device failure changes the amplifier's output impedance arbitrarily, the proposed solutions for improvement of combiner efficiency are very complex, targeting on isolating the failed amplifiers and replacing them by well defined terminations. The new approach in this thesis is to intentionally deactivate high power amplifiers in a reproducible manner to obtain well defined almost reactive impedances, which increase combining efficiency, without need of additional circuitry.

Three-port combiner structures

Following, this concept is shown at the example of a reciprocal three port power combiner network, as depicted in Fig. 3.16 (a). Two states are analyzed. In state 1 both amplifiers are turned ON, in state 2 the amplifier at port 2 is turned OFF. The scattering matrix approach is chosen to obtain general equations for combining efficiency and mismatch. These will help to draw general conclusions on properties of possible highly efficient implementations.

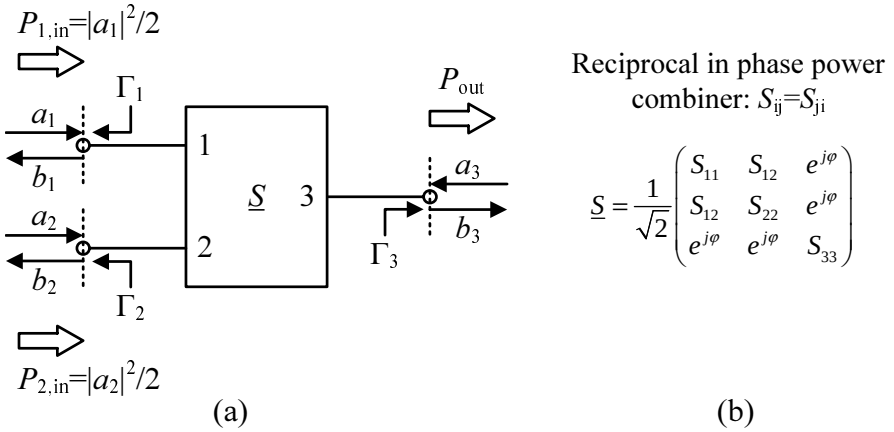


Figure 3.16: (a) Block Schematic of a general three port power combining network (b) Scattering matrix.

Combining efficiency

In a power combiner network the total voltage at the output is the vectorial sum of the individual voltage components resulting from each amplifier at the input. Since the voltage components are proportional to the square root of the power components $P_{i,\text{out}}$, the total output power P_{tot} for coherent summation can be calculated as:

$$P_{\text{tot}} = \left(\sum_{i=1}^{n} \sqrt{P_{i,\text{out}}} \right)^2 , \tag{3.14}$$

where n is the number of active amplifiers. This equation is only valid, if the input signals are properly phased to get in-phase summation through the combiner at the output. Input signals with phase differences or combiners with different phase relations between input and output ports cannot be analyzed this way.

The individual power contributions $P_{i,\text{out}}$ are calculated with help of the transducer gain:

$$P_{i,\text{out}} = P_{i,\text{in}} \cdot G_{\text{T},3i} \tag{3.15}$$

The transducer gain is a function of the reflection coefficients Γ_i at each port and is always smaller than 1 for a passive network. Here it is calculated for a source at port 1 and output at port 3:

$$\begin{aligned} G_{\text{T},31} &= \frac{\text{Power delivered to the load}}{\text{Power available from the source}} = \frac{\frac{|b_s|^2}{2} \cdot \left|\frac{b_3}{b_s}\right|^2 (1-|\Gamma_3|^2)}{\frac{|b_s|^2}{2}\frac{1}{1-|\Gamma_1|^2}} \\ &= \left|\frac{b_3}{b_s}\right|^2 \cdot \left(1-|\Gamma_3|^2\right)\left(1-|\Gamma_1|^2\right) \end{aligned} \tag{3.16}$$

Due to symmetry of the scattering matrix this is equal to $G_{\text{T},32}$ between port 2 and port 3 for appropriate Γ_i. To find a general expression for the transmission factor

$\frac{b_3}{b_s}$, Mason's loop rule is applied (see Appendix B.2).

$$\frac{b_3}{b_s} = \frac{1}{\Delta}\frac{e^{j\varphi}}{\sqrt{2}}\left(\Gamma_2\left(S_{12} - S_{22}\right) + 1\right), \tag{3.17}$$

with

$$\begin{aligned}\Delta =& \Gamma_3\left\{\Gamma_1\left[\Gamma_2\left(e^{j\,2\varphi}\left(\frac{S_{22}+S_{11}}{2} - S_{12}\right) + S_{12}^2 S_{33} - S_{11}\,S_{22}\,S_{33}\right) + \right.\right. \\ & \left.\left. S_{11}\,S_{33} - \frac{e^{j\,2\varphi}}{2}\right] + \Gamma_2\left(S_{22}\,S_{33} - \frac{e^{j\,2\varphi}}{2}\right) - S_{33}\right\} + \\ & 1 - S_{22}\,\Gamma_2 + \Gamma_1\left(\Gamma_2\left(S_{11}\,S_{22} - S_{12}^2\right) - S_{11}\right)\end{aligned} \tag{3.18}$$

The *combining efficiency* can be defined as:

$$\eta_{\mathrm{C}} = \frac{\left(\sum_{i=1}^{n}\sqrt{P_{i,\mathrm{out}}}\right)^2}{\sum_{i=1}^{n} P_{i,\mathrm{in}}} \tag{3.19}$$

Due to the network symmetry and with equal input powers this becomes:

$$\eta_{\mathrm{C}} = \frac{P_{i,\mathrm{in}}\left(\sum_{i=1}^{n}\sqrt{G_{\mathrm{T},3i}}\right)^2}{n\,P_{i,\mathrm{in}}} = n\,G_{\mathrm{T},31} \tag{3.20}$$

For the two operation cases with one or two active amplifiers it is assumed that amplifier port 1 and output port 3 are always matched with the reference impedance ($\Gamma_1 = \Gamma_3 = 0$). Only port 2 has a variable impedance, depending on the amplifier state (ON/OFF). On these conditions equation (3.18) simplifies and the combining efficiency can be written as:

$$\eta_{\mathrm{C}} = \frac{n}{2}\left|\frac{1 - \Gamma_2\left(S_{22} - S_{12}\right)}{1 - \Gamma_2\,S_{22}}\right|^2 \tag{3.21}$$

If both amplifiers are turned ON ($n = 2,\ \Gamma_2 = 0$), the ideal network has a combining efficiency of 100 %. If amplifier 2 is turned OFF ($n = 1$), its output impedance has strong influence on the combining efficiency, as long as $S_{12} \neq 0$, otherwise $\eta_{\mathrm{C}} = 50$ %. This means that an improvement of combining efficiency can only be obtained, if the combining network has no isolation between its input ports 1 and 2.

Mismatch and load pulling

For calculation of the combining efficiency it was assumed that the available power from the working amplifiers is constant. That would be the case, if the amplifiers are linear. If they are nonlinear, the so called mismatch or load pulling effect changes the available power from the working amplifiers by additional load impedance changes. To keep the additionally introduced shifts in power and efficiency small, the deviation from the nominal load impedance for each working amplifier has to be small as well. The nominal load impedance $Z_{\mathrm{nom},i}$ is here defined as the impedance seen into port i of the combiner network at maximum output power, i.e. both amplifiers are ON. This

impedance is not necessarily 50 Ω and depends on the type of combiner network. The amount of mismatch can best be described as relative measure ΔZ_i, which describes the normalized vector distance between the actual impedance Z_i and $Z_{\mathrm{nom},i}$:

$$\Delta Z_i = \frac{|Z_i - Z_{\mathrm{nom},i}|}{|Z_{\mathrm{nom},i}|} \quad \text{with} \quad Z_i = \frac{1+\Gamma_{i,\mathrm{in}}}{1-\Gamma_{i,\mathrm{in}}}. \tag{3.22}$$

This definition considers that the PA is not necessarily matched to 50 Ω in the nominal state and in this concern is more meaningful than the definition of VSWR. The input reflection coefficient at port 1 is given by $\Gamma_{1,\mathrm{in}} = b_1/a_1$. Mason's loop rule is used again to obtain a general expression for the three port power combiner:

$$\Gamma_{1,\mathrm{in}} = S_{11} + \frac{\Gamma_3 \left\{ \Gamma_2 \left[\frac{e^{j\,2\varphi}}{2} \left(S_{22} - 2\,S_{12}\right) + S_{12}^2 S_{33} \right] - \frac{e^{j\,2\varphi}}{2} \right\} - S_{12}^2 \Gamma_2}{\Gamma_3 \left[\Gamma_2 \left(\frac{e^{j\,2\varphi}}{2} - S_{22}\,S_{33} \right) + S_{33} \right] + S_{22}\,\Gamma_2 - 1} \tag{3.23}$$

In case of a matched output port ($\Gamma_3 = 0$) this expression simplifies to:

$$\Gamma_{1,\mathrm{in}} = S_{11} + \frac{S_{12}^2 \Gamma_2}{1 - S_{22}\,\Gamma_2} \tag{3.24}$$

The input impedance is constant and independent of Γ_2, only if the two input ports are isolated from each other ($S_{12} = 0$). This conflicts with the demand of obtaining maximum combining efficiency ($S_{12} \neq 0$). Hence, during the design of the combiner network a compromise for both conditions has to be found.

Quantitative analysis

The derived compact formulas allow a fast comparison of different combining networks for their potential use with the efficiency enhancement concept. Only the scattering matrix has to be known. Without loss of generality the combining efficiency and mismatch are calculated exemplarily for a Wilkinson combiner (subscript W) and a lossless transmission line combiner (subscript L) with scattering matrices (see Appendix A.3 and Appendix A.4):

$$S_{\mathrm{W}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & -j \\ 0 & 0 & -j \\ -j & -j & 0 \end{pmatrix} \quad \text{and} \quad S_{\mathrm{L}} = \frac{1}{2} \begin{pmatrix} 1 & -1 & -j\sqrt{2} \\ -1 & 1 & -j\sqrt{2} \\ -j\sqrt{2} & -j\sqrt{2} & 0 \end{pmatrix}. \tag{3.25}$$

The amplifier at port 2 is considered to be OFF with arbitrary impedance. Using (3.21) the combining efficiencies equate to:

$$\eta_{\mathrm{W}} = 50\ \% \quad \text{and} \quad \eta_{\mathrm{L,max}} = 2\,\frac{1 - 2|\Gamma_2|\cos(\varphi_2) + |\Gamma_2|^2}{4 - 4|\Gamma_2|\cos(\varphi_2) + |\Gamma_2|^2}\bigg|_{\Gamma_2=-1} = \frac{8}{9} = 89\ \% \tag{3.26}$$

The mismatch, related to the nominal impedance follows from (3.22) as:

$$\Delta Z_{\mathrm{W},1} = 0\ \% \quad \text{and} \quad \Delta Z_{\mathrm{L},1} = \left| \frac{2\,\Gamma_2}{3(1-\Gamma_2)} \right| \Bigg|_{\Gamma_2=-1} = \frac{1}{3} = 33\ \% \tag{3.27}$$

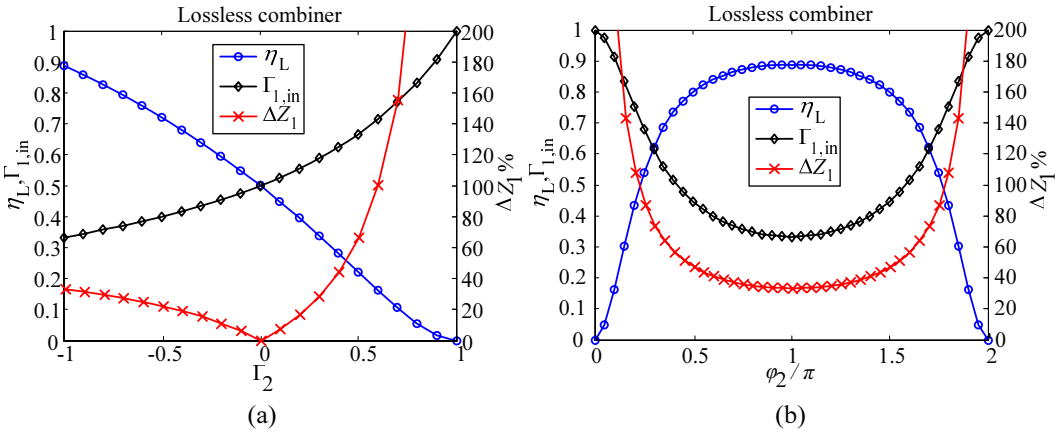


Figure 3.17: Lossless 3-port power combiner with one active amplifier: (a) Γ_2 sweep with $\varphi_2 = 0$, (b) φ_2 sweep with $|\Gamma_2| = 1$.

For the wilkinson combiner all parameters are constant due to isolation between ports 2 and 3. In contrast, these ports are strongly coupled in the lossless transmission line combiner, so it has varying parameters, as expected. Fig. 3.17 displays the varying performance versus Γ_2. An improvement of combining efficiency to a maximum of 89 % is achieved but at the expense of 33 % load mismatch ($\Gamma_2 = -1$). About this optimum the sensitivity to the phase of Γ_2 is smaller than to the magnitude, which eases the implementation. Nevertheless, the PA plus matching network at port 2 must have almost zero losses with about the phase of a SHORT when turned OFF. The phase can easily be adjusted with a transmission line, while the losses are basically predetermined by the technology in use.

These calculations for standard combiner circuits show that the impact of the inactive PA on the mismatch is considerable. If Γ_2 is matched to the maximum of η_C, the altering load degrades the efficiency and output power of the active PA according to its load pull contours. In worst case it even cancels out the improvements in combining efficiency, deteriorating the overall efficiency. A performance optimum, considering both effects has to be found for specific realizations as there exists no simple general solution including the load-pull effect on the amplifier itself.

Benefit of non-reciprocal combiner network
The possibility to adjust combining efficiency without altering the input impedance is offered by a *non-reciprocal* combiner network. Applying the same calculations as above but using $S_{12} \neq S_{21}$ leads to following equations, if $n = 1$ and Γ_1, $\Gamma_3 = 0$:

$$\eta_C = \frac{1}{2} \left| \frac{1 - \Gamma_2 \, (S_{22} - S_{21})}{1 - \Gamma_2 \, S_{22}} \right|^2 \quad \text{and} \quad \Gamma_1 = S_{11} + \frac{S_{12} \, S_{21} \Gamma_2}{1 - S_{22} \, \Gamma_2}. \tag{3.28}$$

An ideal network with $S_{12} = 0$ leads to the desired parameter independence. A combiner with such properties is obtained by adding an isolator at port 1 in Fig. 3.16 (a) to terminate the backscattered wave there. The scattering matrix has following form:

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 \\ S_{21} & S_{22} & e^{j\,\varphi} \\ e^{j\,\varphi} & e^{j\,\varphi} & S_{33} \end{pmatrix} \tag{3.29}$$

Although the performance could be substantially improved, it is not possible to implement such a non-reciprocal network with isotropic, passive components. For integration into semiconductor chips this solution would be too large and costly, thus it is not further considered.

Key design trade-offs
Summing up, following design compromises have to be considered for a 3-port combiner network to improve the PA efficiency at more than 3 dB back-off:

- Matched ports are non-essential ($S_{ii} \neq 0$) as long as the unity PAs can be matched to their optimum impedance.
- In a reciprocal network the coupling between input ports has to be small ($|S_{21}| = 0$) to avoid load pull effects (small ΔZ value).
- In a reciprocal network the coupling between input ports has to be large ($|S_{21}| = 1/\sqrt{2}$) to improve coupling efficiency (large η_C value).
- Small isolation between input ports (large $|S_{21}|$) might increase the possibility of spurious oscillations.
- A lossless network improves combining efficiency in back-off but causes the largest load pull effect.
- With a non-reciprocal network η_C can be improved while ΔZ is constant.

Extension to N-Port combiner structures
To improve efficiency at back-off, a properly designed three port combiner network offers the possibility for one switching level at which the PA reconfigures. This switching level can be 3 dB below the peak power or even lower, if a combiner with unequal split ratio is used. For multi-mode operation it is of importance to have more than one high efficient BO state. Therefore an extension to an N-port combiner structure is desired. Without proof it is stated that above findings are

accordingly valid for N-port in-phase combiner networks with $n = N - 1$ input ports and one output port.

Two main groups can be distinguished, namely the parallel (or n-way) combining structure and the binary (or tree, corporate) cascaded combining structure. Their differences concerning combining efficiency will be discussed in the following.

N-port parallel combining structure
This group of networks subsumes topologies where all voltage contributions are added at one node in the network. Examples are the T-junction, the transmission line, the Wilkinson and the transformer based combiner. The design equations together with calculations of maximum possible combining efficiency and smallest relative load mismatch are presented in Appendix A. Here, these results are compared in Table 3.3 for the special case of $N = 5$, i.e. $n = 4$ parallel amplifiers. The calculations were done for the assumption that three amplifiers are turned off, because this case shows the maximum performance improvement at smallest back-off power P_{BO}. All networks, except the Wilkinson combiner are lossless in theory. Loaded with proper impedances, they have equal combining efficiencies. The Wilkinson combiner has large isolation between its ports. Hence, its combining efficiency is worse than for the other networks. The power is lost in the isolation resistors. Its relative mismatch is zero for the same reason. For the lossless networks the smallest ΔZ is obtained for the transmission line and transformer based combiners. They provide the best compromise between η_{C} and ΔZ for usage within this concept. As the transformer based combiner also has advantages for integration on chip due to smaller area, larger bandwidth, immunity to interferers and suppression of even order harmonics, this will be preferred for an implementation. An efficiency enhanced power amplifier which uses the transformer based combiner network was reported in literature [Haldi 08, Liu 06, Liu 08]. In these publications one can also find calculations of combining efficiency and mismatch but they are based on network equations of the implemented circuit. A generalized analysis with scattering parameters, as shown here, was not included and could not be found in other literature. This generalized analysis offers the possibility of fast and convenient comparison with other networks to find the one with best properties for the application.

Combiner type	P_{BO}	η_{C}	ΔZ	Γ_{ON}	Γ_{OFF}	$\Gamma_{N,\mathrm{opt}}$
T-junction	-7.96 dB	64%	75%	0	1	$-3/5$
Transmission line	-7.96 dB	64%	42.86%	0	-1	0
Wilkinson	-12.04 dB	25%	0%	0	-1	0
Transformer based	-7.96 dB	64%	42.86%	0	-1	0

Table 3.3: Comparison of N-port parallel combining structures for $N = 5$ and one active input amplifier. Γ_{ON}, Γ_{OFF}: reflection coefficients, seen into the amplifiers in ON and OFF state. $\Gamma_{N,\mathrm{opt}}$: optimum load reflection coefficient at the output for maximum power transfer, if all amplifiers are ON.

N-port binary cascaded combiner structure
A simple way to extend the number of active amplifier stages is to use identical 3-port power combiners cascaded as shown in Fig. 3.18 (a). Consequently, the number

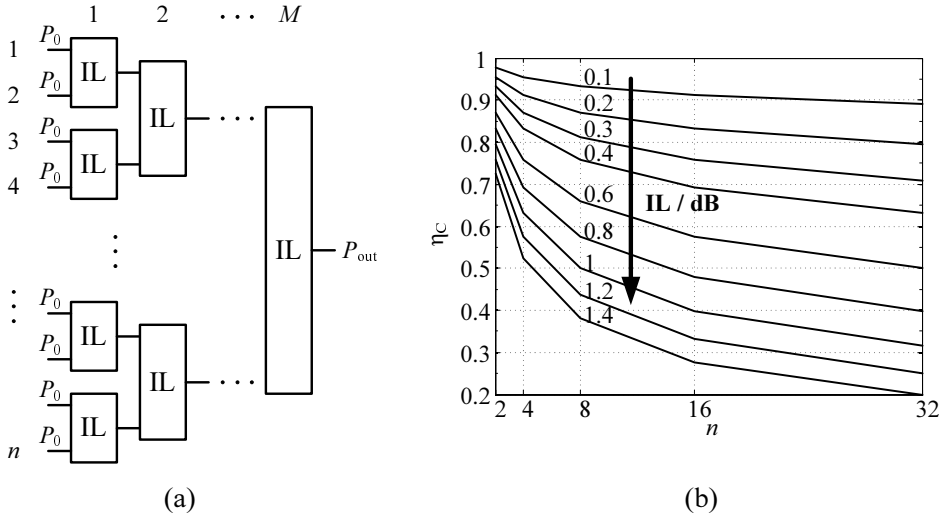


Figure 3.18: Binary cascaded combiner structure: (a) Block diagram (b) efficiency degradation per cascaded stage.

of input ports n can only be a multiple of 2, that is why this structure is called a binary cascaded combiner. With the loss per stage IL the total output power equates to:

$$P_{\text{out}} = 2\,(\,2\,(\,2\,(\,\ldots)\text{IL}\,)\text{IL}\,)\,P_0\,\text{IL} = (2\ \text{IL})^M P_0, \tag{3.30}$$

where $M = \log_2 n$ is the number of cascaded stages. Related to the input power this gives the combining efficiency for the binary-tree combiner structure:

$$\eta_{\text{C}} = \text{IL}^{\log_2 n} \tag{3.31}$$

Each additional stage decreases the combining efficiency by IL. Fig. 3.18 (b) shows a plot versus n for various values of IL. One can see that the network becomes very inefficient for a large number of amplifiers, due to the exponential increase.

In a theoretical comparison, for an N-port *parallel* combiner, the maximum combining efficiency versus N is constant IL. Thus, the parallel combiner would be preferred over the binary combiner structure. But in practice, losses usually increase with the number of ports N, due to e.g. space constraints or shortcomings of the technology in use. The decrease cannot be described in general, it may be linear or even of higher order. Taking this into consideration, the value η_{C} in a binary combiner, especially for small N, is not necessarily smaller than in a parallel combiner topology. Therefore, if the number of input amplifiers is predetermined, the choice of combiner group with largest η_{C} has to be made by comparison.

Apart from the combiner losses at maximum output power, the characteristics of η_{C} or ΔZ in the reconfigured PA at power back-off are a further decision support for the

group of combiner topology. In general, the characteristics will not be equal for *binary* or *parallel* combined networks, which use the same basic combiner type. For example, while a T-junction, Wilkinson or transformer based combiner gives the same results as shown in Appendix A, the transmission line combiner has worse characteristics when used in a binary combined topology. The reason for this is the preassumption, which has been made for all calculations. To guarantee symmetry, all amplifiers are assumed to have the same matching networks and therefore have equal output impedances in ON or OFF state. For the transmission line combiner a $\Gamma_{\mathrm{OFF}} = -1$ for all amplifiers, is not appropriate anymore. It has to be adapted for each state and each amplifier separately, to obtain similar values for η_{C} and ΔZ in a binary combined structure. This can only be realized with adaptive matching networks, but they have been excluded in this concept, due to the required switches. Consequently, not all lossless 3-port combiners can be used to build up *binary* combined PAs with switch-less efficiency improvement at BO. Combiner networks with an additional port can overcome the problem of using adaptive matching networks. For example in a 4-port quadrature hybrid, the unused port can be loaded with proper impedances to tune the Γ_{OFF} values and therefore improve η_{C}. This application will be described in more details in later chapters where it was used to realize the multi-mode power amplifier. It is worth noting that this kind of tuning can be generalized to N-port power combiner networks with $n > N - 1$ input ports and 1 output port. The remaining $N - (n + 1)$ passive ports have to be isolated from the input and output and are used to adapt the combiner, when reconfigured in back-off states.

The discussion of efficiency enhancement with symmetrical power combining topologies has shown several possibilities of implementation without need of switches. Upper bounds for performance improvement have been derived, which indicate that essential increase of efficiency can be gained within a back-off power range of ~ 6 dB. At BO power > 10 dB an efficiency improvement is still present, but at a smaller amount. The complexity and costs of this concept are moderate, in addition to the PA a proper designed power combiner network and logic in the biasing circuitry are sufficient - no additional active RF components like switches are necessary. The performance of this concept and the potential for application in a multi-mode PA is comparable to the LINC, Doherty or supply variation concept, where a considerable efficiency improvement is achieved only in the upper 6 dB back-off power range. However, the advantage of this concept over the others is its easy and cheap implementation and the high potential for integration.

3.2.7 Comparison of Concepts

In the following section a comparison of the previously presented efficiency improvement concepts is made, to put them in order of relevance for multi-mode applications. This ranking is based on calculations of average DC current consumption for GSM/EDGE and WCDMA and contrasted to the effort in a practical realization. So far, there is no reported study comparing head to head all the previously described concepts. Based on a compilation of the results presented in open literature

in handset PA contexts, an attempt to fairly compare the different concepts is done. As the results are likely to be device dependant, idealized nonlinear models for the dynamic efficiency curves are used for all concepts.

The average DC current consumption is found by weighting of the current consumption characteristics versus output power with specific probability density functions (PDF) of the transmit standards. For this purpose the current consumption can easily be derived from the collector efficiency by fixing a certain supply voltage. Analytical expressions for $\eta(P_{\text{out}})$ together with preassumptions made for each PA type are listed in the following:

Linear class AB
This is the reference PA for the comparison, as it describes a classical single ended PA, operated in class AB to assure linear operation for WCDMA signals. The maximum collector efficiency $\eta_{\max}$ is chosen to be 70 %. The relations are given by:

$$\eta = \eta_{\max}\sqrt{\frac{P_{\text{out}}}{P_{\max}}}\,, \tag{3.32}$$

with P_{out} being the instantaneous output power and $P_{\max}$ being the maximum transmission power.

Doherty PA
The derivation of collector efficiency for a 2-branch Doherty PA is given in [Srirattana 05]. It includes the effect of impedance changes over output power but is based on ideal assumptions:

$$\eta = \begin{cases} \eta_{\max}\dfrac{k_1\frac{P_{\text{out}}}{P_{\max}}}{(k_1+1)\sqrt{\frac{P_{\text{out}}}{P_{\max}}}-1} & \text{if} \quad \dfrac{P_{\max}}{k_1^2} < P_{\text{out}} < P_{\max} \\ \eta_{\max}\sqrt{\dfrac{P_{\text{out}}}{P_{\max}}}\,k_1 & \text{if} \quad 0 < P_{\text{out}} < \dfrac{P_{\max}}{k_1^2} \end{cases} \tag{3.33}$$

The constant k_1 describes the back-off power level at which the peaking amplifier becomes active. It is adjusted by the size ratio of carrier and peaking amplifier. Here, $k_1 = 1.4$ is assumed.

Outphasing PA (LINC)
In [Hakala 05], a closed form equation is derived, based on the combining efficiency of an ideal LC Chireix power combiner. The back-off power level for the amplifier's roll off is defined by the constant k_2. It is adjusted by the L and C values of the combiner network and has a value of $k_1 = 0.9$ for the current investigations.

$$\eta = \frac{\eta_{\max}}{\sqrt{1+\frac{1}{4}\left(\frac{\sin 2\theta - k_2}{\sin^2\theta}\right)^2}}, \quad \text{with } \theta = \arcsin\left(\sqrt{\frac{P_{\text{out}}}{P_{\max}}}\right) \tag{3.34}$$

Supply variation PA
An empirical equation can be defined, which is based on a comparison to simulations

of the characteristic [Bakalski 08,a]:

$$\eta = \begin{cases} \dfrac{\eta_{\max}}{k_3 \left(\sqrt{P_{\max}} - \sqrt{P_{\text{out}}}\right)^{k_4} + 1} & \text{if} \quad P_{\text{b}} < P_{\text{out}} < P_{\max} \\ \eta_{\max} \sqrt{\dfrac{P_{\text{out}}}{P_{\max}}}\, k_5 & \text{if} \quad 0 < P_{\text{out}} < P_{\text{b}} \end{cases} \tag{3.35}$$

Depending on the transistor type, the constants k_3, k_4 and k_5 have to be adjusted to match the shape of the curve. $P_{\text{b}} = f(P_{\max}, \eta_{\max}, k_3, k_4, k_5)$ is the intersection point of the two η branches, described by a nonlinear, implicit function which can only be solved numerically. The efficiency of the DCDC converter is included in $\eta_{\max} = \eta_{\text{DCDC}}\, \eta_{\text{PA}}$.

Linearizing
This amplifier type is based on a single ended standard class B PA, which has a maximum collector efficiency of 78.5 %. It is assumed that, due to linearization, also EDGE and WCDMA signals can be transmitted without changing the class of operation. So, as an approximation, the formula for η is the same as for the linear class AB PA with a larger value for $\eta_{\max}$.

Switched-combined PA
This indicates a PA, consisting of three amplifiers, which can reconfigure in a way to improve efficiency at two power back-off levels. Two amplifiers are combined with a transformer, like in the power combining concept and the third one, which is smaller than the others, is added similar to the stage size- and load adaption concept. The latter amplifier is only active in the lowest power range. In the first back-off state one of the transformer combined amplifiers is switched OFF to benefit from the power combining concept. In the second back-off state only the smallest amplifier is activated to get a significant improvement of efficiency. The characteristic in each part of the η curve follows the formula of the linear class AB PA. The values for $\eta_{\max}$ in descending order are: 78.5 %, 69.8 % and 35 %.

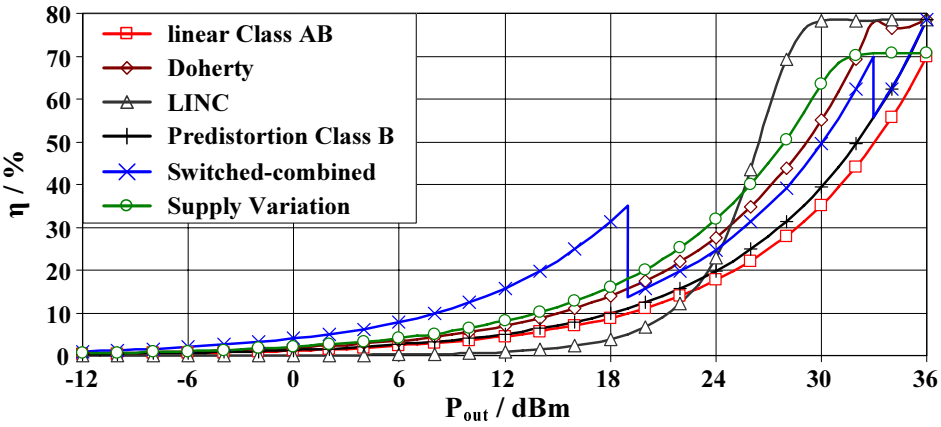


Figure 3.19: Comparison of theoretically achievable collector efficiency curves.

Fig. 3.19 shows a plot of all collector efficiency curves versus output power in one diagram. The only concept which has remarkable efficiency improvement at more than 15 dB back-off power is the switched-combined type.

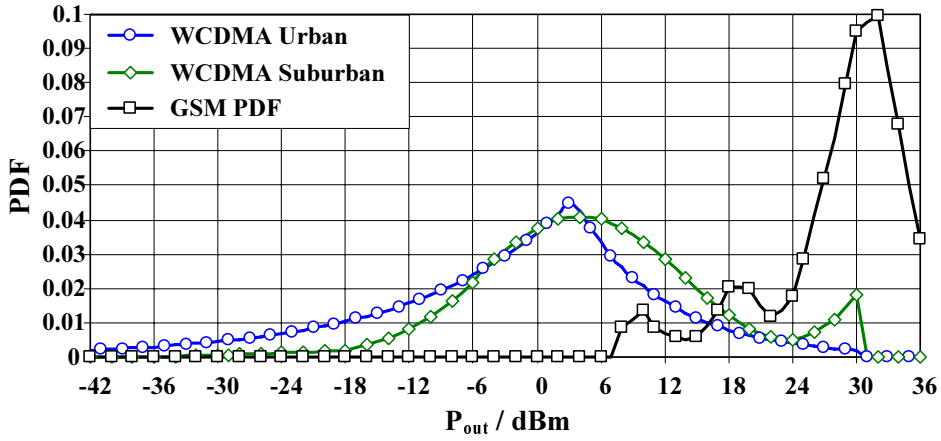


Figure 3.20: Probability density functions (PDF) of output power for WCDMA urban, suburban and GSM [Egolf 07].

However, for a meaningful comparison an average current consumption has to be defined. This includes an average over the signal power statistics and a time average. For the first average, the actual DC current $I = P_{\text{out}}/\left(V_{\text{CC}} \cdot \eta\right)$ at a typical supply voltage of $V_{\text{CC}} = 3.5$ V is weighted with empirical data of the PDF, shown in Fig. 3.20, and integrated over the operation power range:

$$\overline{I} = \int \left(I\left(P_{\text{out}}, \eta\left(P_{\text{out}}\right)\right) \cdot \text{PDF}(P_{\text{out}})\right)\, dP_{\text{out}} \tag{3.36}$$

For the second average, $\overline{I}$ is integrated over the operation duration of the considered transmission standards. In this regard the usage of GSM/EDGE or WCDMA over time is hard to predict and depends amongst others on the user profile, costs and availability of services. Users, which use their phone for voice will have long GSM operation time, while laptop users with data traffic will have long WCDMA operation time. To define conditions, which are as general as possible, it is assumed that the multi-mode amplifier is operated in each standard for the same duration. So, for this general case, a mean value $\overline{I_{\text{mean}}}$, which is the arithmetic average of all $\overline{I}$, is calculated:

$$\overline{I_{\text{mean}}} = (\underbrace{D_{\text{GSM}}\, \overline{I}_{\text{GSM}}}_{\overline{I}'_{\text{GSM}}} + D_{\text{WCDMA}}\, \overline{I}_{\text{WCDMAurb}} + D_{\text{WCDMA}}\, \overline{I}_{\text{WCDMAsub}})/3 \tag{3.37}$$

The factors D are the duty-cycles of the respective standards. GSM/EDGE has a pulsed operation and the PA is active for only $t_{\text{ON}} = T/8$ of the transmit time T, hence $D_{\text{GSM}} = t_{\text{ON}}/T = 12.5$ %. WCDMA has no pulsed operation, hence $D_{\text{WCDMA}} = 100$ %.

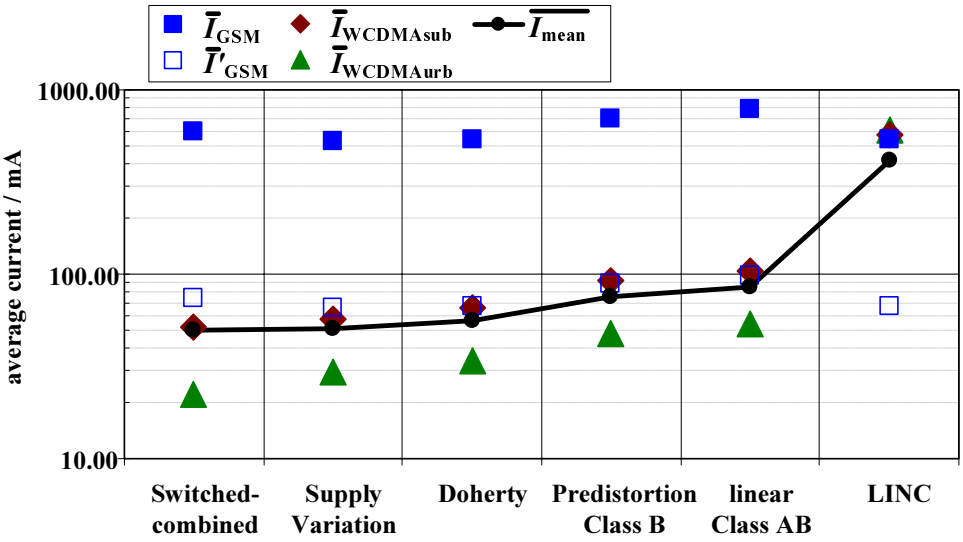


Figure 3.21: Average DC supply current weighted with WCDMA and GSM PDF.

Fig. 3.21 shows the mean supply current values for power and time average. Although the probability for large output power and therefore large current consumption is higher in GSM, the time average value is in the same order of magnitude as for WCDMA. This is due to the smaller duty-cycle of GSM. If the duty cycle was not considered, GSM would have an approximately 10 times larger current consumption and therefore dominate the concept ranking. As a consequence, concepts with high peak power efficiency would be preferred over the others and the need for improved back-off power efficiency would not be given. The concepts which are plotted on the x-axis of the figure are ranked by the values of $\overline{I_{\text{mean}}}$, including the duty-cycle. In this case it is clearly seen that the lowest average current is achieved by concepts which have an improved efficiency at small power levels. The values for peak power efficiency are therefore less relevant for the talk time of a GSM/EDGE/WCDMA multi-mode cell phone. For example the LINC concept has highest efficiency above 26 dBm output power, but smallest efficiency below 22 dBm. In the evaluation of average current it ranks last due to its inefficient operation at low power.

In a real system the average current values would be higher than presented here due to several additional sources of power loss. These include, amongst others, losses in non ideal PA transistors, losses in passive elements, bias circuit power loss, power loss in driver stages and peripheral circuits e.g. power detector. While the results presented in Fig. 3.21 are optimistic, they form a good starting point for the concept classification and reflect the general importance of considering efficiency and current consumption at power back-off for transmitter design.

A combination of the presented concepts can further enhance efficiency over the output power range. Not every combination is useful or makes sense. In Table 3.4 advantages and disadvantages of the individual concepts are summarized, based on discussions in the previous sections. Appropriate methods for a possible compensa-

tion of disadvantages are filtered in a separate column. The effort and complexity for the combination has to be contrasted with the benefits of reducing power consumption. It was tried to illustrate this trade-off in Fig. 3.22. The basis for the order in x-direction is mainly the previously calculated $\overline{I_{\text{mean}}}$. The order in y-direction is based on considerations on system complexity and sensitivity in mass production, which partly reflect the manufacturing costs. These include the number of additional active and passive building blocks (on-chip and off-chip), further required control units and flexibility to counteract technology parameter drifts which affect frequency, bandwidth, power and gain. The basic concepts are distinguished by different colors which are shared with the related extensions (additional concepts) marked with a "+". The widespread oval areas which overlap between many concepts show that a distinct boundary cannot be defined in general. This uncertainty is e.g. caused by the technology. For example using standard RF CMOS the complexity for DMPD is lower than for a Doherty design, which requires a low loss RF technology. On the other hand implementing both in an RF GaAs technology, the complexity for DMPD is higher as an additional digital control chip is necessary.

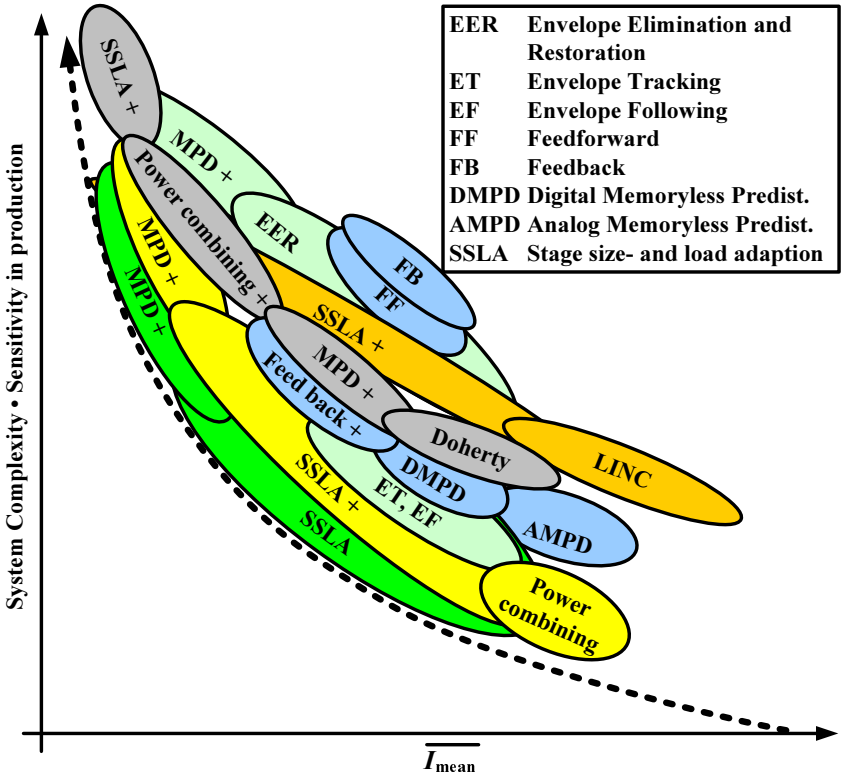


Figure 3.22: Comparison of efficiency enhancement techniques based on system complexity and mean current consumption.

One example for an improvement with additional low complexity is the power combining concept together with the SSLA concept. The RF switches can reduce the sensitivity to load variations in power back-off or avoid self biasing of deactivated transistors. This further reduces current consumption, as the load adjustment will

be more accurate and isolation between stages increases. Depending on the number of additional switches the complexity can vary in a wide range. A further extension with higher circuit complexity can be made by adding the memoryless predistortion concept. With this, the output power range for linear operation is expanded, which allows the use of smaller transistors and therefore increases efficiency for WCDMA. The additional current consumption of the predistortion circuit may partly compensate the benefit of efficiency improvement. This especially in the low power region, where the transistors bias currents will be in the range of the predistortion circuit's current consumption. Consequently, the additional efforts of this approach will give only a marginal overall enhancement.

A combination with much higher complexity is e.g. the LINC concept together with the SSLA. Switching to a bypass amplifier at low power levels, the low efficiency of LINC at large BO levels can be compensated. The reduction of current consumption is comparable to the combination SSLA and power combining. Although LINC and power combining both require several amplifier paths and a combiner network, the special signal generation for LINC causes the much higher complexity.

Overall, a trend can be formulated: Further reduction of current consumption below a certain point will cause excessive effort. This is indicated by the dashed line in Fig. 3.22.

From Fig. 3.22 and the underlying considerations it turns out that the SSLA and the power combining concepts provide best compromise between current consumption and implementation costs. Supply variation (ET and EF), predistortion and Doherty concepts can also form the basis for a relatively cheap and efficient design. As their main benefit is the efficiency improvement in the upper power range, attempts towards cellular multi-mode handset PAs are rare.

An important factor for saving manufacturing costs is the level of on-chip integration, i.e. integration of power transistor and periphery on the same substrate. Clearly, not every function can be implemented on one die, because the technology would have to offer a lot of features which again increase process complexity. A general comparison between HBT and CMOS technology is drawn in Fig. 3.23. For some key concepts it is shown how a transfer from HBT to CMOS can increase the level of on-chip integration and how the current consumption is influenced by this. CMOS technologies offer efficient integration of digital processing with small area, while HBT technologies tend to provide higher quality factor of passive elements. Hence, to increase the level of on-chip integration without degrading efficiency, CMOS technologies are favorable for concepts based on elaborate signal processing like digital predistortion, LINC and EER. Concepts with mainly critical analog parts like analog predistortion, SSLA, ET and EF can have a higher level of integration in HBT technologies. In general, the circuits tend to have higher current consumption when more periphery is collected on the same die, which is indicated by the grey arrow in the figure.

The attempts for higher integration are driven by the single-chip radio approach where the PA and the transceiver share the same technology. Since CMOS PAs

Technique	**Advantages / Disadvantages**	**Improvements by combination with other methods**
Doherty	⊕ Two PAE peaks in high power range	
	⊖ Narrowband	⇒SSLA
	⊖ Sensitive to load impedance	
	⊖ Hard to achieve linearity	⇒Linearization
	⊖ Poor PAE in low power range	⇒Power combining; SSLA
LINC	⊕ PAE improvement in high power range	
	⊕ Nonlinear amplifiers sufficient	
	⊖ Special phase signal generation	
	⊖ Narrowband	⇒SSLA
	⊖ Sensitive to load impedance	
	⊖ Poorer PAE in low power range compared to single stage	⇒SSLA
Supply Variation	⊕ Approx. constant PAE in high power range	
	⊕ Wideband matching networks possible	
	⊕ Tolerance to load impedance	
	⊖ High efficient DC-DC converter	
	⊖ Broad band DC-DC converter	⇒Envelope Tracking
	⊖ Amplitude and phase path time alignment	
	⊖ Amplitude and Phase distortion due to supply voltage variation	⇒Linearization
	⊖ Larger distortion at small supply voltages due to transistor's cut-off	
Linearization (Memoryless predistortion)	⊕ PAE improvement up to saturated power	
	⊕ Implementation in base band processing	
	⊖ Accurate device characterization (Memory effects, device aging, process variations...)	⇒Feedback Loop
SSLA	⊕ PAE improvement at arbitrary BO levels	
	⊕ Multiple frequency bands	
	⊖ Linear, low loss and high isolation switches necessary	⇒Power Combining; Linearization
Power Combining	⊕ PAE improvement at high power levels	
	⊕ no switches necessary	
	⊖ sensitive to load impedance in back-off	⇒SSLA; Linearization
	⊖ Transistor losses dominate PAE at back-off	

Table 3.4: Major efficiency enhancement techniques, their advantages and disadvantages and possibility for improvement by combination with other techniques.

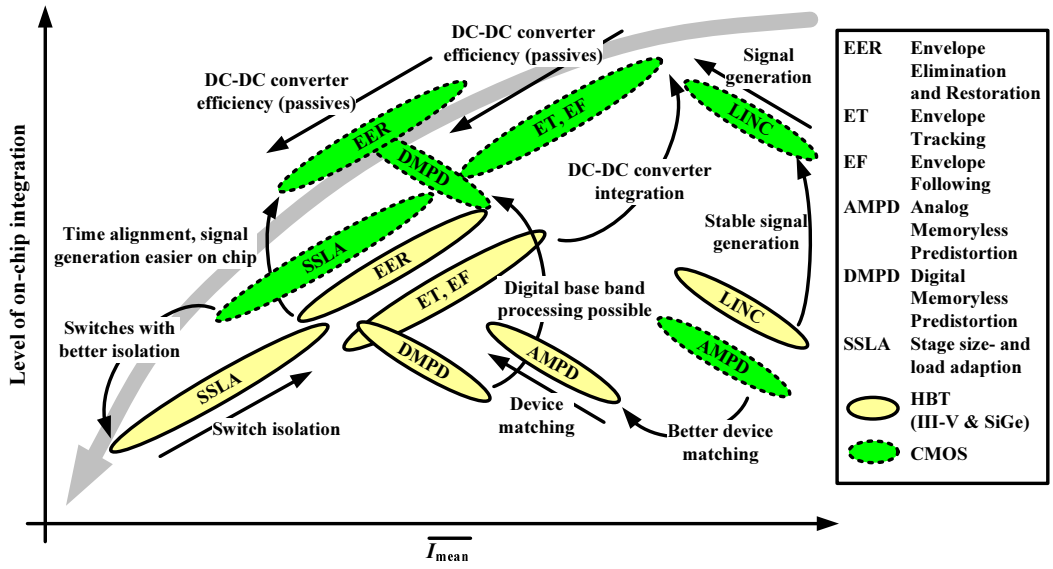


Figure 3.23: Comparison of efficiency enhancement techniques based on level of on-chip integration and mean current for CMOS and HBT technologies.

have in general worse performance than HBT or LDMOS, enhancement techniques are inevitable for being competitive. Nevertheless for a performance comparison of multi-mode PA concepts to state of the art PAs a less demanding prototype design which is produced with more technologies will give more information on conceptual feasibility and potential application to handset PAs. For this reason, in this work it was decided to use a cheap but efficient SiGe HBT technology with low loss off-chip power combiners and switches to test a novel multi-mode PA approach using the SSLA and power combining concept. The results of this work can then be understood as a lower edge benchmark which can be extrapolated to estimate efficiency reduction in case of higher integration.

3.3 Circuits for reducing impact of VSWR

It was assumed so far that the PA delivers its output power into a specified constant load impedance, usually 50 Ω. This condition is eligible for a laboratory test setting but when mounted into a handset the PA is directly exposed to impedance mismatches caused by the antenna. These variations are induced by user motion, environmental impacts in the near field of the antenna like proximity of metal objects, or in an extreme case even disconnection of the antenna during operation [Toftgard 93, Ogawa 99]. Depending on the order of the mismatch, these scenarios may cause three adverse effects on the PA:

- **Performance degradation:** A light load impedance shift causes a variation of output power, efficiency, linearity and gain according to the load pull contours of the PA. A reduced link-quality and even interruption of the connection

to the base station are likely. Hence, limits for power reduction at a certain VSWR at the antenna are set in the PA specification.

- **Oscillation:** For best performance trade-off a PA is usually designed at the limits of the technology and therefore only achieves conditional stability. Therefore there exist passive loads which lead to oscillations. Under antenna mismatch these can in best case only cause performance degradation and spurious emissions but in worst case lead to failure of the device.

- **Reduced reliability:** The reflected power from the antenna can create, under certain conditions, constructive interference with the incident power at the collector of the PA transistor. This can lead to high collector voltage, if the load becomes very high-ohmic or it can increase the collector current, if the load has low-impedance. These excessive stresses can lead to irreversible damage of the PA output transistor. The robustness limits are the collector emitter breakdown voltage V_{CE0} to avoid punch-through or carrier multiplication effects, the maximum collector current I_{C0} to avoid a secondary breakdown and hot-spot formation, and the maximum power dissipation to avoid thermal overload. In addition to interference from the antenna, also an increased battery voltage has to be considered for the transistor ruggedness. An oversupply condition occurs for example, if the handset is charged with a fast battery charger which provides a supply voltage as high as 5 V. For this reason, the ruggedness specification is usually expressed in terms of a maximum tolerable output VSWR under a specified oversupply condition. Typical data sheets of commercial power amplifiers guarantee that no permanent damage is caused by a load VSWR of 10 : 1 at a supply voltage of 5 V.

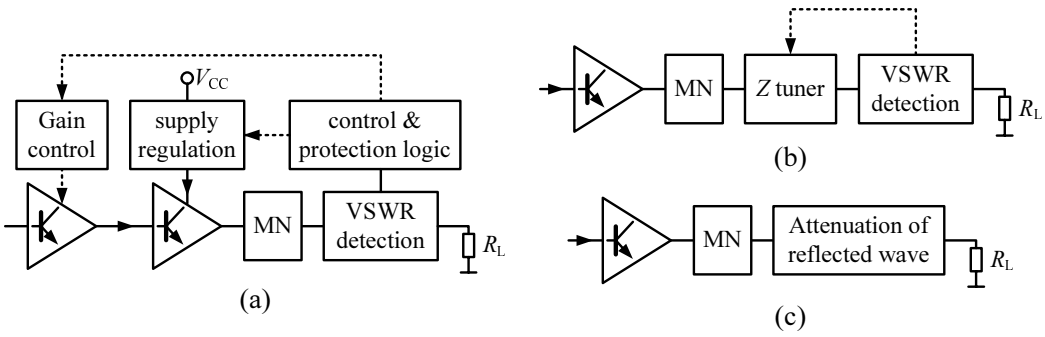


Figure 3.24: Protection against antenna impedance deviation. (a) Transistor overvoltage and thermal overload protection (b) Retuning of antenna impedance (c) Termination of reflected power.

Commonly used strategies for protection against load variations can be divided into two major groups, namely first the protection of the transistor itself and second the direct reduction of load reflections. Some practical realizations are aimed at only one of these strategies while others combine both at once, depending on cost and size

aspects. In the following a short description of potential techniques is given without going deep into circuit implementation details.

- **Overhead design:** The straightest way is to design a PA being sufficiently robust to deliver the output power irrespective of the VSWR value. The transistor has to be sufficiently rugged to withstand the stresses. In this regard III-V technologies have an advantage over SiGe due to the higher V_{CE0}. A robustness improvement in SiGe technology (e.g. resistive emitter/collector degradation) will also lead to a reduced performance. In general this design approach wastes space and power and leads to worse PAE performance and more expensive PA modules.

- **Overload protection:** As shown in Fig. 3.24 (a) a detection circuit (e.g. a directional coupler, a voltage sampler or a peak detector) is used to extrapolate a quantity which is proportional to the load VSWR. A protection logic can then effectively control the voltage or current from the supply to a safe level. Herewith voltage and current peaks are clamped at the collector of the power transistor. The drive power from first amplifier stages can also be controlled to reduce dissipated power in the final stage. These usability of these methods is limited by the response time of the feedback regulation loop. Additionally, the use of clamping protection circuits may result in linearity reduction and RF performance degradation, e.g. prevention of input overdrive which is crucial for high PAE. Nevertheless, assuming knowledge of the transistor's safe operating limits and breakdown characteristics, these techniques are usually easy to implement together with the PA on one die and consume smallest chip area [Scuderi 05, King 00, Bezooijen 05].

- **Load tuning:** An impedance tuner between the PA and the load (see Fig. 3.24 (b)) can compensate for VSWR changes of the antenna. For integration into a handset it typically consist of an inductor or a transmission line with two digitally tunable binary combined capacitors in a PI- or T-configuration [Mingo 04, Fidler 92, Qiao 05]. Although this retuning of the load would help for protection of the transistor as well as for avoiding an RF performance degradation, a practical design has to be rather elaborate. Issues which make a realization that complex are tuning speed, tuning range, bandwidth, sensitivity, linearity and power loss of the tuner. Compensation over the whole smith chart is only possible with component tuning values between zero and infinity but due to size constraints variable components can be adjusted only in a limited range. Therefore the actual matching domain of a practical network only covers a small part of the theoretical matching domain. Consequently very high VSWR values will be reduced only by compromising an increased insertion loss. Another important issue for a load tuning system is the strategy of iterating to the VSWR minimum. First a circuit to determine the actual load VSWR is inevitable and second an algorithmic approach for fast reduction of reflected power has to be developed. These problems are addressed in [Mileusnic 97, Meng 06]. As not only the amplitude but also the phase of

the load reflection coefficient has severe influence on the performance other more compact load tuning implementations which adapt the load angle are conceivable [Keerti 05].

- **Termination of reflections:** By reducing the reflected wave from the mismatched antenna and terminating it at least partially, the effective VSWR, seen by the PA, is improved. This helps for stabilization and improves reliability of the PA transistor but mismatch losses which cause efficiency and output power reduction are inevitable. Fig. 3.24 (c) shows the configuration with a black-box for the attenuation circuit. This can actually be an isolator which most effectively minimizes impedance variations without substantially attenuating the incident wave. Disadvantages are its relatively large size and high cost in production. The use of anisotropic materials makes it difficult to integrate it onto the same substrate together with the PA. Moreover its frequency response is narrow-band when designed for smallest insertion loss and largest isolation with a power capability of more than 2 W. For these reasons an isolator is not used in low cost mobile phones. A VSWR improvement is directly achieved by an attenuator or some other circuitry which adds a certain amount of insertion loss between the PA and the antenna. The mismatch at the output of the attenuator is improved by an amount equal to twice the value of the IL in dB. This is expressed with the following formula where VSWR' at the PA is calculated from the insertion loss IL of the circuit and the return loss RL at the antenna, derived from the VSWR at the antenna (see Fig. 3.25(a)):

$$\mathrm{VSWR}' = \frac{1 + 10^{-(\mathrm{RL}_{[\mathrm{dB}]} + 2\,\mathrm{IL}_{[\mathrm{dB}]})/20}}{1 - 10^{-(\mathrm{RL}_{[\mathrm{dB}]} + 2\,\mathrm{IL}_{[\mathrm{dB}]})/20}}\ ,\ \text{with} \tag{3.38}$$

$$\mathrm{RL}_{[\mathrm{dB}]} = -20\ \lg \left| \frac{\mathrm{VSWR} - 1}{\mathrm{VSWR} + 1} \right|$$

Fig. 3.25(b) shows the reduction of mismatch in double logarithmic scale over the antenna VSWR for several insertion loss values. An IL of about 1 dB can improve the VSWR from e.g. 10 : 1 to 4.7 : 1. The front end parts such as filter, duplexer, directional coupler and antenna switch have IL values which typically sum up to more than 1 dB. Hence, without providing an additional attenuator, a considerable protection against mismatch is already included in the front end anyway. Nevertheless, by aiming at a small current consumption, the development of more efficient front ends is intended and the consequential efficiency improvements should not be intentionally wasted by an attenuator for VSWR reduction. Therefore a VSWR improvement with an attenuator circuit is a sub-optimal solution. The most promising approach is the use of a 90° hybrid combiner. This circuit will be introduced in the following chapter and has the advantage of reducing mismatch by about 50% without namable reduction of output power (IL $\sim$ 0.3 dB). This special 4-port power combiner is a reciprocal passive structure which can be implemented in low Q standard semiconductor technologies. The reflected wave is primarily terminated at a

proper resistor, therefore, no current consuming VSWR detection or regulation circuitry is required.

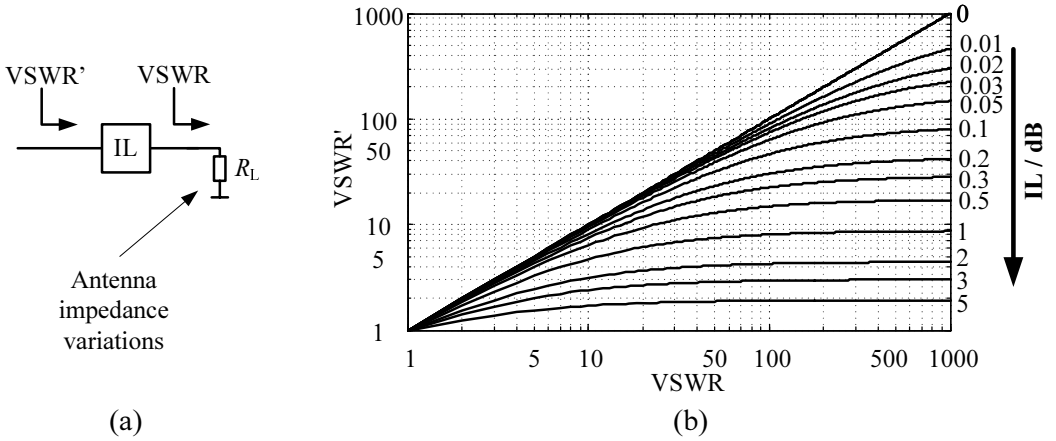


Figure 3.25: Effect of insertion loss on antenna VSWR changes: (a) Explanation of variables, (b) Parameter plot of input versus output VSWR.

The presented options of VSWR protection in circuit design have many advantages and disadvantages which have to be balanced in the particular application. Under the aspect of simultaneously reducing the average current consumption and antenna mismatch, a PA topology which combines both is preferable. A proper concept for these requirements is a balanced amplifier. It consists of 90° hybrids, which inherently reduce antenna mismatch. Additionally it is a power combining topology, allowing an improvement of efficiency at back-off. In the following section the balanced amplifier concept is introduced. Some novel extensions are shown which enable a complete and almost load independent multi-mode PA realization.

3.4 Balanced amplifier theory

This section describes the balanced amplifier operation principle, starting with idealized assumptions. The quadrature hybrids are considered to be lossless and symmetric and the amplifiers are linear and matched at all ports. Non-ideal influences such as imbalance and mismatch will be included afterwards and the degradation of gain and return loss are discussed.

3.4.1 State of the art concept

The general theory behind a balanced amplifier and its benefits of operation were first published in the mid 20th century by K. Kurokawa [Kurokawa 65,a]. This concept enables very broadband and stable RF amplifier designs with small return loss at input and output. Compared to single ended topologies the load sensitivity is

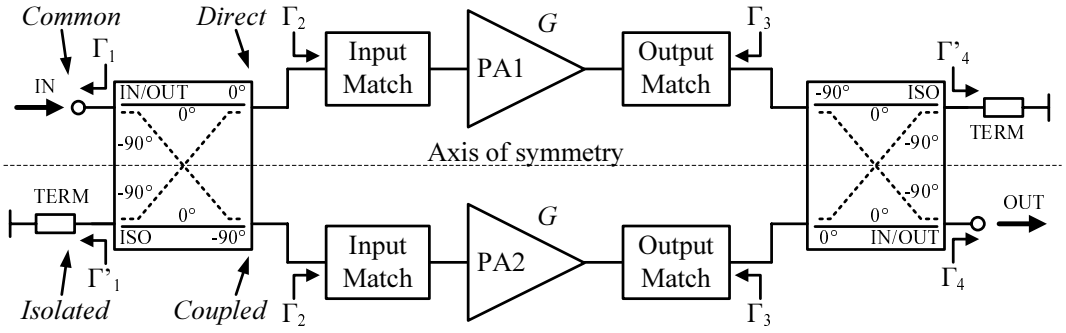


Figure 3.26: Conceptual block diagram of the balanced power amplifier.

improved by about 50 %. The classical balanced amplifier configuration is a special kind of a 2-way power combining topology making use of 90° hybrids, also called quadrature couplers/combiners[4] (see Fig. 3.26). This coupler consists of 4 ports with the general convention that the input or output is called the *common port*, the 0° or −90° ports are called the *direct and coupled ports* respectively and the remaining one is the *isolated port*. Since the hybrid has two electrical symmetry axes (see also Chapter 4.2) it does not make any difference which port is the common port as long as the phase relation between the ports remains. The operating principle can basically be described as follows. An incident signal is split into two signals of equal power but 90° phase difference. These are amplified by two identical PAs with the same gain G, equal input impedance Γ_2 and output impedance Γ_3 and combined again at the output where the initial phase shift is compensated. The voltage of the two signals being in quadrature ($0°/-90°$) sum up in phase ($-90°/-90°$) at the common output port and cancel each other ($0°/-180°$) at the isolated port where they are out of phase. Hence, for the ideal case the input signal is amplified in total with the gain G of the individual stages and no power is lost in the isolation resistors. Assumptions for this ideal case are a perfect electrical symmetry of the two PA paths and no gain and phase errors in the coupler structures as illustrated in Fig. 3.26. The main advantage of the balanced structure becomes apparent when considering reflected power from a load mismatch ($\Gamma_4' \neq \Gamma_3$). At the common port this reflection experiences a $0°/-90°$ phase shift through the hybrid and additional reflections at the PA outputs. The power being already attenuated by Γ_3 reenters the hybrid at the direct and coupled port and splits up to the isolated and common port. A summation happens at the isolated port ($-90°/-90°$) and cancelation at the common port ($0°/-180°$). Consequently, reflected power is absorbed in the termination resistance and the VSWR seen into the power amplifier is 1 : 1, regardless of Γ_3. In addition the VSWR of a varying load is improved due to the power splitting property of the coupler. Reflected power from the load is reduced by 3 dB at each PA, having the same effect on VSWR as a circuit with 3 dB insertion loss (see Fig. 3.25). The difference to a 2-port attenuator is that output power and efficiency

[4]When it is obvious from the context, the terms 90° and quadrature will be omitted for convenience in the following text.

are not essentially influenced by the hybrid circuit, similar as for an isolator circuit. It has to be pointed out that the impedances seen from the 0°/90° ports must be equal (Γ_3) but a complete match of the hybrid with $\Gamma_3 = \Gamma_4$ is not necessary to maintain the functionality.

The termination resistor at the isolated port can be seen as internal part of the quadrature coupler and the network reduces to a 3-port power combiner with the scattering matrix[5]

$$\underline{S}_{\mathrm{BC}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -j \\ 1 & -j & 0 \end{pmatrix}. \tag{3.39}$$

In general, the combining network in a balanced amplifier has to realize this scattering matrix for the ideal operation principle. Although in literature the quadrature coupler is always considered as essential part of the balanced amplifier, it can be seen as special implementation out of other possibilities. For example a 3-port wilkinson power combiner with $-90°$ phase shift at one of the dividing ports also has appropriate electrical characteristics. Nevertheless, for realization of the multi-mode PA, the isolated port of the hybrid has other useful advantages which will be explained and used in later chapters.

To complete the description of the balanced amplifier architecture the main advantages and disadvantages are summarized:

Advantages

- Impedances seen into the balanced amplifier at the input and the output are determined by the termination at the hybrids isolated port. Assuming a 50 Ω system and symmetry of the amplifier a termination resistance of 50 Ω provides zero return loss of the complete amplifier. This is independent of the actual impedances seen into the single amplifier paths which do not necessarily have to be matched.

- Due to the inherent matching property, the balanced amplifier is easy to cascade. It represents a constant and well defined load to a driver stage. This improves stability and eases specification of available driver power over process variations.

- Isolation between the two device sides is very high, indicated by $S_{12} = S_{21} = 0$ in (3.39). Therefore, spurious inner loop oscillations are suppressed, which improves overall amplifier stability.

- One amplifier side needs to deliver only half of the overall output power. Therefore smaller devices with smaller power handling capability can be used. The heat sink dimensions for each side are relaxed as well.

- With duplication of the PA the load-line impedance transformation at the output is simplified. The target load value increases by a factor of two which

[5]The port order is: 1 direct port, 2 coupled port, 3 common port.

improves bandwidth, efficiency and insensitivity to device parameter variations of the matching network.

- This architecture also provides some protection against amplifier device failure in that if one device breaks down, the amplifier still works but with reduced output power.
- Odd order harmonics are canceled and second order harmonics are reduced by 3 dB due to the quadrature hybrid [Shumaker 02]. In combination with the typical low pass characteristic and the second harmonic trap of the matching network of each PA, very low harmonic content is obtained.

Disadvantages

- No virtual ground connection is existent. Unlike the push-pull topology, the balanced amplifier has the same drawbacks concerning the ground connection as the single ended topology.
- Compared to a single ended design the part count is increased. The matching networks as well as the bias circuitry have to be doubled and two additional combining networks are necessary. These additional components lead to higher power loss, larger area and higher cost of the balanced amplifier. In a multi-stage amplifier the part count can be reduced by only operating the output stage in a balanced configuration. The input gain stages can remain single ended.
- The termination resistance has to withstand the power reflected from a mismatched load. In worst case it has to absorb the total output power delivered from the two amplifier branches. Proper power handling capability and heat sinking are required for a reliable design. When integrating the resistance on chip together with the amplifiers the increased chip-area and occurring parasitics due to e.g. bond wires and pads have to be considered.
- The bandwidth of the design is mostly limited by the couplers. These add losses at the output which also reduce efficiency.

3.4.2 Effect of non-idealities

For the following analysis the quadrature hybrids are replaced by the more general 3-port power combiner representation from equation (3.39). In Fig. 3.27 the balanced architecture is shown with additional errors inherent to a practical design. These include imperfect symmetry of the quadrature couplers with gain/phase errors ε/β and only finite isolation $S_{12\mathrm{A}}/S_{12\mathrm{D}}$. As the couplers at input and output have different power handling criteria and may be fabricated in different technologies, it has to be assumed that these errors are not equal. The return loss at each coupler port can be easily reduced by additional matching why all S_{ii} are assumed to be

zero. Furthermore the two amplifier branches are expected not to be symmetrical regarding gain S_{21} and input/output matching S_{11}/S_{22}. As the amplifiers consist of several stages for higher gain the coupling from output to input is supposed to be negligible, hence $S_{12\mathrm{B}} = S_{12\mathrm{C}} = 0$. The amplifier's overall properties in dependence

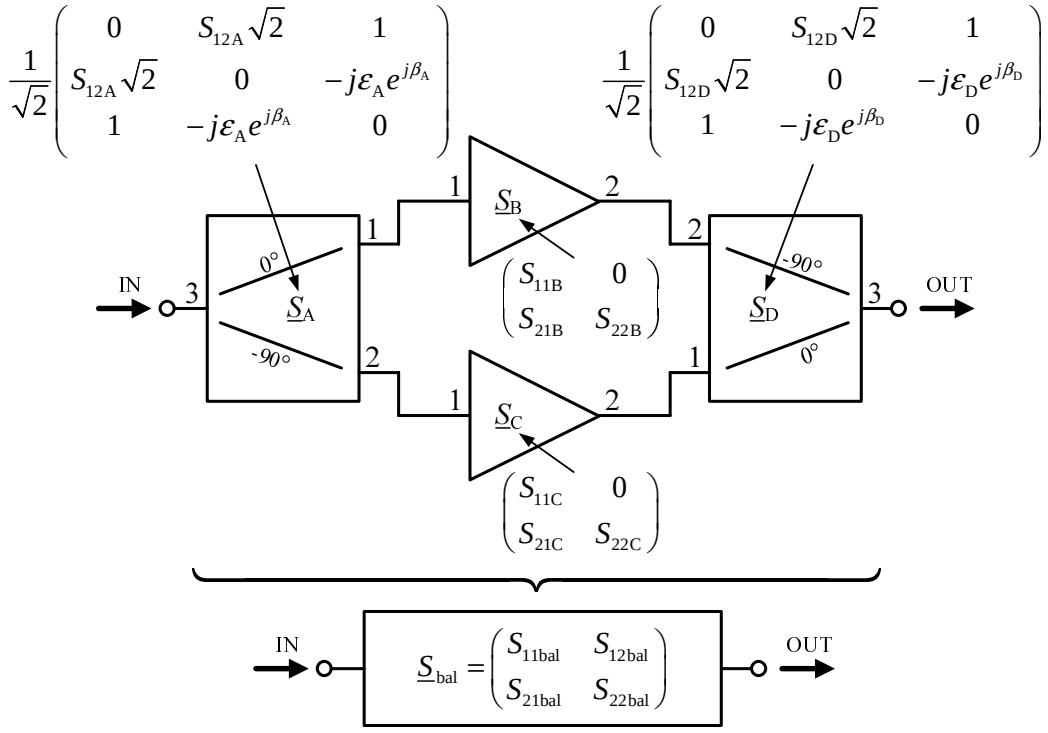


Figure 3.27: Complete two port scattering matrix representation of the balanced power amplifier including imbalance, gain, phase, isolation and matching errors.

of these errors can be found by derivation of the representative two port scattering matrix $\underline{S}_{\mathrm{bal}}$. This matrix includes all partial scattering matrices, as illustrated in Fig. 3.27, and is calculated by using the multi port network connection equations (see Appendix D). The four matrix elements are:

$$S_{11\mathrm{bal}} = \frac{1}{2}\,\frac{\left(2\,j\,\varepsilon_{\mathrm{A}}\,e^{j\beta_{\mathrm{A}}} S_{12\mathrm{A}}\,S_{11\mathrm{B}} + \varepsilon_{\mathrm{A}}^2\,e^{j2\,\beta_{\mathrm{A}}}\right) S_{11\mathrm{C}} - S_{11\mathrm{B}}}{S_{12\mathrm{A}}^2 S_{11\mathrm{B}}\,S_{11\mathrm{C}} - 1} \tag{3.40}$$

$$S_{22\mathrm{bal}} = \frac{1}{2}\,\frac{\left(2\,j\,\varepsilon_{\mathrm{D}}\,e^{j\beta_{\mathrm{D}}} S_{12\mathrm{D}}\,S_{22\mathrm{C}} + \varepsilon_{\mathrm{D}}^2\,e^{j2\,\beta_{\mathrm{D}}}\right) S_{22\mathrm{B}} - S_{22\mathrm{C}}}{S_{12\mathrm{D}}^2 S_{22\mathrm{B}}\,S_{22\mathrm{C}} - 1} \tag{3.41}$$

$$S_{12\mathrm{bal}} = 0 \tag{3.42}$$

$$S_{21\mathrm{bal}} = \frac{1}{2}\,\frac{A}{\left(1 - S_{12\mathrm{A}}^2 S_{11\mathrm{B}}\,S_{11\mathrm{C}}\right)\left(1 - S_{12\mathrm{D}}^2 S_{22\mathrm{B}}\,S_{22\mathrm{C}}\right)} \tag{3.43}$$

$$\begin{aligned} A = {} & \left(\varepsilon_{\mathrm{A}}\,\varepsilon_{\mathrm{D}}\,e^{j(\beta_{\mathrm{A}}+\beta_{\mathrm{D}})} S_{21\mathrm{C}}\,S_{22\mathrm{B}} - S_{21\mathrm{B}}\,S_{22\mathrm{C}}\right) S_{12\mathrm{D}} + \\ & \left(\varepsilon_{\mathrm{A}}\,\varepsilon_{\mathrm{D}}\,e^{j(\beta_{\mathrm{A}}+\beta_{\mathrm{D}})} S_{21\mathrm{B}}\,S_{11\mathrm{C}} - S_{21\mathrm{C}}\,S_{11\mathrm{B}}\right) S_{12\mathrm{A}} + \left(\varepsilon_{\mathrm{A}}\,e^{j\beta_{\mathrm{A}}} S_{21\mathrm{B}}\,S_{22\mathrm{C}}\,S_{11\mathrm{C}} + \right. \\ & \left. \varepsilon_{\mathrm{D}}\,e^{j\beta_{\mathrm{D}}} S_{21\mathrm{C}}\,S_{22\mathrm{B}}\,S_{11\mathrm{B}}\right) j\,S_{12\mathrm{A}}\,S_{12\mathrm{D}} + j\varepsilon_{\mathrm{D}}\,e^{j\beta_{\mathrm{D}}} S_{21\mathrm{B}} + j\varepsilon_{\mathrm{A}}\,e^{j\beta_{\mathrm{A}}} S_{21\mathrm{C}} \end{aligned}$$

Mismatch

A balanced design is supposed to have better input and output matching than a single ended design, but as indicated by (3.40) and (3.41) a non ideal coupler causes matching errors which decrease this theoretical benefit. For PA applications the effects at the input and the output are not equal. While the input of each of the two combined PAs is usually conjugately matched (typically $|S_{11\mathrm{B}}|$, $|S_{11\mathrm{C}}| < 0.2$, 1 % difference), the output has higher return loss due to the load line match for maximum power (typically $|S_{22\mathrm{B}}|$, $|S_{22\mathrm{C}}| \sim 0.7$, 1 % difference). To show the difference with an example, a coupler with typical errors of $\varepsilon = 0.5$ dB, $\beta = 10°$ and $|S_{12\mathrm{A}}| = |S_{12\mathrm{B}}| = 0.1$ is applied at the input and the output. The resulting input reflection coefficient is $|S_{11\mathrm{bal}}| = 0.04$, while at the output $|S_{22\mathrm{bal}}| = 0.19$ is obtained. Improved matching at the input ($S_{11\mathrm{bal}} = 0$) can be achieved, if the two amplifier branches are matched ($S_{11\mathrm{B}} = S_{11\mathrm{C}} = 0$), despite the gain and phase error of the input combiner network. In a similar way this is also true for the output match ($S_{22\mathrm{bal}} = 0$), but a conjugate output match of power amplifier branches is contradicting with the load line match for maximum power and therefore not possible in general. For this reason the errors of the output hybrid cannot be compensated by matching purposes without degrading output power. For $S_{22\mathrm{bal}} = 0$ the hybrid itself has to offer almost ideal properties and the PA outputs have to be symmetrical as well ($S_{22\mathrm{B}} = S_{22\mathrm{C}}$). Concluding, for a conjugate input and output match of the balanced power amplifier the output coupler has to be much more accurate than the input hybrid.

Gain

If the output hybrid is ideal and the amplifier paths are symmetrical and matched at the input, then $S_{21\mathrm{bal}}$ from (3.43) reduces to:

$$S_{21\mathrm{bal}} = -\frac{1}{2}\, j \left(S_{21\mathrm{B}} + \varepsilon_{\mathrm{A}}\, S_{21\mathrm{C}}\, e^{j\beta_{\mathrm{A}}} \right) \tag{3.44}$$

The deviation from the ideal case is then caused by amplitude and gain imbalance of the input coupler. It has to be noted that with these assumptions, the input coupler's isolation has no influence on the overall gain. Again, this indicates that the design of an input coupler has more relaxed requirements than that of an output coupler. For part count reduction the input coupler can be realized on the same chip together with the power transistors where due to substrate loss and coupling mechanisms a high isolation between the two branches is difficult to achieve.

The realization of an almost errorless output coupler may be possible for one frequency but not over a broad operation bandwidth. Hence, gain and phase errors will be inevitable. A reduced sensitivity of the design is gained, if the coupler still offers high isolation ($S_{12\mathrm{D}} \sim 0$). Together with a good input match ($S_{11\mathrm{B}} = S_{11\mathrm{C}} = 0$) the overall gain only depends on amplitude and phase imbalance of the two couplers:

$$S_{21\mathrm{bal}} = -\frac{1}{2}\, j \left(\varepsilon_{\mathrm{D}}\, S_{21\mathrm{B}}\, e^{i\beta_{\mathrm{D}}} + \varepsilon_{\mathrm{A}}\, S_{21\mathrm{C}}\, e^{j\beta_{\mathrm{A}}} \right) \tag{3.45}$$

The same result is obtained, if the isolation of both couplers is assumed to be very high ($S_{12\mathrm{A}} = S_{12\mathrm{D}} \sim 0$), but as already mentioned, this tightens specifications at the input and makes an implementation on chip more difficult.

The overall balanced amplifier gain is proportional to $|S_{21\mathrm{bal}}|^2$. Implying symmetry $S_{21\mathrm{B}} = S_{21\mathrm{C}} = |G|$, (3.45) equates to:

$$|S_{21\mathrm{bal}}|^2 = \frac{G^2}{4}\left(\varepsilon_\mathrm{A}^2 + \varepsilon_\mathrm{D}^2 + 2\,\varepsilon_\mathrm{A}\,\varepsilon_\mathrm{D}\,\cos(\underbrace{\beta_\mathrm{A} - \beta_\mathrm{D}}_{\Delta\beta})\right) \tag{3.46}$$

A phase imbalance of one coupler can be completely compensated by an inverse imbalance of the other coupler. By contrast, amplitude imbalances cannot be compensated, they even add up with quadratic proportionality. Fig. 3.28 shows a plot of (3.46) to illustrate the sensitivity of gain versus amplitude and phase imbalance.

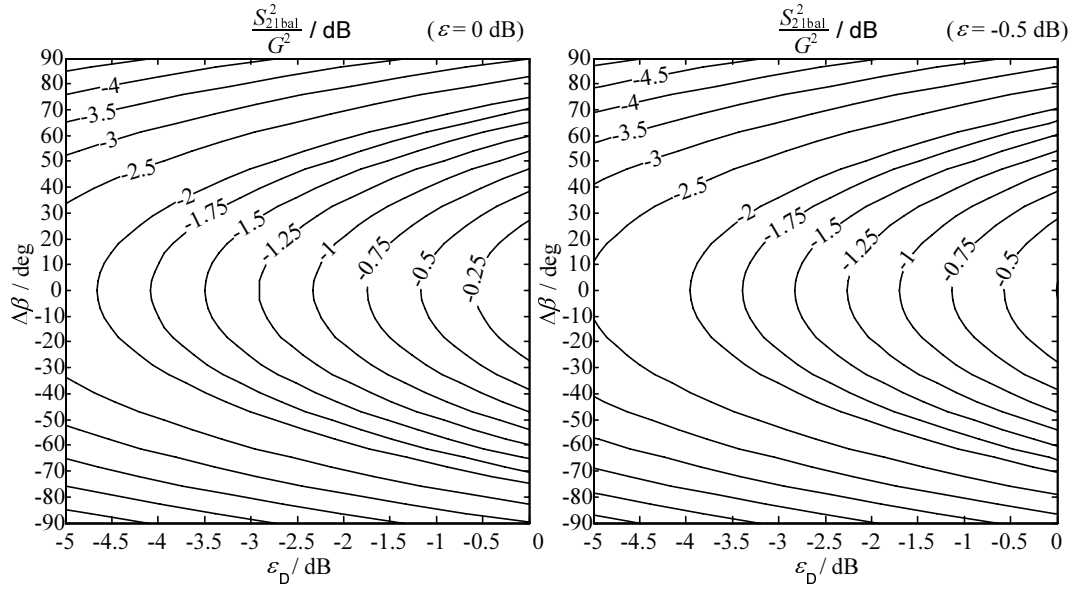


Figure 3.28: Effect of gain and phase errors of the combiner networks on the overall balanced amplifier gain.

For low return loss and high gain of the balanced amplifier the following can be concluded:

- The output coupler has to have high isolation between the coupled and direct port.
- The isolation of the input coupler's coupled and direct port has to be high, only if the two amplifier branches are not matched.
- The amplitude error of both couplers has to be as low as possible.
- The phase error of both couplers have to compensate in a way that $\Delta\beta = \beta_\mathrm{A} - \beta_\mathrm{D} \sim 0$. For maintaining the advantage of reduced load mismatch sensitivity, the absolute phase of the hybrids should not exceed the 90° too much.

- Matching at the input or output can be improved by a reduced phase error of the respective coupler.
- Matching symmetry of the two amplifier branches is necessary for improvements of performance.

3.5 Switch-less multi-mode PA – Balanced switched path PA

This section introduces the concept of a PA with efficiency enhancement at far back-off power levels. Its characteristic feature is a purely bias controlled PA reconfiguration without RF switches. The limits of efficiency improvement are derived and discussed with regard to multi-mode operation. Some aspects and design examples of this concept are published in [Sogl 09] and [Grebennikov 07].

3.5.1 Block diagram

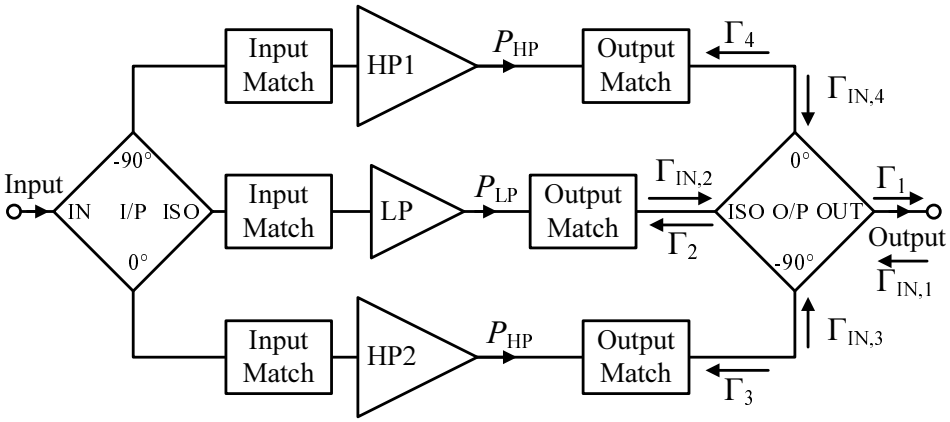


Figure 3.29: Block diagram of a balanced switched path power amplifier.

The basic topology is shown in Fig. 3.29. It is based on a balanced PA architecture. An additional separate low power amplifying path is connected between the isolated ports of the 90° hybrid combiners. In case of symmetry the isolation from the ISO to the 0°/90° ports is very high, hence the hybrid acts as an isolator between high power (HP) and low power (LP) amplifier paths. This enables two almost independent high efficiency modes of operation (Table 3.5).

In high power mode a bias control unit provides proper biasing of the HP1 and HP2 amplifiers and switches off bias currents in the LP amplifier. The output power from both active paths combine through the output hybrid without being influenced by the LP path, thereby providing high efficiency at high output power.

Bias settings	**HP**	**LP**
HP1, HP2	ON	OFF
LP	OFF	ON
$P_{\text{out,max}}$	$2\,P_{\text{HP}}$	P_{LP}
VSWR protection	✓	–

Table 3.5: Balanced switched path PA operation modes

In low power mode the LP amplifier is biased properly while the HP1 and HP2 amplifiers are turned off via bias control and therefore provide high reflections at input and output. Incident power at the input hybrid is reflected at the HP1/HP2 paths and sums up at the ISO port where it enters the LP amplifier path. At the output of the LP path the amplified signal enters the output hybrid and is reflected at the HP1/HP2 path outputs before it is delivered to the load at the hybrid's OUT port. Hence, high efficiency at lower output power levels can be achieved without any additional switches to isolate HP and LP paths.

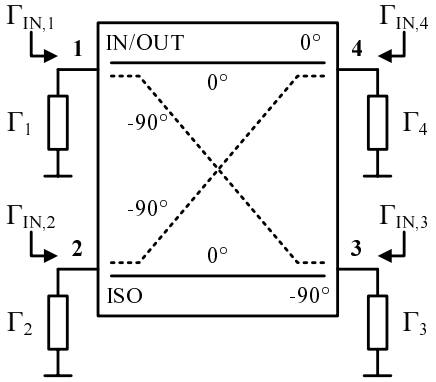
One of the main advantages is that each of the three amplifier branches can be optimized almost separately for best compromise between output power, efficiency and linearity, therefore guaranteeing fast prototyping. But as for the power combining concept a reconfiguring PA which is based only on load impedance changes without use of switches has inherent limits. Therefore it is necessary to quantify these limits with respect to output power, combining efficiency and load sensitivity.

3.5.2 Performance limits

The analysis of the ideal classical balanced amplifier in Chapter 3.4 was based on the assumption that impedances, seen into and from the ports of the quadrature hybrids are constant during operation and fulfill specific properties for regular operation. Now, due to the additional LP path which replaces the resistors at the ISO ports, the impedances change, depending on the operation mode and will not fulfill the specific requirements anymore. In HP mode the reflection coefficient, seen into the ISO port will not be 0, as in the ideal case. In LP mode the impedances, seen into the turned off HP paths have to be considered. The performance in both modes therefore depends strongly on the power transfer properties of the quadrature hybrid under different load impedance conditions.

The compromises and performance limits resulting from non-ideal loads at the quadrature hybrid will be calculated for the output hybrid. This, due to the severe influence of power loss and mismatch on the overall PAE. Losses at the input reduce the overall gain and have less impact on PAE, as long as the gain of the individual paths is larger than 10 dB (see Fig. 3.14). Nevertheless, the calculations are valid for the input as well and can be applied in an analog manner.

For the following analysis the amplifier paths will be replaced by their characteristic



$$\underline{S} = \begin{pmatrix} 0 & 0 & \frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \\ \frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} & 0 & 0 \end{pmatrix}$$

(b)

Figure 3.30: Quadrature hybrid: (a) Schematic diagram and port naming convention. (b) Scattering parameters.

small signal reflection coefficients Γ_i at the ports of the quadrature hybrid, as shown in Fig. 3.30 (a). The impedances seen into the hybrid will be denoted as $\Gamma_{\mathrm{IN},i}$. The influence of Γ_i on the power summation, power transfer and reflections $\Gamma_{\mathrm{IN},i}$ can be calculated with any of the introduced algorithms in Appendix B. For a compact notation the generalized scattering parameters are used here. In difference to the commonly used scattering parameters (Fig. 3.30 (b)), which are related to the same, mostly real reference impedance at each port (e.g. 50 Ω), the generalized scattering parameters can be related to complex impedances, which are different at each port (Appendix B.3). With knowledge of the generalized scattering parameters the load conditions can be optimized for a better power transfer and smaller losses of the circuit.

The hybrid's generalized scattering matrix is:

$$\underline{S'} = \underline{\Gamma}_{\mathrm{abs}} \cdot \begin{pmatrix} \frac{N_{2341}}{\left(1-|\Gamma_1|^2\right) D}\,\frac{A_1^*}{A_1} & \frac{j(\Gamma_3+\Gamma_4)}{D}\,\frac{A_2^*}{A_1} & \frac{j\sqrt{2}(\Gamma_2\,\Gamma_4+1)}{D}\,\frac{A_3^*}{A_1} & \frac{\sqrt{2}(\Gamma_2\,\Gamma_3-1)}{D}\,\frac{A_4^*}{A_1} \\ \frac{j(\Gamma_3+\Gamma_4)}{D}\,\frac{A_1^*}{A_2} & \frac{N_{1432}}{\left(1-|\Gamma_2|^2\right) D}\,\frac{A_2^*}{A_2} & \frac{\sqrt{2}(\Gamma_1\,\Gamma_4-1)}{D}\,\frac{A_3^*}{A_2} & \frac{j\sqrt{2}(\Gamma_1\,\Gamma_3+1)}{D}\,\frac{A_4^*}{A_2} \\ \frac{j\sqrt{2}(\Gamma_2\,\Gamma_4+1)}{D}\,\frac{A_1^*}{A_3} & \frac{\sqrt{2}(\Gamma_1\,\Gamma_4-1)}{D}\,\frac{A_2^*}{A_3} & \frac{N_{4123}}{\left(1-|\Gamma_3|^2\right) D}\,\frac{A_3^*}{A_3} & \frac{j(\Gamma_1+\Gamma_2)}{D}\,\frac{A_4^*}{A_3} \\ \frac{\sqrt{2}(\Gamma_2\,\Gamma_3-1)}{D}\,\frac{A_1^*}{A_4} & \frac{j\sqrt{2}(\Gamma_1\,\Gamma_3+1)}{D}\,\frac{A_2^*}{A_4} & \frac{j(\Gamma_1+\Gamma_2)}{D}\,\frac{A_3^*}{A_4} & \frac{N_{3214}}{\left(1-|\Gamma_4|^2\right) D}\,\frac{A_4^*}{A_4} \end{pmatrix} \tag{3.47}$$

$$A_i = \frac{1-\Gamma_i^*}{|1-\Gamma_i|}\sqrt{1-|\Gamma_i|^2}, \quad \underline{\Gamma}_{\mathrm{abs}} = \begin{pmatrix} |\Gamma_1|^2-1 \\ |\Gamma_2|^2-1 \\ |\Gamma_3|^2-1 \\ |\Gamma_4|^2-1 \end{pmatrix}^{\mathrm{T}},$$

$$D = (\Gamma_1\,\Gamma_3+1)(\Gamma_2\,\Gamma_4+1) + (\Gamma_1\,\Gamma_4-1)(\Gamma_2\,\Gamma_3-1),$$

$$N_{ijkl} = (\Gamma_i\,\Gamma_j-1)(\Gamma_k-\Gamma_l^*) + (\Gamma_i\,\Gamma_k+1)(\Gamma_j+\Gamma_l^*)$$

High Power mode

In high power operation mode the total output power P_{out} at port 1 has to be the perfect sum of the two 90° phase shifted input signals from port 3 and 4 ($P_{\text{in}} = P_3 + P_4$). This can be written as:

$$P_{\text{out}} = \frac{|b_1'|^2}{2} = \frac{1}{2}\,|S_{13}'\,a_3' + S_{14}'\,a_4'|^2 \quad \text{with} \quad a_i' = \sqrt{2\,P_i}\,e^{j\varphi_i} \tag{3.48}$$

$$P_{\text{in}} = \frac{1}{2}\left(|a_3'|^2 - |b_3'|^2 + |a_4'|^2 - |b_4'|^2\right), \tag{3.49}$$

where the indices of incident and reflected waves a_i and b_i refer to the port numbers in Fig. 3.30. Taking account of symmetry ($P_3 = P_4$) and phase shift ($\varphi_3 = 0°$, $\varphi_4 = -90°$) of the input signals, the combining efficiency in terms of Γ_i is given by:

$$\eta_{\text{c}} = \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{|S_{13}'\,a_3' + S_{14}'\,a_4'|^2}{|a_3'|^2 - |S_{33}'\,a_3' + S_{34}'\,a_4'|^2 + |a_4'|^2 - |S_{43}'\,a_3' + S_{44}'\,a_4'|^2} \tag{3.50}$$

With substitution of a_i from (3.48) and S_{ij}' from (3.47) the η_{C} is given in terms of the variables Γ_i. After simplification it turns out that this fraction is always 1, if $\Gamma_3 = \Gamma_4$ and $\Gamma_1 = 0$. The power sum at the output is therefore independent of Γ_2. In other words, if symmetry of the HP amplifiers is maintained ($\Gamma_3 = \Gamma_4$), the switched off LP amplifier path has no influence on the power summation.

Nevertheless, an output power or efficiency reduction occurs, if the load for each HP amplifier changes and the optimum load line match is affected. The amount of mismatch, caused by the LP amplifier's output impedance has to be determined in this regard. The relation between the mismatch level at port 2 (Γ_2) and the resulting impedance changes at ports 3 and 4 ($\Gamma_{\text{IN},3}$ and $\Gamma_{\text{IN},4}$) can be expressed with the generalized scattering parameters (3.47):

$$\Gamma_{\text{IN},i} = S_{ii}'\Big|_{\Gamma_i=0} \Rightarrow \begin{cases} \Gamma_{\text{IN},3} = \dfrac{\Gamma_1 - \Gamma_2 + 2\,\Gamma_1\Gamma_2\Gamma_4}{\Gamma_4(\Gamma_1 - \Gamma_2) - 2} \\ \\ \Gamma_{\text{IN},4} = \dfrac{\Gamma_1 - \Gamma_2 - 2\,\Gamma_1\Gamma_2\Gamma_3}{\Gamma_3(\Gamma_1 - \Gamma_2) + 2} \end{cases} \quad \text{with } \Gamma_3 = \Gamma_4 \tag{3.51}$$

Two cases of mismatch have to be distinguished. First case, the HP amplifiers experience the same load shift ($\Gamma_{\text{IN},3} = \Gamma_{\text{IN},4}$). This shift can be compensated by the matching networks, without influencing symmetry between the two HP paths. Equation (3.51) shows that this only occurs, if $\Gamma_1 = \Gamma_2$, as for the classical balanced amplifier. This condition has to be aimed at also in presence of the LP path. Second case, the two HP amplifiers experience a different load shift ($\Gamma_{\text{IN},3} \neq \Gamma_{\text{IN},4}$, if $\Gamma_1 \neq \Gamma_2$). For fully compensation the HP matching networks would have to be different $\Gamma_3 \neq \Gamma_4$), which causes unbalance and therefore reduces combining efficiency (3.50). To keep η_c high, it is better to maintain symmetry and to use the same matching networks ($\Gamma_3 = \Gamma_4$). As a consequence the output power contribution of

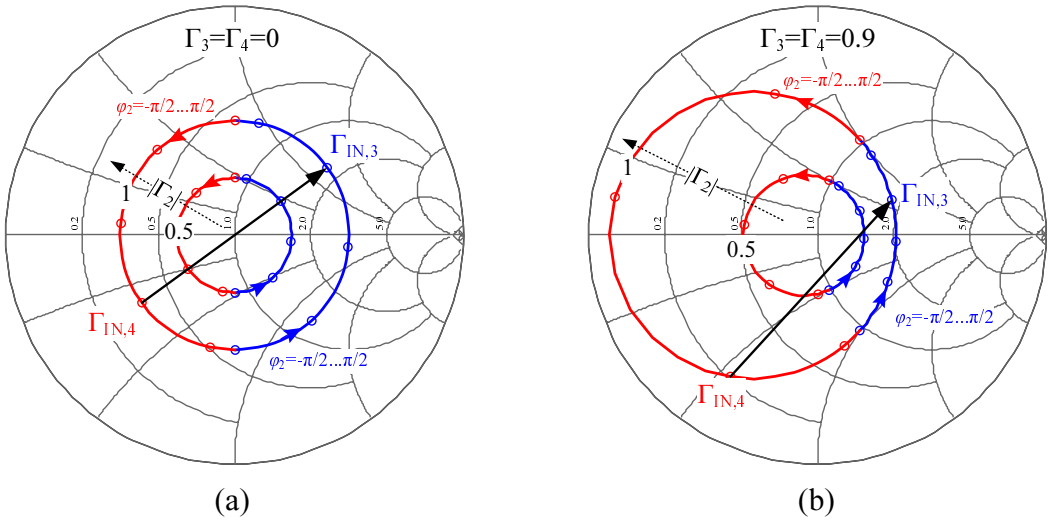


Figure 3.31: Smith chart plot of $\Gamma_{\mathrm{IN},3}$ and $\Gamma_{\mathrm{IN},4}$ contours versus Γ_2 at (a) $\Gamma_3 = \Gamma_4 = 0$ and (b) $\Gamma_3 = \Gamma_4 = 0.9$.

each HP amplifier will be different. To find an upper bound on Γ_2 for a tolerable amount of unbalance, equation (3.51) will be further analyzed.

Fig. 3.31 illustrates (3.51) with variation over $\Gamma_2 = |\Gamma_2|\, e^{j\varphi_2}$ in a smith chart. The same assumptions which lead to maximum combining efficiency are made here as well: $\Gamma_1 = 0$ and $\Gamma_3 = \Gamma_4$. As the HP PAs reflection coefficients Γ_3, Γ_4 during active operation can take on arbitrary values, they are not further constrained here. Exemplarily Fig. 3.31 shows the influence for two extreme cases, 0 and 0.9. A phase shift of Γ_3, Γ_4 would only rotate the drawn contours for Γ_2 about the center. The shape would remain unchanged. The first conclusion, which can be drawn from the figure is that imbalance occurs for every $\Gamma_2 > 0$. The second important conclusion is that the phase difference between $\Gamma_{\mathrm{IN},3}$ and $\Gamma_{\mathrm{IN},4}$ is almost constant (about 180°) while the amplitude imbalance increases with the magnitudes of $\Gamma_3 = \Gamma_4$ and Γ_2. This allows simplification of further analysis by considering only the magnitudes of $\Gamma_{\mathrm{IN},3}$ and $\Gamma_{\mathrm{IN},4}$. Substitution of $\Gamma_i = |\Gamma_i|\, e^{j\varphi_i}$ and $\Delta\varphi_{23} = \varphi_2 + \varphi_3$ into (3.51), results in:

$$|\Gamma_{\mathrm{IN},3}| = \frac{|\Gamma_2|}{\sqrt{|\Gamma_2\,\Gamma_4|^2 + 4 + 4\,|\Gamma_2\,\Gamma_4|\,\cos(\Delta\varphi_{23})}} \tag{3.52}$$

$$|\Gamma_{\mathrm{IN},4}| = \frac{|\Gamma_2|}{\sqrt{|\Gamma_2\,\Gamma_3|^2 + 4 - 4\,|\Gamma_2\,\Gamma_3|\,\cos(\Delta\varphi_{23})}} \tag{3.53}$$

The magnitude imbalance $|\Gamma_{\mathrm{IN},3}| - |\Gamma_{\mathrm{IN},4}|$ strongly depends on the cosine function in the denominator. For $\cos(\Delta\varphi_{23}) = \pm 1$ it is maximum and for $\cos(\Delta\varphi_{23}) = 0$ it is zero. These extreme cases are illustrated as contour plots in Fig. 3.32. For other $\Delta\varphi_{23}$ the results are lying in between. The figures shows that for small values of $|\Gamma_2| \leq 0.2$ there is little variation in the magnitudes, even for high reflection coefficients at ports 3 and 4. The resulting worst case magnitudes $|\Gamma_{\mathrm{IN},3}| \sim |\Gamma_{\mathrm{IN},4}|$ are about 0.1, which equals a VSWR of ~ 1.22 or $\sim 22\%$ load shift at the output of

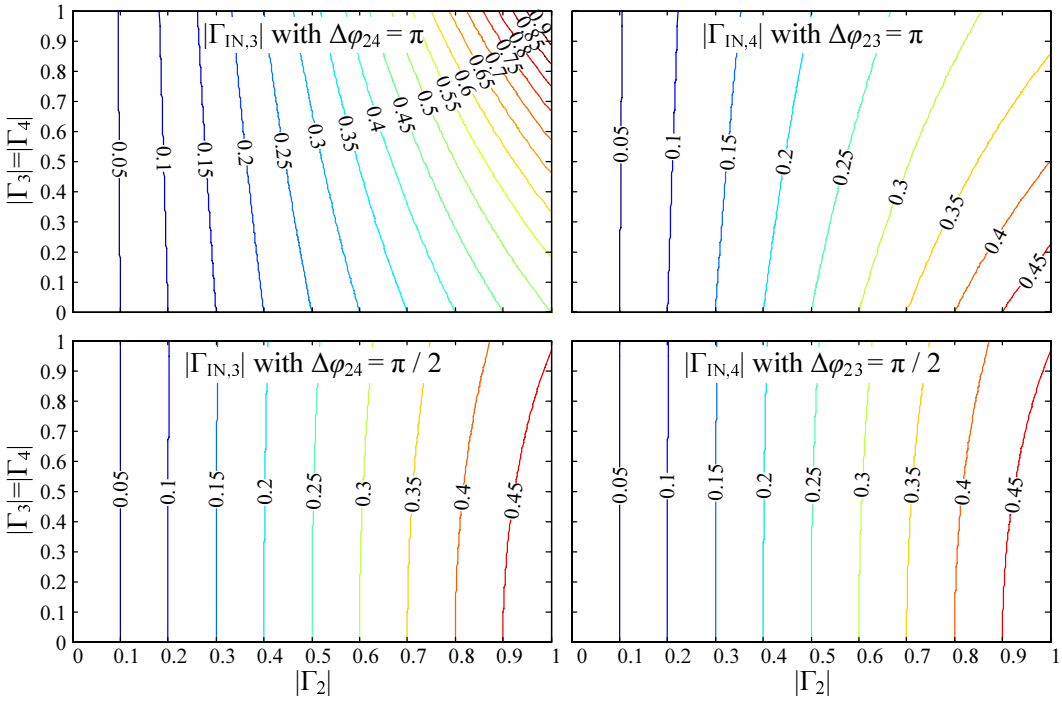


Figure 3.32: Contour plots of $|\Gamma_{\mathrm{IN},3}|$ and $|\Gamma_{\mathrm{IN},4}|$ with matched load $\Gamma_1 = 0$.

each PA. A further reduction of $|\Gamma_2|$ will reduce performance in LP operation mode, as will be shown later.

For the design of the HP paths following conclusions can be drawn. The two HP transistors have to be matched with the same network. The paths have to be capable to deliver the desired output power when loaded with $|\Gamma_{\mathrm{IN},3}| = |\Gamma_{\mathrm{IN},4}| = 0.1$ over all phase conditions. This requires an overhead design, being the main difference to the classical balanced amplifier. The amount of overhead is technology dependent and will further be influenced by errors of the 90° hybrid. Optimization with computer aided simulations and measurements will be inevitable to obtain an appropriate implementation. The analysis presented here is the basis for understanding the mechanisms, when the LP path is connected to the isolated port of the coupler.

Low Power mode

The next case to be analyzed is the LP operation mode. In this mode most power should be delivered from the LP amplifier at port 2 (see Fig. 3.29) to the load at port 1, while the HP amplifiers are turned off and present passive loads to ports 3 and 4. The power gain $G_{\mathrm{P},12}$ between port 2 and 1 ($\Gamma_1 = 0$) has to be maximized. Using the generalized scattering parameters this is written as:

$$G_{\mathrm{P},12} = \frac{|S'_{12}|^2}{1-|S'_{22}|^2} = \frac{|\Gamma_3+\Gamma_4|^2}{4-|\Gamma_3-\Gamma_4|^2} \xrightarrow{\Gamma_3 \sim \Gamma_4} \sim |\Gamma_3|^2 \tag{3.54}$$

From the previous discussion and for simplification, it can be assumed that the reflection coefficients seen into the HP paths are almost equal. The power gain is then proportional to the squared magnitude of the HP path reflection coefficients. There is no dependence on phase, as long as symmetry is maintained. Therefore, reflection coefficient magnitudes $|\Gamma_3| = |\Gamma_4|$ close to 1 result in high power transfer in LP mode. This condition is usually obtained for a high efficient HP amplifier path. The HP transistor's impedance in OFF state is highly reflective, it is equivalent to a large parasitic parallel capacitance and a small parallel conductance. The matching network between the HP transistor and the hybrid has low losses, why the OFF state reflection coefficient is primarily shifted in phase without being reduced in magnitude. Hence, the switched OFF HP path chain may be considered as almost lossless, presenting an arbitrary phase, but high reflective load to the hybrid, which increases $G_{\mathrm{P},12}$ close to 1.

Although attenuation by the hybrid can be kept small with low loss HP matching networks, the overall efficiency in LP mode will substantially be degraded by losses of the LP matching network. These losses are added intentionally to fulfill matching conditions in HP and LP mode with a single passive network, as illustrated in Fig. 3.33. The boundary conditions and their impact are analyzed next.

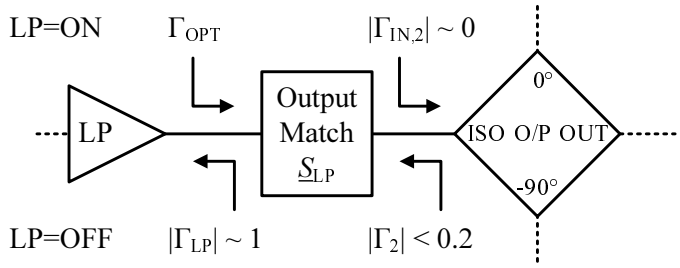


Figure 3.33: Boundary conditions for impedance transformation in LP ON and OFF state.

If the LP amplifier is switched ON, the matching network has to transform $\Gamma_{\mathrm{IN},2}$ into Γ_{OPT} with lowest possible loss, to guarantee operation at high efficiency. $\Gamma_{\mathrm{IN},2}$ is the load for the LP path seen into port 2 of the hybrid, as shown in Fig. 3.33. It can be calculated using (3.47):

$$\Gamma_{\mathrm{IN},2} = \frac{1}{2}\left(\Gamma_3 - \Gamma_4\right) \xrightarrow{\Gamma_3 \sim \Gamma_4} \sim 0, \tag{3.55}$$

hence, for symmetry it equals the load impedance $\Gamma_1 = 0$. The optimum impedance for the LP transistor Γ_{OPT} depends on the output power level and operation class.

If the LP amplifier is switched OFF, the same network has to transform the high reflective transistor output $\Gamma_{\mathrm{LP}} \sim 1$ to $|\Gamma_2| < 0.2$. Remember, that the upper bound for $|\Gamma_2|$ was found during analysis of the HP mode to improve symmetry and thus guarantee high efficient operation there.

In general the transformations in LP=ON and LP=OFF state are not simultaneously realizable with a linear, reciprocal and low loss network. In order to fulfill both

conditions and maintain small losses in LP mode, some tuning or switching elements, e.g. varactor or pin-diode switches could be used. These elements increase complexity of the network and degrade linearity. Another method to enforce matching in ON and OFF state is to design a network with a certain tolerable loss. This is untypical for PAs but sufficient for the LP mode being already much more efficient at lower power levels than the HP mode due to its lower bias current.

For design of the LP path a compromise between HP and LP mode performance has to be found, mainly determined by losses in the LP matching network. Higher losses, forcing smaller $|\Gamma_2|$ increase symmetry and efficiency in HP mode. Smaller losses resulting in larger $|\Gamma_2|$ increase efficiency in LP mode. If a certain minimum efficiency in HP mode has to be achieved, an upper bound target for $|\Gamma_2|$ has to be specified, as already shown. With this value it is possible to define a theoretical lower bound for losses in the LP matching network. This lower bound is necessary to estimate potential efficiency improvement in LP mode and therefore rate the benefit of this concept. An analytical method to find this lower bound is explained below.

Synthesis of LP matching network with lowest possible loss

For the synthesis of the LP path matching network, it is represented by its 2-port scattering matrix $\underline{S}_{\mathrm{LP}}$. With reference to Fig. 3.33 port 1 is connected to the amplifier and port 2 is connected to the hybrid.

$$\underline{S}_{\mathrm{LP}} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{3.56}$$

This scattering matrix approach eases investigations, as no restrictions to a specific circuit realization or topology of the matching network have to be made. The network synthesis is then reduced to computation of the four unknown scattering parameters, while certain conditions have to be fulfilled:

(i) If port 2 is terminated with $\Gamma_{\mathrm{IN},2}$, the reflection coefficient seen into port 1 has to be Γ_{OPT}.

(ii) If port 1 is terminated with Γ_{LP}, the reflection coefficient seen into port 2 has to be Γ_2.

(iii) If the LP amplifier is active (condition (i)), losses of the network have to be as low as possible, i.e. the operating power gain $G_{\mathrm{P,LP}}$ has to be maximized.

(iv) The network has to be realizable with real and passive components such as capacitors, inductors, transmission lines and resistors. Therefore the scattering matrix has to satisfy the realizability properties. It has to be linear, reciprocal and passive.

Conditions (i) and (ii) are equated as described in Appendix B.2:

$$\Gamma_{\mathrm{OPT}} = \frac{b_1}{a_1} = S_{11} + \frac{S_{12}S_{21}\Gamma_{\mathrm{IN},2}}{1 - S_{22}\Gamma_{\mathrm{IN},2}} \tag{3.57}$$

$$\Gamma_2 = \frac{b_2}{a_2} = S_{22} + \frac{S_{12}S_{21}\Gamma_{\mathrm{LP}}}{1 - S_{11}\Gamma_{\mathrm{LP}}} \tag{3.58}$$

The operating power gain for the lowest possible loss condition (iii) can be written as (definition see Appendix C):

$$G_{\mathrm{P}} = \underbrace{\frac{|S_{21}|^2}{|1 - \Gamma_{\mathrm{IN},2}\, S_{22}|^2}}_{\text{maximize over } \mathrm{S}_{21}, \mathrm{S}_{22}} \cdot \underbrace{\frac{1 - |\Gamma_{\mathrm{IN},2}|^2}{1 - |\Gamma_{\mathrm{OPT}}|^2}}_{\text{constant}} \tag{3.59}$$

The proof for a network being reciprocal, as requested in condition (iv) is shown, if its scattering matrix is symmetric along the main diagonal:

$$\underline{S}_{\mathrm{LP}} = \underline{S}_{\mathrm{LP}}^{\mathrm{T}} \tag{3.60}$$

The condition for a network being passive can be derived from the power conservation theorem, as in [Carlin 56]. The derivation leads to a quadratic form for the absorbed power in the network, $P_{\mathrm{loss}} = \underline{V}^{*T}\, \underline{Q}\, \underline{V}$. The port voltages are collected in the vector $\underline{V}$ and the hermitian matrix $\underline{Q}$ can be expressed by the network scattering matrix $\underline{S}_{\mathrm{LP}}$ and the unity matrix $\underline{U}$:

$$\underline{Q} = \underline{U} - \underline{S}_{\mathrm{LP}}^{*\mathrm{T}} \cdot \underline{S}_{\mathrm{LP}} \tag{3.61}$$

Losses in the network can be identified by testing certain properties of $\underline{Q}$:

$$\underline{Q} \text{ is } \begin{cases} \underline{0}, & \text{if power absorbed in the network} = 0 \text{ (lossless);} \\ \text{positive definite,} & \text{if power absorbed in the network} > 0; \\ \text{positive semidefinite,} & \text{if power absorbed in the network} \geq 0. \end{cases}$$

$\underline{Q}$ is *positive definite*, if all *leading principal minors* are nonnegative. A leading principal minor of $\underline{Q}$ of order r is obtained by deleting all but the first r rows and the first r columns of $\underline{Q}$, then taking the determinant of the resulting $r \times r$ matrix. A matrix of dimension $n \times n$ has n leading principal minors.

$\underline{Q}$ is *positive semidefinite*, if all *principal minors* are nonnegative. A principal minor of $\underline{Q}$ of order r is obtained by deleting all but r rows and the corresponding r columns of $\underline{Q}$, then taking the determinant of the resulting $r \times r$ matrix.

The LP path matching network is not necessarily lossy, hence, the power absorbed is assumed ≥ 0 and positive semidefiniteness has to be proven for $\underline{Q}$:

$$Q = \begin{pmatrix} 1 - \left(|S_{11}|^2 + |S_{21}|^2\right) & S_{11}^* S_{12} + S_{21}^* S_{22} \\ S_{12}^* S_{11} + S_{22}^* S_{21} & 1 - \left(|S_{22}|^2 + |S_{12}|^2\right) \end{pmatrix} \tag{3.62}$$

$$\Rightarrow \text{principal minor determinants} \begin{cases} |S_{11}|^2 + |S_{21}|^2 \leq 1 \\ |S_{22}|^2 + |S_{12}|^2 \leq 1 \\ \det\left(\underline{Q}\right) \geq 0 \end{cases} \tag{3.63}$$

The system of complex equations (3.57), (3.58), (3.59), (3.60), (3.63) has to be solved for the four unknown S-parameters. Due to inequalities (3.63), there will be an infinite number of solutions, which can be made finite by maximizing equation (3.59). An explicit solution cannot be obtained in general, because the system of equations is complex and nonlinear. The standard solution procedure such as the successive variable eliminating algorithm has to be adapted by a numerical approximation.

First with (3.60) ($S_{12} = S_{21}$) the number of unknown complex variables can be reduced to three: S_{11}, S_{22} and S_{21}.

After that, equations (3.57) and (3.58) are separated for S_{11} and S_{22}:

$$S_{11}^2 - S_{11}\underbrace{\frac{(1+\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}})}{\Gamma_{\mathrm{LP}}}}_{A_1} + \underbrace{\frac{\Gamma_{\mathrm{OPT}}}{\Gamma_{\mathrm{LP}}}}_{B_1} = S_{21}^2\underbrace{\frac{\Gamma_{\mathrm{IN},2}}{\Gamma_{\mathrm{LP}}}\frac{(1-\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}})}{(1-\Gamma_2\Gamma_{\mathrm{IN},2})}}_{C_1} \tag{3.64}$$

$$S_{22}^2 - S_{22}\underbrace{\frac{(1+\Gamma_2\Gamma_{\mathrm{IN},2})}{\Gamma_{\mathrm{IN},2}}}_{A_2} + \underbrace{\frac{\Gamma_2}{\Gamma_{\mathrm{IN},2}}}_{B_2} = S_{21}^2\underbrace{\frac{\Gamma_{\mathrm{LP}}}{\Gamma_{\mathrm{IN},2}}\frac{(1-\Gamma_2\Gamma_{\mathrm{IN},2})}{(1-\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}})}}_{C_2} \tag{3.65}$$

By taking the absolute value, the equations can be identified as *Cassini Ovals* [Lawrence 72] in the complex S_{ii} plane with focus points M_{i1}, M_{i2} and constant D_i ($i = 1, 2$):

$$\left|S_{ii}^2 - S_{ii}A_i + B_i\right| = |C_i|\ |S_{21}|^2 \tag{3.66}$$

$$\Rightarrow\ \left|S_{ii}^2 - M_{i1}\right| \cdot \left|S_{ii}^2 - M_{i2}\right| = D_i, \tag{3.67}$$

$$M_{11} = \frac{1}{\Gamma_{\mathrm{LP}}}\ ,\ M_{12} = \Gamma_{\mathrm{OPT}}\ ,\ M_{21} = \frac{1}{\Gamma_{\mathrm{IN},2}}\ ,\ M_{22} = \Gamma_2 \text{ and } D_i = |C_i|\ |S_{21}|^2 \tag{3.68}$$

Since only passive loads ($|\Gamma| < 1$) are considered, the foci M_{i1} lie outside the unity cycle, while M_{i2} lie inside. The appearance of the Cassini oval is determined by the parameter F:

$$F = \frac{D_i}{|M_{i1} - M_{i2}|} = 2|S_{21}|\sqrt{\frac{\Gamma_{\mathrm{LP}}\Gamma_{\mathrm{IN},2}}{|1-\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}}|\,|1-\Gamma_2\Gamma_{\mathrm{IN},2}|}} \tag{3.69}$$

$F > 1$: there is one loop comprising both foci

$F = 1$: the two loops meet between the foci (leminiscate)

$F < 1$: there are two separate loops, each comprising its corresponding focus

The two loops correspond to the two roots (subscript a and b), obtained as solution of equations (3.64) and (3.65):

$$S_{11\mathrm{a,b}} = \frac{(1-\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}})}{2\Gamma_{\mathrm{LP}}} \cdot \left[\frac{(1+\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}})}{(1-\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}})} \pm \sqrt{1+S_{21}^2 \frac{4\Gamma_{\mathrm{LP}}\Gamma_{\mathrm{IN},2}}{(1-\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}})\,(1-\Gamma_2\Gamma_{\mathrm{IN},2})}}\right] \tag{3.70}$$

$$S_{22\mathrm{a,b}} = \frac{(1-\Gamma_2\Gamma_{\mathrm{IN},2})}{2\Gamma_{\mathrm{IN},2}} \cdot \left[\frac{(1+\Gamma_2\Gamma_{\mathrm{IN},2})}{(1-\Gamma_2\Gamma_{\mathrm{IN},2})} \pm \sqrt{1+S_{21}^2 \frac{4\Gamma_{\mathrm{LP}}\Gamma_{\mathrm{IN},2}}{(1-\Gamma_{\mathrm{OPT}}\Gamma_{\mathrm{LP}})\,(1-\Gamma_2\Gamma_{\mathrm{IN},2})}}\right] \tag{3.71}$$

Only solution pairs (S_{11a}, S_{22a}) and (S_{11b}, S_{22b}) satisfy (3.57) and (3.58). A graphical representation of the solutions, which have the shape of Cassini Ovals, is shown in Fig. 3.34. With these ovals the first estimation on minimum necessary losses in the network can be made. In order to fulfill the first two inequalities of (3.63), $|S_{ii}|$ have

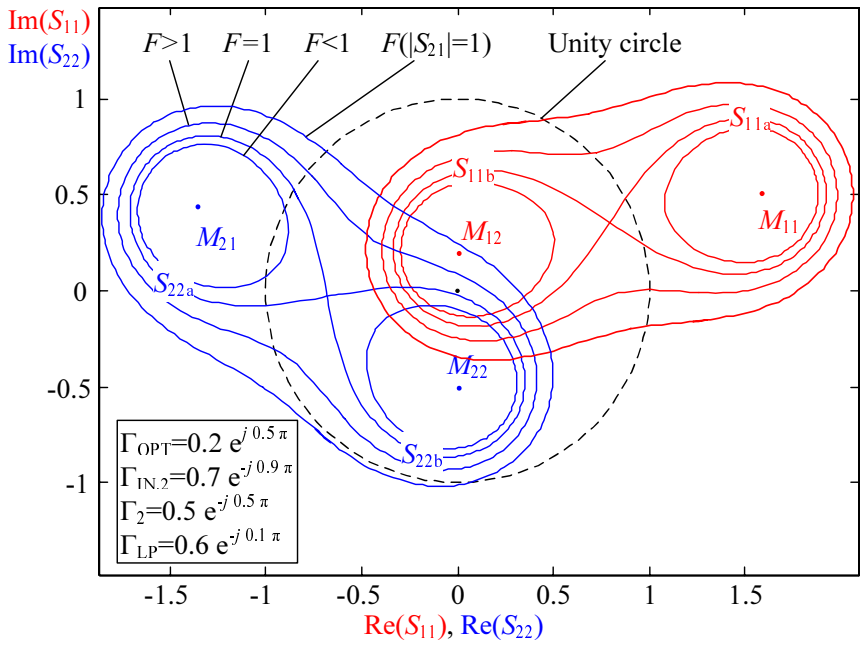


Figure 3.34: Cassini ovals of S_{11} and S_{22} in the complex plane.

to be ≤ 1 and therefore only values inside the unity circle lead to potential solutions. Satisfying the inequalities for maximum gain ($|S_{21}| \sim 1$) demands a $|S_{ii}|$ close to zero. Therefore the ovals with largest $F(|S_{21}| = 1)$ have to cross the center. If not, the network will not be lossless. To keep losses low, the loops have to be as close to zero as possible. The closer they are the higher the realizable $|S_{21}|$ and therefore G_{P}. In the example of Fig. 3.34 the oval $F(|S_{21}| = 1)$ does not cross the center, but the oval branches, which correspond to the solution pair (S_{11b}, S_{22b}) are very close to zero. Therefore the network will not be lossless, but can potentially have "small" losses, when using solutions (S_{11b}, S_{22b}). A first evaluation on basis of the position

and shape of the Cassini ovals can only be a rough estimation for excluding cases which lead to networks with very high attenuation. In the example the solutions (S_{11a}, S_{22a}) can be excluded. Nevertheless, with help of the Cassini ovals effort can be saved, which would be necessary for a time consuming computation of an exact solution. The exact quantitative evaluation has to be done numerically.

The numerical solution is based on (3.70) and (3.71). Substituting $S_{22a,b}$ into the operating power gain equation (3.59), only S_{21} remains as variable. The range of S_{21} is $|S_{21}| = 0 \ldots 1$ and $\varphi_{21} = -\pi \ldots \pi$, as only passive networks are considered. A computer aided sweep over the whole range can be performed to get contour plots of $G_{\mathrm{P}a}$ and $G_{\mathrm{P}b}$ versus $|S_{21}|$ and φ_{21}, as shown in Fig. 3.35.

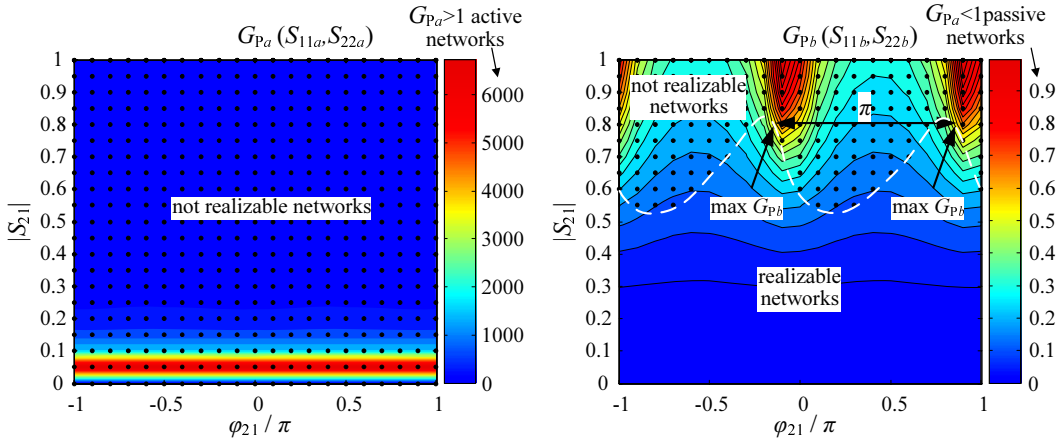


Figure 3.35: Contour plots of $G_{\mathrm{P}a,b}$ versus S_{21} obtained by elimination of (S_{11a}, S_{22a}) and (S_{11b}, S_{22b}), with $\Gamma_{\mathrm{OPT}} = 0.2\,e^{j0.5\pi}$, $\Gamma_{\mathrm{IN},2} = 0.7\,e^{-j0.9\pi}$, $\Gamma_2 = 0.5\,e^{-j0.5\pi}$ and $\Gamma_{\mathrm{LP}} = 0.6\,e^{-j0.1\pi}$. Areas for non realizable networks are marked by black circles.

Not every value on these plots leads to a realizable network. To find the S_{21} for realizable networks, (3.63) has to be evaluated. Therefore $S_{11a,b}$ and $S_{22a,b}$ are substituted in inequalities (3.63), which then also depend on S_{21} only. For each S_{21} in the sweep range of the $G_{\mathrm{P}a,b}$ plot, the inequalities are checked. The S_{21} values which do not fulfill the inequalities are marked by circles in the plot. This way two areas can be identified. The area filled with circles, which corresponds to not realizable active networks and the circle free area which gives the solutions for realizable passive networks. The border between these areas is marked with a dashed line in Fig. 3.35. As already shown with the Cassini Ovals, the solution pair (S_{11a}, S_{22a}) does not lead to realizable networks and the plot for $G_{\mathrm{P}a}$ will not be further investigated.

Now, the problem, which was set at the beginning to find a network with minimum loss, when the LP path is ON can be solved with help of the $G_{\mathrm{P}b}$ plot. Within the area of realizable networks the maximum values for $G_{\mathrm{P}b}$ can be found by inspection. The $|S_{21}|$, φ_{21} values, corresponding to max$G_{\mathrm{P}b}$ are read off and inserted to (3.70) and (3.71). This finally gives the complete $\underline{S}_{\mathrm{LP}}$ matrix with the desired properties. Accuracy of the solution depends on step size of the S_{21} sweep. The higher the resolution, the more exact the edge of the decision areas in the plot and therefore the

maximum $G_{\mathrm{P}b}$. For the example a maximum of $G_{\mathrm{P}b} = 0.36 \sim -4.4$ dB is obtained at $S_{21} = 0.76\, e^{-j0.13\pi}$ or $S_{21} = 0.76\, e^{j0.87\pi}$ (phase shifted by π) and the resulting scattering matrix is:

$$\underline{S}_{\mathrm{LP}} = \begin{pmatrix} 0.41e^{j0.019\pi} & 0.76e^{j-0.13\pi} \\ 0.76e^{j-0.13\pi} & 0.18e^{j-0.86\pi} \end{pmatrix} \text{ or } \begin{pmatrix} 0.41e^{j0.019\pi} & 0.76e^{j0.87\pi} \\ 0.76e^{j0.87\pi} & 0.18e^{j-0.86\pi} \end{pmatrix} \tag{3.72}$$

In both cases S_{11} and S_{22} are the same, due to the quadratic dependence of S_{21} in (3.70) and (3.71), which eliminates a phase shift by π.

A lumped or distributed passive element matching network, corresponding to the generated scattering matrix has to be found as next step. As no restrictions were made at the beginning the resulting network will be of arbitrary complexity. This means that its number of elements may be infinite and its topology may not be suitable for a PA output matching (e.g. containing a DC path to ground). The calculated solution with minimum loss is therefore mostly of theoretical interest, as it provides an upper bound for losses in the matching network. On the one hand it shows if a realization is possible in principle and how much losses have to be contained. On the other hand the losses of an implemented network can be compared to this ideal solution to see, how close the realization is to the optimum.

A more practical approach for synthesis of the LP network is to predefine its topology. The advantage is that decoupling of RF and supply can be considered (bias T) and the matching be made appropriate for PA applications (low path characteristic). Consequently the network may have higher attenuation, due to less topological degrees of freedom. To find the network with smallest performance degradation within the defined topology, the two port Z- or Y-matrix is calculated for every S_{21} sweep value. Both matrices are relevant, because for some networks either of them does not exist:

$$\underline{Z}_{\mathrm{LP}} = Z_0(\underline{U} + \underline{S}_{\mathrm{LP}})(\underline{U} - \underline{S}_{\mathrm{LP}})^{-1} = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \tag{3.73}$$

$$\underline{Y}_{\mathrm{LP}} = \frac{1}{Z_0}(\underline{U} - \underline{S}_{\mathrm{LP}})(\underline{U} + \underline{S}_{\mathrm{LP}})^{-1} = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} \tag{3.74}$$

Also a symbolic expression of Z- or Y-matrix, representing the topology, has to be calculated. For illustration of the method a simple Π- and T-network consisting of

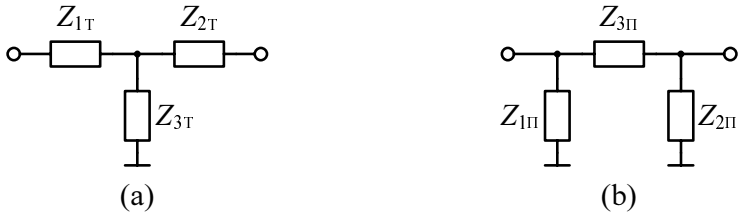


Figure 3.36: (a) T, (b) Π circuit as basis for minimum loss network realization.

three complex impedances (Fig. 3.36) are chosen. The symbolic representation of

the matrices are:

$$\underline{Z}_{\mathrm{LPT}} = \begin{pmatrix} Z_{1\mathrm{T}} + Z_{3\mathrm{T}} & Z_{3\mathrm{T}} \\ Z_{3\mathrm{T}} & Z_{2\mathrm{T}} + Z_{3\mathrm{T}} \end{pmatrix}, \quad \underline{Y}_{\mathrm{LP\Pi}} = \begin{pmatrix} \frac{1}{Z_{1\Pi}} + \frac{1}{Z_{3\Pi}} & -\frac{1}{Z_{3\Pi}} \\ -\frac{1}{Z_{3\Pi}} & \frac{1}{Z_{2\Pi}} + \frac{1}{Z_{3\Pi}} \end{pmatrix}$$

By solving $\underline{Z}_{\mathrm{LP}} = \underline{Z}_{\mathrm{LPT}}$ or $\underline{Y}_{\mathrm{LP}} = \underline{Y}_{\mathrm{LP\Pi}}$ the network elements are obtained:

$$\begin{aligned} Z_{1\mathrm{T}} &= Z_{11} - Z_{21} & Z_{1\Pi} &= \frac{1}{Y_{11}+Y_{21}} \\ Z_{2\mathrm{T}} &= Z_{22} - Z_{21} & Z_{2\Pi} &= \frac{1}{Y_{22}+Y_{21}} \\ Z_{3\mathrm{T}} &= Z_{21} & Z_{3\Pi} &= -\frac{1}{Y_{21}} \end{aligned}$$

The $Z_{i\mathrm{T}}$ & $Z_{i\Pi}$ are complex impedances, realized as serial or parallel connection of R and L/C. Therefore the real part of each impedance has to be ≥ 0, to make it realizable. Depending on the type of connection (parallel or serial) and component (C or L), the imaginary part can be positive or negative. A similar analysis as for the realizability condition can be performed. The S_{21} is swept in amplitude and phase, for each value all $Z_{i\mathrm{T}}$ or $Z_{i\Pi}$ are calculated from the Z- or Y-matrix and the sign of the real part is checked. If one of the real parts is negative, a mark is plotted into the corresponding contour plots of $G_{\mathrm{P}a,b}$. Only if all real parts are positive the mark is omitted. As a result the contour plot can again be divided into realizable and not realizable areas and minimum loss is found by inspection (see Fig. 3.37). With the resulting S_{21}, the $Z_{i\mathrm{T}}$ or $Z_{i\Pi}$ values are determined and from these the lumped component values and type of connection. This completes the synthesis of a minimum loss network. For this network synthesis it is required that the network impedances can be expressed in terms of $\underline{Z}$- or $\underline{Y}$-matrix elements. If an explicit analytical expression cannot be found, a numerical analysis has to be done, which increases computation time and reduces accuracy.

Fig. 3.37 shows the results for both network types. The attenuation of the T-network (-4.8 dB) is very close to the theoretical minimum of -4.4 dB. A much higher attenuation of -6.7 dB is obtained for the Π-network in this particular case. The reduction of drain efficiency can directly be calculated by multiplying η obtained with a loss less network or from load pull measurements with the corresponding G_{P}. Assuming a LP amplifier with e.g. $\eta = 50$ %, the drain efficiency is reduced to 18 % with the theoretical optimum network. A T-network gives at best 16.5 % and a Π-network gives at best 10.5 %. For this example the loss in efficiency is remarkable. Nevertheless, compared to the HP mode the drain efficiency at the same output power levels may still be larger in LP mode. In a practical design the losses in LP mode can be reduced by compromising efficiency in HP mode. Therefore a trade-off will have to be found based on optimization of the average current consumption.

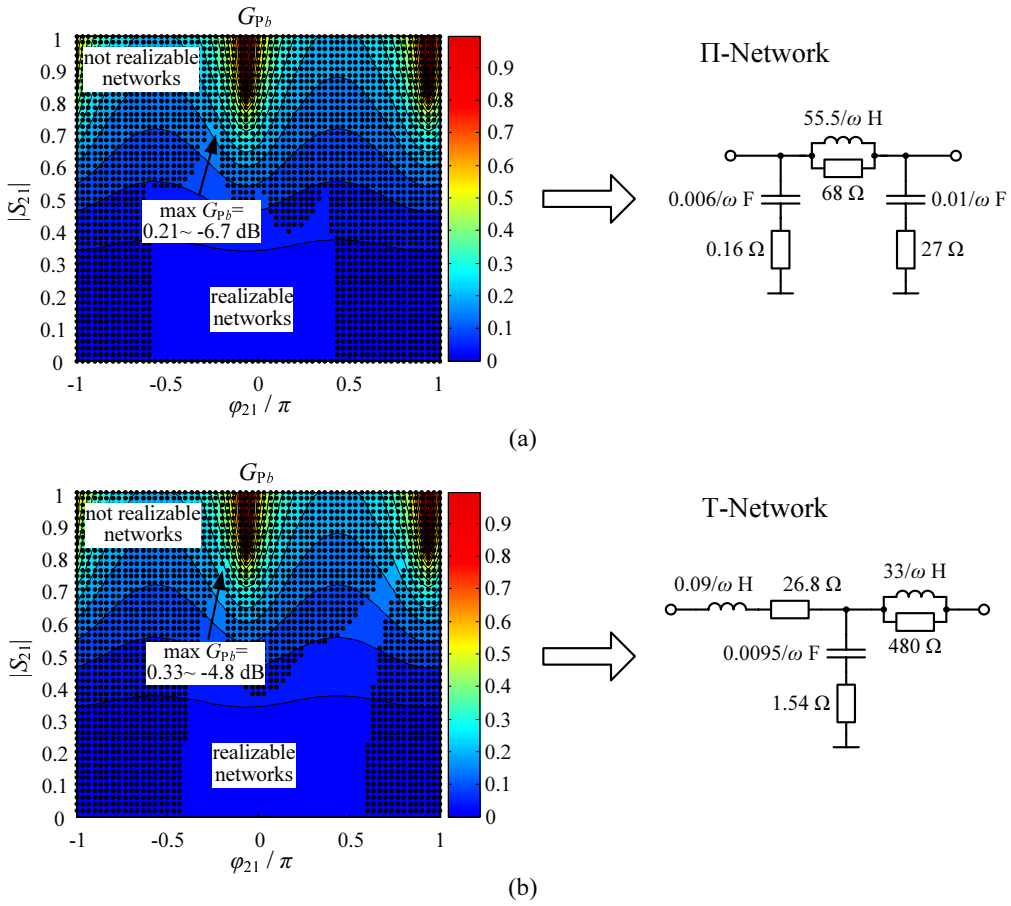


Figure 3.37: G_{Pb} contour plots with topology dependent realizability areas and resulting minimum loss networks for (a) Π-, (b) T-topology.

3.5.3 Load Insensitivity

The balanced switched path amplifier topology has inherent mechanisms to reduce variations of the antenna impedance in HP and LP operation mode, without need of additional circuitry.

High Power mode

Application of the output hybrid helps to improve insensitivity against load variations at the output port. Some amount of reflected power is terminated by the lossy LP path at the isolated port. Therefore Γ_1 variations appear with reduced magnitude at direct ($\Gamma_{IN,4}$) and coupled ($\Gamma_{IN,3}$) port. The dependence between these reflection coefficients is given by equation (3.51). In HP mode the reflection coefficients seen into the HP amplifiers are in worst case $|\Gamma_3| = |\Gamma_4| \sim 0.9$. This corresponds to high reflection of the transistor (close to 1) and ~ 0.5 dB attenuation of the HP matching networks. The maximum reflection coefficient seen into the LP path is $|\Gamma_2| \leq 0.2$, as deduced from Fig. 3.32. With these conditions the magnitudes of

$\Gamma_{IN,3}$ and $\Gamma_{IN,4}$ are bounded, which allows conclusions on load insensitivity. For the special case $|\Gamma_2| = 0$, the contour plots from Fig. 3.32 can be used with exchanged subscripts $1 \leftrightarrow 2$ and $3 \leftrightarrow 4$ to determine the maximum variation. In all other cases an additional dependence of φ_2 has to be considered, which makes it difficult to illustrate in a contour plot. Therefore a sweep over all angles φ_1, φ_2 and $\varphi_3 = \varphi_4$ at fixed magnitudes Γ_2 and $\Gamma_3 = \Gamma_4$ is performed and maximum and minimum are determined for each $|\Gamma_1|$. The result is shown in Fig. 3.38 where the dotted area shows a boundary for load variance. For very small load variations $|\Gamma_1| < 0.05$ there

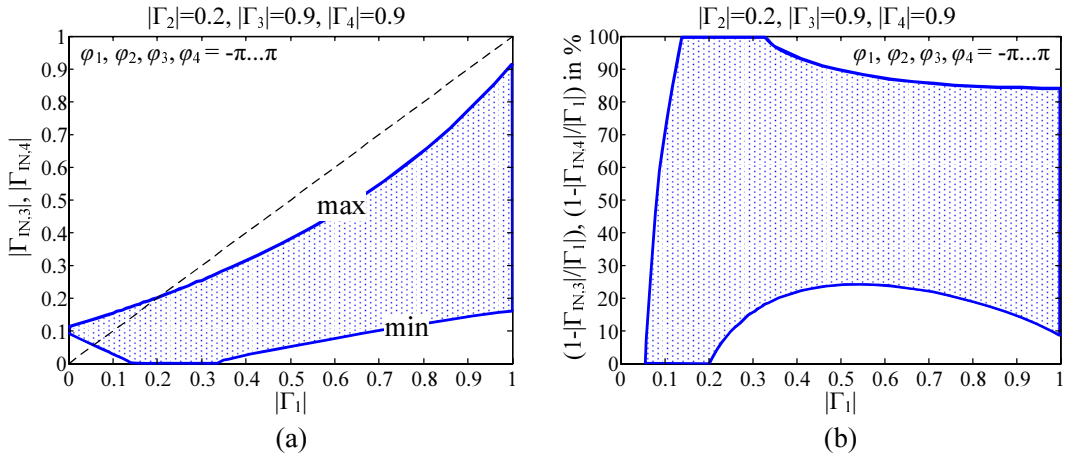


Figure 3.38: Reduction of antenna load variations in HP mode due to the hybrid. (a) absolute values (b) relative improvement in percent, over antenna load mismatch $|\Gamma_1|$.

is no improvement, but amplifier performance parameters (e.g. output power, efficiency,...) are not much influenced in that range. In the range $0.05 < |\Gamma_1| < 0.3$ the load variation remains unchanged or can even be canceled out. For load variations up to the maximum $|\Gamma_1| = 1$, the magnitude is always reduced and improvement can be between $10 \ldots 100$ %. This analysis shows that there exists no general statement for the amount of load insensitivity. The amount depends strongly on phase relations between impedances at all ports of the hybrid. Most critical for operation in HP mode are high reflection coefficients at the load, because these can lead to breakdown of the amplifier. With a careful and accurate design load variations can be kept below 0.17 even for $|\Gamma_1| = 1$, which is an improvement by 83 % for at least one HP path.

Low Power mode

Also in LP mode antenna impedance variations are reduced. This is due to the imperfect $|\Gamma_3| = |\Gamma_4| < 1$ which causes attenuation of the reflected wave. With a similar worst case assumption of ~ 0.5 dB loss, as for the HP mode, the insertion loss between LP matching network and load is $G_{P,12} = 0.9^2 \triangleq -0.92$ dB (see equation (3.54)). The return loss seen into the isolated port is therefore improved by at least

$2\,G_{\mathrm{P},12_{[\mathrm{dB}]}} = -1.84$ dB. In contrast to HP mode this reduction of load variation by 19 % is constant over all amplitudes and phases of Γ_1.

Following the same principle, another improvement is obtained by the intended losses of the LP matching network. Typically these will be $1 \ldots 2$ dB, why an overall load insensitivity of about $36 \ldots 49$ % can usually be achieved in LP mode.

3.6 Multi-mode PA with improved load insensitivity – Double balanced switched path PA

The previously introduced switch-less multi-mode PA concept allows efficiency improvement at one arbitrary power back-off level. For pure WCDMA operation it can be used to reduce the average current consumption significantly, by acitvating a low power path at more than 16 dB below the maximum output power. The approach of this section is to introduce the concept for multi-mode operation with several efficiency improved back-off power levels, based on the previously introduced operation principle. Depending on the communication standard (GSM, EDGE, WCDMA) and the actual transmission power, the multi-mode PA then offers the ability to reconfigure for an optimum compromise between efficiency and linearity. In comparison to the balanced switched path architecture the operation modes will offer inherent mechanisms to further reduce impact of antenna impedance variations.

3.6.1 Block diagram

The topology of this amplifier is shown in Fig. 3.39. It consists of two identical balanced amplifier paths (PATH1 and PATH2) building an overall balanced amplifier with additional quadrature hybrids at input and output. The performance and design of PATH1 and PATH2 were already discussed in Chapter 3.5. By doubling the balanced amplifier paths the number of possibilities for switching amplifier stages ON and OFF is extended and therefore the number of potential efficiency improvement steps. Due to further operation modes, additional switches SW1, SW2 and SW3 are optional for a performance improvement in some of these modes.

There exist 9 different modes of operation, which are summarized in Table 3.6: Three high power modes (HPa, HPb, HPc) for GSM/EDGE, four medium power modes (MPa, MPb, MPc, MPd) for GSM/EDGE/WCDMA and two low power modes (LPa, LPb) for EDGE/WCDMA in power back-off. The corresponding maximum output power of each mode can be calculated by summing up the contributions of each activated HP or LP path. Only in HPb, HPc and MPc mode the output hybrid 1 acts as equal power combiner with unequal input signals. Instead of a simple summation the total output power has to be calculated using equation (A.3) from Appendix A.1. All operation modes, except the MPd mode provide improved load insensitivity due to attenuation of the reflected wave in case of mismatch.

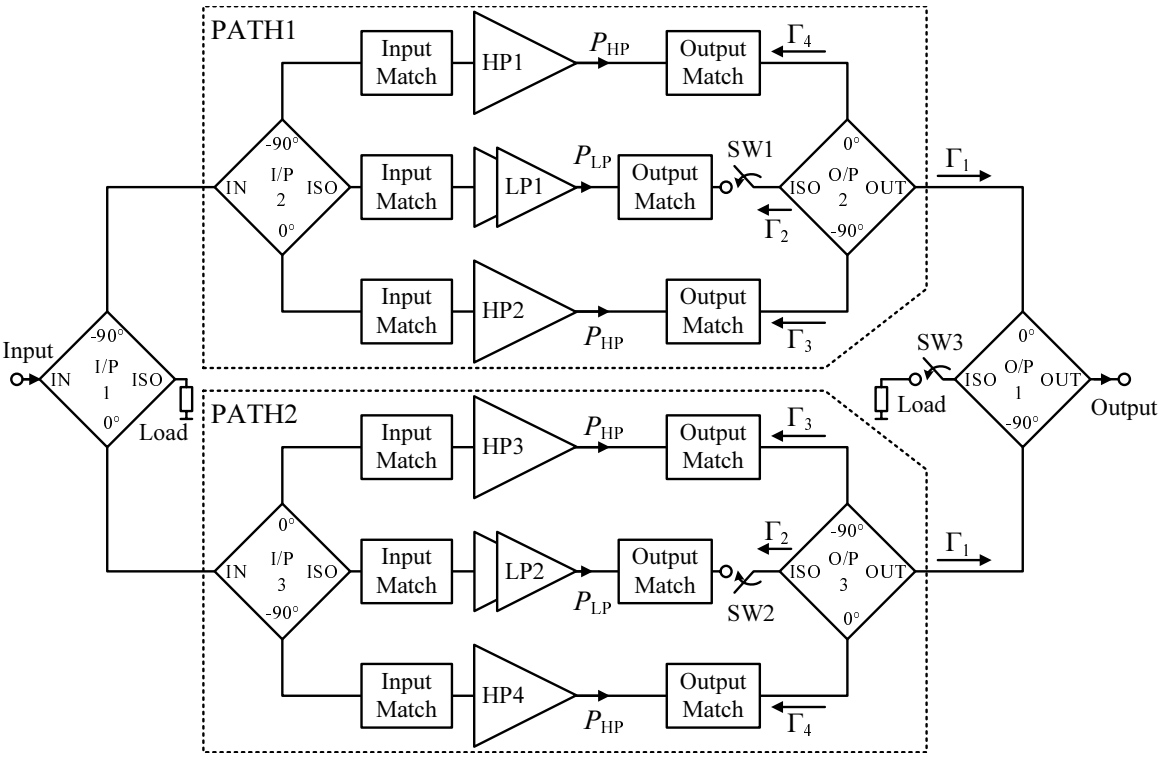


Figure 3.39: Double balanced switched path power amplifier.

Settings	**HPa**	**HPb**	**HPc**	**MPa**	**MPb**	**MPc**	**MPd**	**LPa**	**LPb**
HP1	1	1	1/0	1/0	1/0	1/0	1/0	0	0
HP2	1	1	1/0	0/1	1/0	0/1	0/1	0	0
HP3	1	0	0/1	0/1	0/1	0	0	0	0
HP4	1	1	0/1	1/0	0/1	0	0	0	0
LP1	0	0	0/1	0	0	0	0	1	1/0
LP2	0	0	1/0	0	0	1	0	1	0/1
SW1	1	1	1	0	1	0	0	1	1/0
SW2	1	0	1	0	0	1	0	1	0/1
SW3	1	1	1	1	0	1	0	1	0
$P_{\text{out,max}}$	$4\,P_{HP}$	$2.91\,P_{HP}$	$\frac{(\sqrt{2\,P_{HP}}+\sqrt{P_{LP}})^2}{2}$	$2\,P_{HP}$	$2\,P_{HP}$	$\frac{(\sqrt{P_{HP}}+\sqrt{P_{LP}})^2}{2}$	P_{HP}	$2\,P_{LP}$	P_{LP}
VSWR prot.	✓	✓	✓	✓	✓	✓	–	✓	✓
GSM	✓	✓	✓	✓	✓	✓	✓	–	–
EDGE	✓	✓	✓	✓	✓	✓	✓	✓	✓
WCDMA	–	–	–	✓	✓	✓	✓	✓	✓

Table 3.6: Double balanced switched path PA operation modes.
(Switches SW: 1=CLOSED, 0=OPEN; Amplifiers HP/LP: 1=ON, 0=OFF)

3.6.2 Operation modes

HPa mode

In this mode the PA delivers maximum output power $P_{\max} = 4\,P_{\mathrm{HP}}$. The three hybrids at input and output act as ideal power dividers/combiners with high isolation to the LP paths. Symmetry at direct and coupled ports of the hybrid $\Gamma_3 = \Gamma_4$ as well as small return loss at the isolated port $\Gamma_2 \leq 0.2$ are necessary for high PAE and P_{out} values. A more detailed description on the operation conditions in presence of the LP path was already given in Chapter 3.5.2.

In this high power mode maximum load insensitivity is achieved by the double balanced topology. In Chapter 3.5.3 the behavior in case of mismatch is described for PATH1 and PATH2. Application of the output hybrid 1 brings an additional improvement of about 50 %. As an ideal termination is applied at the isolated port ($\Gamma_2 = 0$), the exact value for this additional improvement can be taken from Fig. 3.32, where subscripts $1 \leftrightarrow 2$ and $3 \leftrightarrow 4$ have to be exchanged. The y-axis of this figure then contains the reflection coefficients seen into PATH1 and PATH2 at the output. These reflection coefficients are equal to $-\Gamma_2\,\Gamma_3\,\Gamma_4$. As $|\Gamma_3| = |\Gamma_4| \leq 1$ and $\Gamma_2 = 0.2$, the product is ≤ 0.2. Hence, for that range the magnitudes of reflection coefficients seen into the coupled and isolated port of hybrid 1 ($\Gamma_{\mathrm{IN},3}$, $\Gamma_{\mathrm{IN},4}$ in Fig. 3.32) are about 50 % of all possible reflection coefficient magnitudes and phase conditions at the output port. Considering also the improvement due to hybrid 2 and 3 (Fig. 3.38), an overall improvement between $55\ldots100$ % can be obtained for the whole PA.

HPb mode

In this mode PATH1 works as a balanced amplifier, while in PATH2 only one amplifier is active. The ports of input and output hybrid 3 have to be terminated in a way that maximum power transfer is possible from the coupled/direct to the output/input port. Depending on whether HP3 or HP4 is active, the power gain $G_{\mathrm{P},13}$ or $G_{\mathrm{P},14}$ has to be maximized. In terms of generalized scattering parameters the gains are:

$$G_{\mathrm{P},14} = \frac{|S'_{14}|^2}{1-|S'_{44}|^2}, \quad G_{\mathrm{P},13} = \frac{|S'_{13}|^2}{1-|S'_{33}|^2} \tag{3.75}$$

Inserting (3.47) and after some simplification steps this can be written as

$$G_{\mathrm{P},14} = 2\,\frac{1-2\,|\Gamma_2\Gamma_3|\,\cos(\Delta\varphi_{23})+|\Gamma_2\Gamma_3|^2}{4-4\,|\Gamma_2\Gamma_3|\,\cos(\Delta\varphi_{23})+|\Gamma_2\Gamma_3|^2-|\Gamma_2|^2}, \tag{3.76}$$

$$G_{\mathrm{P},13} = 2\,\frac{1+2\,|\Gamma_2\Gamma_4|\,\cos(\Delta\varphi_{24})+|\Gamma_2\Gamma_4|^2}{4+4\,|\Gamma_2\Gamma_4|\,\cos(\Delta\varphi_{24})+|\Gamma_2\Gamma_4|^2-|\Gamma_2|^2}, \tag{3.77}$$

with $\Delta\varphi_{23} = \varphi_2 + \varphi_3$ and $\Delta\varphi_{24} = \varphi_2 + \varphi_4$.

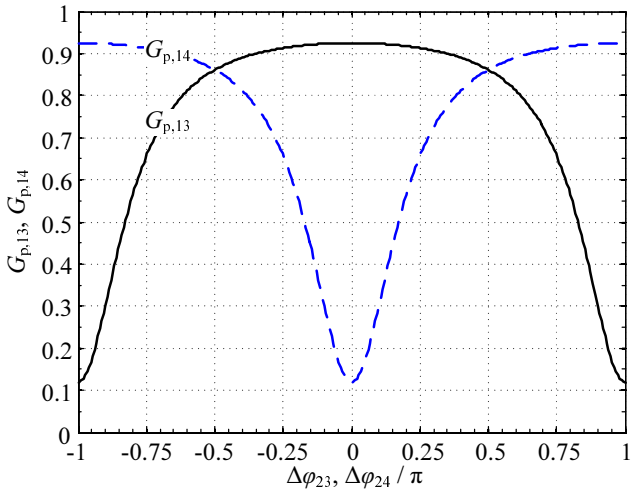


Figure 3.40: Power gain between direct/coupled and output port with loading conditions: $|\Gamma_2| = |\Gamma_3| = 0.9$. The maxima and minima along $\Delta\varphi_{23}$ and $\Delta\varphi_{24}$ are the same for other $|\Gamma_2|$ and $|\Gamma_3|$.

Inspecting $G_{\mathrm{P},14}$ exemplarily[6], the gain is 1, if all power is reflected at the isolated and coupled port. Therefore, $|\Gamma_2|$ and $|\Gamma_3|$ would have to be 1 with arbitrary phase. As these perfect load conditions cannot be achieved in a real circuit, $G_{\mathrm{P},14}$ is further analyzed for $|\Gamma_2| \leq 1$ and $|\Gamma_3| \leq 1$. Fig. 3.40 shows the dependence on $\Delta\varphi_{23}$. The maximum value is obtained at $\Delta\varphi_{23} = \pi$ and due to the small slope in this point the sensitivity to phase is not very high. This makes an implementation much easier because the phase difference between the reflection coefficient of the switched OFF amplifier and the switch SW2 has to be realized only with small accuracy ($\pm\pi/4$). Here it becomes clear, that the impedance of SW2 in OPEN state depends on the design of HP3. The switch does not necessarily have to realize an OPEN. Any parasitic capacitance, which rotates the reflection coefficient of SW2 in the smith chart can be even useful to achieve $\Delta\varphi_{23} = \pi$. Hence, the switch can potentially have high reactive parasitics in OPEN state, which usually reduces costs of this component. More details on the requirements for the switch in a real implementation will be given in Chapter 4.4.4.

The variation of Γ_2 and Γ_3 for achieving maximum $G_{\mathrm{P},14}$ also changes the reflection coefficient seen into the direct port $\Gamma_{\mathrm{IN},4}$. This is quantified with equation (3.51):

$$\Gamma_{\mathrm{IN},4}\Big|_{\Gamma_1=0} = \frac{\Gamma_2}{\Gamma_3\Gamma_2 - 2}, \tag{3.78}$$

where the condition $\Gamma_1 = 0$ results from the fact that the impedances at the OUT and ISO port of the output hybrid 1 are assumed to be equal in this operation case. The equation shows that the desired minimum load shift ($\Gamma_{\mathrm{IN},4} = 0$) can only be obtained with $\Gamma_2 = 0$. In this case half of the output power would be lost at the ISO

[6]The same results are valid for $G_{\mathrm{P},13}$, including an additional offset factor of π for $\Delta\varphi_{23}$.

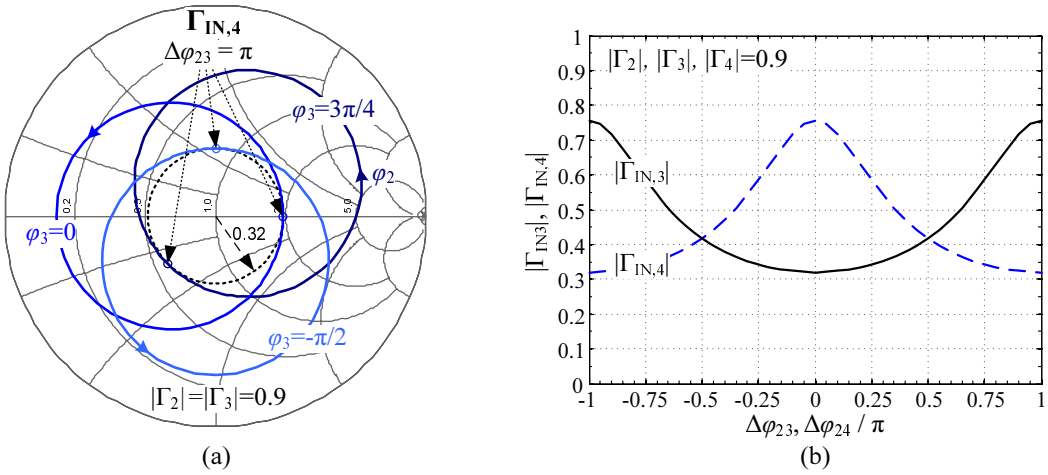


Figure 3.41: (a) Variation of reflection coefficient at direct port versus φ_2 and φ_3. (b) Reflection coefficient magnitudes versus φ_2.

port of hybrid 3: $G_{P,14} = -3$ dB. Consequently, it is assumed, that $|\Gamma_2| = |\Gamma_3| = 0.9$ in best case. Fig. 3.41 (a)[7] shows, that values inside a circle of radius 0.32 cannot be taken on, just by varying the phase. More general, the radius of this circle is

$$|\Gamma_{IN,4,min}| = \frac{|\Gamma_2|}{|\Gamma_2\,\Gamma_3| + 2}. \tag{3.79}$$

The relation for the magnitudes is illustrated in Fig. 3.41 (b), which for completeness also shows the variation of $|\Gamma_{IN,3}|$ in case of maximizing $G_{P,13}$. Hence, when optimizing for minimum attenuation of the hybrid by increasing $|\Gamma_2|$ (e.g. using switch SW2), the load of the active PA is pulled towards $|\Gamma_{IN,4}| \sim 1/3$. Therefore, the output matching network, designed for $\Gamma_{IN,4} \sim 0$, does not anymore transform to the optimum load for the amplifier. A compromise has to be found between these two opposing effects. Larger $|\Gamma_2|$ values decrease performance due to load pulling, while smaller $|\Gamma_2|$ values decrease performance due to power dissipation in the LP path. The weight of each effect depends on the active amplifier's load pull contours and the return loss at the output of the switched OFF amplifiers. The performance of the switch may therefore be further relaxed, allowing reduced return loss (larger resistive parasitics) in OPEN state. Similar to the investigations for $G_{P,14}$, the sensitivity to phase variations is very low (see Fig. 3.41(b)).

Further PAE improvement in HPb mode is achieved, when reducing performance in HPa mode. Fig. 3.42 shows exemplary load pull contours of an active HP amplifier. These contours do not change in HPa and HPb mode. For simultaneous maximum P_{out}/PAE, the reflection coefficient Γ_a has to be realized in HPa mode. Due to the load pulling effect in HPb mode this is shifted to Γ_b, which leads to reduced simultaneous P_{out}/PAE performance. The performance can be held constant by proper

[7] The result for the coupled port is the same, when $\Delta\varphi_{23} = 0$

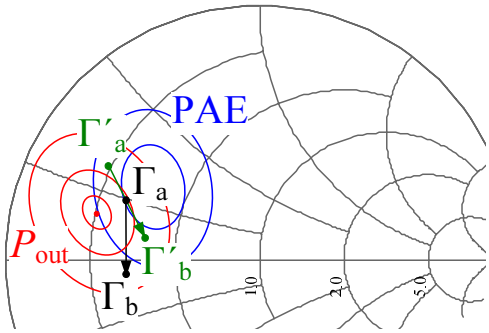


Figure 3.42: Exemplary constant P_{out} and PAE contours and load pull effect, when switching from HPa to HPb mode. Γ_{a}, Γ'_{a} ... realized load reflection coefficients in HPa mode. Γ_{b}, Γ'_{b} ... shifted load reflection coefficients in HPb mode.

design of the matching network and realization of Γ'_{a} in HPa mode. This will then result in Γ'_{b} in HPb mode and similar P_{out}/PAE values in both modes. HPb mode performance can therefore be improved at cost of HPa mode performance. Nevertheless, PAE in HPb mode will always be lower than in HPa mode, due to increased power dissipation in SW1 and in the switched OFF HP amplifier path.

The amount of load insensitivity in HPb mode is not equal for all amplifiers. For HP1 and HP2 any variation of antenna impedance is reduced by about 55 . . . 100 % (similar as in HPa mode). The active amplifier in PATH2 experiences 50 % reduction, due to output hybrid 1.

HPc mode

This mode of operation is similar to HPa mode, except that only the low power amplifier LP2 in PATH2 is active. For improvement of efficiency the total output power P_{out} has to be larger than $2\,P_{\text{HP}}$. Otherwise, LP2 increases current consumption without contributing to output power. Hence, there exists a threshold level for $P_{\text{HP}}/P_{\text{LP}}$ at which efficiency is reduced instead of increased. This is calculated using (A.3):

$$P_{\text{out}} = \frac{1}{2}(\sqrt{2\,P_{\text{HP}}} + \sqrt{P_{\text{LP}}})^2 > 2\,P_{\text{HP}} \quad \longrightarrow \quad P_{\text{LP}} > \underbrace{2(\sqrt{2}-1)^2}_{-4.6\text{ dB}}\,P_{\text{HP}}$$

If LP2 delivers 4.6 dB less power than HP1 or HP2, most power is dissipated in the termination resistor of output hybrid 1 due to the strong asymmetric power levels. As for a multi-mode approach the LP2 amplifier is designed for very large back-off power levels > 17 dB (WCDMA/EDGE), this mode of operation has no advantage over the others.

MPa mode

This operation mode is symmetric in that input/output hybrids 1 act as quadrature power divider/combiner and hybrids 2 and 3 as lossy bypass networks at the input and the output of the active amplifiers. The design of the output matching networks of the active high power amplifiers is similar to that described for PATH2 in HPb mode, including all compromises with performance in other modes. The complete amplifier has similar properties as a single balanced amplifier. Therefore, antenna load variations are reduced by at least 50 %, due to output hybrid 1 and due to increased dissipation in the hybrids, which act as bypass networks.

MPb mode

The output power range in this mode is approximately the same as in MPa mode. Power is delivered only from PATH1 (or PATH2), the other path is switched OFF and has to be bypassed. This is achieved by opening SW3 at output hybrid 1. As described for the HPb mode, if the output return loss of the deactivated path reduces, losses of the bypassing hybrid increase. If PATH2 is deactivated, the return loss is mainly determined by losses of HP3, HP4 and SW2 in OFF state. Assuming a reflection coefficient magnitude of 0.9 each, the overall reflection coefficient seen into PATH2 is $|\Gamma_2\,\Gamma_3\,\Gamma_4| = 0.729$. The attenuation through the output hybrid is therefore $G_{\mathrm{p},14} = 0.878$ (use (3.77) with $|\Gamma_2| = 0.9$, $|\Gamma_3| = 0.729$ and $\Delta\varphi_{23} = \pi$). Compared to MPa mode, where $G_{\mathrm{p},14} = 0.924$ (use (3.77) with $|\Gamma_2| = 0.9$, $|\Gamma_3| = 0.9$ and $\Delta\varphi_{23} = \pi$), the output power in MPb mode will always be lower. The difference becomes larger, if using semiconductor technology, which causes higher return loss in the switched OFF paths. In this case MPb mode has no benefit over the other modes and will therefore not be implemented. Regarding load insensitivity, a minimum of 50 % improvement is gained by the output hybrid 2 and by losses of PATH2.

MPc mode

As in HPc mode, the benefit in PAE improvement depends on the power level difference between the active high power and the active low power amplifier. The threshold in this mode is twice as large: $P_{\mathrm{LP}} > \underbrace{(\sqrt{2}-1)^2}_{-7.7\,\mathrm{dB}}\,P_{\mathrm{HP}}$. Related to the maximum output power of $4\,P_{\mathrm{HP}}$, the maximum back-off power level for the low power amplifier is 13.7 dB. For improvements in multi-mode operation this level is too high, why this mode of operation can be omitted.

MPd mode

Only one high power amplifier is active and all hybrids act as bypass networks. Assuming output reflection coefficient magnitudes of 0.9 at all deactivated amplifier paths, the total output power is reduced by at least $0.878 \cdot 0.924 = 0.8123$, due to

losses in the hybrids (see $G_{\mathrm{p},14}$ calculations for MPb mode). In difference to HPb mode, here the load pulling effect can be completely canceled, without changing the losses in the hybrids. To achieve this cancelation, the phase of the reflection coefficients seen into SW1, SW2 and SW3 (φ_{SW1}, φ_{SW2}, φ_{SW3}) have to be properly adjusted, related to the reflection coefficient phase seen into the switched OFF HP amplifiers (φ_{HP}). The phase conditions can be calculated in general with help of equation (3.51). As the derivation is long, only the results are shown in Table 3.7.

active amplifier:	HP1	HP2	HP3	HP4
$\varphi_{\mathrm{SW1}} + \varphi_{\mathrm{HP}}$	π	0	π	0
$\varphi_{\mathrm{SW2}} + \varphi_{\mathrm{HP}}$	0	π	0	π
$\varphi_{\mathrm{SW3}} + \varphi_{\mathrm{HP}}$	0	π	0	π

Table 3.7: Reflection coefficient phase seen into SW1, SW2 and SW3 for minimum attenuation and load pulling.

Since all hybrids act as bypass networks, the only mechanism for reduction of antenna impedance variations is attenuation of the reflected wave in the switches and the switched OFF HP amplifiers. This enables an improvement of 34 %, if the magnitudes of reflection coefficients seen into the passive paths are all 0.9, as already assumed. Hence, there is some load insensitivity in this operation mode but it is worse than in all other modes.

LPa mode

The switched OFF high power amplifiers in PATH1 and PATH2 are bypassed by the active low power amplifiers, and input/output hybrids 1 enable operation as a single balanced amplifier. The design of the matching networks is equal to the description in Chapter 3.5.2.

Load insensitivity is higher than in a classical balanced amplifier, because of the LP amplifiers lossy matching networks. Applying the results from Chapter 3.5.3, antenna impedance variations are reduced by about 68 . . . 74 %, at cost of PAE.

LPb mode

In this mode only one LP amplifier is active and all HP amplifiers are deactivated. Assuming, e.g. that LP2 is turned ON, the phases of SW1 and SW3 have to be adjusted for maximizing power transfer through hybrid 1 (see Table 3.7). Consequently the load to the active LP amplifier changes from its nominal value, similar as in all other operation modes where switches are opened to operate the hybrids as bypass networks. Assuming reflection coefficient magnitudes of 0.9, seen into the deactivated HP amplifiers and opened switches, the resulting load shift can be determined in 3 steps. First the reflection coefficient seen into hybrid 2 is calculated:

$|\Gamma_2\Gamma_3\Gamma_4| = 0.729$. This is used in the second step to determine Γ_1 (see Fig. 3.39) at the coupled port of hybrid 1. Using (3.79) gives $\frac{0.9}{0.9 \cdot 0.729+2} = 0.339$. In the last step the shift by hybrid 3 is calculated: $|\Gamma_{\mathrm{IN},2}| = |0.339\,\Gamma_3\Gamma_4| = 0.274$. Compared to HPb or MPb mode (worst case ~ 0.32) the load pull effect is slightly smaller in this operation mode. The overall power loss is the product of 0.775 attenuation by hybrid 1 (obtained using generalized scattering parameters, similar to (3.54), but $\Gamma_1 = 0.339$) and 0.878 attenuation by hybrid 3 (obtained using (3.77)). In total the output power is reduced by 0.681, which is more than in HPb or MPb mode. This shows again, that there is always a trade-off between the amount of load shift and attenuation, determined by the reflection coefficients seen into the deactivated amplifier ports. Hence, a compromise with performance in the other modes has to be found.

Antenna variations in LPb mode are reduced due to losses in the bypass hybrids (0.681) and the low power matching network ($0.6 \ldots 0.8$). An amount of about $46 \ldots 57$ % can be achieved.

In this section it was shown that from a conceptual point of view, a double balanced amplifier has potential to be used in multi-mode applications. The aim of improved efficiency can be achieved by optimization of operation modes, based on proper matching network design and, during operation, tuning of the quadrature hybrids isolated port impedances. In this regard optimization of the output hybrids is more important than that of the input hybrids. Imperfect matching at the input causes smaller gain of the overall amplifier, which has less effect on PAE than smaller drain efficiency. Therefore, switches at the input are omitted, but if necessary, can also be used when gain of the amplifier is a critical issue. Due to symmetry of the quadrature hybrid the switch optimization is then similar to that at the output. The advantage of omitting switches at the input is reduction of die size and components. Due to a small input power the input hybrids together with matching networks can be implemented on chip in the same technology as the PA paths. This enables the possibility of a more compact implementation where only output matching networks and output hybrids are off-chip.

Chapter 4

Multi-Mode PA Circuit Design

4.1 SiGe:C bipolar technology and modeling

This section gives an overview of the silicon germanium (SiGe) bipolar technology used for implementation of the multi-mode power amplifier. The technology belongs to Infineon's 0.35 μm bipolar process family B7HFx, where x denotes adaptations for specific applications. For each process option the transistors are optimized for an appropriate trade-off between the maximum transit frequency f_T and the collector emitter breakdown voltage V_{CE0}. For cellular communication transceivers an f_T = 75 GHz and V_{CE0} = 2.5 V [Klein 99] form a good compromise. For radar applications the f_T is increased to 200 GHz at cost of V_{CE0} = 1.7 V [Infineon 05]. For front end PA designs the f_T is reduced to improve V_{CE0} for better device ruggedness. This technology option is called B7HFP. The devices and key parameters are shown in Table 4.1. Among the different devices the following brief description emphasizes only on the active and passive devices which are used in the design.

0.35 μm bipolar SiGe:C technology				
Metal stack and contacting				
3-layer Aluminium metallization (2.8 μm/0.6 μm/0.4 μm)				
Sinker: low resistive contact to highly conductive substrate ($h \cdot \varrho = 4600\ \Omega\mu\text{m}^2$)				
Active devices	f_T/GHz	V_{CE0}/V	**Passive devices**	
NPN high power HBT	33	5.5	MIM capacitor	$C' = 1$ fF/μm^2
NPN low power HBT	35	5.0	MIS capacitor	$C' = 3$ fF/μm^2
VPNP low power	3.5	6.5	p,n Poly resistor	$R' = 65\ \Omega/\square, 1$ k$\Omega/\square$
ESD structures	1 kV(HBM[1])		Inductor	$Q \sim 6$ @ 2 GHz

[1] ESD hardness for Human Body Model ($R = 1.5$ kΩ, $C = 100$ pF)

Table 4.1: B7HFP technology overview.

4.1.1 High Power heterojunction bipolar transistor (HBT)

The key device of the PA is the high power transistor, which has to be optimized for high gain, small parasitics, high RF cut-off frequency and sufficient ruggedness. Devices with very good properties for PA applications are fabricated with III/V semiconductors, such as GaAs, which offer up to three times higher maximum carrier velocity than silicon and a higher breakdown voltage. The disadvantages are high costs of fabrication and incompatibility to standard CMOS technologies for highly integrated transceivers. Using silicon as basis brings an advantage in fabrication costs and even offers the possibility of combination with CMOS in a so called BiCMOS process. This way the good RF properties of bipolar devices are combined with the good down scaling options of CMOS.

Improvements of RF performance over standard Si bipolar junction transistors (BJT) were brought by invention of SiGe heterojunction bipolar transistors (HBT) [Kroemer 57]. The HBT is fabricated using heterostructures of compound semiconductors. This device has a narrow bandgap SiGe base, and both emitter and collector regions of wider bandgap Si. In an NPN HBT, the hole current flowing from base to emitter is suppressed by the potential barrier originating from the bandgap difference between the emitter and the base, which implies that a higher current injection efficiency is obtained compared with that of the homojunction bipolar transistor. This allows the base to be more heavily doped than the emitter, leading to a low base resistance and emitter-base capacitance. In contrast, a heavily doped base with the homojunction bipolar transistor degrades the emitter injection efficiency. These three features - high emitter injection efficiency, low base resistance and low emitter-base capacitance - are the essential points in obtaining high current gain at high frequencies. The speed of the transistor can also be increased by a reduced base region transit time, being improved by a reduced base width and graded Ge content across the base. Problems arise here, because after doping of the SiGe base region with boron (B), the subsequent high temperature annealing leads to diffusion of B into emitter and collector areas. This broadens the base region, and slows down the transistor. To counteract this effect, carbon (C) is added to the SiGe base, which significantly decreases the B diffusion coefficient over a wide temperature range [1].

Fig. 4.1 shows a cross section picture of the fabricated double-polysilicon self-aligned high power transistor. The term double-polysilicon means that both, emitter and base contact are realized with polysilicon. Self-aligned refers to the emitter-base isolation which is realized by thin dielectric layers, called 'spacers'. These are manufactured using vertical anisotropic etching of a SiO_2 layer, where at steep stages (such as the emitter opening) a very thin isolating spacer remains (see Fig. 4.1(a)). This allows the construction of structures smaller than the lithographic limits and therefore reduces costs of manufacturing the process masks [Reisch 03].

[1] This addition of carbon is indicated in the technology name by suffixing ':C' to 'SiGe'.

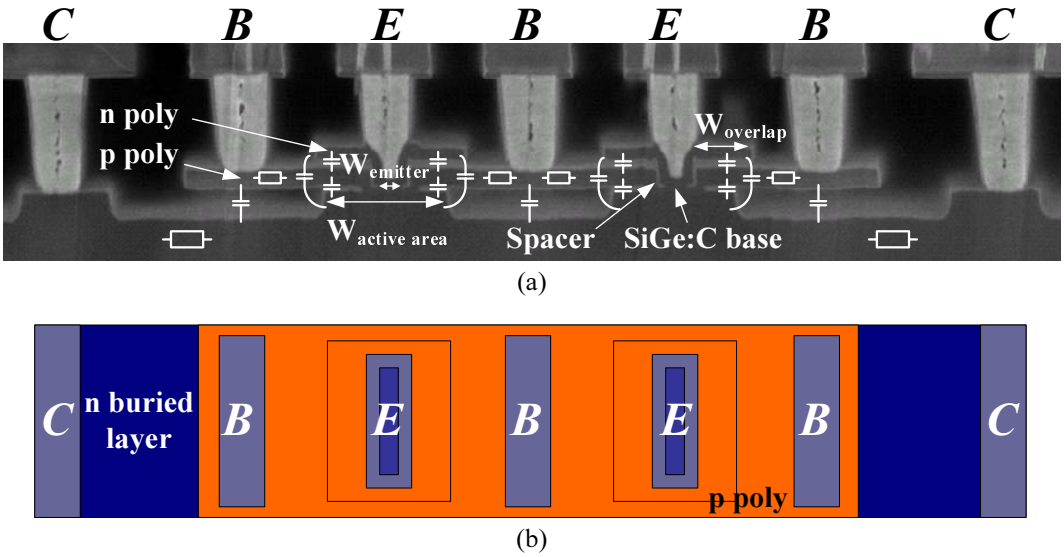


Figure 4.1: Double emitter HBT with self aligned base-emitter formation: (a) Cross section micrograph with indicated intrinsic resistive and capacitive parasitics. (b) Top view layout schematic of mask areas.

Transistor frequency characteristic

Important parameters for description of a transistor's frequency characteristic are the *transit frequency* f_{T} and the *maximum oscillation frequency* f_{max}. The transit frequency is the frequency where the *small signal forward current gain* of an output AC shorted transistor in emitter circuit is equal to unity. The small signal forward current gain is the gradient of the collector current versus the base current at constant collector emitter voltage:

$$\beta_{\mathrm{AC}} = \left.\frac{dI_{\mathrm{C}}}{dI_{\mathrm{B}}}\right|_{V_{\mathrm{CE}}=const.} \tag{4.1}$$

The frequency characteristic of $|\beta_{\mathrm{AC}}|$ can be approximated by a first order low pass with slope of -20 dB/decade in logarithmic scale (Fig. 4.2(a)). As β_{AC} depends on the DC bias conditions, the deduced value f_{T} also changes with the collector current I_{C}. There exists a maximum of f_{T}, when sweeping over V_{CB} and I_{C}. This global maximum is usually written into the data sheet as characteristic value for the frequency behavior. This value, specified for one elementary transistor with emitter area $A_{\mathrm{E,b}}$, can be used for a rough estimate of the total emitter area for the PA, consisting of many elementary transistors in parallel. For this the maximum current through the transistor is approximated by summation of the mean RF current[2] $\bar{I}_{\mathrm{RF}} = \sqrt{2}\,P_{\mathrm{RF}}/V_{\mathrm{DC}}$ and the DC bias current I_{DC}. The total emitter area for largest small signal gain at the desired output power can be calculated from

$$A_{\mathrm{E}} = (\sqrt{2}\,P_{\mathrm{RF}}/V_{\mathrm{DC}} + I_{\mathrm{DC}}) \cdot (A_{\mathrm{E,b}}/I_{\mathrm{C}}). \tag{4.2}$$

[2]The RF power P_{RF} is related to the output power at the antenna P_{out} with the collector efficiency η of the operation class, which includes losses of the matching network: $P_{\mathrm{RF}} = P_{\mathrm{out}}/\eta$

For fixed P_{RF} this simplified relation shows that for a linear PA, which needs higher I_{DC}, the active area has to be larger, compared to a nonlinear PA.

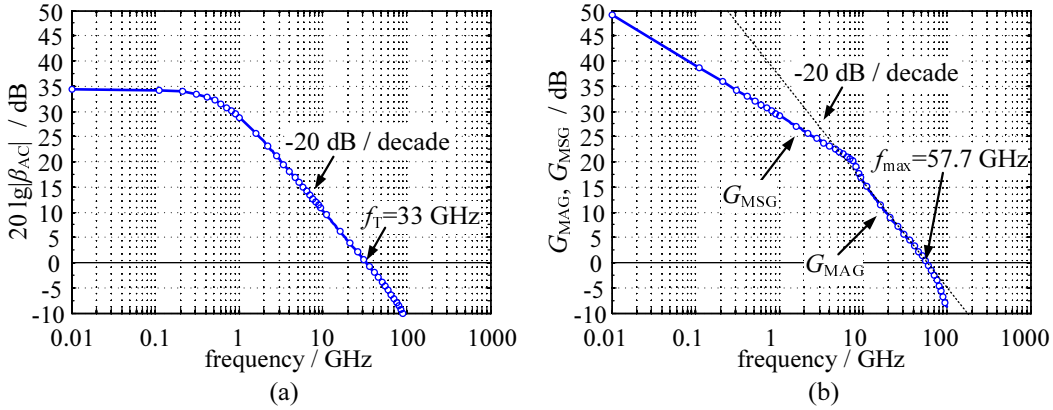


Figure 4.2: Model simulations of (a) small signal forward current gain, (b) maximum stable gain G_{MSG} and maximum available power gain G_{MAG} versus frequency ($V_{\mathrm{CB}} = 1$ V, $I_{\mathrm{C}} = 36$ mA, $A_{\mathrm{E}} = 2 \times 40\ \mu\mathrm{m} \times 1.55\ \mu\mathrm{m}$).

Another figure which describes the transistor's frequency behavior is based on the power gain. This is more important for the amplifier design, as it combines voltage and current gain. f_{max} is defined as the frequency where the power gain reduces to unity. Attention has to be paid on the exact definition of power gain. One possibility is to use the maximum available power gain G_{MAG}, where the input and output of the amplifier are assumed to be conjugately matched. A typical amplifier is not unconditional stable over all frequencies and the matched source and load impedances are not defined for a potentially unstable device. Consequently, G_{MAG} can not be calculated and is replaced by an approximation which is known as the maximum stable gain G_{MSG} [Medley 93], which presumes the device being critically stable. Fig. 4.2(b) illustrates these two regions above and below 9 GHz. The potentially unstable region is typically found at low frequencies, where the device has higher gain and therefore more likely fulfills the oscillating condition due to feedback. To avoid stability problems in measurements the classical definition of f_{max} is based on the unilateral power gain G_{U} [Zinke 87, Sze 81]. The transistor is made unilateral by adjusting an appropriate feedback network around the transistor. Independent of the definition a simple relation between f_{T} and f_{max} is given by [Reisch 03]:

$$f_{\mathrm{max}} \approx \sqrt{\frac{f_{\mathrm{T}}}{8\pi R_{\mathrm{B}} C_{\mathrm{BC}}}} \tag{4.3}$$

Hence, f_{max} is inverse proportional to the base collector capacitance C_{BC} and the base resistance R_{B}. One possibility to reduce R_{B} is the implementation of multi base fingers, as shown in Fig. 4.1. This reduces R_{B} at cost of an increased C_{BC}. A technology optimization via geometry therefore also includes a trade-off of these two parameters.

Breakdown behavior

The breakdown effects in a bipolar transistor can be divided into two categories. First the breakdown due to overvoltage (voltage breakdown) and second the breakdown due to thermal overload (second breakdown) [Tholl 76].

- **Voltage breakdown:** The minimum breakdown voltage between two terminals of the transistor is always obtained with the remaining terminal left open.
  - **Emitter-base breakdown:** The maximum E-B reverse voltage V_{EB0} is determined by the breakdown voltage of the E-B junction. The limiting mechanism in this context is the Zener effect occurring at the left and right sidewalls of the E-B junction. There the electric field strength has its maximum, but as long as the the maximum power dissipation of the material is not exceeded, the device is not destroyed. This voltage is typically in the range of a few volts.
  - **Collector-base breakdown:** Avalanche breakdown of the B-C junction defines the maximum reverse voltage that may be applied between collector and base V_{CB0}. As soon as the maximum allowed voltage is exceeded, the device is destroyed. It is not reversible as the Zener effect. Compared to V_{EB0} the values are relatively large (~ 10 V) and can be even increased with reduced collector doping.
  - **Collector-emitter breakdown:** For values of V_{CE} larger than the maximum breakdown voltage V_{CE0} the collector current is no more controlled by the base current. The responsible physical mechanisms behind are the punch-through effect or avalanche breakdown in the B-C junction[3]. The maximum allowed V_{CE} is influenced by the external circuitry. Worst case occurs for an open base ($R_{\mathrm{B}} = \infty$). If the base source resistance is reduced the safe operation collector-emitter voltage can be much larger than V_{CE0} ($R_{\mathrm{B}} \leq 10\ \mathrm{k}\Omega$ is already sufficient to see an effect).
- **Second breakdown:** This special thermal breakdown occurs only at large values of V_{CE}. It is characterized by an abrupt decrease in device voltage followed by a destructive increase of current. A description of the physical breakdown process was found in [Tholl 76]. There it is stated that similar to the E-B breakdown the electrical field strength is concentrated at the left and right sidewalls of the E-B junction. The current is therefore concentrated at these regions, where the substrate is heated up. This leads to increased electrical conductivity in this area and therefore further increase of current. This positive feedback process builds up until the temperature in these areas is large enough to fuse and destroy the crystal. These very small areas between

[3] **Punch-through effect**: Occurs for very thin base regions, where due to high C-E voltage the B-C space charge region reaches through the E-B space-charge layer. The reduced potential barrier causes a significant increase in current. **Avalanche effect**: Carrier multiplication due to impact ionization. For more detailed explanations see [Reisch 03].

emitter and base, where the current density increases more and more are called *hot spots*. The second breakdown can therefore destroy the device, although the maximum allowed power dissipation (collector emitter voltage × collector current) is not exceeded.

To improve device reliability some technological enhancements were implemented for the B7HFP high power transistor:

- **Lightly doped N buried layer (LDNBL):** The so called 'buried layer' (BL) is a heavily n^+-doped region below the active transistor area, which forms a highly conductive collector connection around the transistor (see Fig. 4.8). The doping is reduced towards the active area to increase the width of the space charge region between collector and base. This increases the collector-emitter breakdown voltage without changing the maximum allowed electrical field strength of the substrate. f_T reduces due to the increased width of the base region resulting in an increased carrier transit time. f_{max} changes marginally as C_{BC} reduces and thus the fraction f_T/C_{BC} in (4.3) may remain constant.
- **Base and emitter ballasting:** A large power transistor cell consists of multiple elementary cells in parallel. Power dissipation leads to increase of temperature, especially in the center of the cell. Due to the positive temperature/collector current feedback of the bipolar device more and more current will flow through the elementary cells in this center region until thermal breakdown occurs. To counteract this effect a negative feedback resistor in series to the base or emitter connection can be applied. For B7HFP such a stabilizing resistance was realized by increasing the n poly silicon series resistance for the two edgewise emitter contacts (see Fig. 4.8). This equalizes temperature distribution along the transistor cell at cost of an increased power dissipation, reducing efficiency of the transistor.
- **Finger geometry:** Close to device breakdown the current distribution along the cross section of the emitter is not equal. Instead of using one wide emitter, the emitter is made narrow and e.g. doubled.The same current is therefore distributed over a larger effective emitter area and hot spot formation is less likely to occur. In addition the number of base connections has to be increased as well, improving dissipation of heat. Nevertheless, multi emitter configurations lead to increase of parasitic capacitances which slow down the transistor.

Transistor model

The high power HBT transistor model used for simulations is composed of an *internal model*, describing the intrinsic transistor in the silicon substrate and an *external model*, subsuming extrinsic parameters of interconnections in the three metal layers above the substrate.

For the *internal model*, provided through the technology's simulation library, the so called HIgh CUrrent Model (HICUM) is used [Schröter 05]. It is a physical

based model, developed for an accurate description of a bipolar transistor's high speed large-signal characteristic. The large signal HICUM/Level2 equivalent circuit is shown in Fig. 4.3. It includes modeling of the high current operation region, including avalanche and tunneling currents, modeling of the external base collector region (parasitic PNP), modeling of emitter periphery injection and charge storage and modeling of self heating effects (thermal network). The illustrated equivalent circuit shows only the main parameters of the model, most of them described by further equations with additional parameters. The number of model parameters depends on the desired accuracy of the model and can easily exceed 90. For a more detailed description of the model equations and methods for parameter extraction from measurements it is referred to [Berkner 02]. Although the transistor model for large signal circuit analysis can be made very accurate, it shows long simulation time and severe convergence problems. Simulations, including statistics of process variations (monte carlo simulation) are therefore either very time consuming or not possible. As a consequence these uncertainties have to be compensated in a realization by tunability of the RF matching and DC biasing circuitry. Some possible methods will be discussed in Chapter 4.3 and Chapter 4.4.2.

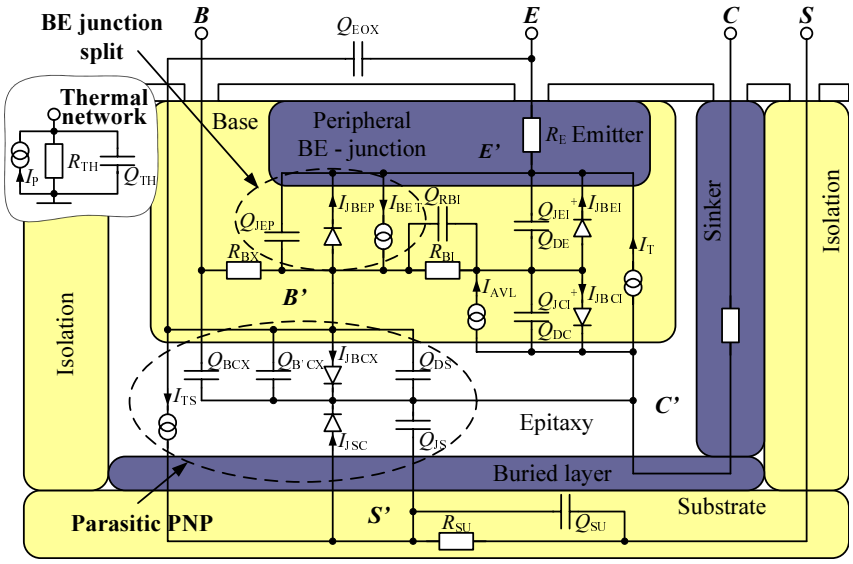


Figure 4.3: Intrinsic large signal transistor model: HICUM/Level2 [Berkner 02]

The large power capability of the output transistor is achieved by parallel circuit of many elementary cells. The interconnections of these elementary cells in the metal layers above the intrinsic transistor increase capacitive coupling as well as series resistance and inductance. These additional parasitic effects have severe impact on the transistor parameters, such as input impedance, output impedance, isolation or stability. Therefore they have to be included in the model. The resulting lumped element *external model* is illustrated in Fig. 4.4. Typically at low frequencies the influence of coupling capacitors (C_{BC}, C_{BE}, C_{CE}) and contact resistors (R_{B}, R_{C}, R_{E}) will dominate over the other parasitics. At high frequencies also losses particularly to

substrate have to be considered. These are modeled by the series circuit of capacitor and resistor at each contact of the transistor. This lumped element model provides

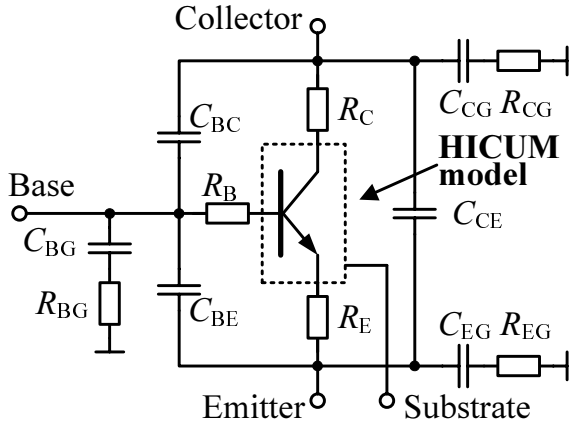


Figure 4.4: Transistor model including extrinsic interconnect parasitics.

an estimation for the main parasitics, but does not include distributed effects such as an asymmetric current distribution along the transistor blocks. These effects can only be modeled by partition of the whole transistor block into elementary extraction cells. Fig. 4.5 shows this partition at the example of the implemented HBT high power layout cell. The elementary extraction cell consists of two transistor cells. The interconnect metallization structure of this cell is simulated using the electromagnetic (EM) field solver Sonnet and the result is written into an S-parameter file. The distributed network can then be built by directly using the S-parameters or using the lumped element equivalent network of Fig. 4.4 after a parameter fit to the S-parameters. The smaller the elementary extraction cell, the faster the EM simulation. The disadvantage is that coupling between connection lines, which are further apart, is not included. For this work a good compromise between simulation time and coupling was found by using two transistor cells.

By knowledge of parasitic effects around the internal HBT transistor, simulations can be done to optimize the transistor towards high efficient operation. Following parameters can be improved by such sumulations:

- Small saturation voltage ($V_{CE,sat}$): This is achieved by reducing the collector and emitter resistance. One possibility is to place more transistor cells in parallel but at the same time capacitive coupling is increased.
- Small resistive losses (R_B, R_C, R_E): They can be reduced by using broad and short connection lines for interconnection.
- Small capacitive coupling (C_{BE}, C_{BC}, C_{CE}): The capacitive coupling directly influences speed and gain of the transistor. It can be reduced by placing less transistor cells in parallel and using thin and short connection lines. In this context the geometry of the elementary cell plays an important role. By using

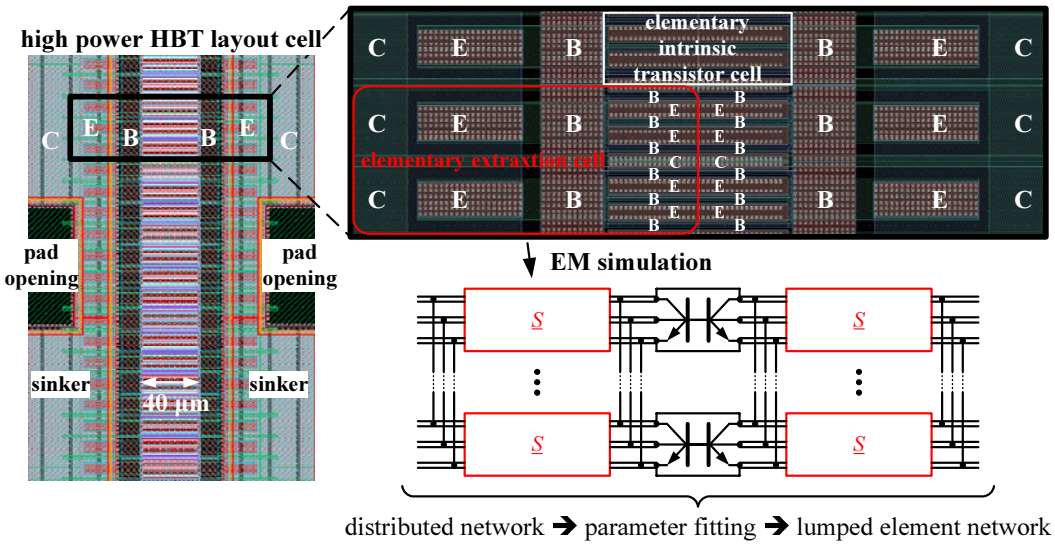


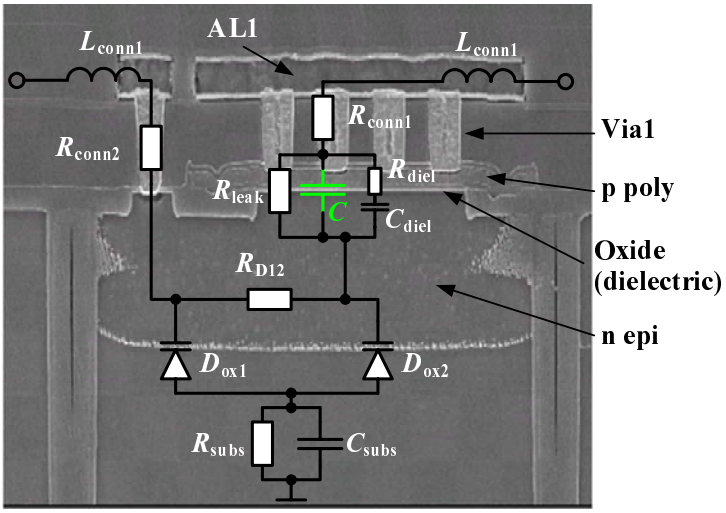
Figure 4.5: Parasitic extraction of the high power HBT transistor layout cell.

wide emitter stripes the current density per cell is larger, why the total transistor consists of less cells and therefore additional capacitive parasitics due to top wiring are reduced. On the other hand the breakdown voltage decreases and hot spot formation is more likely to occur.

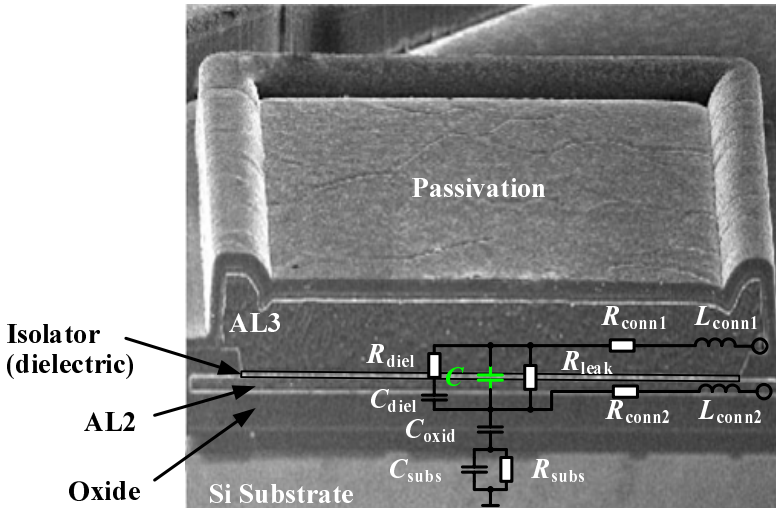
To find a balance between all effects is a challenging task during development of the technology. For B7HFP the elementary cell dimensions (emitter width, length) are mainly optimized for increased reliability, while the top layout interconnections are optimized for symmetry of the large cells and small R and C values to improve efficiency.

4.1.2 Capacitors

An on chip capacitor can in general be created at positions, where a dielectric material lies between two conducting materials. This constellation is possible between two aluminium plates in the metal stack, forming a so called MIM (Metal Isolator Metal) capacitor, or between aluminium and silicon substrate, forming a so called MIS (Metal Isolator Substrate) capacitor. The capacitance per area, quality factor ($Q = \frac{\mathrm{Im}(Z_C)}{\mathrm{Re}(Z_C)}$) and maximum stress voltage are benchmarks to compare both types of capacitors. The maximum stress voltage depends on the applied voltage form (AC or DC), capacitor area and temperature conditions. For B7HFP the values are related to a worst case scenario with a DC voltage applied to a capacitor area of 0.5 mm^2 at 125 °C junction temperature.



(a)



(b)

Figure 4.6: Cross section micrograph of B7HFP (a) MIS-capacitor and (b) MIM-capacitor, including lumped element model.

Metal isolator substrate (MIS) capacitor

Fig. 4.6 (a) shows the cross section of a MIS capacitor. The top plate is formed by a heavily doped p polysilicon layer and the bottom plate by an n type silicon (n epi). Between these electrodes the dielectric layer is made of silicon oxide/nitride. As this structure is very similar to a MOS transistor, the capacitor value is voltage dependent, potentially producing non linearity in a circuit. The quality factor is low ($Q_{\mathrm{MIS}} \sim 7 \ldots 50$) because of the large connection resistance through vias and poly ($R_{\mathrm{conn1,2}}$) and due to the large coupling to substrate (D1 and D2) and the losses encountered there (R_{subs}, C_{subs}). Due to the relatively large dielectric constant of the

oxide/nitride isolator ($\varepsilon_r \sim 3.9/7.5$) and the small thickness of the dielectric layer the capacitance per unit area is high (3 fF/μm^2), which allows compact capacitor dimensions. The maximum allowed stress voltage is about 5.5 V_{DC} DC. Due to the small quality factor the application of this capacitor type is constrained to DC and low frequency applications such as decoupling (blocking) of bias feeding lines of the PA circuit.

Metal isolator metal (MIM) capacitor

A cross section of a MIM capacitor is shown in Fig. 4.6 (b). It is fabricated between the two upper aluminium layers (AL2 and AL3). For an increase of capacitance per unit area the distance between these layers is reduced in an additional process step. At the location of the capacitor the isolating silicon oxide is etched away and replaced by a thin dielectric layer of silicon nitride. For processing reasons the thickness of this dielectric layer is larger than for the MIS capacitor, why the capacitance per unit area is 1 fF/μm^2. The advantage of the MIM capacitor is a larger stress voltage (9 V_{DC}) and a higher quality factor ($Q_{MIM} > 100$). The increase in quality factor has its reason in the smaller coupling capacitance (larger distance) to substrate and the larger conductivity of the metal plates. This type of capacitor is mainly used for input and interstage impedance matching. For DC blocking the MIM capacitor is placed above the MIS capacitor to increase the capacitance value per area.

4.1.3 Inductors

The matching networks and bias feeds of the integrated PA circuits require the integration of inductive elements on the chip. Amongst many possibilities, monolithic integrated printed inductors and bond wire inductors are used in this work. The main design target is the inductance value L_s, but there are three more figures of merit which critically affect the PA design. They are the quality factor Q, the self-resonant frequency f_s and the maximum current I_{max}. The Q factor relates to loss mechanisms in the inductor, f_s defines the inductor's maximum operation frequency range and I_{max} sets a limit on the current carrying capacity, e.g. when used as bias feeding choke.

The *monolithic integrated printed inductor* has at best a spiral form, because this exhibits less metal resistance for a given value of inductance and metal wire width. As B7HFP design rules do not permit circular geometries, the integrated inductors are approximated to this optimum by rectangular or octagonal shapes. The inductors use the top metal layer for the main part of the windings and the connection to the center is made by crossing under the top metal using AL1 and AL2. Fig. 4.7 shows the principal geometry of a rectangular inductor above the silicon substrate.

Aided by a three dimensional EM field solver, the inductor can be characterized by calculating its scattering matrix and using it for the circuit simulation process. A full EM simulation is usually unpractical for optimization of the inductor, since

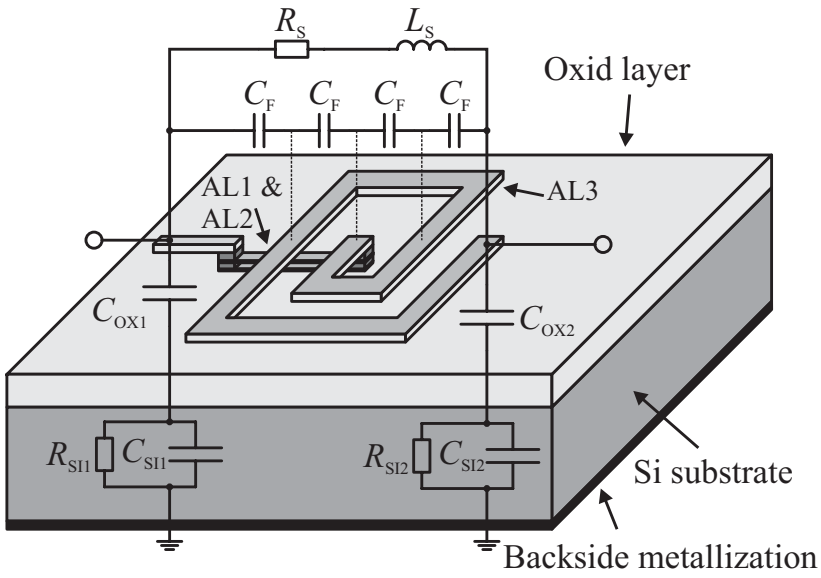


Figure 4.7: Schematic three dimensional view of a rectangular shaped printed inductor and lumped element model.

it is computationally intensive and very time consuming (up to several hours simulation time). A high level of accuracy is only obtained with small meshing size, resulting in an increased number of numerical matrix manipulations and therefore long calculation time.

A more practical oriented approach is to use a physical based model of the inductor with geometry dependent values. The geometry can then be rapidly optimized until the desired L_s, Q, f_s and I_{max} are obtained. The final iteration is a complete EM field simulation for fine tuning of the geometry. The time spent for simulation depends on the accuracy of the geometry dependent model. Simple techniques for analysis include the Greenhouse method [Greenhouse 74], the Wheeler formula [Wheeler 28] or a data-fitted empirical expression [Mohan 99]. Methods for more accurate modeling and optimization of monolithic integrated inductors can be found in [Niknejad 00]. The simplest lumped element model is shown in Fig. 4.7. It gives only a crude estimate of the performance but can be used to explain the principle physical mechanisms, which affect the inductor's characteristic.

The main parameter is the inductance value L_s. It is proportional to the diameter of the inductor D and to the square of the number of turns n ($L_s \propto n^2\, D$).

The fringing capacitance C_F describes the coupling between the turns and between the turns and the underpass. It has an effect on f_s and can be reduced by larger winding to winding distance. The overall capacitance value across the terminals of the inductor is usually negligible as it is a series connection of all C_F, as long as the underpass is kept small in width.

Two main loss mechanisms are included in the model of Fig. 4.7. The first loss mechanism is found in the series resistance R_S. It is directly proportional to the

metal resistance, which in turn depends on metal cross section area D and frequency (Skin effect[4] and current crowding or proximity effect[5]). In this concern the top metal layer is used primarily because of the large cross section area and therefore low sheet resistance. For a given value of L_{S}, the series resistance can be held low by increasing n, instead of D, since L_{S} increases quadratically and R_{S} linearly with the number of turns.

The second loss mechanism is found in energy coupled into the low resistive substrate ($\rho = 1.5 \cdot 10^{-5}$ Ωm). The amount of loss is modeled by the substrate resistance R_{SI}. The mechanisms behind can be divided into capacitive and magnetic coupling. Capacitive coupling is modeled by C_{OX} and C_{SI} which are directly proportional to the inductor's windings area and indirect proportional to their distance from the substrate. The inductive coupling causes image currents induced in the substrate beneath the inductor, counteracting the primary magnetic field, thus lowering L_{s}. Both effects have influence on Q and f_{s} and can be reduced by implementing the inductor's windings in the top metal layer, farther away from the substrate.

For a fully monolithic integrated inductor the maximum current I_{max} is determined by the maximum current density of the metal layers together with the effective width of the windings. Here frequency dependent current crowding effects have to be taken into account at the frequency of operation. For B7HFP the maximum current densities per metal width in AL1 and AL2 are about 2.5 times smaller than for AL3. This shows that the limiting factor for I_{max} is the under crossing in the lower metal layers. A trade-off between low parasitic effects and high power capability is also a concern of optimization for matching and especially bias feeding inductors.

Inductors integrated in B7HFP technology ($L_{\mathrm{s}} \sim 2$ nH) have following typical figures of merit: $Q < 5$ at $1 \ldots 2$ GHz; $f_{\mathrm{s}} \sim 9$ GHz; $I_{\mathrm{max}} < 200$ mA.

The small quality factor of printed inductors is the main reason for an alternative realization of the desired inductance with help of *bond wires*. The used bond wires are made of aluminium with a diameter of 25 μm. Compared to printed inductors they have larger surface area per length (less resistive loss) and can be placed well above conductive planes (reduced capacitive and magnetic coupling) improving the overall $Q > 100$ and f_{s}. The disadvantage is a limited range of the desired inductance L_s. As a first approximation the inductance can be calculated by [Lee 98]

$$L_s = 2 \cdot 10^{-7}\, l \left(\ln\left(\frac{2l}{r}\right) - 0.75 \right) \sim 1 \text{ nH/mm} \tag{4.4}$$

An inductance of 2 nH would therefore require a bond wire of 2 mm length, usually much longer than the overall chip dimension. Therefore, only inductors in the range

[4]The current density in a conductor with finite resistivity ρ and magnetic permeability μ decreases exponentially with depth d from the surface ($J \propto e^{-d/\delta}$, with the skin depth $\delta = \sqrt{\frac{2\rho}{2\pi f \mu}}$). Hence with higher frequency f the current is crowded to the edge of the conductor, increasing the series resistance.

[5]The magnetic field generated by nearby lines changes the current distribution and results in an increased current density at the edges of the line. The available cross section area for conduction is reduced and so the series resistance.

of 0.5 . . . 1.5 nH can be realized with bond wires. The necessary bonding pads introduce some coupling to the substrate in the order of several hundred fF, thus being usually negligible. The implementation of bond wires gives an additional opportunity of tuning the matching circuits. The length and shape of the bond wire can be adjusted to optimize for example the interstage matching between two amplifiers and therefore compensate for process variations during fabrication (see Chapter 4.4.2). This is not possible using a printed inductor. However, bond wires have much higher tolerances than printed inductors.

4.1.4 Sinker

Mounting the chip onto a printed circuit board (PCB) requires a low resistive and low inductive connection of the chip's and the PCB's reference ground. One possibility is to use many short bond wires at the edge of the chip. The additional bond pads and the required low resistive on-chip routing of the ground connections with broad lines results in an increased chip area.

The B7HFP technology offers a so called sinker which basically is a low impedance connection between the silicon surface and the p substrate (see Fig. 4.8). Together with the low resistivity of the substrate ($\rho = 1.5 \cdot 10^{-5}$ Ωm), a reduced wafer height (< 100 μm) and metallization of the chip backside, a low resistive path through the chip is obtained. A sinker resistance of 0.25 Ω at each emitter of the implemented power cells (see Fig. 4.27) leads to a voltage drop of 0.1 V, thus reducing drain efficiency by only 2.8 %. The advantage compared to bond wires is that the sinker can be placed directly at the emitter of the power transistor, leading to short on chip ground connection lines and thus negligible inductance in the operating frequency range (< 100 pF).

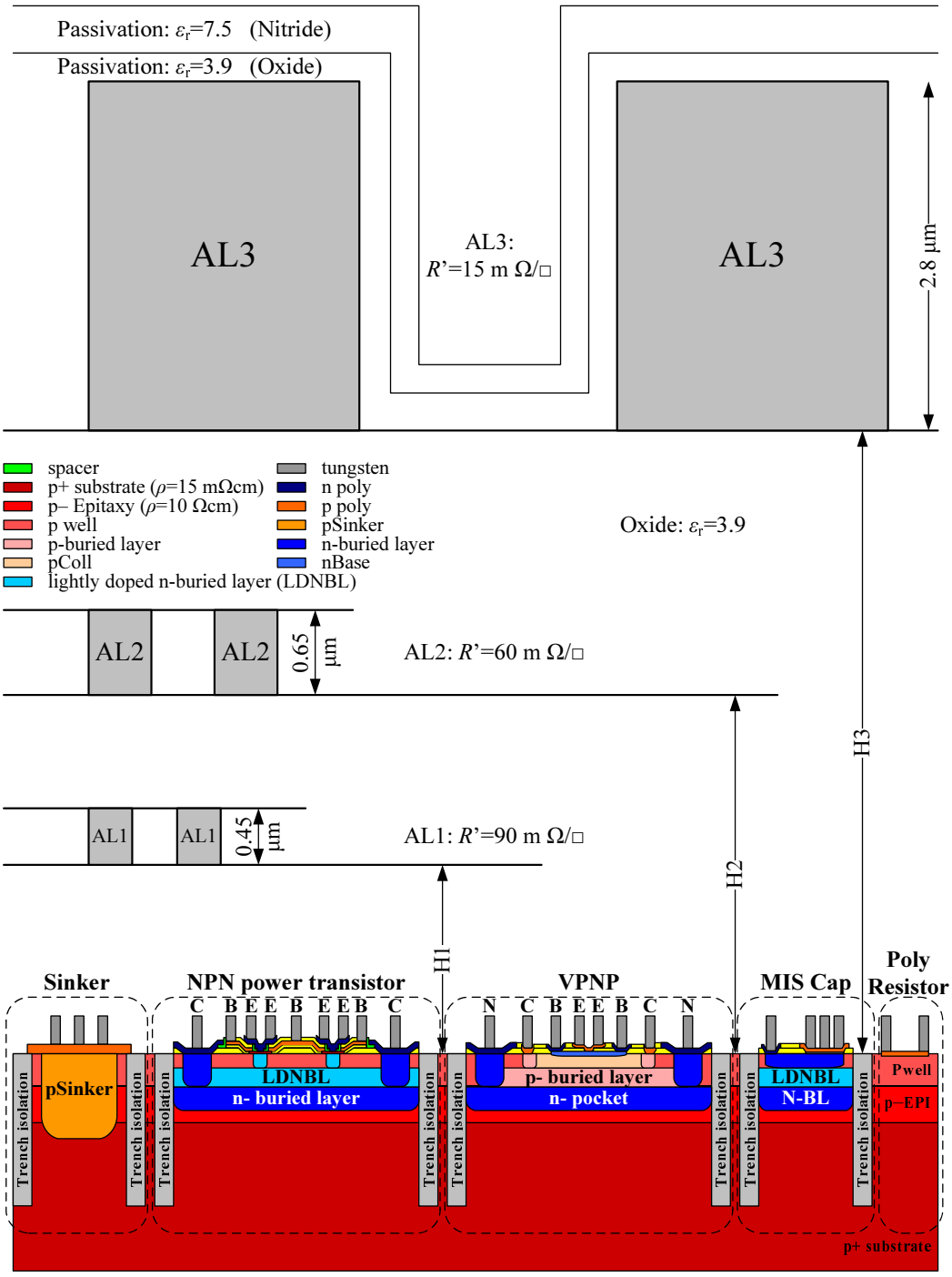


Figure 4.8: Schematic cross section of B7HFP technology and aluminium metal layer stack

4.2 Integration of quadrature Hybrids

The quadrature hybrid so far was characterized by its idealized 4-port scattering matrix, without further assumptions on the underlying network. In this section possibilities for realization of the quadrature hybrids are discussed and weighted for their characteristics versus bandwidth. After this analysis implementation examples in B7HFP technology are presented. The simulation and measurement results for one and two stage inductively coupled hybrids will be compared to choose an adequate solution for the selected GSM/EDGE/WCDMA operation band.

The starting point for derivation of realizable quadrature hybrid networks is the general scattering matrix $\underline{S}$ of an ideal hybrid. This is written as:

$$\underline{S} = \frac{e^{j\varphi}}{\sqrt{2}} \begin{pmatrix} 0 & 0 & -j & 1 \\ 0 & 0 & 1 & -j \\ -j & 1 & 0 & 0 \\ 1 & -j & 0 & 0 \end{pmatrix}, \tag{4.5}$$

with φ being an arbitrary but common phase shift at every port of the hybrid. A lossless finite element network can in theory fulfill this scattering matrix at one frequency point (center frequency f_0) but will deviate above and below this frequency from the ideal characteristic [Andrews 06]. The bandwidth Δf of the coupler can therefore be defined as the frequency range in which the hybrid's gain error ($\varepsilon = \frac{|S_{31}|}{|S_{41}|}$) and phase error ($\beta = |\varphi_{31} - \varphi_{41}| - 90°$) are within a certain limit. Hence, over frequency any hybrid network can only approximate $\underline{S}$, except at f_0, and the bandwidth of the approximation depends on the sensitivity of the application to these errors.

In Chapter 3.4.2 it was shown that in a balanced amplifier a certain amount of gain and phase errors are tolerated. This imbalance of the input hybrid as well as of the output hybrid reduce the overall gain of the amplifier (see equation (3.46)). The amplitude errors sum up while the influence of the phase errors may cancel out, if they have opposite sign. A value for the maximum tolerable errors of each hybrid can therefore only be given by defining a maximum gain reduction of the overall amplifier (including input and output hybrid) within the operation band. For implementation of the multi-mode PA this gain flatness is chosen to be < 0.5 dB. Besides the gain and phase error, also the overall loss ($ATT = \frac{1}{2}\left(|S_{31}| + |S_{41}|\right)^2$)[6] of the hybrid has influence on the performance. Losses at the output of the amplifier have stronger influence on PAE than losses at the input (see Fig. 3.14). Hence, at the output an external low loss ceramic hybrid from Murata is used. This component together with its frequency characteristic is shown in Fig. 4.9. The errors within the operation band 824 ... 915 MHz (low band) are $ATT > -0.3$ dB, $|\varepsilon| < 0.1$ dB and $|\beta| < 0.5°$. With these values, the maximum tolerable errors of the input hybrids, implemented on chip, can be calculated from equation (3.46). The results are shown in Table 4.2, which is the basis for further analysis.

[6] Defined here as: $ATT = \frac{P_{\text{out at Port 1}}}{P_{\text{source at Port 3}} + P_{\text{source at Port 4}}} = \frac{1}{2}\left|j \cdot S_{31} + S_{41}\right|^2 \approx \frac{1}{2}\left(|S_{31}| + |S_{41}|\right)^2$, if $|\beta| < 20°$ and $|\varepsilon| < 10$ dB.

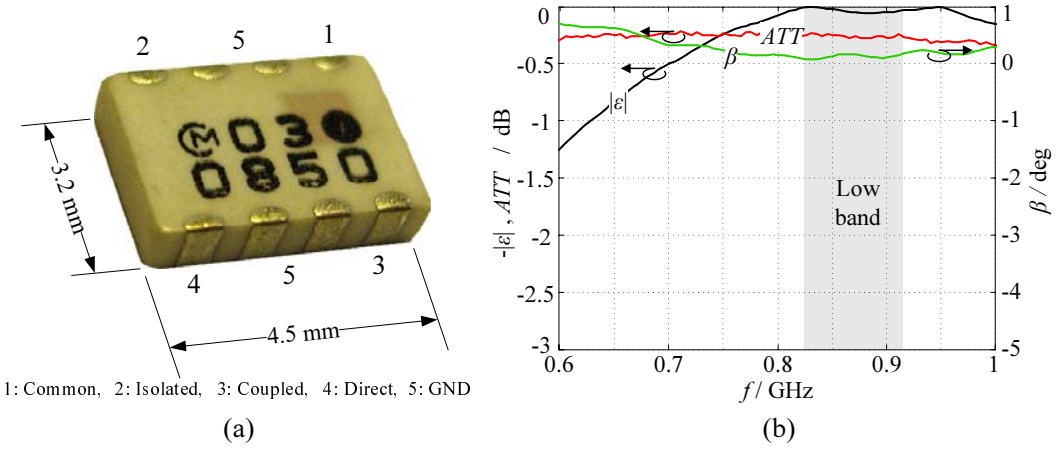


Figure 4.9: External output ceramic hybrid for low band [Murata 08]. (a) Photograph, (b) frequency characteristic

frequency band GHz	relative bandwidth $\frac{f_{max}-f_{min}}{f_0}$	$\lvert\varepsilon\rvert$ dB	$\lvert\beta\rvert$ deg
0.824 . . . 0.915	10 %	< 0.75	< 20

Table 4.2: Maximum tolerable errors for the input quadrature hybrids for an overall gain flatness of 0.5 dB.

By knowledge of the imbalance limits a quadrature hybrid circuit has to be found, fulfilling these constraints. Possible implementations of quadrature hybrids are divided into two main groups: *distributed* and *lumped element* hybrids. Classical examples for distributed hybrids are the rat-race ring hybrid, branch-line coupler and Lange coupler, all being described in [Pozar 93]. For microwave integrated circuits (MMIC) up to 2 GHz the overall size of the distributed hybrids with quarterwave transmission lines is too large. Therefore it is attractive to replace the transmission lines with a combination of inductors, capacitors and transformers to obtain a lumped element hybrid.

To find appropriate lumped element topologies, the admittance matrix $\underline{Y} = \left(\frac{1}{Z_0}\underline{U} + \underline{S}\right)^{-1} \cdot \left(\frac{1}{Z_0}\underline{U} - \underline{S}\right)$ of the idealized hybrid from equation (4.5) is considered. This is given by

$$\underline{Y} = \begin{pmatrix} Y_{11} & Y_{12} & Y_{13} & Y_{14} \\ Y_{12} & Y_{11} & Y_{14} & Y_{13} \\ Y_{13} & Y_{14} & Y_{11} & Y_{12} \\ Y_{14} & Y_{13} & Y_{12} & Y_{11} \end{pmatrix}, \text{ with } \begin{cases} Y_{11} &= \frac{-j\,\tan(2\,\varphi)}{Z_0} \\ Y_{12} &= \frac{-j}{Z_0\,\cos(2\,\varphi)} \\ Y_{13} &= \frac{j\sqrt{2}\cos(\varphi)}{Z_0\,\cos(2\,\varphi)} \\ Y_{14} &= \frac{j\sqrt{2}\sin(\varphi)}{Z_0\,\cos(2\,\varphi)} \end{cases} \tag{4.6}$$

and Z_0 being the common reference impedance at every port. This matrix has two axes of symmetry along Y_{11} and Y_{14}. Consequently also the physical network can

have two axes of symmetry. This leads to the topologies shown in Fig. 4.10, where the symmetry axes are marked with dotted lines. These circuits will be analyzed in the following.

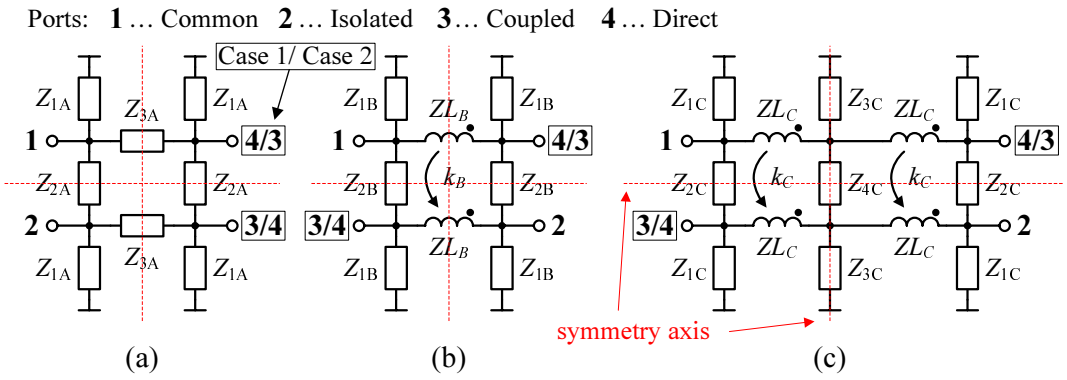


Figure 4.10: Physical implementation of quadrature hybrids: (a) uncoupled lumped element hybrid (b) inductively coupled lumped element hybrid (c) two section inductively coupled lumped element hybrid.

4.2.1 Uncoupled lumped element hybrid

The lumped element hybrid of Fig. 4.10 (a) consists of three different impedances $Z_{i\mathrm{A}}$. As a lossless hybrid is assumed for the current analysis, the impedances are purely imaginary, representing inductors L and capacitors C. The admittance matrix is calculated with linear network theory:

$$\underline{Y}_{\mathrm{A}} = \begin{pmatrix} Y_{11\mathrm{A}} & Y_{12\mathrm{A}} & Y_{13\mathrm{A}} & Y_{14\mathrm{A}} \\ Y_{12\mathrm{A}} & Y_{11\mathrm{A}} & Y_{14\mathrm{A}} & Y_{13\mathrm{A}} \\ Y_{13\mathrm{A}} & Y_{14\mathrm{A}} & Y_{11\mathrm{A}} & Y_{12\mathrm{A}} \\ Y_{14\mathrm{A}} & Y_{13\mathrm{A}} & Y_{12\mathrm{A}} & Y_{11\mathrm{A}} \end{pmatrix}, \text{ with } \begin{cases} Y_{11\mathrm{A}} = \left(\frac{1}{Z_{1\mathrm{A}}} + \frac{1}{Z_{2\mathrm{A}}} + \frac{1}{Z_{3\mathrm{A}}}\right) \\ Y_{12\mathrm{A}} = \left(-\frac{1}{Z_{2\mathrm{A}}}\right) \\ Y_{13\mathrm{A}} = 0 \text{ or } \left(-\frac{1}{Z_{3\mathrm{A}}}\right) \\ Y_{14\mathrm{A}} = \left(-\frac{1}{Z_{3\mathrm{A}}}\right) \text{ or } 0 \end{cases} \tag{4.7}$$

The second solution for $Y_{13\mathrm{A}}$ and $Y_{14\mathrm{A}}$ corresponds to the case where the direct port 3 and coupled port 4 are swapped ($3 \rightarrow 4$, $4 \rightarrow 3$). The values for $Z_{1\mathrm{A}}$, $Z_{2\mathrm{A}}$, $Z_{3\mathrm{A}}$ and φ are calculated by equating the corresponding Y-parameters in (4.6) and (4.7). The results are shown in Table 4.3.There are four different possibilities of arranging L ($\mathrm{Im}(Z_{\mathrm{iA}}) > 0$) and C ($\mathrm{Im}(Z_{\mathrm{iA}}) < 0$). Inductors consume a lot of chip area and have poorer quality factor than capacitors. Therefore only networks with a minimum number of inductors are relevant for on chip integration. Among the solutions just Case 2 with $\mathrm{Im}(Z_{3\mathrm{A}}) > 0$ gives a network with two L and four C elements. The element values are: $C_{1A} = \frac{1}{\omega_0 Z_0}$, $C_{2A} = \frac{1}{\omega_0 Z_0 (1+\sqrt{2})}$, $L_{3A} = \frac{Z_0}{\omega_0 \sqrt{2}}$, where $\omega_0 = 2\pi f_0$

Variable	Case 1	Case 2
φ	$\frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$	$0, \pi, 2\pi, \ldots$
Z_{1A}	$jZ_0(1 \pm \sqrt{2})$	$-jZ_0(1 \pm \sqrt{2})$
Z_{2A}	jZ_0	$-jZ_0$
Z_{3A}	$\mp jZ_0 \frac{1}{\sqrt{2}}$	$\pm jZ_0 \frac{1}{\sqrt{2}}$

Table 4.3: Component values for the lumped element hybrid network.

defines the center frequency of the hybrid. These equations indicate that there is no degree of freedom to vary L and C values (and therefore the size of the hybrid) for a given center frequency and reference impedance. The amplitude and phase imbalance over normalized frequency is shown in Fig. 4.11. The relative bandwidth is ruled by the amplitude imbalance and is about 4.34 % for low band operation. This value is too small for the current application, why this type of hybrid is not further investigated.

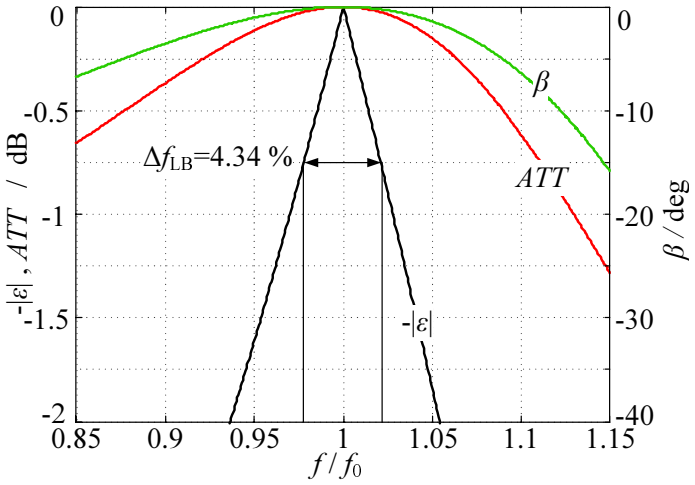


Figure 4.11: Frequency characteristic of the lumped element hybrid for determination of the relative bandwidth.

One effect, visible in Fig. 4.11 should be mentioned here, as it is common for all types of hybrids. Only in the vicinity of f_0 the hybrid has no loss, $ATT = 0$ dB. Although the elements of the hybrid are lossless, the ATT takes on values < 0 dB versus frequency, which may be interpreted as an increase of insertion loss. The explanation is that ATT includes network properties as mismatch (S_{11}, S_{22}, S_{33}, S_{44}) and coupling to the isolated port (S_{21}, S_{23}, S_{24})[7]. With increasing distance from the center frequency these values differ more and more from the ideal ones. On the one hand this leads to increased mismatch and therefore reduction of delivered power to the load. On the other hand the increased unbalance of coupling to the isolated port

[7]In a realizable passive and reciprocal network S_{31} and S_{41} are not independent of the other scattering parameters

leads to power dissipation in the isolation resistor, thus also lowering the delivered power. Therefore, the frequency characteristic of ATT is a further figure of merit describing out of band mismatch and isolation errors of the hybrid. Nevertheless, the influence in the desired frequency bands is usually negligible, compared to the amplitude imbalance.

4.2.2 Inductively coupled lumped element hybrid

A lumped element coupler can also include *capacitive* or *inductive* coupling. Capacitive coupling is obtained by swapping ports 1 and $\boxed{4/3}$ in Fig. 4.10 (a). The bandwidth properties are about the same as for the uncoupled lumped element hybrid. The phase error is increased in the range of a few percent, but the overall bandwidth is still ruled by the amplitude imbalance, being as large as for the uncoupled lumped element hybrid. Substantial improvement of bandwidth is obtained by inductively coupled lumped element hybrids. This topology makes use of a 1:1 transformer with coupling coefficient k and tuning impedances $Z_{i\mathrm{B}}$ in a configuration as shown in Fig. 4.10 (b). The derivation of the component values is similar to that of the uncoupled lumped element hybrid. First the admittance matrix is calculated with linear network theory:

$$\underline{Y} = \begin{pmatrix} Y_{11\mathrm{B}} & Y_{12\mathrm{B}} & Y_{13\mathrm{B}} & Y_{14\mathrm{B}} \\ Y_{12\mathrm{B}} & Y_{11\mathrm{B}} & Y_{14\mathrm{B}} & Y_{13\mathrm{B}} \\ Y_{13\mathrm{B}} & Y_{14\mathrm{B}} & Y_{11\mathrm{B}} & Y_{12\mathrm{B}} \\ Y_{14\mathrm{B}} & Y_{13\mathrm{B}} & Y_{12\mathrm{B}} & Y_{11\mathrm{B}} \end{pmatrix}, \text{ with } \begin{cases} Y_{11\mathrm{B}} = \left(\frac{1}{Z_{1\mathrm{B}}} + \frac{1}{Z_{2\mathrm{B}}} + \frac{1}{ZL_{\mathrm{B}}(1-k^2)} \right) \\ Y_{12\mathrm{B}} = \left(\frac{k}{ZL_{\mathrm{B}}(1-k^2)} \right) \\ Y_{13\mathrm{B}} = \left(-\frac{1}{Z_{2\mathrm{B}}} - \frac{k}{ZL_{\mathrm{B}}(1-k^2)} \right) \text{ or } \left(-\frac{1}{ZL_{\mathrm{B}}(1-k^2)} \right) \\ Y_{14\mathrm{B}} = \left(-\frac{1}{ZL_{\mathrm{B}}(1-k^2)} \right) \text{ or } \left(-\frac{1}{Z_{2\mathrm{B}}} - \frac{k}{ZL_{\mathrm{B}}(1-k^2)} \right) \end{cases} \tag{4.8}$$

Corresponding to the two different port configurations in Fig. 4.10 (b) there are two different solutions for the matrix elements $Y_{13\mathrm{B}}$ and $Y_{14\mathrm{B}}$. Equating (4.6) with (4.8) and solving for $Z_{1\mathrm{B}}$, $Z_{2\mathrm{B}}$, Z_{LB} and k delivers equations to calculate all component values in dependence of the arbitrary phase shift φ. As a real transformer with positive mutual inductance can only realize coupling coefficients $0 \leq k \leq 1$ the possible impedance values are constrained by this condition. The resulting bounds for φ together with the equations are summarized in Table 4.4.

The impedances $Z_{i\mathrm{B}}$ have to be capacitive for reducing size of the hybrid. This is only obtained for Case 2, which is illustrated in Fig. 4.12 (a).

Two rules for a realization can be read of the figure. First, the coupling coefficient of the transformer is not necessarily 1 to realize a quadrature hybrid, but it hast to be $> \frac{1}{\sqrt{2}}$. Any coupling coefficient between $\frac{1}{\sqrt{2}} \ldots 1$ can be compensated by the inductance and capacitance values. Second, $|Z_{1\mathrm{B}}|$ can be made infinite in the vicinity of $k = 1$ and $\varphi = \frac{\pi}{4}$. The capacitors to ground can therefore be omitted. Also $C_{2\mathrm{B}}$ and L_{B} are minimum in this range. Hence, with increasing k the size of the overall hybrid can be reduced for given Z_0 and f_0.

Variable	Case 1	Case 2
k	$\frac{1}{\sqrt{2}\sin(\varphi)}$	$\frac{1}{\sqrt{2}\cos(\varphi)}$
$0 \leq k \leq 1 \rightarrow \varphi$	$\frac{\pi}{4} \leq \varphi - 2n\pi \leq \frac{3\pi}{4}$	$-\frac{\pi}{4} \leq \varphi - 2n\pi \leq \frac{\pi}{4}$
$Z_{1\mathrm{B}}$	$\frac{jZ_0(1-2\sin^2(\varphi))}{(1-\sqrt{2}\sin(\varphi))\,(1-\sqrt{2}\cos(\varphi))}$	$\frac{-jZ_0(1-2\cos^2(\varphi))}{(1-\sqrt{2}\sin(\varphi))\,(1-\sqrt{2}\cos(\varphi))}$
$Z_{2\mathrm{B}}$	$jZ_0\left(1+\sqrt{2}\cos\left(\varphi\right)\right)$	$-jZ_0\left(1+\sqrt{2}\sin\left(\varphi\right)\right)$
ZL_{B}	$-jZ_0\sqrt{2}\sin\left(\varphi\right)$	$jZ_0\sqrt{2}\cos\left(\varphi\right)$

Table 4.4: Component values for the inductively coupled lumped element hybrid network.

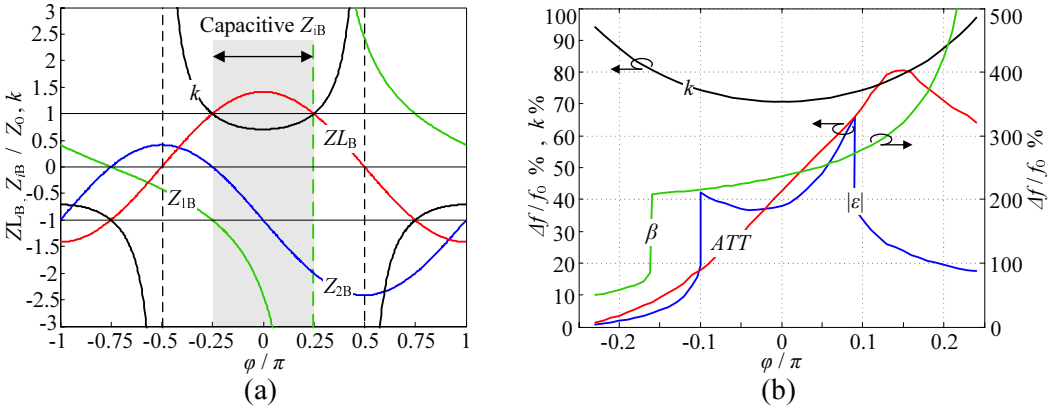


Figure 4.12: Inductively coupled lumped element hybrid with purely capacitive $Z_{i\mathrm{B}}$ values (Case 2): (a) Impedance values (normalized to Z_0) versus φ. (b) Relative bandwidth for attenuation $ATT > -0.1$ dB, gain error $|\varepsilon| < 0.75$ dB and phase error $|\beta| < 20°$.

Further, the impact of $k < 1$ on the bandwidth is examined. For each value of φ the network elements and scattering parameters are calculated and swept versus frequency. Thereof the maximum bandwidths for $ATT > -0.1$ dB, $|\varepsilon| < 0.75$ dB and $|\beta| < 20°$ can be determined. Continuing this procedure for the whole range of $|\varphi| < \frac{\pi}{4}$ (marked grey in Fig. 4.12 (a)) leads to the illustration in Fig. 4.12 (b). To keep the hybrid's size small, φ has to be > 0. In this range the overall bandwidth is limited by the bandwidth of $|\varepsilon|$. This bandwidth reduces when k approaches 1, but is still above the desired 10 %. The maximum bandwidth is obtained, if $k \sim 0.72$. For this coupling coefficient the size of the hybrid is larger than the smallest possible size. Hence, smallest size and largest bandwidth cannot be obtained simultaneously by this topology.

From Fig. 4.12 it can be concluded, that size and bandwidth of the hybrid are mainly determined by the coupling coefficient. Larger k reduces size and smaller k increases bandwidth. In a real implementation a trade-off between both has to be found, including further parasitic effects (e.g. resistive losses of the inductors), not included in this simple examination. Nevertheless a good explanation for the compromise between bandwidth and size in the hybrid design is given by this analysis.

4.2.3 Two section inductively coupled lumped element hybrid

The bandwidth of an inductively coupled lumped element hybrid was shown to be substantially increased compared to a hybrid without coupling. Further increase is obtained by cascading two of the inductively coupled hybrids. This leads to a structure as shown in Fig. 4.10 (c). The derivation of the admittance matrix is done by calculating the analytical S-parameters of a one section transformer based hybrid from equation (4.8). Two of these networks are cascaded, using the multi port network connection algorithm (Appendix D). In the resulting 4×4 S-matrix substitutions in the following order are made: $\frac{2}{Z_{1\mathrm{B}}} \rightarrow \frac{1}{Z_{3\mathrm{C}}}$, $\frac{2}{Z_{2\mathrm{B}}} \rightarrow \frac{1}{Z_{4\mathrm{C}}}$, $Z_{\mathrm{iB}} \rightarrow Z_{\mathrm{iC}}$ and $\mathrm{ZL_B} \rightarrow \mathrm{ZL_C}$. Finally the S-matrix is again transformed into an admittance matrix resulting in:

$$\underline{Y} = \begin{pmatrix} Y_{11\mathrm{C}} & Y_{12\mathrm{C}} & Y_{13\mathrm{C}} & Y_{14\mathrm{C}} \\ Y_{12\mathrm{C}} & Y_{11\mathrm{C}} & Y_{14\mathrm{C}} & Y_{13\mathrm{C}} \\ Y_{13\mathrm{C}} & Y_{14\mathrm{C}} & Y_{11\mathrm{C}} & Y_{12\mathrm{C}} \\ Y_{14\mathrm{C}} & Y_{13\mathrm{C}} & Y_{12\mathrm{C}} & Y_{11\mathrm{C}} \end{pmatrix}, \text{ with } \begin{cases} Y_{11\mathrm{C}} = \left(\frac{1}{Z_{1\mathrm{C}}} + \frac{1}{Z_{2\mathrm{C}}} + A + B + C\right) \\ Y_{12\mathrm{C}} = (A + kB - C) \\ Y_{13\mathrm{C}} = (A - B + C) \text{ or } \left(-\frac{1}{Z_{2\mathrm{C}}} + A - kB - C\right) \\ Y_{14\mathrm{C}} = \left(-\frac{1}{Z_{2\mathrm{C}}} + A - kB - C\right) \text{ or } (A - B + C) \end{cases}$$

$$A = \frac{1/4}{2Z_{3\mathrm{C}} + ZL_{\mathrm{C}}(1+k)}, \quad B = \frac{1/2}{ZL_{\mathrm{C}}(1-k^2)}, \quad C = \frac{1/4}{\frac{1}{\frac{1}{2Z_{3\mathrm{C}}} + \frac{1}{Z_{4\mathrm{C}}}} + ZL_{\mathrm{C}}(1-k)} \tag{4.9}$$

In Fig. 4.10 (c) two different port configurations (Case 1 and Case 2) are shown, and thus also (4.9) has two possible solutions for $Y_{13\mathrm{C}}$ and $Y_{14\mathrm{C}}$. The impedance values are determined by equating (4.9) with (4.6) for each of the two cases separately. The resulting system of equations contains seven different variables $Z_{1\mathrm{C}} \ldots Z_{4\mathrm{C}}$, ZL_{C}, k and φ. Four of these variables can be eliminated with the other three remaining independent. Similar to the single section inductively coupled hybrid, the coupling factor k is bounded between $0 \ldots 1$ and the impedances to ground $Z_{1\mathrm{C}}$ and $Z_{3\mathrm{C}}$ should be negative and their magnitude as large as possible to reduce size. Hence, expressing the other variables with these variables would give direct insight into the interdependence. The problem is, that the phase shift φ appears only inside trigonometric functions and therefore cannot be expressed in an explicit symbolic form. In order to obtain compact closed form expressions, the variables $Z_{1\mathrm{C}}$, k and φ are chosen as independent variables and the system of equations is solved for the remaining impedances. The results for Case 1 and Case 2 are shown in Table 4.5.

Compared to the single stage hybrid there are, in addition to φ, two more degrees of freedom, which can be used to find a better trade-off between size and bandwidth. For a lossless hybrid all impedances have to be purely imaginary. To obtain small dimensions and high quality factors the only inductive element should be ZL_{C}. The remaining elements should be capacitances with negative imaginary parts. Hence, among all combinations of $Z'_{1\mathrm{C}}$, k and φ only solutions with the intended signs ($\mathrm{Im}(Z_{i\mathrm{C}}) < 0$ and $\mathrm{Im}(ZL_{\mathrm{C}}) > 0$) are relevant. The impedances are all directly proportional to Z_0 (assumed to be real). The dimensions of capacitances therefore

Variable	Case 1 and Case 2
$Z'_{1C} = Z_{1C}/Z_0,\ k,\ \varphi$	independent variables
Z_{2C}	$Z_0 \cdot \dfrac{Z'_{1C}\,(1-k)}{k-(k\,R_{A1,2}+R_{B1,2})\,Z'_{1C}}$
Z_{3C}	$Z_0 \cdot \dfrac{Z'^2_{1C}\,R_{D2,1}}{(1-(R_{A1,2}+R_{B2,1})\,Z'_{1C})\cdot(Z'_{1C}\,(R_{C1,2}+R_{D1,2})-1)}$
Z_{4C}	$Z_0 \cdot \dfrac{Z'^2_{1C}\,(1-k)^2\,R_{B2,1}R_{D2,1}}{(1-(R_{A1,2}+R_{B1,2})\,Z'_{1C})\cdot R_{E1,2}}$
ZL_C	$Z_0 \cdot \dfrac{Z'_{1C}}{(1+k)\,((R_{A1,2}+R_{B1,2})\,Z'_{1C}-1)}$

Substitutions with $Y'_{ij} = Z_0\,Y_{ij}$ taken from (4.6):
$R_{A1} = Y'_{11} - Y'_{13},\ R_{B1} = Y'_{14} - Y'_{12},\ R_{C1} = Y'_{11} + Y'_{13},\ R_{D1} = Y'_{14} + Y'_{12}$
$R_{A2} = Y'_{11} - Y'_{14},\ R_{B2} = Y'_{13} - Y'_{12},\ R_{C2} = Y'_{11} + Y'_{14},\ R_{D2} = Y'_{13} + Y'_{12}$
$R_{E1,2} = Y'_{12}\cdot\left(\left(R_{B1,2}+k^2R_{D1,2}\right)Z'_{1C} + \left(1+k^2\right)(R_{A1,2}\,Z'_{1C}-1)\right) +$
$+2\,k\cdot\left(Y'_{13,14}\,\left(\left(R_{B1,2}+Y'_{11}-kY'_{13,14}\right)Z'_{1C}-1\right) - Y'^2_{12}Z'_{1C}\right)$

Table 4.5: Component values for the two section inductively coupled lumped element hybrid network.

reduce, and the dimensions of inductances increase with the reference impedance. Normalizing all impedances to the reference impedance ($Z'_{iC} = Z_{iC}/Z_0$ and $ZL_C{}' = ZL_C/Z_0$) eliminates this dependence for the moment.

With the equations for the normalized impedance values the relative bandwidth of ε, β and ATT can be calculated in dependence of Z'_{1C}, k and φ. The aim is a graph, similar to Fig. 4.12 (b) to identify regions with maximum bandwidth. As a plot over three independent variables cannot be illustrated, a further assumption has to be made. The bandwidth of the amplitude imbalance is most critical for the design. For each combination of Z'_{1C} and k there exists a φ_{max} where the bandwidth of ε is maximum. Taking only these maximum values eliminates the independent variable φ, which is a free design variable, anyway. The bandwidth and impedance values can then be illustrated in a contour plot versus Z'_{1C} and k. For that purpose a script is written in Matlab, to automate following operations:

- Calculation of Z'_{iC} and ZL'_C for $Z'_{1C} = 0.1\ldots10^4$, $k = 0\ldots1$ and $\varphi = -\pi\ldots\pi$.
- Elimination of all results which do not fulfill the constraints: $\mathrm{Im}(Z'_{iC}) < 0$ and $\mathrm{Im}(ZL'_C) > 0$.
- Calculation of normalized bandwidths ε, β and ATT for all remaining results.
- For each combination of Z'_{1C} and k the φ_{max} is determined, where the bandwidth of ε is maximum.
- Drawing contour plots of ε, β, ATT, Z'_{iC} and ZL'_C versus Z'_{1C} and k for φ_{max}.

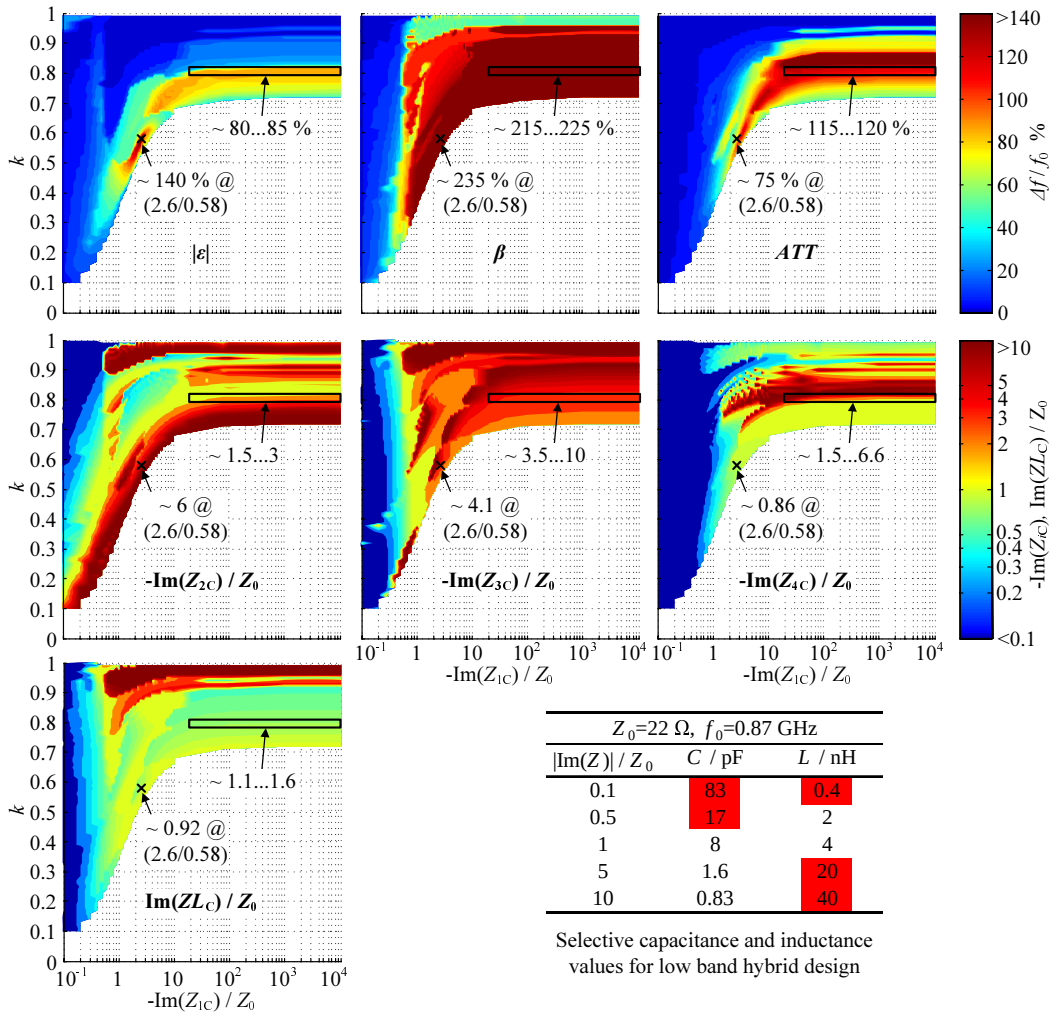
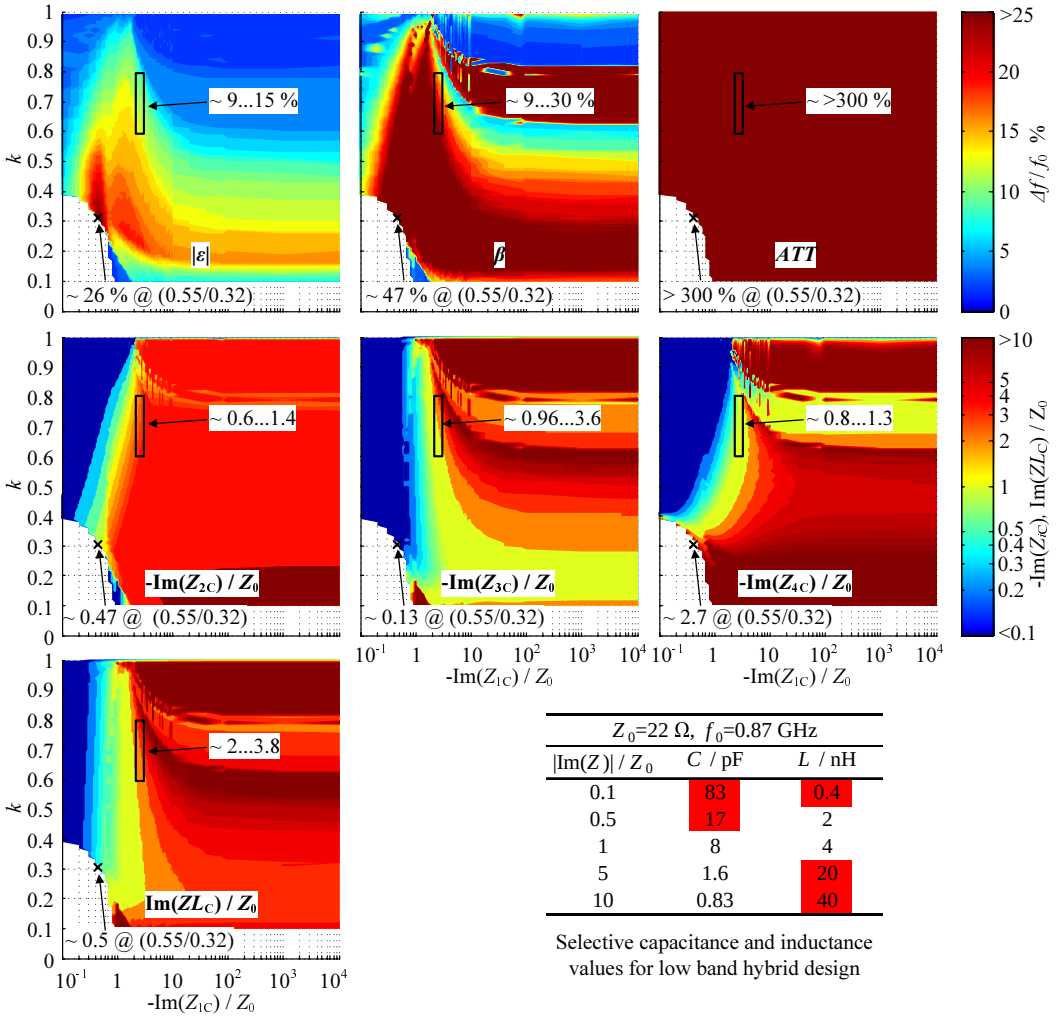


Figure 4.13: Contour plots of relative bandwidths and normalized impedance values versus shunt impedance Z'_{1C} and coupling factor k for φ_{max} and Case 1 port orientation.

The final plots for both cases are shown in Fig. 4.13 and Fig. 4.14, which refer to error constraints in low band operation ($|\varepsilon| < 0.75$, $|\beta| < 20°$ and $ATT > -0.1$ dB, see Table 4.2). Analyzing Case 1, the maximum achievable bandwidth of $|\varepsilon|$ is 140 %, which is more than 2 times that of a single section inductively coupled hybrid. This point is marked with a cross in all contour plots of the figure. While the phase imbalance bandwidth is very high (235 %), the ATT bandwidth is only 75 %. This implies that the attenuation is above 0.1 dB, when operating up to 140 % bandwidth. Attenuations due to mismatch or dissipation are unavoidable for a hybrid but for a balanced amplifier they can be tolerated at the input, being less critical there than at the output. Hence, the smaller ATT bandwidth does not exclude the application of a hybrid with maximum ε bandwidth. One reason why such a hybrid is hardly realizable can be the overall size. The figure shows selective L and C values for low band operation within the illustrated normalized impedance range. The red marked values are too large or too small for an integration on chip, when using



Z_0=22 Ω, f_0=0.87 GHz		
|Im(Z)| / Z_0	C / pF	L / nH
0.1	83	0.4
0.5	17	2
1	8	4
5	1.6	20
10	0.83	40

Selective capacitance and inductance values for low band hybrid design

Figure 4.14: Contour plots of relative bandwidths and normalized impedance values versus shunt impedance Z'_{1C} and coupling factor k for φ_{max} and Case 2 port orientation.

MIM capacitors or spiral coils. For the point with maximum ε bandwidth the single component values are within limits realizable on chip, but the overall sum is large. Capacitors with totally 35 pF and four coupled inductors with 3.7 nH, each are needed. Also the capacitors to ground cannot be omitted. The size of the hybrid can be reduced using values within the zone surrounded by a rectangle in Fig. 4.13. The bandwidth of amplitude imbalance is then reduced to 80 %, but C_{1C} and C_{3C} are small enough, so that parasitic capacitances coming from the spiral inductor and wiring are sufficient. As these parasitics scale in a nonlinear way with the inductor value, the final values and therefore the size of the hybrid strongly depend on the specific layout. There exist other than the marked regions in the plot, with smaller lumped element values, but this only at cost of further bandwidth reduction.

The plots for the port configuration of Case 2 (Fig. 4.14) show that the bandwidth of $|\varepsilon|$ is at best 26 %. This is five times smaller than the maximum achievable

bandwidth for the port configuration of Case 1. Also the inductance and capacitance values are larger than for the first case. The only advantage for Case 2 is that the ATT bandwidth is > 300 % over the whole range of values. As already explained, this is less relevant for the input hybrid of a balanced amplifier, because a larger ATT only leads to a reduced gain, which slightly reduces PAE, but does not affect the principle functionality of the amplifier. The areas in the figure, which are marked by the rectangle, point out designs where all lumped element values are within a range, which is realizable on chip. The resulting bandwidth is at best 15 %. Although this would be sufficient for the low band design, there is only few margin, when losses, parasitic effects and tolerances are included. Therefore it can be concluded, that this port configuration is less relevant for the design of a small size and broadband quadrature hybrid.

The analysis shows the main tendency for bandwidth and size, and the potential of the two stage inductively coupled lumped element hybrid over the single stage hybrid. It also shows, that among the two different port configurations only Case 1 has relevance for the balanced amplifier. This information is important for the physical design process to reduce the number of optimization steps and therefore simulation time. The main difference between this ideal analysis and a final real hybrid, which contains non-ideal elements, is the remarkable part of resistance in the inductors. Without proof it is stated, that this leads to a reduced bandwidth of $|\varepsilon|$ and an increased overall ATT. At the center frequency also a significant phase deviation from ideal quadrature 90° difference will be obtained. Despite these differences, the present calculation is valuable as it provides initial values for the design iterations. A circuit simulation software like ADS can be used for fine-tuning of the components, including parasitic effects. An efficient way is to run an automatic numerical optimization simulation with the initial element values taken from the calculation. As target values the calculated bandwidths are entered, to guarantee better and faster convergence of the simulator. Without the analytical calculation the size/bandwidth trade-off is not known and goals which are close to the optimum are hard to define.

4.2.4 Final hybrid design

The previous discussions have shown that largest bandwidth of the lumped element hybrid can be obtained with a transformer based topology (Fig. 4.10 (b), (c)). This enables the integration of an on-chip hybrid with a minimum of 10 % bandwidth, as necessary for low band operation. The one and two section topologies were realized in the already introduced B7HFP technology.

Beginning with the *single section* hybrid, the design procedure starts with a proper guess of the element values. This is given by the previously calculated values for the ideal hybrid. To fulfill the aim of a compact broadband design, it has been shown that the coupling coefficient of the hybrid has to be about 0.72. A transformer with such medium coupling coefficients can be integrated as a bifilar spiral structure of so called Frlan type [Frlan 89]. This uses two closely spaced intertwined parallel windings, as illustrated in Fig. 4.15 (a). The transformer is symmetric and the main

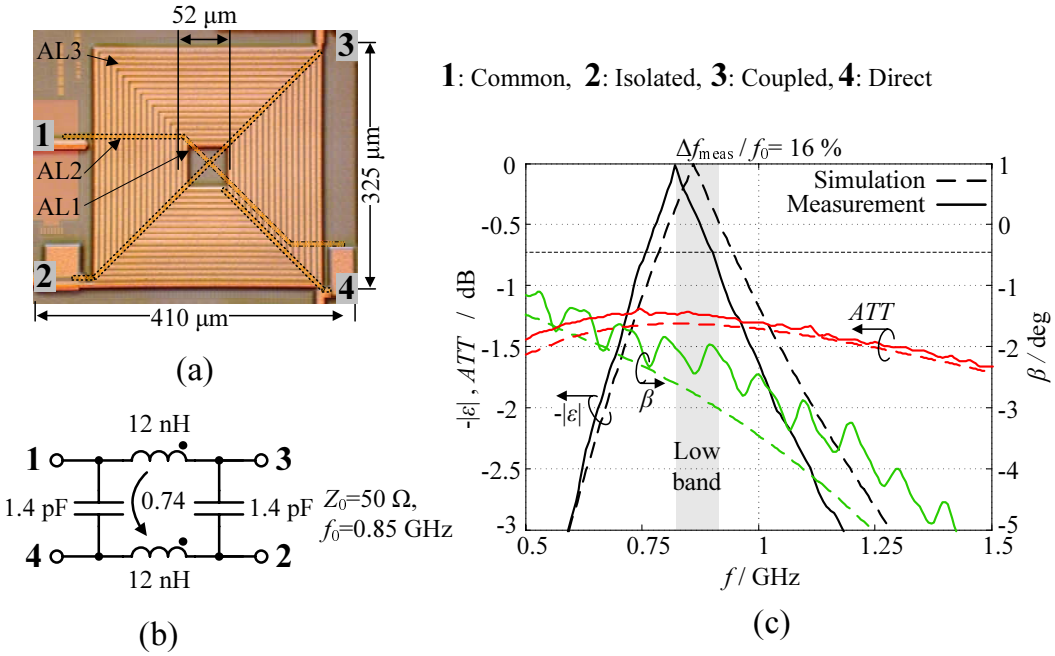


Figure 4.15: Realized hybrid Low Band: (a) Layout micrograph, (b) lumped element equivalent circuit, (c) measurement and simulation results.

part (the two coils) can be made in the low loss top metal layer. Cross-overs are realized in lower metal layers. The resulting low insertion loss of the transformer, due to low series resistance and substrate capacitance of the coils, is the main advantage of the bifilar structure. The disadvantage therein is that the transformer uses a lot of silicon area. This, because the tracks of the two coils have to be placed side by side. In comparison to a structure where primary and secondary are realized in two metal layers which lie on top of each other, the Frlan transformer needs about twice the area. Another disadvantage is the small coupling coefficient. The magnetic fluxes are coupled between parallel lines through the center of the two coils. The mutual coupling is therefore reduced for the Frlan coupler, because the primary and secondary coils are not as closely placed, as if two different metal layers would be used. The minimum spacing is constrained by the design rules for the fabrication process. Nevertheless, the coupling coefficient, which can be achieved for such a structure with reasonable quality factor in B7HFP is at best $k < 0.8$.

The main mechanism to vary the coupling coefficient in a Frlan transformer is variation of the ratio of inner diameter to outer diameter. With an increase of this ratio more lines are surround by the mutual magnetic flux, which increases the coupling coefficient. In turn the size, as well as the metal resistance will increase, when inductance is kept constant. The metal resistance can be reduced again by increasing the line width. This has a negative effect on the parasitic capacitance to substrate and the coupling coefficient, as the ratio of inner and outer diameter reduces again. The tuning of only one parameter of the transformer is therefore impossible, without influencing the others.

The main tuning mechanisms for the transformer design can be summarized in the following items:

- Adjustment of k by variation of ratio between inner diameter and outer diameter and spacing of windings.
- Reduction of series resistance by increasing widths of windings in the top metal and cross-overs in the bottom metal layers.
- Reduction of substrate coupling by small line widths.
- Variation of shunt capacitance by variation of winding and cross-over line widths.

The analysis of the ideal hybrid has shown that for $k < 1$, some amount of capacitance to ground is needed (C_{1B}), but it can be much smaller, compared to C_{2B}. With the tuning possibilities of the transformer layout, k can be adjusted to obtain a C_{1B} value, sufficiently small, to be already covered by the unavoidable parasitic substrate capacitance. This way no additional MIM capacitor to ground is needed. The layout optimization of the two transformer coils was done by electromagnetic simulation, using Sonnet. The extracted scattering parameters were imported to an ADS simulation setup to determine the values of the remaining coupling capacitors at both sides of the transformer. The final hybrid is shown in Fig. 4.15 (a) together with the simplified equivalent circuit in Fig. 4.15 (b), containing the main values. The ideal hybrid analysis assumes that the shunt and coupling capacitances (C_{1B} and C_{2B}) are lumped elements. As the transformer already includes some distributed inter winding, coupling and substrate capacitance, the final values of the remaining lumped elements obviously deviate from the initial guess. Fig. 4.15 (c) shows the simulation and measurement results of the hybrid. The $|\varepsilon|$ characteristic of the measured hybrid is shifted by about -30 MHz (-4 %) compared to the simulation. This is mainly due to additional capacitive coupling to substrate. As the small tuning capacitors to ground are realized by the intrinsic wire capacitance of the transformer, the hybrid is sensitive to process variations of the distance between AL3 and substrate. In addition the measurement was done on waver level, using additional pads for landing of the GSG (ground signal ground) probes. The capacitance of these pads (several 100 fF) was not deembedded, which also explains the frequency shift. The other parameters φ and ATT show less sensitivity to the additional parasitics. Concluding, the hybrid has an overall bandwidth of 16 % and its center frequency is shifted to the lower edge of the desired frequency band. The attenuation of -1.25 dB is due to the transformer's wire resistance ($\sim 10\ \Omega$) and substrate losses, but low enough to use the hybrid at the input of the amplifier.

To achieve a further improved bandwidth a *two section* hybrid was implemented. The reference impedance was chosen $Z_0 = 22\ \Omega$ for three reasons. First the area can be made smaller, because the inductor values of the transformer reduce. The transformer represents the majority of the complete hybrid. Second the sensitivity to parasitic shunt capacitances is smaller, because the absolute capacitor values become

larger. Third the matching network to the high power transistor will have smaller transformation ratio and therefore larger bandwidth. The input impedance of the large transistor array is $\sim 2\ \Omega$ (see Chapter 4.4) and an on-chip matching network to 50 Ω would consume a lot of area and be narrow-banded, thus constraining the overall bandwidth by the matching network. All together this smaller Z_0 value brings a reduced overall size, less sensitivity to tolerances and an increase in bandwidth.

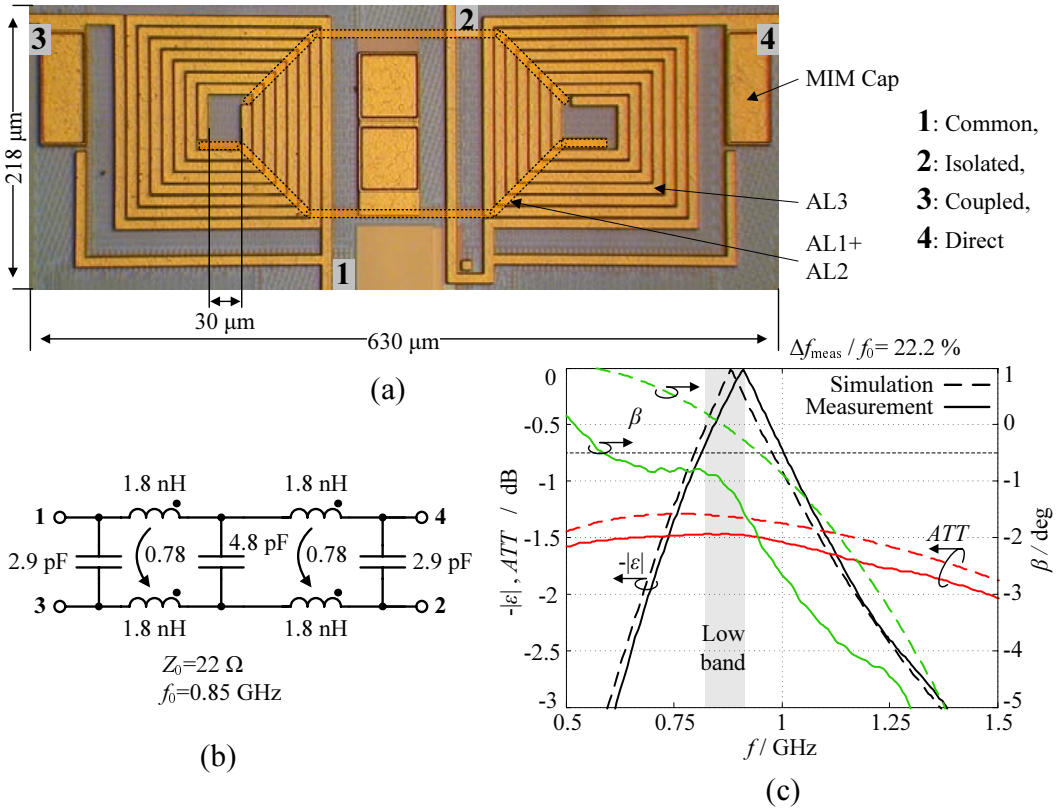


Figure 4.16: Realized hybrid Low Band: (a) Layout micrograph, (b) lumped element equivalent circuit, (c) measurement and simulation results.

Similarly to the single stage hybrid, the design procedure starts with values from the ideal analysis. The initial values are taken from the region marked with the rectangle in Fig. 4.13 ($-\mathrm{Im}(Z_{1C})/Z_0 = 11$, $k = 0.78$). Based on these values, the layout of the transformer is optimized using EM field simulation, in order to find a compromise between inductance value, coupling factor and substrate capacitance. The resulting series resistance of one spiral coil is 3 Ω. Compared to the impedance of the inductor $Z = 9.6\ \Omega$ @ 0.85 MHz it is not negligible and causes remarkable difference to the initial values from the ideal calculation. The simulated scattering parameters of the transformer were imported to ADS. All remaining coupling capacitances were determined by a numerical optimizer simulation using as target the expected ε, β and ATT values from the previous calculation with idealized assumptions. The final hybrid design is shown in Fig. 4.16 (a),(b). The design process was more elaborate

than for the single section hybrid, because of an additional inductive coupling between the two transformers. The distance between the transformers had therefore influence on center frequency and bandwidth. This influence could be reduced by applying the capacitor C_{4B} as long wide block in the gap between the transformers.

The measurement results in Fig. 4.16 (c) are in good agreement to the simulations. The amplitude imbalance is slightly shifted to higher frequency, but still within the bound of 0.75 dB. Within the operation band the phase shift is only 0.5° and the attenuation is 1.48 dB. The overall bandwidth, where $|\varepsilon| < 0.75$ dB is 22.2 %.

A comparison to the single section hybrid shows, that the attenuation is only 0.23 dB larger in the two section design. For the same values of amplitude imbalance the single section hybrid in Fig. 4.15 has a bandwidth of 16 %. The bandwidth is therefore improved by the two section hybrid by 39 %. The area of the two section hybrid is only about 3 % more than for the single section hybrid. In both designs most area is consumed by the transformers, which scale in proportion to their nominal inductance $L \propto Z_0/\omega_0$. The center frequency is the same for both designs, but the reference impedance differs by a factor of 2. Designing the single section hybrid for the same $Z_0 = 22\ \Omega$ would show a size reduction of less than 30 % (L scales with the square of the number of turns and C increases with Z_0). It can be concluded that the two section hybrid consumes more chip area, but increases the overall bandwidth remarkably while keeping about the same level of insertion loss. For the design of the multi-mode PA prototype it was necessary to have larger bandwidth and more stability over process variations, than fulfilling size constraints. Therefore the final amplifier contains a two section inductively coupled hybrid.

4.3 On board performance optimization

The design of a power amplifier, delivering more than 30 dBm output power from a battery supply of 3.5 V requires realization of a load smaller than 6 Ω, seen from the active device. This load has to be matched at a very high accuracy to meet all specifications. Classical design procedures aim at accurate modeling of the transistor and the matching network, requiring costly measurement equipment and a time consuming process. This section is reserved for describing a method, developed during work on the multi-mode PAs, which allows quick and direct optimization of the lumped element matching networks for the final design. Due to the use of external SMD (Surface Mount Device) components, in the following, this method is termed SMD Load Pulling.

4.3.1 Conventional Load Pull measurement procedures

The principle technique of load pull analysis is a well known method to measure a PA's RF and DC parameters as function of load impedance. The intentions behind range from characterization under real operating conditions to modeling for simulation or optimization of the device under test (DUT). Adapted to the purpose a

different set of input and output parameters is recorded. The typically measured output parameters are output power P_{out} (at fundamental and higher harmonics), gain G, PAE, supply current I_{DC} and ACLR / EVM as measure of linearity. The input parameters for one sweep over the complex valued impedance are bias voltage V_{B}, supply voltage V_{CC}, input power P_{in} and frequency f. These can also be swept, e.g. for development of behavioral models to improve accuracy and to extend the scope of the model.

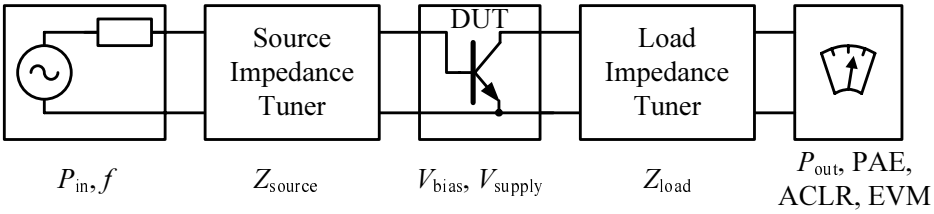


Figure 4.17: Load Pull setup.

The simplest load-pull setup for such an analysis is shown in Fig. 4.17. It basically consists of a calibrated power source, output power meter and a calibrated impedance tuner. The tuner is the key element of the setup and recently great effort is spent to improve its tuning range, reproducibility and frequency response. Tuners are divided into *active* and *passive* systems. Examples for both are depicted in Fig. 4.18.

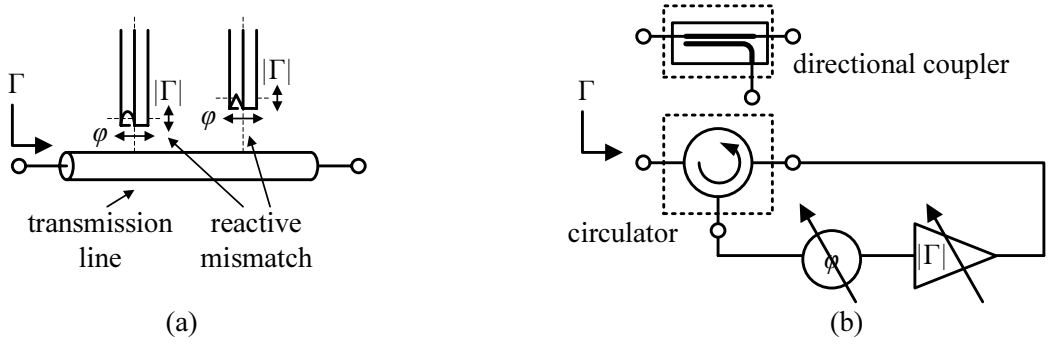


Figure 4.18: Load Pull tuner examples: (a) passive slidescrew tuner, (b) active closed loop tuner

Passive tuners consist of mechanically or electronically adjustable passive structures to vary magnitude and phase of the complex reflection coefficient $\Gamma = |\Gamma|e^{j(\varphi)}$, seen into either of the ports, with the other terminated in a fixed load. A various number of possible implementations can be found in [Mauri microw 08, Focus microw 08]. The drawback of passive tuner systems is that they can only realize $|\Gamma| < 1$ at their ports. The usable range is further reduced by dissipative losses between the DUT and the tuner. For connectors, cables and test fixtures these losses can easily add up to -1 dB. Best passive tuner systems can realize reflection coefficient of 0.9, which

would result in an effective reflection coefficient at the DUT of $0.9 \cdot 10^{-1/10} = 0.71$. The tuning range $8.3 \ldots 295\ \Omega$ is too small for measurement of high power devices, which require a load smaller than $6\ \Omega$. There exist possibilities for improvement by usage of a pre-matching network near the DUT, but this requires a special test fixture, characterization of this network and overall reduces tuning possibilities at higher harmonics.

Active tuners can overcome these problems as they can realize $|\Gamma| > 1$. The reflected wave is generated by a separate amplifier and controlled in amplitude and phase. The maximum $|\Gamma|$ is then only limited by the output power of the load amplifier, which generates the backward traveling wave. The example for an active tuner system, shown in Fig. 4.18 (b), uses a loop to align amplitude and phase of forward and backward traveling waves. This setup has two main drawbacks. First, only narrowband signals can be measured because of delay times of the active load. Second, the system shows tendency to oscillate, because a small portion of the loop amplifier's output power is fed back into its input through the imperfect circulator (directional coupler). A loop free implementation with a separate path for the reflected wave can enhance the properties of an active load pull system. This in turn requires a larger number of measurement equipment and high control synchronicity of the two paths. Inherent phase-drifts of clock signals limit the frequency range of the system, especially when higher harmonics are tuned as well.

The applicability of load pull measurement results for the design of an output matching network is strongly dependent on the kind of DUT. The measurement conditions as well as layout and size of the DUT should be very similar to the final PA transistor. Performing the measurements *on-wafer* has some basic drawbacks:

- The maximum measurable output power is limited by the maximum safe operating current of the GSG probes. Probes with higher power capability increase acquisition costs.
- The contact resistance between the transistor pads and the GSG probe tips can be in the order of the load impedance, when measuring large transistor arrays. This reduces accuracy of the measured values.
- The size and output power of the measured DUT is therefore smaller than for the final PA. A scalable model has to be developed, which is a time consuming process. Even the best models still contain some inaccuracy due to differences between the on-wafer measurement conditions and the final design. Amongst others, the main differences are:
  - Bias conditions and current distribution along the transistor array.
  - GND reference plane and GND contact impedance.
  - Source impedance, seen at the input of the transistor.

To avoid development of a scalable model from on-wafer measurements, the final transistor layout can also be measured on a separate *test fixture*. This has some disadvantages, as well:

- Transmission lines and connectors of the test fixture have to be deembedded. This requires calibration standards (such as reflections and terminations) on a separate calibration board.
- The bond wires cannot be deembedded, except with a separate chip, containing suitable calibration standards. To obtain proper results, the calibration process is very time consuming. Even, if the actual load impedance is determined very accurate, the tolerance of the bond wire inductance is usually very large, so that the final results and the reference plane for the DUT are not exact.

It can be concluded that Load Pull setups, which deliver precise results require a large number of additional costly measurement equipment. There are areas where this effort can be justified, as for the already mentioned behavioral modeling, but for an optimization process on board a cheaper and more appropriate approach is desired. The SMD Load Pull method, which will be described next, can be carried out with standard RF measurement equipment, already available to trace an input power sweep of the PA.

4.3.2 SMD Load Pulling

Algorithm

The basic idea behind this optimization procedure is to measure the PA's characteristic versus load impedance directly under the final chip and board layout conditions. The matching network on board is tuned and the resulting load impedance, seen by the PA is calculated with an appropriate model. For every change in the matching network, the parameters of interest are recorded. After some iterations the load pull contours can be plotted and from these, tendencies for an optimum load can be deduced. If a satisfying compromise between all parameters is found, the matching network is optimum and the procedure can be finished. Fig. 4.19 shows a flow chart

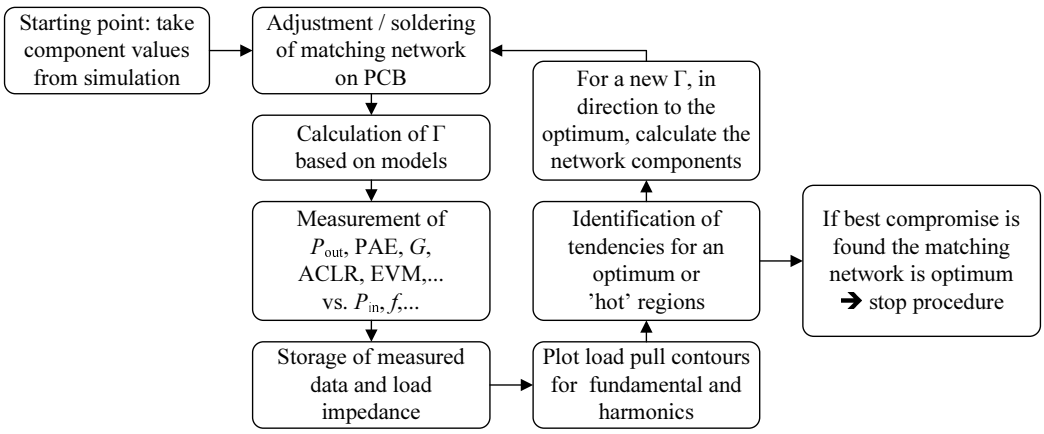


Figure 4.19: Flow chart for SMD Load Pulling optimization procedure.

of the procedure. For matching of the PA not only Γ at the fundamental, but also at higher harmonics has influence on the performance. If the matching network model is suitable to calculate Γ at higher harmonics, the procedure can also be used to optimize the influence of the impedance at these harmonics.

Tuneable Matching Network

In order to apply this procedure the output matching network must offer tuning possibilities in an appropriate range of the smith chart. The range of Γ will be about the expected peak values of the considered parameters (P_{out}, PAE, $G, \ldots$). Variability can be achieved by using a series of lumped capacitors and inductors with different values, or distributed elements, like transmission lines with tuneable lengths. Besides flexibility, the matching network also has to provide a layout, which can be modeled, to calculate Γ.

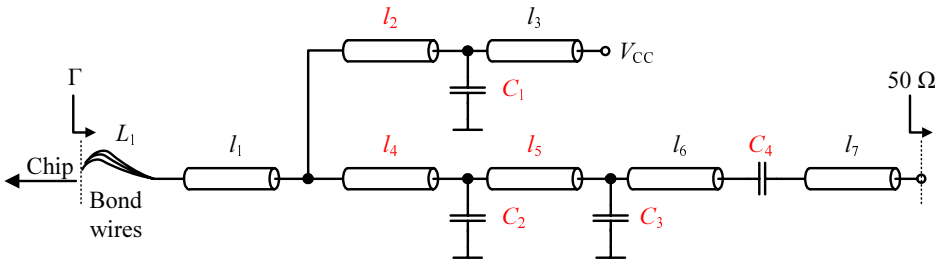


Figure 4.20: Tunable matching network consisting of bond wires, transmission lines and lumped capacitors.

A matching network, fulfilling these requirements and frequently used in PA design is shown in Fig. 4.20. The bias feed is realized with a transmission line to V_{CC}. Its length is tuned with a shunt capacitor. In order to reduce dimensions, an inductor can be used instead of the line. Along the output line two other lumped capacitors can be soldered at different positions, to tune the impedance at the fundamental and harmonics. Main function of the series capacitor C_4 is DC blocking, but it can also be used for tuning purposes. In theory the whole smith chart can be covered with this similar Π-type network. The range of Γ is only reduced by dissipative losses, the set of available SMD component values and the maximum line lengths. The variables in Fig. 4.20 with red font color indicate the degrees of freedom. These values can be changed during optimization.

Matching Network Models

To calculate Γ in dependence of the variable L, C and l values, a realistic model of the network has to be defined. For automation of the procedure in a stand alone Matlab program, every element is defined by its chain matrix ($ABCD$). This matrix is convenient, because cascading of elements is done by multiplication of the

corresponding matrices. This way the topology can be built by multiplication of the matrices in the same order as the order of components in the network. The resulting overall *ABCD* matrix is then converted to a scattering matrix, being normalized to the load impedance (50 Ω). Finally, $\Gamma = S_{11}$ is obtained. The models, used for calculation of the *ABCD* matrix are summarized in Fig. 4.21.

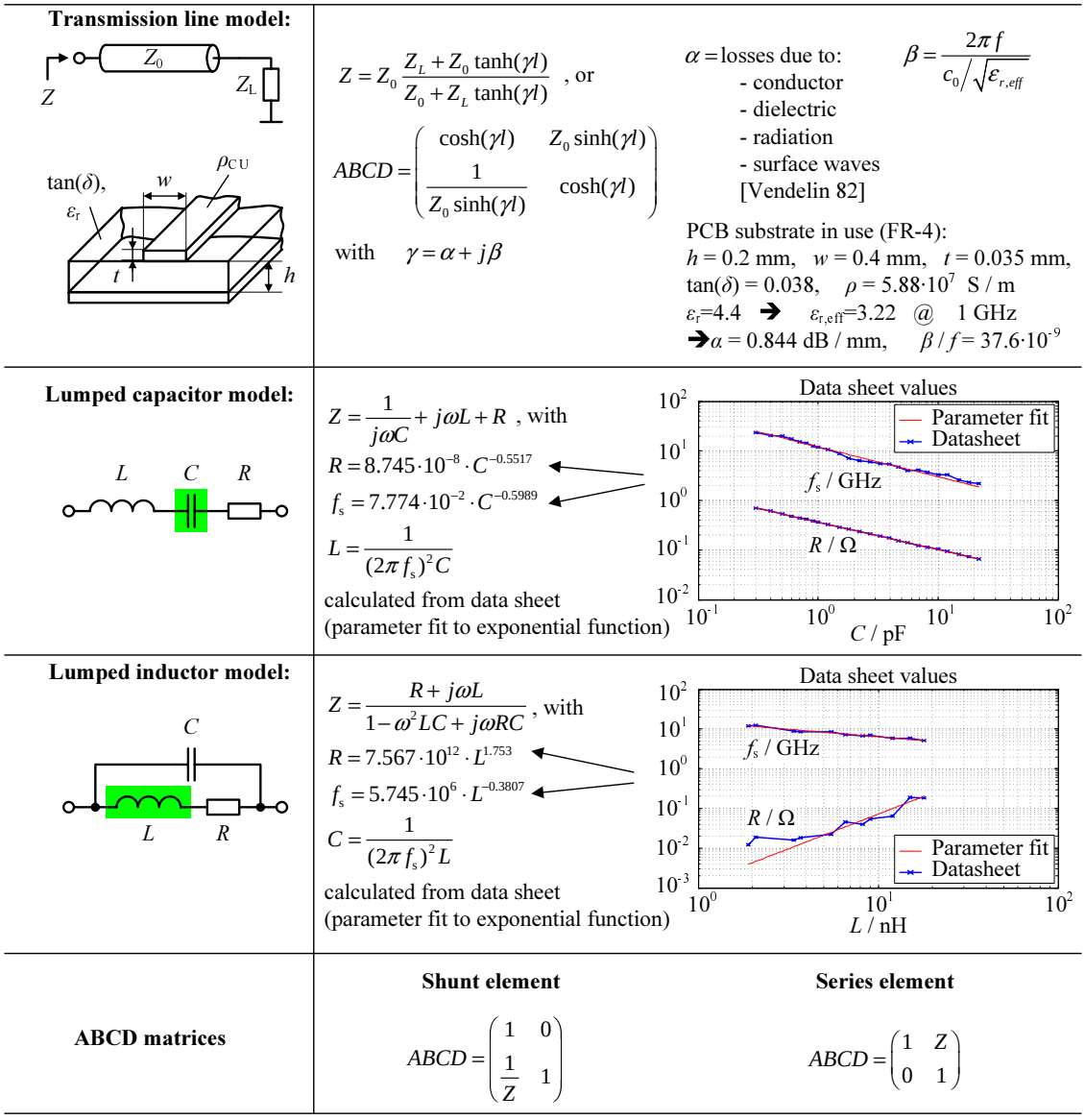


Figure 4.21: Models for calculation of Z and *ABCD* matrix. Capacitors in use: Multilayer ceramic capacitors [Epcos 08]; Inductors in use: Coilcraft chip inductors [Coilcraft 08]

The chain matrix of the *transmission line* contains the characteristic impedance Z_0 and the complex propagation constant γ as absolute terms. The line length l is the tuneable variable. The PCBs in this work use FR-4 as dielectric and copper as conductor material. The realized transmission line is a microstrip line with charac-

teristic impedance 50 Ω. The dimensions are given in Fig. 4.21. The other values in the figure, quoted from the data sheet, are used to calculate the constants α and β with formulas taken from [Vendelin 82].

The models used for the *lumped capacitor* and *lumped inductor* are well known in literature. The self-resonant frequency f_s and series resistance R vary with the component values. These values are provided in the data sheets usually as plot or table. In order to use a function for calculation of the parasitics, the data sheet specifications are fitted to an exponential form ae^b, where a and b have to be determined. The resulting functions together with a comparison of the data sheet values and the parameter fit are also shown in Fig. 4.21. The parameter fit matches very well, only the series resistance R of the inductor for small L values is too optimistic. Nevertheless, the function can be applied as these L values are not in use here. Only large inductor values are used, applied as RF choke in the DC bias feed ($L > 10$ nH $\triangleq 63\,\Omega$ @ 1 GHz).

For completeness also the *ABCD* matrices for shunt and series connection of the lumped components are given in Fig. 4.21.

Matlab Program

In order to automate calculations and to speed up the optimization process a program was written in Matlab. Its flow chart is shown in Fig. 4.22. Two main functions

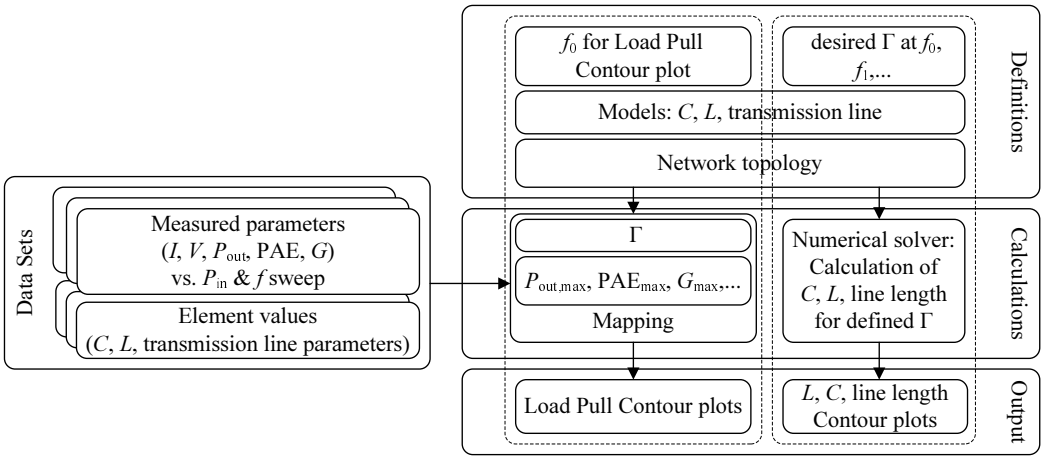


Figure 4.22: Flow chart for the SMD Load Pulling program.

are implemented, tagged by the dotted boxes:

- Calculation of the load pull contour plots based on mapping of the measured data sets and element values. The graphical output is used to determine a new Γ, which is closer to the optimum.
- Calculation of the L, C and l values, which correspond to the new Γ. These element values are used for the next iteration step.

Both functions use the same *ABCD* matrix definition of network topology and lumped element models, to guarantee consistency.

For drawing of the load pull contours the fundamental frequency f_0 has to be defined. This must also be available in the measured datasets, because no frequency interpolation is done. Two datasets are generated during measurement. The first one contains the recorded parameters (P_{out}, PAE, ...), swept over Γ, input power and frequency. For each Γ and the defined f_0 the maximum values over input power are read out ($P_{\text{out,max}}$, PAE_{max}, ...). The second dataset is a table, containing the lumped element values, used for calculation of Γ. The connection between both datasets is done via an index number. During measurement a unique number is given to the matrix of the measured data and the same number is entered in the table for the actual L, C, l combination. This way the program can perform a mapping between the datasets. The unique number is then plotted at the position of the Γ value, to easily reconstruct the matching network, by looking into the table.

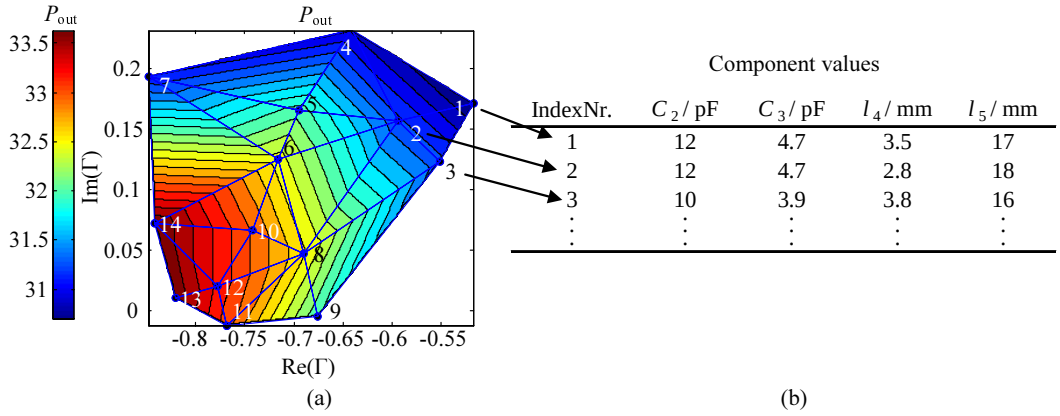


Figure 4.23: Program output of the first function: (a) Load Pull contours with triangular mesh grid and index numbers, (b) table of component values.

Fig. 4.23 shows an example for the component values table and the resulting load pull contour plot, containing mapping numbers from the table. The data points collected from the measurements are irregularly spaced. To draw smooth contours, the non-uniform, sparse data samples have to be fitted to a smooth surface. The solution requires scattered data interpolation to determine the value at any arbitrary position within the measured region. The technique used here for interpolation is based on triangulation. A triangular mesh between measured data points is generated, spanning a three dimensional surface, made up of triangular patches. Within each triangle the data is linearly interpolated over a regularly spaced fine grid. Thus, load pull contours at user defined levels can be plotted at the resolution of the new grid. Fig. 4.23 shows the triangular mesh, overlayed the measured data points. The triangulation approach is computationally intensive and smoothness of the contours generated by this approach suffer due to piecewise linear interpolation within triangles. Nevertheless, this technique is especially attractive for the generation of load pull contours from measured data, as it does not make any assumption regarding

the shape of the contours. This shape can be circular, elliptic or another general oval, varying from one measured parameter to another. To improve smoothness the individual contours can be further post processed by polynomial curve fitting. This is not done here, as it would not improve the accuracy for the geometrical process of finding an optimum load impedance.

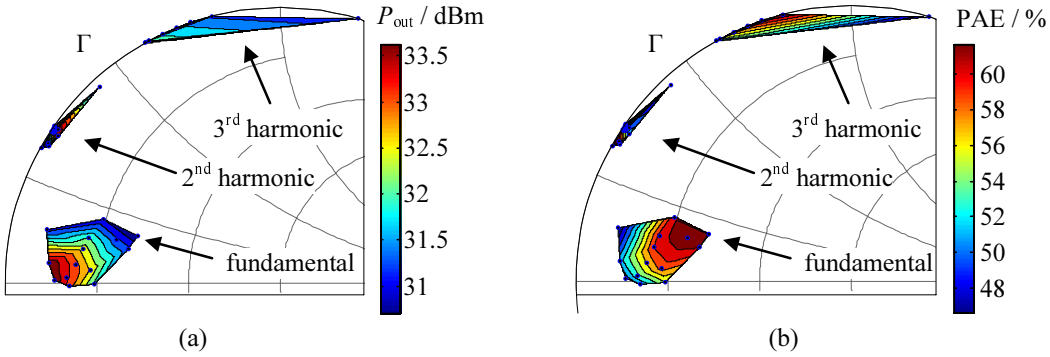


Figure 4.24: Load Pull contours of output power and PAE at fundamental, 2nd and 3rd harmonic.

The first function of the program also generates load pull contours at the harmonic frequencies. For this the Γ values are calculated at multiples of the fundamental frequency $f_1 = 2f_0$, $f_2 = 3f_0$,... and contours are plotted for the parameters of interest (P_{out}, PAE,...), which are measured at f_0. Fig. 4.24 shows an example, containing the fundamental frequency and the first two harmonics f_1 and f_2. As the values of the parameters (P_{out}, PAE) are the same for the three contours, the same contour levels (color codes) should be visible as well. Due to the small resolution of the figure the red areas for 2nd and 3rd harmonic are not clearly visible. Illustrating the harmonics in different figures with higher resolution would allow to identify related Γ values by the color code. This plot is necessary to identify the harmonic termination during measurement. Unlike conventional load pull systems, where harmonics can be filtered and tuned separately, here the control is limited by the topology and the specific layout. When setting a specific Γ_{f_0} it happens that Γ_{f_1} and Γ_{f_2} can not be set to the desired value due to e.g. missing values in the capacitor series or constraints on length of the transmission lines. By taking available component values, next to the calculated, the harmonics will vary about the desired harmonic termination, as illustrated in the figure. If the variation becomes too large, the contours at Γ_{f_0} will be deformed, showing notches and non-oval shapes. With help of the plots the points with large deviation from the desired harmonic terminations can be identified. Therefore it can be decided whether a distortion of the contours comes from a measurement error or from a wrong harmonic termination.

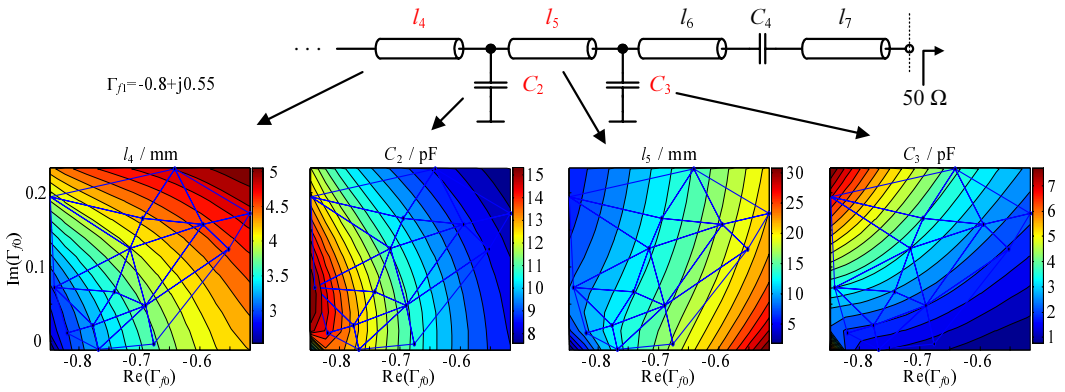


Figure 4.25: Program output of the second function: Values of 4 tuneable components, calculated for $\Gamma_{f_0} = -(0.85 \ldots 0.51) + j(0 \ldots 0.23)$ and $\Gamma_{f_1} = -0.8 + j0.55$

The second function of the program provides an instrument to accomplish the reverse mapping between Γ and the variable element values, which can be collected in the vector $\underline{V} = (\underline{L}, \underline{C}, \underline{l})^{\mathrm{T}}$. From the *ABCD* matrix a nonlinear complex valued equation is deduced to calculate $\underline{\Gamma}(\underline{V})$ at different frequencies. Specifying the reflection coefficient at fundamental and harmonic frequencies ($\underline{\Gamma}_{f_i}$) results in a nonlinear equation system, which can be numerically solved for $\underline{V}$. In matrix notation this is written as:

$$\underline{\Gamma}(\underline{V}) = \underline{\Gamma}_{f_i} \Longrightarrow \underline{V} = \begin{pmatrix} \underline{L}(\Gamma_{f_0}, \Gamma_{f_1}, \ldots, \Gamma_{f_n}) \\ \underline{C}(\Gamma_{f_0}, \Gamma_{f_1}, \ldots, \Gamma_{f_n}) \\ \underline{l}(\Gamma_{f_0}, \Gamma_{f_1}, \ldots, \Gamma_{f_n}) \end{pmatrix}, \text{ with } \Gamma_{f_i} = \begin{pmatrix} \Gamma_{f_0} \\ \Gamma_{f_1} \\ \vdots \\ \Gamma_{f_n} \end{pmatrix} \tag{4.10}$$

It has to be assumed that the topology of the network and the Γ_{f_i} values are appropriate to find a solution (i.e. the roots of the nonlinear equation system exist). Under this assumption, to obtain unique results, the number of specified complex valued Γ_{f_i} has to equal half the number of real valued variables in $\underline{V}$. The definitions in the program therefore depend on the number of tuneable elements in the network. As an example Fig. 4.25 shows the output of the program for a network with 4 tuneable elements. The solutions are calculated for a certain range of Γ_{f_0} and fixed Γ_{f_1}. The algorithm to solve the nonlinear system of equations is a variant of the Powell dogleg method described in [Powell 70, Moré 80]. This procedure is already implemented in Matlab, which reduces efforts in programming. The results are plotted as contours in a certain range of Γ_{f_0} to see the sensitivity to variations. Moreover, appropriate values out of the set of realizable and available component values can be found more easily.

Field of Application

The presented SMD Load Pulling technique can be understood as a convenient practical optimization method but does not claim to accurately measure the DUT, as done with conventional load pull systems. Against this background the fields of application can be summarized as follows:

- Optimization of arbitrary characteristics of the DUT versus load impedance variations under the final layout conditions. All parasitic effects such as resistive/inductive GND connections, bond wires, losses, harmonic terminations,... are included in the process.
- No constraints on the topology of the output matching, as long as it is tuneable and a model can be developed. The accuracy of the optimization depends on the model. This is the bottleneck of the procedure, but is not a drawback, compared to conventional computer aided optimization approaches, using behavioral or scalable transistor models. The output matching network has to be modeled there as well, with similar accuracy.
- Optimization of balanced/push-pull amplifiers with matching networks between the DUT and the hybrid/BALUN. Influence of unbalance, cross coupling, poor isolation,... are included, without need of extra test structures to measure the hybrid/BALUN.
- Optimization of the DUT performance versus input impedance variations. Thus, the gain of a PA can be optimized. One can also think about application to LNA design in order to improve the noise figure.

With SMD Load Pulling a DUT can be optimized by relative measurements. This means, that the role of the output impedance as absolute value becomes irrelevant. The calculated contours can be but don't have to be at the position of the real output impedance, seen by the DUT. As long as the shape of the contours remains the same. The calculated output impedance is then only a variable, describing tendencies, when changing matching network parameters. This way the influence of e.g. bond wires becomes smaller, as a certain inductance is assumed and held constant throughout the whole relative measurement process, causing only a shift of the contours in the smith chart.

4.4 Realization of the double balanced switched path PA

This chapter presents the final realization of the multi-mode PA, based on the concept introduced in Chapter 3.6. The primary target of the realization was to proof feasibility in a low cost bipolar technology (presented in Chapter 4.1) and to identify possible technological hurdles, when used for transmission of GSM/EDGE/WCDMA signals. The theoretical considerations on the various trade-offs between the different operation modes could be verified by selective experiments.

4.4.1 Design overview

The typical specifications for cellular PAs are summarized in Table 2.1. The double balanced multi-mode PA was designed with a focus on following items:

- Achievement of a maximum output power of 36 dBm in the frequency band 824 . . . 915 MHz and for a supply voltage of 3.5 V.
- Implementation of PA power back-off states, which provide an efficiency improvement as large as possible, at the maximum GSM (36 dBm), EDGE (30 dBm) and WCDMA (27 dBm/17 dBm) transmit power levels.
- Fulfillment of linearity requirements (ACLR/Mod.Spec. and EVM) for GSM, EDGE and WCDMA.
- No specifications on gain of the PA. Intentionally only the final power stage is implemented, as this has most influence on efficiency and linearity.
- Proof of improved load insensitivity due to the double balanced design.
- Flexibility of the output matching circuitry to tune the PA on the PCB with the SMD load pulling procedure.

The complete circuit schematic of the implemented PA is shown in Fig. 4.27. The arrangement of blocks corresponds to the arrangement in the chip and PCB layout (Fig. 4.37 and Fig. 4.38). Due to the high complexity of the architecture, the PA was divided into sub-circuits (X1, X2, . . .) for the design process. This allows a better understanding and faster optimization of the individual PA blocks. The partition of the PA for simulation was done on basis of the circuit symmetry. Two main blocks can be identified, which identically repeat several times in parallel, namely the high power path (HP1, HP2, HP3, HP4) and the low power path (LP1, LP2). These two paths were optimized separately and finally combined with the quadrature hybrids, described in Chapter 4.2.

The final top-level simulations of the complete double balanced amplifier were restricted to small signal S-parameter simulations to identify instabilities. The large

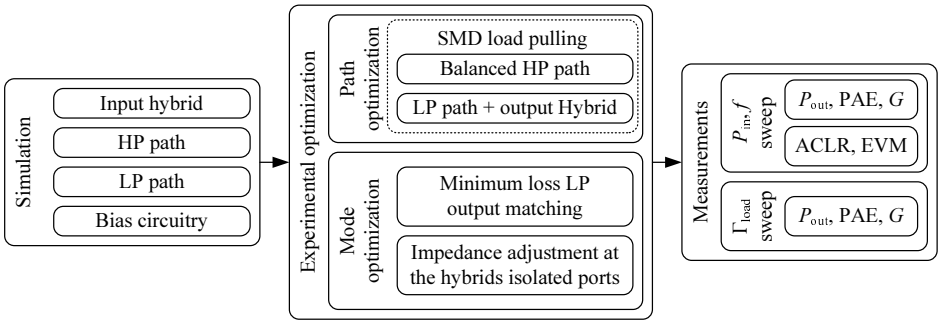


Figure 4.26: PA design flow chart, describing partition into simulation, experimental optimization and measurements.

sinal analysis (input power sweep) with the harmonic balance simulator did not converge. The reason for this bad convergence was identified to be the self biasing of turned OFF amplifiers in the various operation modes. For example for the high power operation in GSM mode all HP amplifiers were turned on and all LP amplifiers turned off via bias. Due to imperfect isolation of the hybrid a small fraction of the input power coupled to the first LP transistor and caused very small self-bias currents ($< 100\ \mu$A). To confirm this as reason for the bad convergence, a separate input power sweep simulation of the single LP path with turned OFF bias was done. The same convergence problem could be shown. Therefore the final optimization was done with focus on experimental tuning on the PCB, which is based on the theory developed in Chapter 3.5, Chapter 3.6 and Chapter 4.3. Fig. 4.26 summarizes the partition of the design process into simulation, experimental optimization and final measurements.

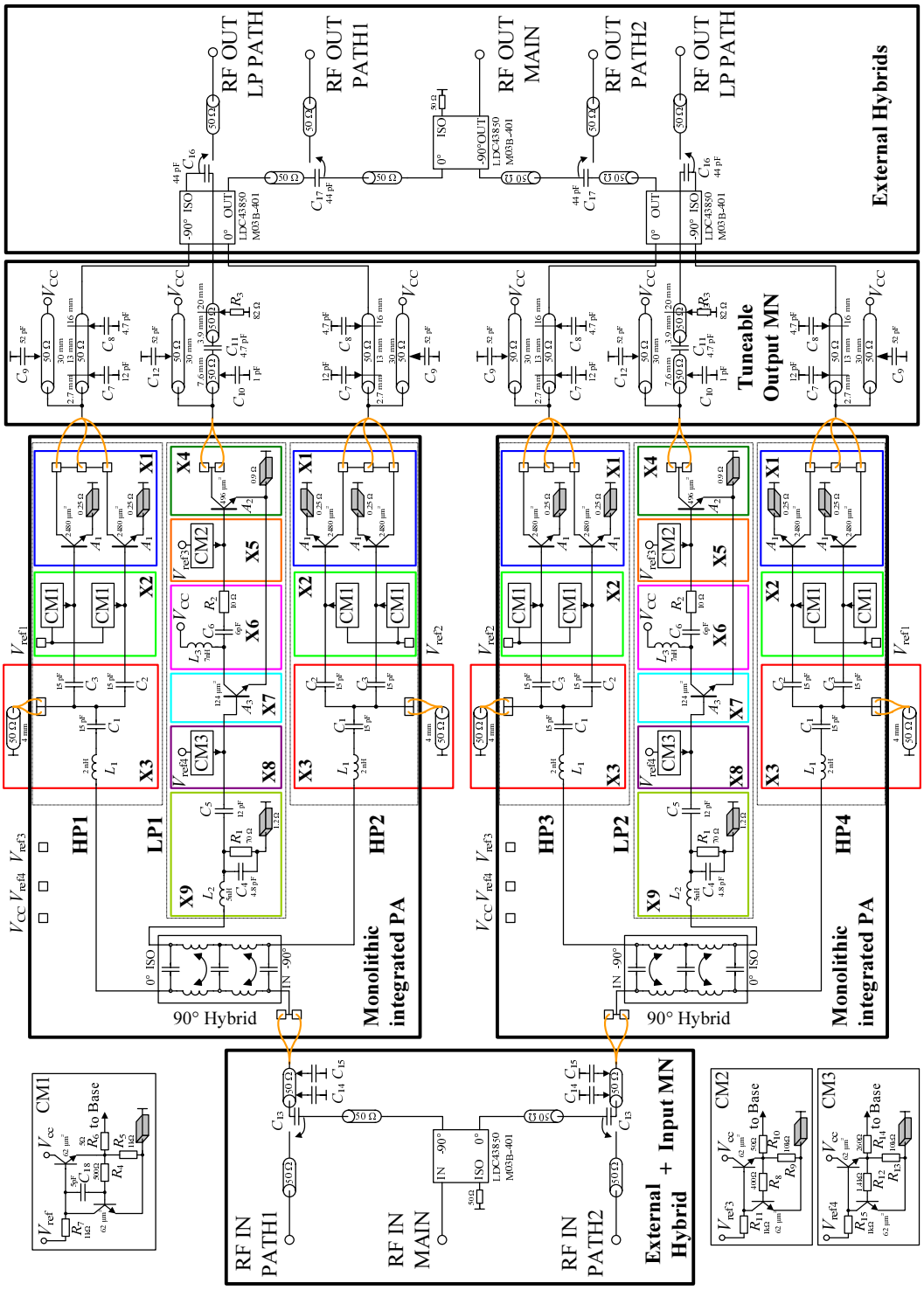


Figure 4.27: Circuit Schematic of the overall double balanced power amplifier, including chip and PCB circuitry.

4.4.2 Simulations

In the following the basic steps and simulation results in the design of the high power paths (HP1, HP2, HP3, HP4) and low power paths (LP1, LP2) are presented.

HP path design

The design of the four identical high power paths requires definitions of their input and output interface:

- Output power: The output power, delivered by each HP path is necessary for estimation of the transistor area. The saturated output power of the overall amplifier has to be 36 dBm at a 50 Ω load with a 3.5 V supply. The double balanced architecture is basically a binary cascaded combiner structure. Therefore, the losses added by the quadrature couplers [Murata 08] (see Fig. 4.9) can be used to recalculate to the output power of each HP path, using equation (3.30). Typical losses of the output matching network were assumed to be 0.5 dB. The resulting estimated power, which has to be delivered from each HP transistor is 31.1 dBm. The complete level plan is shown in Fig. 4.28.

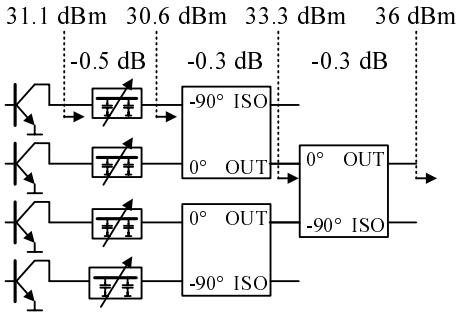


Figure 4.28: Output power level plan for estimation of the HP transistor size.

- Input source impedance: The standard reference impedance for laboratory RF measurement equipment is 50 Ω, hence the amplifier module has to be matched to this impedance at the reference plane of the PCB connectors. This matching is done in two steps, as shown in Fig. 4.29. On-chip the small input impedance (0.5 . . . 2 Ω) of the transistor is transformed to 22 Ω, which is also the design impedance of the monolithically integrated input hybrid. This impedance was chosen for two reasons. First a small impedance transformation ratio allows integration of broadband and low loss matching networks. Second, an optional pre-amplifier, operated in class A or AB, can directly deliver about $P_{\text{preamp}} = 3.5^2/(2 \cdot 22)$ W = 24 dBm to this load, without additional interstage matching. As the PA is intended to be measured without this pre-amplifier, a second tuneable off-chip matching network to 50 Ω was necessary.

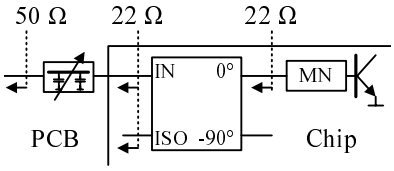


Figure 4.29: Input impedances for chip and PCB.

Power transistor dimensions

With knowledge of the environment definitions, the first step in the design process is the selection of a suitable transistor size for the power stage X1 in Fig. 4.27. This is an iterative process, showing a strong dependence on the layout configuration, the biasing conditions and the input and output impedance. The transistor size is usually characterized by its effective emitter area. As a starting point for the simulation the emitter area can be calculated using equation (4.2). This includes the dependence between output power (31.1 dBm), supply voltage (3.5 V), efficiency ($\sim$ 50%), bias current ($\sim$ 100 mA) and current density at maximum f_{T} (0.28 mA/μm^2). With these values, the starting value for the emitter area equates to 3930 μm^2. Next, the parasitics originated in the layout and thermal effects are taken into account by performing a load pull analysis. Therefore the scalable transistor model, including self-heating effects, is embedded by scalable parasitics of the base, emitter and collector contacts, including wire resistance and coupling capacitances between each other and to substrate. This modeling approach is shown in Fig. 4.4 and Fig. 4.5. The performed load pull analysis is an iterative process, depending on following issues:

- The feedback capacitance C_{BC} causes dependence between the optimum input and output impedance. Different input source impedances lead to significant differences in load contours and maximum output power and efficiency levels. Increasing the area increases the feedback and therefore the dependence.

- The selected bias current influences the PA key parameters (output power, PAE, gain and linearity) and the optimum input and output impedance. A small bias current increases PAE, but decreases output power, gain and linearity. Variation of the transistor area significantly changes the absolute value of the optimum bias current, which has to be traced for each iteration.

- Proper tuning of harmonic terminations influences the achievable performance. While output power shows minor dependence ($\pm$0.5 dB), the PAE can be increased significantly (5 . . . 15% points). The area dependent collector emitter capacitance C_{CE} of the transistor represents a low impedance shunt, especially at higher harmonic frequencies. For high efficient Class F operation all even order harmonics have to be shunted and all odd order harmonics have to be presented open circuit values at the intrinsic collector port. Compensation of C_{CE} for optimizing the odd order harmonic termination with an output resonance circuit (inductance to ground) reduces the operation bandwidth. For maintaining the PA performance over a certain bandwidth the transistor area is upper bounded by the C_{CE} impact.

- To reduce return loss at the input and increase the gain a conjugate match to the area dependent transistor input impedance is desired. For large transistor areas this input impedance can become very small (0.5 Ω) and may not be realizable, taking into account resistive losses of the metallization for the base connection. Moreover, instability is more likely to occur for a large transistor

(large C_{BC}). Stabilization is then done on the input plane, with a resistor, which consequently reduces the overall gain. To avoid instabilities during load pull simulations a source impedance of 5 Ω was chosen.

These reflections show that there are a lot of trade-offs to be considered for the optimum choice of the emitter area. On top of that actions have to be taken for a homogenous temperature distribution along the transistor cell. The HICUM transistor model includes self heating effects of the unity cells, but no mutual heating of adjacent cells, which leads to inhomogeneous temperature distribution. Evidently, the temperature difference between subcells at the edges and in the center increases with the length of the transistor array. Partition into several shorter arrays reduces the temperature difference. For this design the transistor was partitioned into two columns with each 2480 μm^2 total emitter area. Both columns consist of 20 subcells, each having an emitter area of 124 μm^2. Fig. 4.30 shows the layout configuration of the transistor, including three pad openings for bonding and the emitter ground contact through the sinker lying underneath.

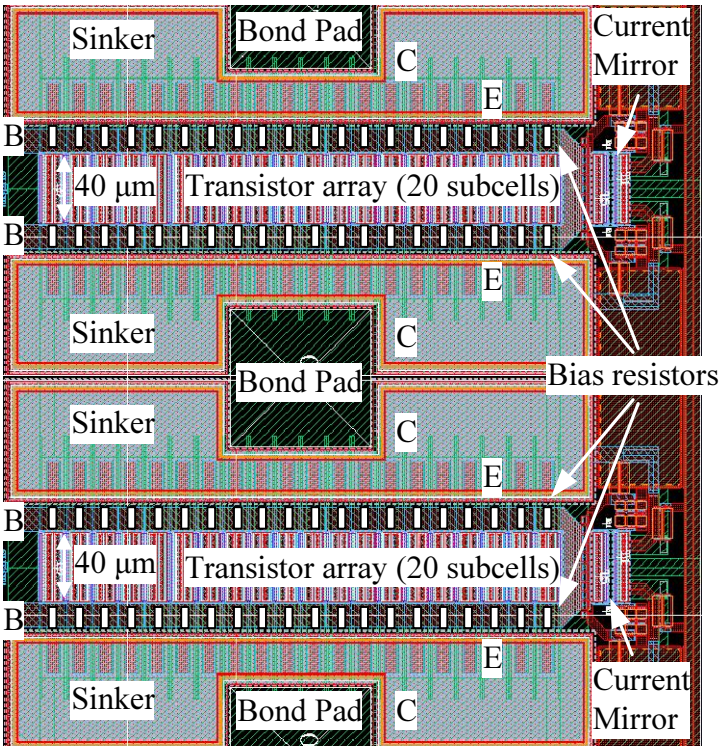


Figure 4.30: High Power transistor cell layout including bond pads, bias and sinker ground contact.

Output matching

Having found an appropriate transistor size and layout, the output matching network is realized with lumped components, including bond wire inductance (0.4 nH) and pad capacitance (120 fF). Small signal S-parameter simulations are performed to

realize the impedances at the fundamental $\Gamma_{f_0} = 0.7|172°$, the second harmonic $\Gamma_{f_1} = 0.97|146°$ and the third harmonic $\Gamma_{f_2} = 0.97|32°$, obtained from the load pull simulation. Once the component values are determined the matching network is further optimized to obtain the desired output power ±0.5 dB tolerance at the two band edges and at the center frequency. For that, power and frequency sweeps of the matching network together with the transistor are performed and the component values optimized with a numerical solver. The final characteristic of the output matching network is shown in Fig. 4.31. The attenuation of the network in the operation band is about 0.39 dB, slightly better than estimated at the beginning.

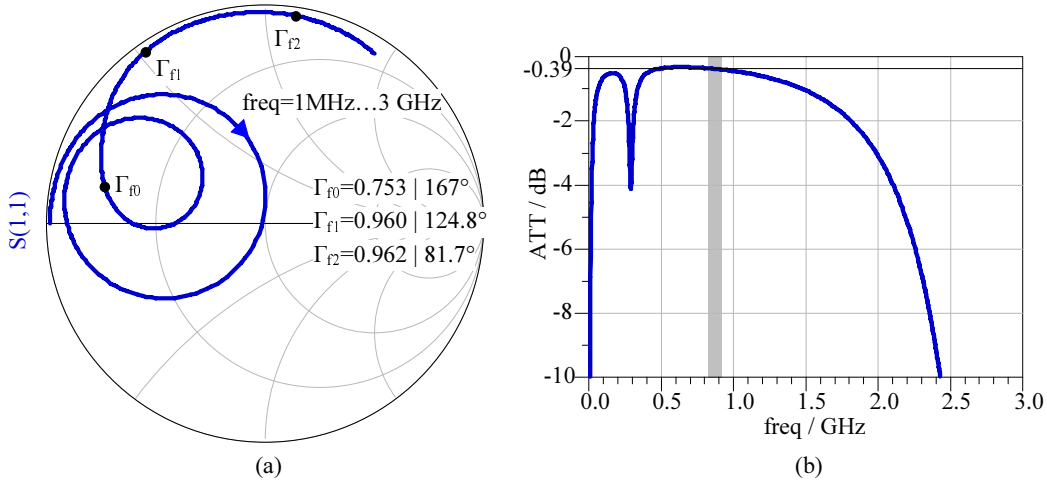


Figure 4.31: Output matching HP path: (a) input reflection coefficient at fundamental and harmonic frequencies, (b) attenuation (maximum available gain).

Bias circuit

The PA operating class is strongly related to the bias point selected in the previous simulations. Providing reliable DC bias is a crucial step in the design process, as any change in the bias point can result in moving into an undesirable region of operation from the point of view of efficiency, linearity or stability. The basic requirements for the transistor bias circuitry of the HP stage are:

- Providing a DC bias current, adjustable with an external voltage.
- Bias circuit current consumption $< 1/10 \ldots 1/100$ times the HP stage DC current.
- Isolation between DC and RF signal.
- Insensitivity to temperature variations.
- Maintaining the self biasing effect for large input drive levels.

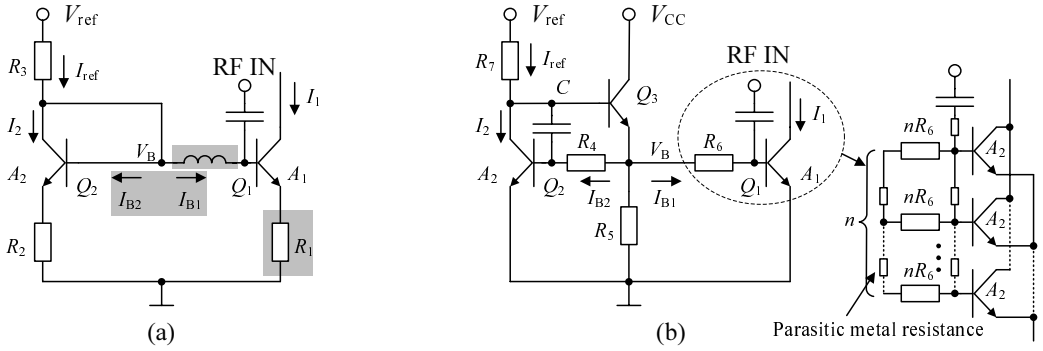


Figure 4.32: Bias circuit: (a) Basic resistor-ratioed current mirror, (b) base current compensated current mirror.

Fig. 4.32 (a) shows a conventional current mirror. The mirror ratio I_1/I_2 is determined by the ratio of the emitter resistors R_2/R_1, known as ballasting resistors. The ratio between their values should be equal to the transistor area ratio A_1/A_2, to improve matching. The larger the voltage drop over R_1 ($> V_T \sim 25$ mV), the more the mirror ratio is determined only by the resistors, why this circuit is also called a resistor-ratioed current mirror. The absolute current value is adjusted by the diode-connected transistor Q_2 together with the reference voltage V_{ref} and the resistors R_2 and R_3: $I_{ref} = \frac{V_{ref}-V_{BE}}{R_2+R_3}$. This applies as long as the base current I_{B1} is negligible. This is often the case in small-signal amplifier circuits. For power amplifiers, this kind of current mirror has at least three basic drawbacks. *First*, for a large power transistor the base current can be in the order of I_{ref} (several mA), substantially influencing the current ratio. *Second*, the ballast resistor R_1 (several Ω) degrades the overall Gain and efficiency of the PA. *Third*, the RF bias choke consumes a lot of chip area and its series resistance distorts the current mirror ratio.

To avoid these drawbacks, an improved current mirror was used, more suited to provide a bias current to the large power transistor. The basic circuit is shown in Fig. 4.32 (b). In difference to the conventional current mirror the transistor Q_3 is added and the ballasting resistors are moved to the base contacts. As the collector currents $I_1 \sim I_{ref}$ and I_2 are proportional to the base currents by the current amplification β, the current mirror ratio is in the first approximation determined by the two base resistors $I_1/I_2 = R_4/R_6$. The transistor Q_3 stabilizes the mirror ratio, as it supplies the base currents I_{b1} and I_{b2}. Only the base current of Q_3 is subtracted from I_{ref}. As Q_3 operates at much lower current levels than Q_1 and Q_2, an additional resistor R_5 is connected to ground, to set a stable bias point for Q_3. Moreover for pulsed supply operation, as in GSM, R_5 increases the bias switching speed. The base charge carriers of Q_1, originated in C_{BC}, have a path to ground and therefore the base voltage V_B approaches zero, much faster, when switched off via V_{ref}.

Another feature of this bias circuit is that there is no need for an additional RF decoupling choke. The resistor R_6 reduces RF leakage into the current mirror, when chosen larger than the input impedance (0.5 Ω) of the power transistor Q_1. Choosing

a value for R_6 is the most critical task in the bias circuit design, as it influences following PA properties:

- Self biasing: With increasing RF base drive level at Q_1 the transistor enters its nonlinear region. The signal is rectified at the base emitter diode, which leads to a desired increase of the DC collector current I_1 and also I_{B1}. This has a positive effect on output power, gain and PAE. In presence of the current mirror the voltage V_B is fixed and determined by V_{ref}, R_7, R_4 and the two base emitter voltages of Q_2 and Q_3. As the current I_{B1} increases, the voltage across R_6 increases as well. Thus, the base emitter voltage $V_{BE,1}$ of Q_1 will decrease, to keep V_B constant. As a result, this counteracts a further increase of I_1 and suppresses the self biasing. The amount of this negative feedback depends on the resistor value R_6. Smaller values cause smaller negative feedback. A lower bound for the resistor value is given by the RF leakage into the current mirror.

- Spurious oscillation: As the bias point of the transistor varies with the RF drive level, also the RF properties change. This can be observed by the varying input and output impedance, mainly caused by the changing intrinsic capacitances. Hence, at certain drive levels the oscillation condition [Gonzales 84] may be fulfilled and the transistor becomes unstable. Suppressing the self biasing by appropriate choice of R_6 can avoid entering these unstable bias points [Vidkjaer 78]. In respect of this concern large R_6 values have to be chosen.

- Linearity: The amount of self biasing has influence on the AM-PM characteristic of the amplifier, depending on the class of operation (A, B, AB,...). Proper choice of R_6 can improve the linearity of the PA.

- Noise: The resistor R_6 is a potential source of thermal noise power. This power is directly proportional to the resistor value. As the resistor is directly connected to the base contact, its noise will be directly amplified by Q_1, without additional filtering. If the resistor value is chosen too large, this may lead to undesired noise power in the receive band of the transceiver. For GSM the noise power at 20 MHz (neighbor channel) offset from the transmit carrier is specified by < -79 dBm/100 kHz.

These observations show that determination of an appropriate bias circuit demands, besides static DC simulations, also large signal, stability and linearity analysis of the complete circuit. Also the layout has to be considered to reduce temperature dependence. The transistors Q_1 and Q_2 have to be placed together as close as possible to bring them on the same junction temperature. Thus, a temperature dependent drop of their base emitter on-voltage $V_{BE,on}$ can be compensated in the current mirror. The temperature of Q_1 tends to increase (with drive level) in the center of the array. A placement of Q_2 in the center would therefore be beneficial to avoid positive temperature feedback, but leads to cross talk between RF and DC interconnect lines, causing ripple at the on-chip supply line to V_{ref}. To reduce cross talk Q_2 is placed at the edge of the array and additionally blocked by the capacitor C. Another layout

issue is the increasing wiring resistance of the Q_1 base contact along the array, causing additional voltage drop and unequal bias points of the base cells. As shown in Fig. 4.32 (b) this influence can be reduced by partition of R_6 into n resistors at each base cell. The metal resistance is thus negligible, compared to the new resistor value nR_6. The layout configuration is shown in Fig. 4.30.

Tuneable input matching circuit

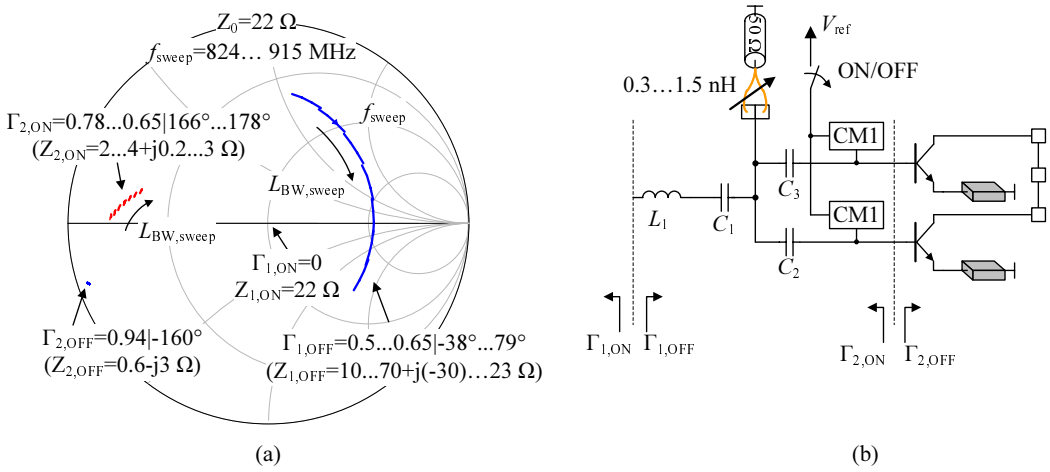


Figure 4.33: HP path input matching network (bond wire tuning range $L_{\mathrm{BW,sweep}} = 0.3 \ldots 1.5$ nH): (a) Input and output reflection coefficients in transistor ON and OFF state, (b) circuit schemtaic.

The final step is the design of an input matching network (X3 in Fig. 4.27). As shown in Fig. 4.33 (b), it has to transform the 22 Ω source impedance of the quadrature hybrid $\Gamma_{1,\mathrm{ON}}$ to an appropriate value at the base of the HP transistor $\Gamma_{2,\mathrm{ON}}$. A stability analysis has shown that a conjugate match leads to oscillation. A source impedance of $Z_{2,\mathrm{ON}} = 2\ \Omega$ at the transistor base was found to avoid this instability. The matching network also has to maintain high reflections $\Gamma_{1,\mathrm{OFF}}$, when the transistor is turned OFF, to increase the power transfer from the input through the hybrid to the LP path. Thus, its attenuation has to be as low as possible.

The high pass topology of Fig. 4.33 (b) was chosen. The inductance to ground is realized with a bond wire and an external shunted 50 Ω transmission line. The advantage over an on-chip spiral inductor is the larger quality factor (100) and the smaller chip area (60×60 μm bond wire pad). With the number of bond wires and variation of their length, the inductance can be tuned in the range of $0.3 \ldots 1.5$ nH. The resulting simulated impedance variations are shown in Fig. 4.33 (a). The real part of $Z_{2,\mathrm{ON}}$ is tuneable between $2 \ldots 4\ \Omega$, hence in the stable operation range of the transistor. In OFF state, the input reflection coefficient $|\Gamma_{1,\mathrm{OFF}}| = 0.5 \ldots 0.65$ causes gain reduction in LP operation mode by $-3.7 \ldots -6$ dB (see (3.54)). The overall attenuation of the network (maximum available gain) is about -2.3 dB.

Power Sweep

The final simulated performance of the HP path is shown in Fig. 4.34. For linearity analysis the AM-AM and AM-PM characteristics were simulated. As rule of thumb the AM-PM phase deviation has to be below 5° at the −1 dB compression point to fulfill ACLR and EVM specifications. This is fulfilled over the whole frequency band.

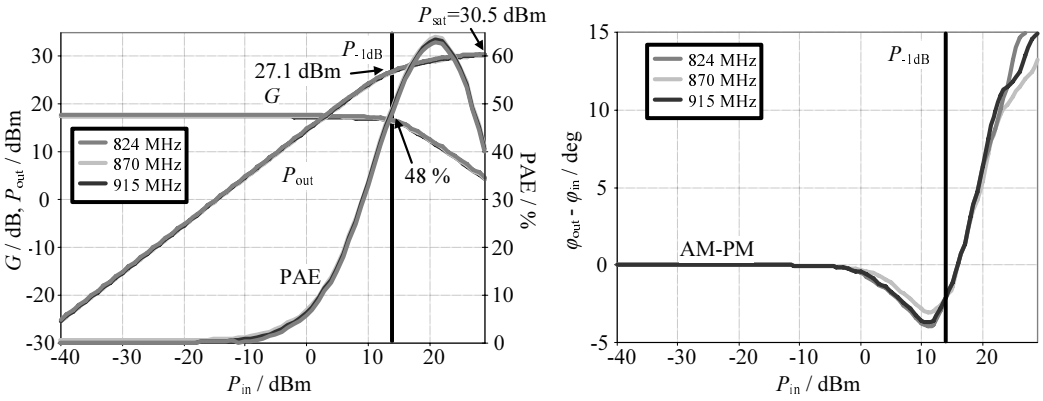


Figure 4.34: Input power sweep simulation results of the HP path. Quiescent current $I_q = 61$ mA

LP path design

The design process of the two identical LP paths is similar to that of the HP path. The input and output interface definitions are given for the ON and OFF state. In ON state the LP path has to deliver about 16 dBm *linear* output power (WCDMA in back-off). In OFF state it has to be matched to the impedances of the hybrids (22 Ω input, 50 Ω output), with lowest possible return loss. A two stage design (driver stage and output stage) was chosen to increase the gain of the LP path and compensate for losses (1 . . . 2 dB) of the input and output matching networks, necessary to fulfill the ON and OFF state conditions (see Chapter 3.5.2).

Output stage design

To estimate the size of the output transistor X4 in Fig. 4.27, the saturated output power was supposed ∼ 4 dB larger than the linear output power (13 dBm for each LP path) to fulfill the ACLR requirements for WCDMA. Losses due to the output matching network and the output hybrid were considered with 3 dB. Hence, the transistor has to deliver at least (13+4+3) dBm = 20 dBm saturated output power. Scaling the transistor size of the HP amplifier to these output power requirements results in a minimum area of (4960 μm^2 · 0.1 W/1.288 W) = 385 μm^2 for the LP output stage. As the area can only be scaled in discrete steps with the unity cell area (124 μm^2), the resulting final transistor area is 496 μm^2.

A harmonic load pull simulation was done to find the optimum load reflection coefficients ($\Gamma_{f_0} = 0.288|87°$, $\Gamma_{f_1} = 0.98|6°$, $\Gamma_{f_2} = 0.99|174°$; PAE = 72 % and $P_{\text{out}} = 22$ dBm). In a first step the output matching network was realized by a low loss Π-network topology, similar as for the HP path. The final realization of the network (including losses to fulfill matching conditions in OFF state) can only be determined with knowledge of the LP amplifier's output impedance in OFF state. As this changes with the transistor's input impedance, this requires prior design of the driver stage.

Driver stage design

The driver stage includes the transistor X7 together with the interstage matching network X6 (Fig. 4.27). The transistor X7 has an area of 124 μm^2 (one unity cell), which is sufficient to drive the output stage. The interstage matching network has a high pass structure to avoid low frequency oscillations. The capacitor C_6 and the inductor L_3 are also used to isolate the DC bias of the two stages. The resistor R_2 at the base of X4 further improves stability of the amplifier. Moreover, by using a resistor as matching element, the size of the overall network is reduced. The dissipation losses (maximum available gain) of the interstage matching network are −2.2 dB, hence tolerable, compared to $\sim$ 15 dB maximum stable small signal gain of each of the transistors X4 and X7. The element values of the interstage matching network, together with the biasing of the first and second stage were tuned to have minimum AM-PM distortion at the 1 dB compression point.

Input and output matching networks

As last step in the design process the LP path core - consisting of driver transistor, interstage matching network and output transistor - has to be matched with passive low loss networks at the input and the output. The target reflection coefficients in transistor ON and OFF state are shown in Fig. 4.35. In Chapter 3.5.2 it is shown that such impedance transformations can be realized with passive matching networks, only if they contain a certain amount of dissipative loss. The method, developed there to find a network with minimum loss, is used here.

At the *input* the matching network has to maximize the LP path gain in ON state, while maintaining a stable operation. With help of the input stability circles [Gonzales 84] $\Gamma'_{2,\text{ON}} = 0.43| - 54°$ is found which has to be realized from $\Gamma'_{1,\text{ON}} = 0.01| - 180°$ (note that the reference impedance is 22 Ω). The same network has to provide minimum return loss ($|\Gamma'_{1,\text{OFF}}| \leq 0.2$) at the hybrid, if the LP path is OFF ($\Gamma'_{2,\text{OFF}} = 0.96| - 12.4°$). To find a network fulfilling these conditions with minimum dissipative losses the maximum available gain together with the realizable solutions for a T-topology are plotted (Fig. 4.35). For several S_{21} the S-matrix and the component values were calculated and compared regarding size and bandwidth. Finally a network with $G_{\text{MAG}} = 0.5$ was implemented. The losses are realized by the resistor R_1 and the metal resistance (4 Ω) of the coil L_2. The series capacitor C_5 avoids DC currents from the driver stage bias to flow through the resistor.

At the *output* the load ($\Gamma_{1,\text{ON}} = 0.288|87°$) for maximum output power and PAE has to be realized in ON state. In OFF state the transistor's output reflection coefficient

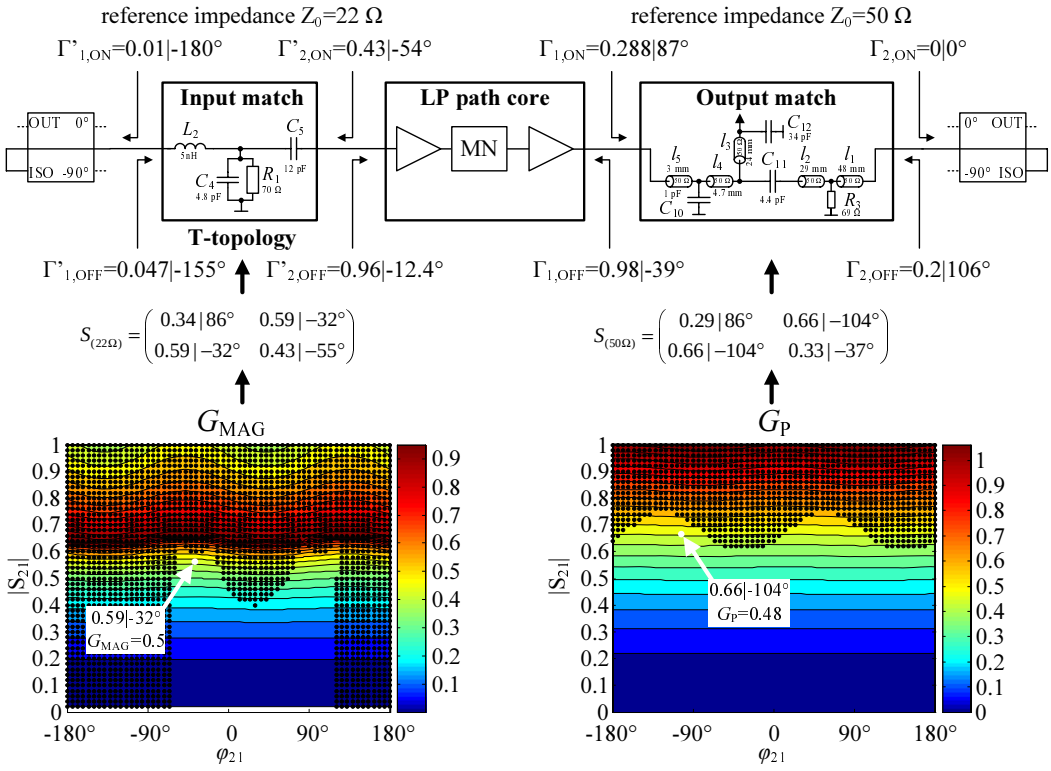


Figure 4.35: Passive minimum loss input and output matching network synthesis for impedance transformation in ON and OFF state.

($\Gamma_{1,\mathrm{OFF}} = 0.98|-39°$) has to be transformed to $|\Gamma_{2,\mathrm{OFF}}| \leq 0.2$ with the same matching network (note that the reference impedance is 50 Ω). Fig. 4.35 shows the theoretical upper limit of power gain G_P for realizable passive networks, which fulfill these matching conditions. The finally realized network basically represents a Π-topology with an additional V_CC feed line. The difficulty here was to find a network, realizable with the available values out of the capacitor series [Epcos 08]. After some iterations the presented network with $G_\mathrm{P} = 0.48$ was found, being very close to the theoretical limit. The dissipative element is the resistor R_3, which is isolated from V_CC by the series capacitor C_{11}.

Power Sweep

Fig. 4.36 displays the simulated power sweep. The maximum saturated output power is about 19 dBm. This is 2 dB larger than estimated at the beginning, mainly because of the larger transistor size and smaller layout parasitics. The PAE is lower compared with the results of the load pull simulations. This is due to the additional driver stage, losses of the output network and different termination of harmonics. Still 31 % are obtained at the −1 dB compression point $P_{-1\,\mathrm{dB}} = 17$ dBm. The AM-PM simulation shows a phase distortion larger than 5° at $P_{-1\,\mathrm{dB}}$. Nevertheless, the usable linear operation range in power back-off will be below 16 dBm, hence below $P_{-1\,\mathrm{dB}}$, where the linearity is sufficient.

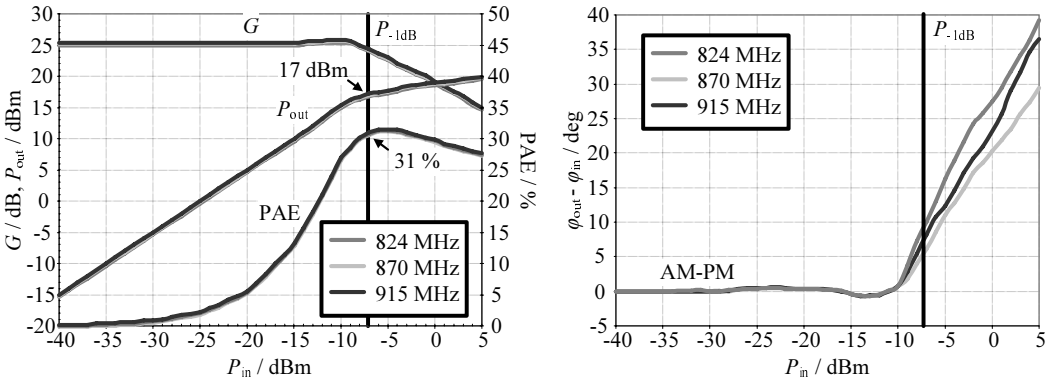


Figure 4.36: Input power sweep simulation results of the LP path. Quiescent current $I_q = 13$ mA

Chip layout considerations

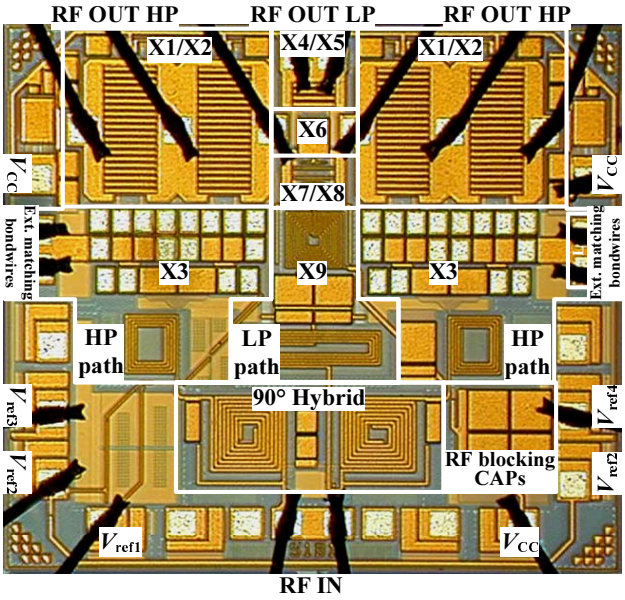


Figure 4.37: Die photograph of the monolithic integrated balanced PA (size 1.49 × 1.32 mm^2).

The chip layout of the monolithic integrated balanced PA is shown in Fig. 4.37. Crosstalk between the 0° and 90° HP paths degrades performance with respect to output power, linearity and load sensitivity, thus, the LP path is placed between the mirror-imaged HP paths. RF input and output pads are placed at opposite chip edges to reduce coupling through bond wires. The V_{CC} supply line of the LP path driver stage is blocked with MIM and MIS capacitors, filling the space besides the 90° hybrid. The remaining biasing lines contain no blocking capacitors, as they would reduce switching speed in pulsed operation. To obtain a symmetrical board

layout of the complete double balanced PA the same chip is fabricated in a mirrored version.

PCB layout considerations

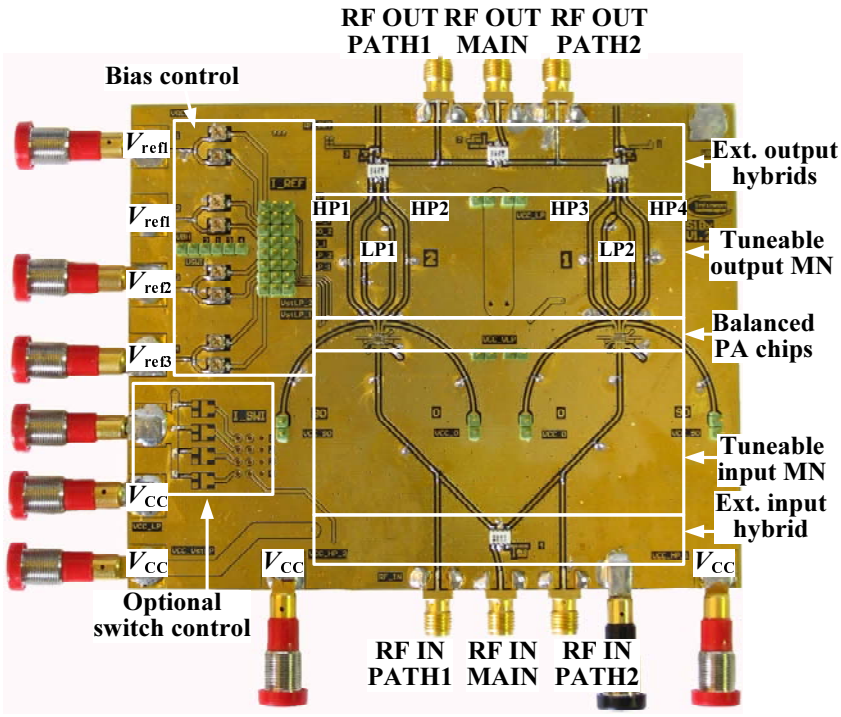


Figure 4.38: Evaluation board (PCB) photograph of the double balanced multi-mode PA (size 13×10 cm^2).

Fig. 4.38 shows the test board for experimental optimization and measurements. For reduced coupling between RF and DC lines it offers three 35 μm thick copper metal layers for wiring. The bottom layer is mainly used for DC bias and supply feeds, the middle layer provides a common RF ground plane and the top layer contains RF transmission lines (6 μm nickel diffusion barrier and 120 nm gold plated for bonding). An epoxy substrate (FR-4) with $\varepsilon_r \cong 4.4$ and $\tan(\delta) \cong 0.038$ is used as dielectric material. The height of the top dielectric layer was chosen 0.2 mm. This way a gap of 0.6 mm between the ground plane and the transmission line (width=0.4 mm) is small enough to solder and shift the SMD capacitors (0402. . .length=1 mm) for tuning purposes (note that the influence of the top ground plane is small so that the transmission line is a microstrip waveguide rather than a grounded coplanar waveguide).

The balanced PA chips with backside metallization are soldered as bare die on the top ground plane and connected with aluminium bond wires (diameter=25 μm). The lengths of the transmission lines at input (50 mm) and output (35 mm) leave enough room for tuning. The PCB features the possibility of measuring the balanced HP and

single ended LP path separately, bypassing the first input and last output hybrid. This is achieved by applying the input and output series decoupling capacitors C_{13}, C_{16} and C_{17} (Fig. 4.27) in a way to support the intended signal transmission. Hence, the PATH1 and PATH2 input/output can optionally be connected to the hybrid or directly to the SMA connector.

The bias control, implemented at the left side of the PCB, consists of potentiometers for adjustment of bias currents and jumpers to set the operation modes. An optional control (potentiometers and jumpers) for series switches at the isolated ports of the hybrids is implemented as well, but not used here.

4.4.3 Experimental optimization

With reference to Fig. 4.26, the next step in the PA design is the optimization of the assembled amplifier module. The basic actions are explained in the following.

Path optimization

The performance of PATH1/PATH2 (no double balanced operation) is optimized separately in high power and low power mode with help of the SMD load pulling procedure (see Chapter 4.3). The intention is to find output matching networks for HP path and LP path, which provide the best compromise between P_{out}, PAE and Gain. In contrast to the simulations the impacts of the on chip input hybrid and off chip output hybrid are included.

High power path

The paths HP1 and HP2 (Fig. 4.27) are combined with input and output hybrid, while the path LP1 is turned OFF via bias. The output of LP1 and the isolated port of the output hybrid are terminated with 50 Ω. The SMD load pull results, measured between RF IN PATH1 and RF OUT PATH1 are depicted in Fig. 4.39 (a). The assigned Γ values are the reflection coefficients seen by one amplifier at the reference plane between the on chip bonding pads and the bond wires. From these contours the best performance was found at an output reflection coefficient $\Gamma_{\text{opt}} = 0.73|170°$, which corresponds to the matching network shown in Fig. 4.39 (b). Compared to the results of the simulation ($\Gamma_{\text{opt}} = 0.753|167°$, from Fig. 4.31), the difference is marginal and can be traced to the tolerances of the bond wires and the lumped components. This output network is further applied to each HP amplifier in PATH2.

Low power path

For this measurement the HP paths are turned OFF via bias and the LP1 path is connected between the isolated port of the input and output hybrid. The performance is measured between RF IN PATH1 and RF OUT PATH1. The SMD load pulling procedure is done with the topology depicted in Fig. 4.41. Losses due to the hybrid at the input ($\sim$ 4 dB) and the output ($\sim$ 1.5 dB) are included in this measurement. The resulting contour plots are shown in Fig. 4.40. To avoid uncertainties in the calculated Γ values due to the bond wire inductance, the reference

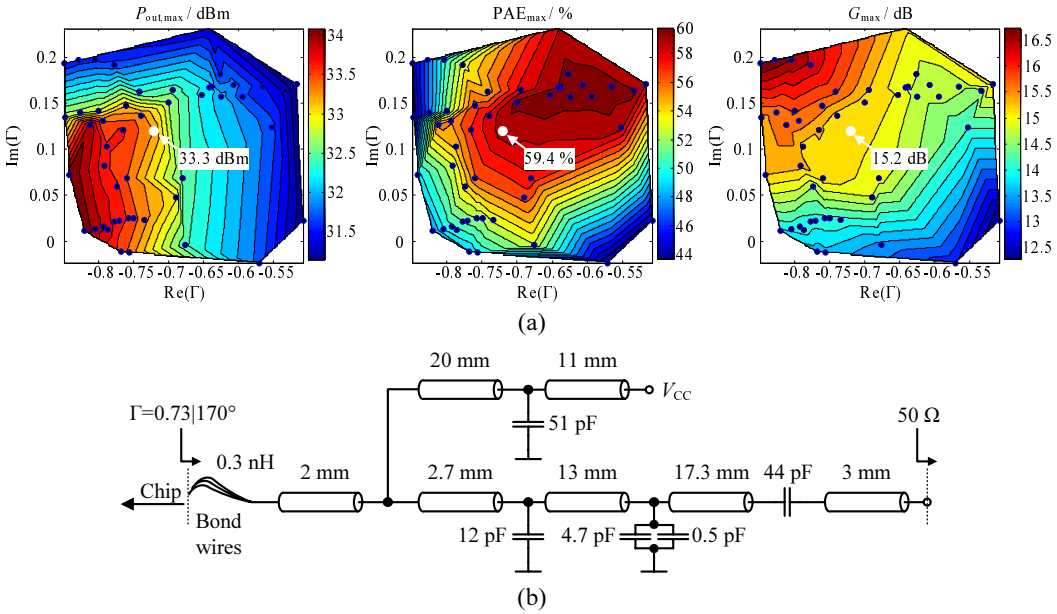


Figure 4.39: (a) Balanced high power path SMD load pull results including input and output hybrid ($f = 840$ MHz and $V_{CC} = 3.5$ V). (b) Output matching network for finally chosen $\Gamma_{opt} = 0.73|170°$.

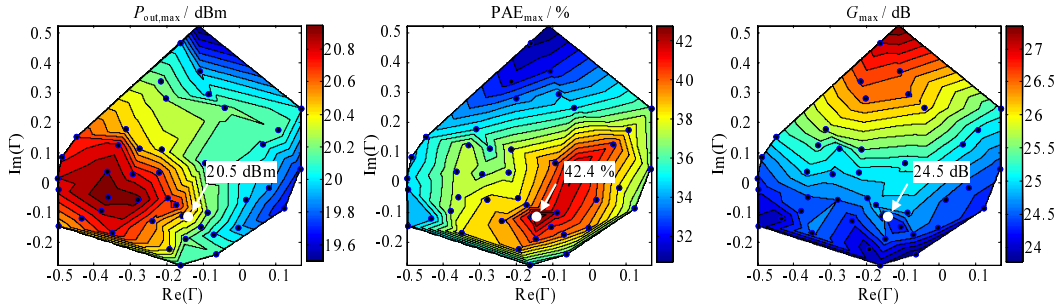


Figure 4.40: Low power path SMD load pull results including input and output hybrid ($f = 840$ MHz and $V_{CC} = 3.5$ V).

plane was chosen at the PCB transmission line. For the LP path, only operated at power back-off, a high PAE is more essential than high output power. Therefore the output reflection coefficient $\Gamma_{opt} = 0.17|-142°$ was chosen as final value, as it shows the highest measured PAE value of 42.4 %.

Mode optimization

Minimum loss LP output matching

For maintaining symmetry of the hybrid in HP mode, the output return loss of the switched OFF LP path has to be below $0.2 \triangleq -13.98$ dB (see Chapter 3.5.2). The matching network obtained with help of the SMD load pull procedure offers a return

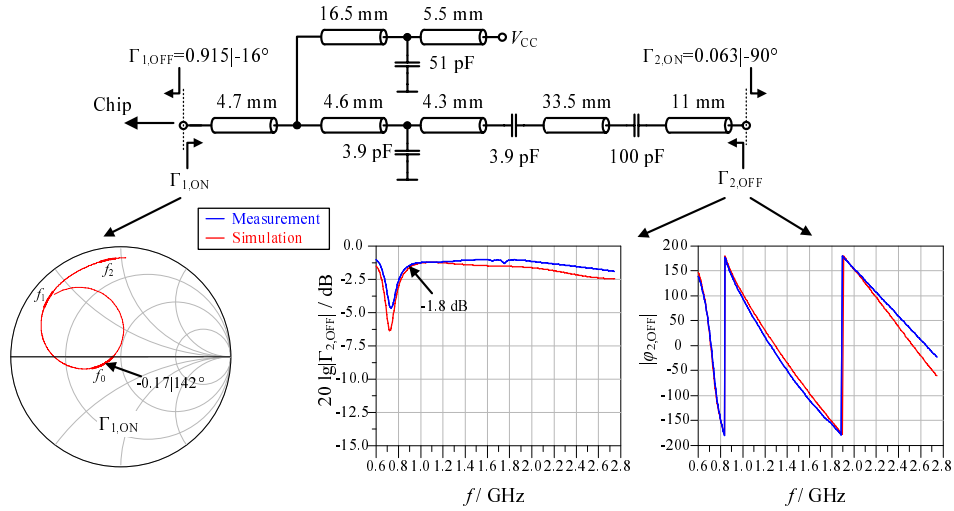


Figure 4.41: Lossless output matching network (LP path). Measured performance: $P_{\text{out,max}} = 20.5$ dBm, $\text{PAE}_{\text{max}} = 42.4$ %, $G_{\text{max}} = 24.5$ dB

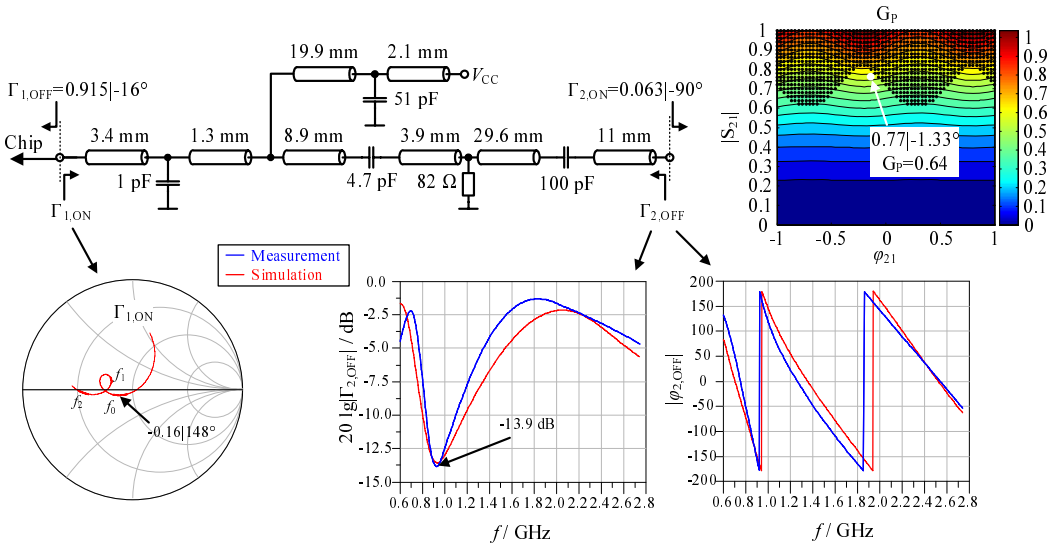


Figure 4.42: Lossy output matching network (LP path). Measured performance: $P_{\text{out,max}} = 18.5$ dBm, $\text{PAE}_{\text{max}} = 26$ %, $G_{\text{max}} = 22.0$ dBm

loss of only −1.8 dB in the operation band (Fig. 4.41). An improvement is achieved by adding losses to the network. Fig. 4.42 shows the theoretical upper limit of G_{p} when realizing the desired reflection coefficients in ON and OFF state. Losses of the network have to be at least $0.69 \mathrel{\hat{=}} -1.6$ dB, which reduces PAE to about 29.1 % (see equation (3.12)). As the topology sets another limit to the maximum realizable G_{p}, the final network has a G_{p} of $0.64 \mathrel{\hat{=}} -1.9$ dB, still very close to the absolute limit. The theoretically expected PAE is about 27 %. A measurement of the LP path with the network of Fig. 4.42 shows a PAE of 26 %. The difference is marginal

and can be explained by the shifted harmonic terminations at f_1 and f_2 (compare the smith charts in Fig. 4.41 and Fig. 4.42).

HPa mode and LPa mode

The LP path output matching network is designed for minimum P_{out} and PAE reduction in HP mode. This causes an output power reduction of 1.9 dB and PAE reduction of 36 % in LPa mode. Higher PAE values in LPa mode can be obtained, but only at cost of reduced symmetry and reduced PAE in HPa mode. Smaller losses at the output of the LP path will also lead to higher $|\Gamma_{2,\text{OFF}}|$ values and therefore higher sensitivity to antenna VSWR changes in HPa mode. Under these circumstances the implemented matching networks represent the best compromise for HPa mode and LPa mode performance.

MPa mode

In this mode only two HP amplifiers are activated via bias, illustrated in Fig. 4.43. As shown in Chapter 3.6 two effects have to be considered, which reduce PAE in this operation mode. Namely the reduced power transfer $G_{\text{p,13}}$ through the output hybrid (see equation (3.77)) and the changed load impedance, seen into the hybrid (see equation(3.78)). Both effects can be influenced by the reflection coefficient Γ_2 at the isolated port of the hybrids 2 and 3. Hybrid 1 still acts as symmetric quadrature power combiner and offers 50 Ω at its ports. The optimization was done in two steps. First the load pull contours of one HP path were measured at the reference plane of the hybrid. For this purpose a separate PCB was assembled, which offers direct access to the HP path output with an SMA connector. The load pull contours were measured with a manual passive slide screw tuner [Maury 09] and deembedded by phase delay and attenuation of the connector and transmission lines, to obtain contours at the reference plane $\Gamma_{\text{IN,3}}$ of the hybrid. The resulting contours for P_{out} and PAE are shown in Fig. 4.43. In a second step the output network between the ports of the hybrids and the output of the PCB was modeled in ADS. The model includes measured scattering parameters of the hybrids and the PCB transmission lines. With this setup the power transfer through hybrid 2 or 3 was plotted versus the phase φ of Γ_2. Also $\Gamma_{\text{IN,3}}$ in dependence of φ is obtained and overlayed the measured load pull contours. With these plots an optimum φ can directly be determined. Here, $\varphi = -80°$ is chosen, as this pulls the load $\Gamma_{\text{IN,3}}$ to a region with PAE = 64.6 % and allows a power transfer $G_{\text{p}} = 0.83 \triangleq -0.8$ dB close to the maximum. By using equation (3.12) and adding −0.3 dB attenuation due to the output hybrid, the expected peak PAE in this operation mode is calculated to 49.7 %. For verification, a measurement was done by applying a shunt capacitor of 2.7 pF at the isolated ports (similar to $\varphi = 80°$ at 840 MHz) and opening the series capacitor connection to the LP path. Alternatively a PIN-diode switch [Infineon 08] could be implemented. Its insertion loss of 0.07 dB (see Table 3.2) reduces $|\Gamma_2|$ from 0.9 to 0.87. This slightly reduces the radius of the $\Gamma_{\text{IN,3}}$ circle, shifting the $\varphi = -80°$ point towards higher PAE values. As $G_{\text{p,13}}$ reduces as well, the resulting PAE will be about the same. This shows that the peak PAE in this design is insensitive to return loss changes at the isolated port, as long as the PAE increases with decreasing $|\Gamma_2|$. The input power sweep is shown in Fig. 4.46, where an overall PAE of 49 % was measured in

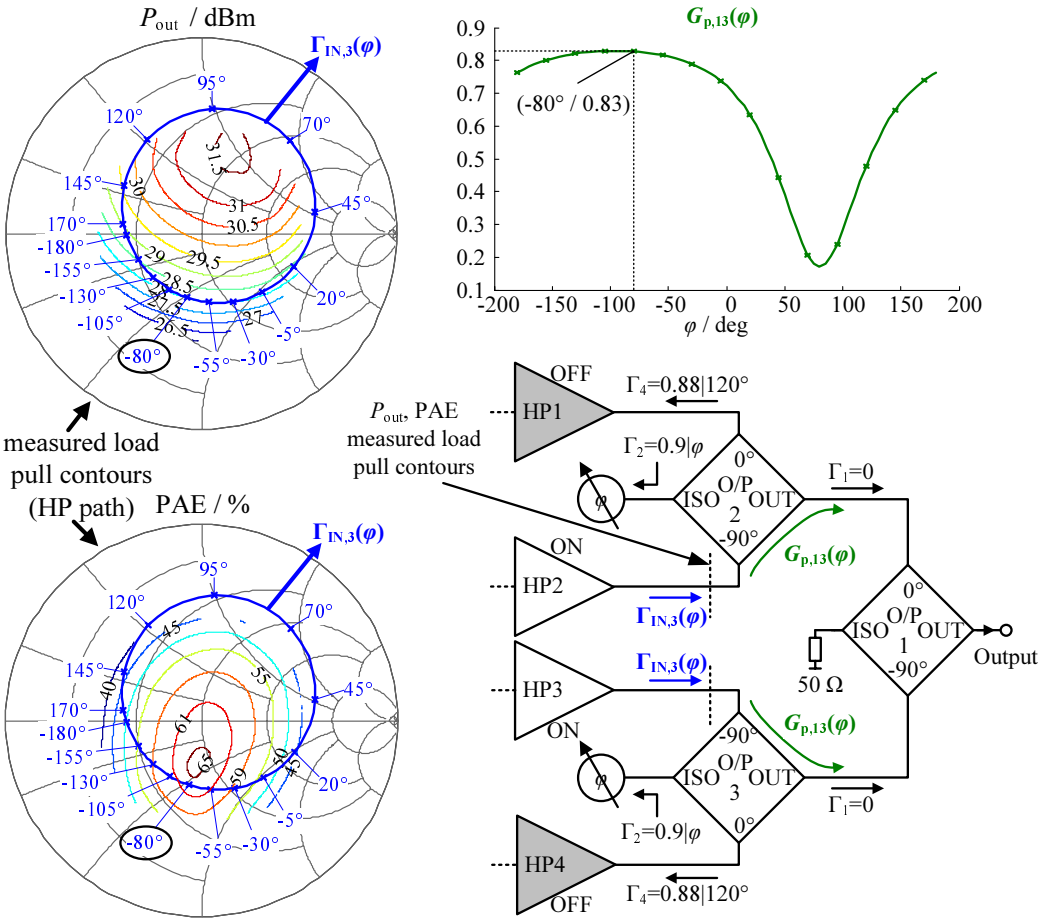


Figure 4.43: MPa mode optimization: Measured HP path load pull contours, simulated $\Gamma_{IN,3}$ and $G_{p,13}$ and schematically illustrated output combiner network.

this operation mode. This deviates from the theoretically expected value by only 0.7 % points.

MPd mode

In this mode only one HP amplifier is active and hybrids 1 and 2 act as bypass networks, as shown in Fig. 4.44. The optimization procedure is similar as for the MPa mode. The same load pull contours for one HP path are plotted and overlayed with the simulated $\Gamma_{IN,3}$ values, obtained by sweeping the phase of Γ_2 at the isolated port of hybrid 1. The isolated port of hybrid 2 is loaded with the same 2.7 pF as in MPa mode. According to Table 3.7 the phase at the isolated port of hybrid 3 has to be shifted by 180°. This condition, calculated for the ideal hybrids, has to be adapted by additional phase shifts due to the interconnect transmission lines. Simulations of the modeled network have shown that 110° are appropriate to maximize $G_{p,O3}$. This phase shift was implemented with a shunt inductor of 6.6 nH. The maximum power transfer is $0.648 \triangleq -1.88$ dB at $\varphi = -80°$. Together with PAE = 59 % for the single HP path a peak PAE of 37.6 % can be estimated for MPd mode. 2.7 pF

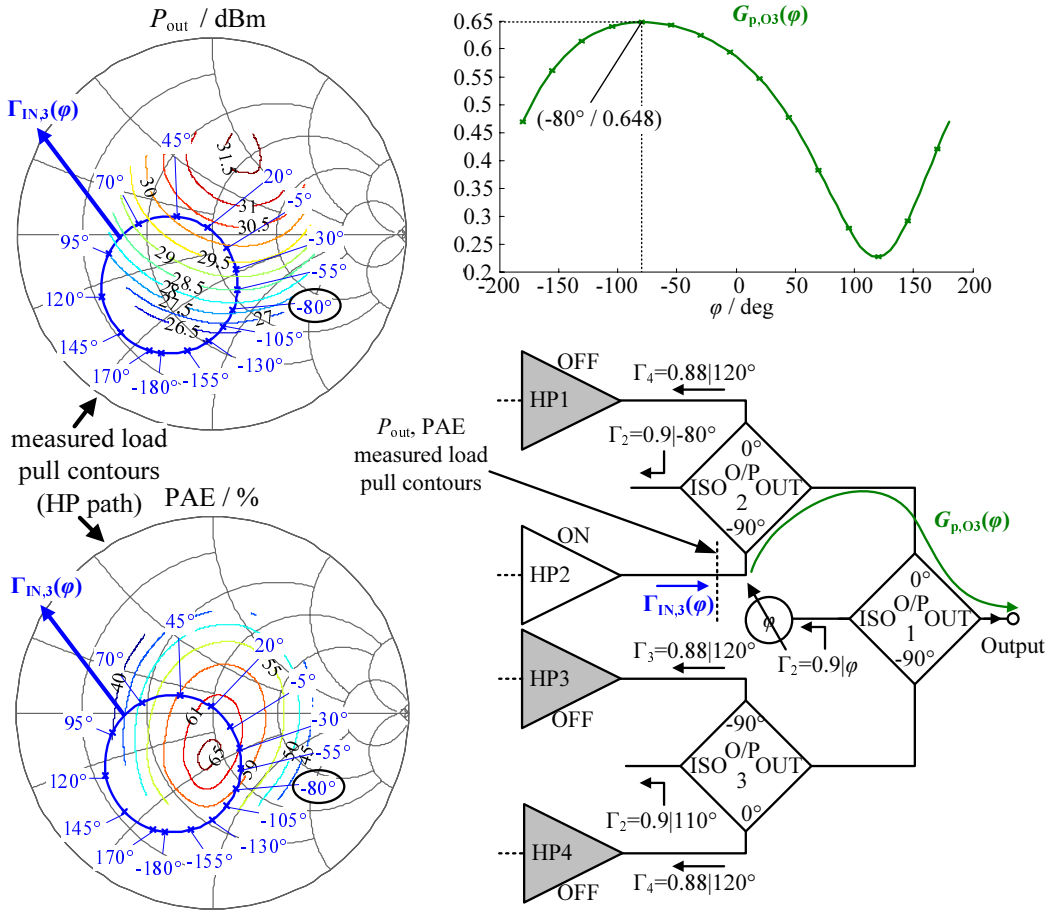


Figure 4.44: MPd mode optimization: Measured HP path load pull contours, simulated $\Gamma_{IN,3}$ and $G_{p,O3}$ and schematically illustrated output combiner network.

were applied to the isolated port of hybrid 1 and a peak PAE of 36 % (see Fig. 4.46) was measured. The difference between measured and calculated PAE results from additional losses at the input hybrids, which cause a further reduction of gain by 3.5 dB.

LPb mode

As depicted in Fig. 4.45, only one LP amplifier is activated via bias. The phase of Γ_2 at the isolated port of hybrid 1 is optimized in a similar way as for the previously described modes. The reference plane for the measured load pull contours is the direct port of hybrid 1. Hence, the measured contours include losses of the hybrid 2 and the deactivated HP paths. The maximum power transfer $G_{p,O1} = 0.786 \triangleq -1$ dB is found at $\varphi = -80°$. The PAE is therefore reduced from 26 % to 21 % in LPb mode. A comparison to the measured PAE = 20.4 % in Fig. 4.46 shows a difference of only 0.6 % points.

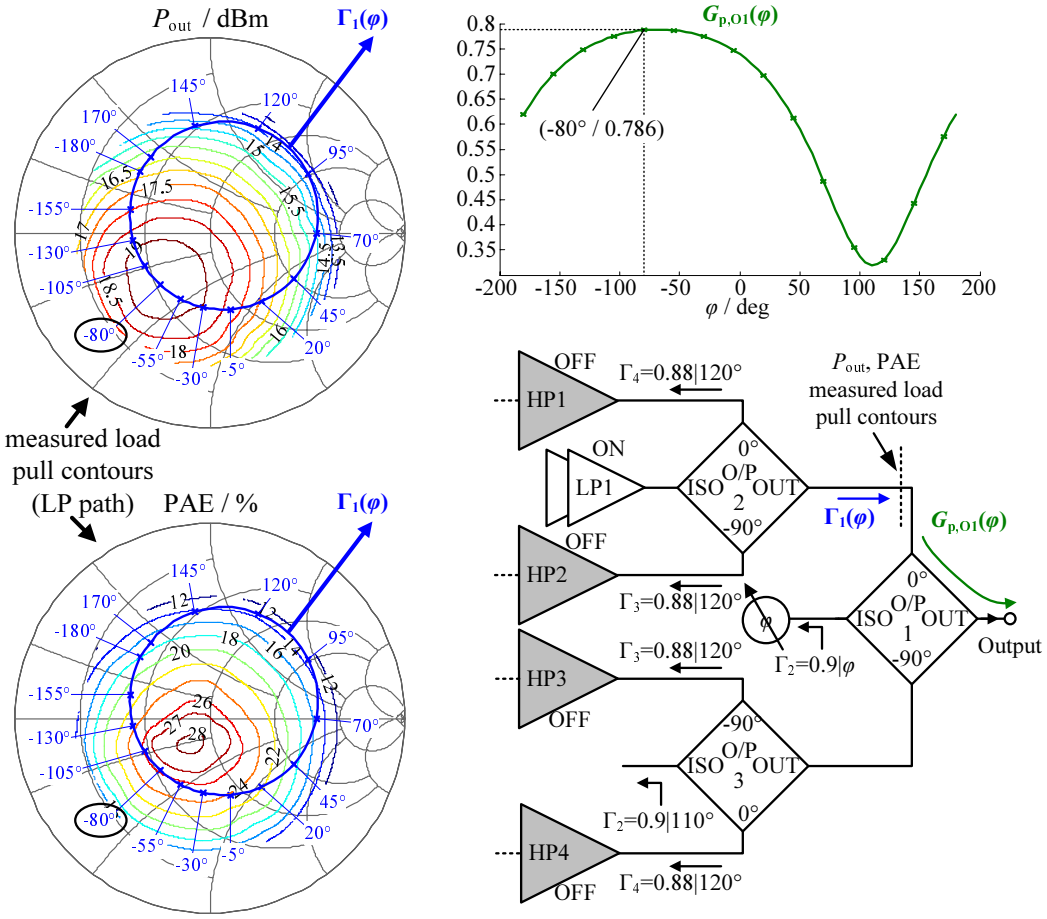


Figure 4.45: LPb mode optimization: Measured LP path load pull contours, simulated Γ_1 and $G_{p,O1}$ and schematically illustrated output combiner network.

4.4.4 Measurement results

For characterization of the multi-mode operation the PA was measured with GSM, EDGE and WCDMA signals in each operation mode. The measurements were done at a supply voltage of 3.5 V, according to nominal conditions for battery operation. The bias currents in each mode were set in a way to achieve best compromise between PAE and linearity of EDGE/WCDMA signals. For the presented measurements the same bias settings were used for measuring EDGE and WCDMA signals. For application in the system different EDGE/WCDMA bias settings can further improve linearity. Proper setting during operation can be done with help of a lookup table.

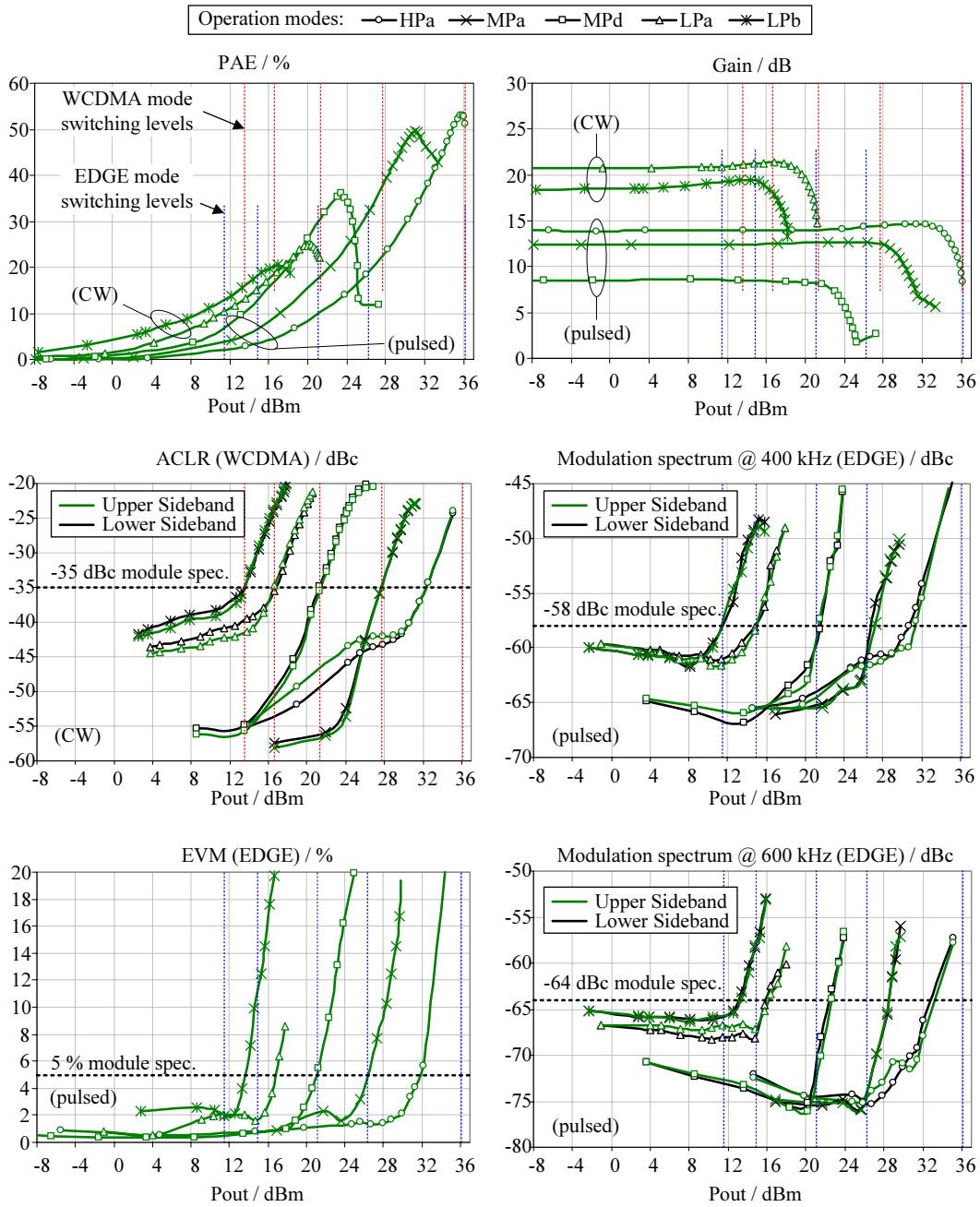


Figure 4.46: Power sweep measurement results at 840 MHz: PAE, G and linearity for WCDMA and EDGE signals versus P_{out}.

Fig. 4.46 shows the input power sweep results measured at 840 MHz. The shown PAE and G characteristics in HPa, MPa and MPd mode were measured with single tone excitation, constant supply voltage and pulsed bias voltage (duty cycle 12 % and pulse width 577 μs). These modes are intended mainly for GSM/EDGE operation. The LPa and LPb mode characteristics were measured under CW conditions (constant bias voltage), as these modes are suggested for WCDMA operation in

power back-off. In HPa mode a maximum output power of 36.1 dBm and a peak PAE of 52.5 % are achieved. At back-off power levels the PAE is reasonably improved. Over an output power range of 28 dB the PAE can be held above 10 % by switching between the operation modes. The switching levels depend on the operation standard (GSM/EDGE/WCDMA) and have to be chosen in accordance with the system and the transceiver. Here, switching levels are suggested, which are chosen such that the linearity measures in each mode fulfill the PA module specification (Table 2.1). These levels are marked with dotted red (WCDMA) and blue (EDGE) lines in the figure. Table 4.6 shows the PAE improvement at these switching levels and summarizes the performance in each operation mode.

Modes	HPa	MPa	MPd	LPa	LPb
quiescent current (mA)	209	103	47	27	14
$P_{\text{out,max}}$ (dBm)	36.1	33.3	27.3	21.2	18.2
G_{max} (dB)	14.8	12.9	9	21.3	19.5
PAE_{max} (%)	52.5	49.7	36	24.8	20.4
PAE improvement (% points) @ WCDMA switching levels	15.0	13.9	2.0	3.0	
PAE improvement (% points) @ EDGE switching levels	14.2	13.7	2.4	3.0	

Table 4.6: Power sweep summary at 840 MHz.

The MPa and MPd mode characteristics in Fig. 4.46 show an increase of PAE and G when driven into compression. If the power is increased the stages, which are only turned OFF via bias start to contribute to the total output power. This self biasing effect is not visible in the LP modes, which is due to the smaller input power levels. These are too small to turn on the deactivated HP and LP paths. The LP path also has higher isolation due to the two stage design, where the first stage acts as buffer. Nevertheless the usable operation range is far below the onset of self biasing, and is determined by the linearity limit.

The frequency sweep measurement results are presented in Fig. 4.47 showing the traced peak P_{out} and peak PAE of all operation modes. Between 700 MHz and 870 MHz the maximum output power of 36.1 dBm drops by only 0.5 dB and peak PAE is above 52 %. The bandwidth of 170 MHz is far above the desired operation bandwidth of 91 MHz, but the center is shifted by 84 MHz to lower frequencies. A similar frequency shift is visible in all operation modes. The MPd mode shows smallest PAE bandwidth, caused by diverging phases of the reflection coefficients (Γ_2) at the two output hybrids 1 and 3. This causes a frequency dependent load pull effect at the active HP amplifier. A similar bandwidth reduction in other back-off modes is visible when comparing HPa and MPa mode or LPa and LPb mode, but the effect is smaller there, as only one hybrid is used as bypass network. The frequency characteristic of $P_{\text{out,max}}$ does not show this reduction, because of the self biasing effect. Due to this the output impedances of the OFF amplifiers change and hence

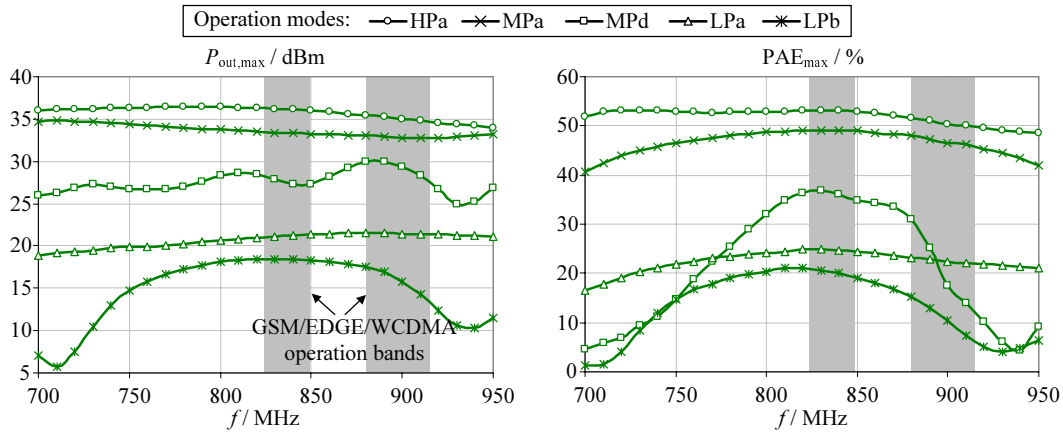
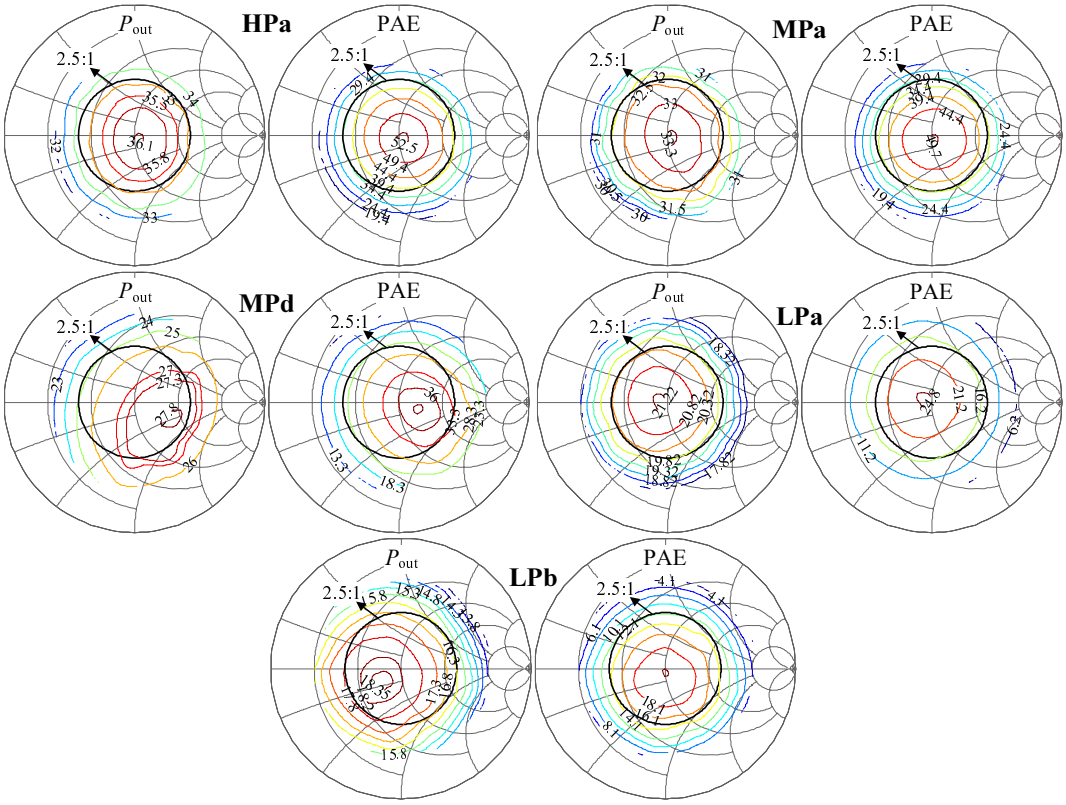


Figure 4.47: Frequency sweep measurement results: peak PAE and peak P_{out}.

pull the load of the ON amplifier when driven into deep compression. Additionally they contribute to the output power. Due to absence of self biasing in LPb mode the $P_{\text{out,max}}$ characteristic drops in a smaller frequency range than in the other modes.

The performance degradation due to load variations was measured with a passive load pull tuner at 840 MHz. The results are shown in Fig. 4.48. For a VSWR of 2.5 : 1 the absolute output power and PAE reduction compared to the nominal 50 Ω load condition ($|P_{\text{VSWR,max}} - P_{50}|$) and the maximum relative variation over phase at constant VSWR ($\Delta P_{\text{VSWR}}(\varphi)$) are listed for each mode.

As predicted in Chapter 3.6 the MPd and LPb modes show highest sensitivity to load variations, because the quadrature hybrids act as bypass networks. The only mechanism, which reduces impedance variations in these modes is the attenuation of reflected waves due to losses in the output matching networks. In this concern the LPb mode shows smaller performance degradation, due to higher intentionally added network losses. Nevertheless, also in MPd mode the maximum output power reduction is only 2.82 dB with a PAE reduction of 16.3 %. This is an improvement of 0.58 dB and 4.4 % points, compared to the single HP path. A comparison of all modes (Fig. 4.48 (a)) to the load pull measurements of the single ended HP path (Fig. 4.43, Fig. 4.44) and LP path (Fig. 4.45) is given in Fig. 4.48 (b). This finally validates the considerable performance improvements due to the application of the quadrature hybrids and the elaborate strategy for switching between the operation modes.



(a)

VSWR=2.5:1	HP path	HPa	MPa	MPd	LP path	LPa	LPb
$\lvert P_{VSWR,max}\text{-}P_{50}\rvert$	3.4 dB	1.63 dB	1.62 dB	2.82 dB	3.3 dB	1.75 dB	2.2 dB
$\Delta P_{VSWR}(\varphi)$	4.7 dB	0.76 dB	0.82 dB	3.35 dB	4 dB	0.91 dB	2.04 dB
$\lvert PAE_{VSWR,max}\text{-}PAE_{50}\rvert$	20.7 %	19.1 %	19.3 %	16.3 %	11 %	10.5 %	9.7 %
$\Delta PAE_{VSWR}(\varphi)$	17 %	6.6 %	5.3 %	15.4 %	9 %	4.5 %	6.18 %

(b)

Figure 4.48: Load sensitivity measurement: (a) Load pull measurement results in all operation modes at 840 MHz. (b) Comparison to single ended HP and LP path at VSWR = 2.5 : 1.

Chapter 5

Conclusion and Outlook

Conclusion

Recent developments in mobile communication systems from second generation (2G) to third generation (3G) show a trend towards multi-mode operation of several cellular standards (GSM, EDGE, WCDMA). The compatibility of the radio system (baseband, transceiver and front end) to the various coexisting standards demands cost effective multi-mode systems with small number of components. The increased shrinkage potential of CMOS technologies together with conceptual improvements in system architectures have paved the way for single multi-mode modules, including baseband and transceiver. Next step is reducing the number of different PAs in the front end. Currently each standard and band uses a separate GaAs-HBT PA. In [Sogl 07,a, Bakalski 08,a] the feasibility of merging GSM/EDGE in a single SiGe-HBT PA module was shown. For operation of WCDMA further conceptual efforts are necessary to reduce power consumption at very large power back-off levels. Additionally sensitivity of PA parameters to antenna mismatch has to be low over the whole output power range. This allows relaxed specifications being reflected in reduced current consumption of the handset in moving environment conditions.

The goal of this work was to develop a single multi-mode PA capable of GSM/EDGE/WCDMA operation in low band. The examination of several techniques has led to a switched path concept with a double balanced PA architecture. A general strategy for optimization of the various correlated operation modes is proposed. Application of this strategy to the fabricated prototype has shown the potential of this concept to considerably reduce the number of different PAs in current front end architectures.

Summarized, the main achievements of this research are:

- PA efficiency enhancement techniques are investigated with the objective to find a suitable architecture, which can be adopted for the multi-mode PA design. A classification of the various concepts is done and advantages and disadvantages of their application are shown. The concepts are weighted by their complexity and compared on basis of their theoretical potential to reduce average current consumption. An extensive and novel description for the use

of power combiners for "switch-less" integration of efficiency enhancement at back-off is included. General requirements for the combiner structures are analytically derived and the achievable performance improvements for symmetrical combiner structures are calculated. As a result it is shown that essential increase of efficiency can be gained within a back-off power range of 6 dB. It is also shown that further reduction of current consumption can be achieved by selective combination of the introduced methods. For the application of these new concepts in handset amplifiers their level of on-chip integration and the resulting system complexity are of major interest, as they reflect the cost of the PA modules. A final comparison regarding these aspects shows that the stage switching and load adaption concept together with the power combining concept provide best compromise between current consumption and implementation costs.

- The classical balanced power amplifier is extended by a low power path, which enables efficiency improvement at far power back-off without need of RF switches in the signal path [Sogl 09, Grebennikov 07]. Due to the quadrature hybrid any load variations at the transistor ports reduce by about 50 % in high power mode. The optimum operation conditions are derived. It is proven that maximum efficiency in HP mode and LP mode can not be optimized independently. A trade-off is necessary, which mainly depends on losses in the LP path matching network. A method is developed to derive matching networks which exhibit the theoretical minimum loss for best compromise between HP and LP mode efficiency.

- A double balanced switched path PA as a novel concept for realization of a multi-mode PA is proposed. In comparison to the single balanced switched path PA the sensitivity to load variations is further improved. Thus, load insensitivity could also be achieved in the several back-off power modes [Sogl 07,b]. A theory for optimization of the various operation modes is developed, showing the potential overall performance and trade-offs.

- A prototype of a double balanced switched path PA is fabricated in a low cost 0.35 μm bipolar SiGe technology [Sogl 09]. At 840 MHz a peak power of 36.1 dBm with 52 % PAE is achieved for GSM operation. In back-off at 27 dBm, 16 dBm and 13 dBm efficiencies of 37 %, 18 % and 15 % for WCDMA signals are measured. For EDGE operation at 30 dBm a peak PAE of 30 % is achieved and held above 10 % over a dynamic output power range of 22 dB, by proper selection of the five amplifier modes. These performance characteristics are comparable to optimized single mode PAs.

- General design equations for quadrature hybrids, based on lumped element, one section and two section inductively coupled topologies are developed. The trade-off between size and bandwidth for on-chip integration could be shown, which helps to reduce design cycles and speed up simulation time with EM simulators. A two stage inductively coupled quadrature hybrid, with attenuation < 1.5 dB, amplitude error < 0.75 dB and phase error $< 2.9°$ over a

relative bandwidth of 22.2 % is realized in SiGe technology. The design fulfills the requirements for the multi-mode PA and is furthermore suitable for on-chip integration.

- A very time efficient "load pull" procedure is developed for experimental optimization of PA matching networks. This avoids complicated deembedding structures.

Fig. 5.1 together with Table 5.1 compare the multi-mode PA prototype of this work to other state of the art single mode WCDMA and dual mode GSM/EDGE PAs. The data is taken from recent product data sheets and scientific publications. The developed amplifier shows average PAE performance for GSM/EDGE operation ($P_{\text{out}} > 26$ dBm) with sufficient margin to the linearity specifications. For WCDMA operation at low power levels ($P_{\text{out}} < 7$ dBm) the PA achieves excellent PAE values, being about two times higher than for actual single mode PAs. Moreover it offers reduced load sensitivity over the whole output power range. The comparison shows the feasibility of the concept to reduce the number of different front end PAs without major performance loss.

Outlook

The good results of this work suggest to head for further integration. The first step could be a system in package (SiP) integration. The input circuitry and the power transistor can be fabricated in the introduced low cost SiGe technology. The output circuitry (hybrids and matching networks) has to be realized on a die with copper metallization and high resistive substrate to achieve higher quality factors and sufficient maximum current carrying capacity. In this regard the thesis has shown possibilities for small size integration of quadrature hybrids with sufficient bandwidth. The size could be further reduced by accomplishing the matching within the hybrids, instead of using a separate matching network.

As present research is driven towards PA integration in CMOS, the next step will be a system on chip (SoC) solution with transceiver and PA on a single die. The input hybrids could be replaced by quadrature signal generation with RC polyphase filters or flip-flops in CMOS, saving chip area and cost. Integration of low loss passive elements for the output could be done in wafer level package, where an additionally processed thick copper interconnect layer can be used for integration of inductors and hybrids. The ruggedness of CMOS transistors and the process stability of wafer level packages will be challenges to be solved before going to mass production.

Another issue to be investigated in future is the "switch-less" efficiency improvement at very large power back-off levels. One possibility may be the combination of large and small transistor stages with asymmetric power combiners. At different power levels the stages can be alternatively activated only via their bias. A trade-off between maximum efficiency of both stages will be necessary, if RF switches are omitted. This problem is similar to the trade-off described in this thesis between high efficiency in HP and LP mode for the balanced switched path architecture. The results obtained here can directly be used for designing minimum loss matching

networks for each amplifier path. To reduce component count for on-chip integration, the combining and matching tasks will have to be merged in one network. The attainable minimum loss of this 3-port networks under changed load conditions will have to be derived to determine limits of this approach. Overall, for applications where high efficiency over a large dynamic range is necessary the power combiner based PA concept is worth for further exploration from a commercial and academic standpoint.

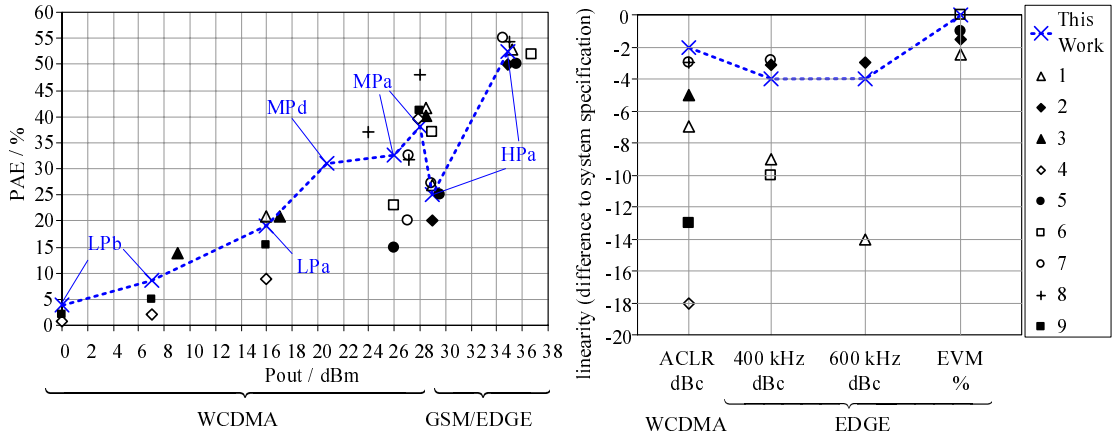


Figure 5.1: Comparison to state of the art GSM/EDGE/WCDMA PAs.

Nr. from Fig. 5.1	Reference	Technology	GSM	EDGE	WCDMA	Comment
1	[Anadigics 09]	InGaP-HBT	LB	LB	HB	separate PAs
2	[RFMD 09]	GaAs-HBT	LB	LB	–	multi-band
3	[Avago 09]	InGaP-HBT	–	–	LB	BO mode
4	[Skyworks 09]	InGaP-HBT	–	–	LB	BO mode
5	[Yamamoto 02]	GaAs-HBT	LB	LB	–	multi-band
6	[Bakalski 08,b]	SiGe-HBT	LB	LB	–	load indep., BO mode
7	[Pusl 01]	SiGe-BiCMOS	LB	LB	LB	separate PAs
8	[Scuderi 08]	SiGe-HBT	LB	LB	LB	load indep., V_{supply} adaption
9	[Han 08]	InGaP-HBT	–	–	LB	BO mode, SSLA
	This Work	SiGe-HBT	LB	LB	LB	multi-mode, load independent

Table 5.1: Comparison to state of the art GSM/EDGE/WCDMA PAs.

Appendix A

N-Port Combiner Networks

A.1 Amplitude and Phase Errors

The characteristics of a power combining network deviate from the ideal characteristics especially when the implementation has to cover large bandwidths. Errors which result from unequal amplitude combining ratios or non-coherent phase summation are derived in the following. Assumed is a general combiner network which has N matched ports with $n = N - 1$ inputs and one output. The forward and backscattered waves are defined by the symmetric scattering matrix with zeros along the main diagonal ($S_{ii} = 0$). The output power is calculated from the backscattered wave at port N:

$$b_N = S_{1N}\, a_1 + S_{2N}\, a_2 + \ldots + S_{nN}\, a_N \tag{A.1}$$

The variables a_i are the forward power waves at each input port. Since any amplitude and phase error of the network (S_{ij}) appears as product with a_i, the error terms can be referred to the signals without substantial loss of information. Hence, the scattering parameters are equal and follow as $S_{iN} = 1/\sqrt{n}$. The individual input signals are described by the nominal power P, each added an amplitude error ε_i and a phase error β_i:

$$a_i = \sqrt{2\,P\,\varepsilon_i} \cdot e^{j\beta_i} \tag{A.2}$$

By substitution and simplification, the total output power ($P_{\mathrm{tot}} = |b_N|^2/2$) results as:

$$P_{\mathrm{tot}} = \frac{P}{n}\left[\left(\sum_{i=1}^{n}\sqrt{\varepsilon_i}\,\cos\beta_i\right)^2 + \left(\sum_{i=1}^{n}\sqrt{\varepsilon_i}\,\sin\beta_i\right)^2\right] \tag{A.3}$$

With this formula the network combining efficiency calculates as:

$$\eta_{\mathrm{C}} = \frac{P_{\mathrm{tot}}}{P_{\mathrm{in}}} = \frac{\left(\sum_{i=1}^{n}\sqrt{\varepsilon_i}\,\cos\beta_i\right)^2 + \left(\sum_{i=1}^{n}\sqrt{\varepsilon_i}\,\sin\beta_i\right)^2}{n\,\sum_{i=1}^{n}\varepsilon_i} \tag{A.4}$$

Small values for ε and β have little impact on the degradation of η_C. Exemplarily, a network with 4 input ports, becoming unbalanced by an amplitude error of 1 dB and a phase error of 10° at each port, still has a combining efficiency of 95%.

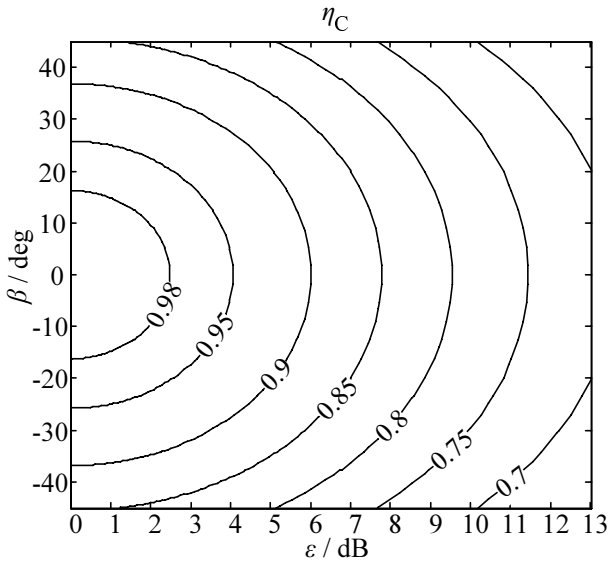
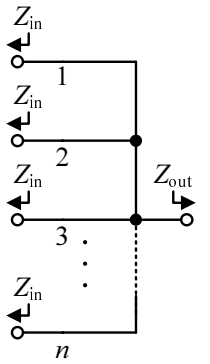


Figure A.1: Power combiner efficiency η_C over Amplitude error ε and phase error β in a three port network.

The low sensitivity to these errors can easily be shown in a contour plot (Fig. A.1) for the case of a 3-port network ($N = 3$). Tolerable combining efficiencies of $< 95\%$ (equal to 0.2 dB insertion loss) are still obtained with about 20° phase error and -3 dB amplitude error. These properties of power combining networks make their implementation with distributed or lumped elements efficient, as manufacturing tolerances cause smaller impact. The following chapters show most important realizations of these kinds of networks. It has to be noted that above calculations have been done for the case of matched ports. Additional degradations due to mismatch caused by amplitude and phase errors were not considered but can be compensated by the device matching networks.

A.2 N-Port T-junction

The simplest power combining network is the lossless T-junction, which is basically a direct connection of n sources to one load. Fig. A.2 shows a simplified circuit diagram and the underlying scattering matrix. The network is not matched at all ports. Hence, for source impedances of $Z_{in} = 50\ \Omega$, the load impedance has to be optimized to obtain maximum power transfer. This load impedance depends on the number of connected sources and can be calculated by maximizing $P_{out}(\Gamma_{out})$. The methods from Appendix B can be used to perform the derivation and to finally obtain $\Gamma_{N,opt} = \frac{2}{N} - 1$. With this load impedance the combining efficiency of the network is 100%. Deactivation of sources leads to a reduced output power. This, due to less input power and due to additional reflections inherent to the N-port, which can be interpreted as additional loss of the network. To characterize the influence, following a general expression for the total output power P_{out} depending on the size N of the network and the number n of activated sources will be derived. It is assumed that all sources deliver the same power P_{in} and ports which are turned OFF change their reflection coefficient from $\Gamma_{ON} = 0$ to $\Gamma_{OFF} = 1$. The total output



$$\begin{pmatrix} \frac{2}{n+1}-1 & \frac{2}{n+1} & \cdots & \frac{2}{n+1} \\ \frac{2}{n+1} & \frac{2}{n+1}-1 & \cdots & \frac{2}{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{2}{n+1} & \frac{2}{n+1} & \cdots & \frac{2}{n+1}-1 \end{pmatrix}$$

(b)

Figure A.2: (a) N-port lossless T-junction (b) scattering matrix.

power is the sum of the individual contributions:

$$\begin{aligned} P_{\text{out}} &= \left(\sum_n \sqrt{P_{\text{in}} \left| \frac{b_N}{b_{\text{s}}} \right|^2 (1 - |\Gamma_N|^2)} \right)^2 \\ &= P_{\text{in}} \left(1 - |\Gamma_N|^2\right) \left(\sum_n \sqrt{\left| \frac{b_N}{b_{\text{s}}} \right|^2} \right)^2 = P_{\text{in}} (1 - |\Gamma_N|^2) n^2 \left| \frac{b_N}{b_{\text{s}}} \right|^2, \end{aligned} \tag{A.5}$$

Due to the network symmetry the term $\left|\frac{b_N}{b_{\text{s}}}\right|$ is equal for all active ports, but varies with the size of the network N and the number of turned ON amplifiers n. A closed form expression can be found empirically by induction. This starts with solving the systems of network equations to obtain $\frac{b_N}{b_{\text{s}}}$ (see Appendix B.1). For different values of N and n it is easy to find the numerical solutions by matrix multiplications. Several cases were evaluated using Matlab and the results up to 7-port networks are listed in Table A.1. From these numerical values it was concluded that:

$$\frac{b_N}{b_{\text{s}}} = \frac{N}{N + n - 1} \tag{A.6}$$

Finally the output power and network efficiency for any number of combined amplifiers can be written in closed form:

$$P_{\text{out}} = P_{\text{in}} \frac{4\,n^2\,(N-1)}{|N+n-1|^2} \quad \text{and} \quad \eta_{\text{C}} = P_{\text{in}} \frac{4\,n\,(N-1)}{|N+n-1|^2} \tag{A.7}$$

The mismatch or load pull effect at the input ports caused by impedance changes of deactivated stages can also be derived in dependence of N and n. It can be calculated in a similar way as the network efficiency by solving the linear network equations for b_1/a_1. A more simple derivation can be done by calculating the equivalent input impedance at one port. As the T-junction doesn't use distributed components, the

Combiner size	Active sources	Port load Γ_i						Transducer Gain b_N/b_s	
N-Port ($\Gamma_{N,opt}$)	n	1	2	3	4	5	6	decimal	fraction
7-Port ($\Gamma_{N,opt} = -5/7$)	1	0	1	1	1	1	1	1	1
	2	0	0	1	1	1	1	0.8750	7/8
	3	0	0	0	1	1	1	0.7778	7/9
	4	0	0	0	0	1	1	0.7000	7/10
	5	0	0	0	0	0	1	0.6364	7/11
	6	0	0	0	0	0	0	0.5833	7/12
6-Port ($\Gamma_{N,opt} = -2/3$)	1	0	1	1	1	1	×	1	1
	2	0	0	1	1	1	×	0.8571	6/7
	3	0	0	0	1	1	×	0.7500	6/8
	4	0	0	0	0	1	×	0.6667	6/9
	5	0	0	0	0	0	×	0.6000	6/10
5-Port ($\Gamma_{N,opt} = -3/5$)	1	0	1	1	1	×	×	1	1
	2	0	0	1	1	×	×	0.8333	5/6
	3	0	0	0	1	×	×	0.7143	5/7
	4	0	0	0	0	×	×	0.6250	5/8
4-Port ($\Gamma_{N,opt} = -1/2$)	1	0	1	1	×	×	×	1	1
	2	0	0	1	×	×	×	0.8000	4/5
	3	0	0	0	×	×	×	0.6667	4/6
3-Port ($\Gamma_{N,opt} = -1/3$)	1	0	1	×	×	×	×	1	1
	2	0	0	×	×	×	×	0.75	3/4
2-Port ($\Gamma_{N,opt} = 0$)	1	0	×	×	×	×	×	1	1

Table A.1: b_N/b_s for different values of N and n. Amplifier states: $\Gamma_i = 0$ ON, $\Gamma_i = 1$ OFF

impedance seen into one port is a simple parallel circuiting of all input impedances (Z_0) and the load impedance ($Z_{opt} = Z_0/(N-1)$).

$$\frac{1}{Z_1} = \frac{n-1}{Z_0} + \frac{1}{Z_{opt}} \Rightarrow Z_1(n,N) = \frac{Z_0}{N+n-2} \tag{A.8}$$

When all amplifiers are ON, the input impedance is $Z_{nom} = Z_1(N-1,N)$. Hence, the relative mismatch can be calculated for any size N and n of the network:

$$\Delta Z = \left| \frac{Z_1(n,N) - Z_1(N-1,N)}{Z_1(N-1,N)} \right| = \ldots = \frac{2N-3}{N+n-2} - 1 \tag{A.9}$$

More insight into the dependence of combining efficiency and mismatch on the number of input amplifiers N and n can be obtained by a contour plot, as shown in Fig. A.3(a) and Fig. A.4(a). An increased number of input ports N has the advantage of a smoother characteristic $\eta_C(n)$ but doesn't alter the absolute values of

combining efficiency. For example a network with $N = 5$ has an efficiency of 88.89 % when half of the input amplifiers is turned OFF ($n = 2$). The same efficiency is obtained for $N = 7$ and $n = 3$, also half of the input amplifiers turned OFF. So, related to the maximum achievable output power for each N, the efficiency at same back-off powers is always the same. A plot of η_C versus the normalized back-off power $P_{out}/P_{max,N}$ for all combinations of N and n confirms this assumption, as all values are on one continuous curve (Fig. A.3(b)). This plot also shows that the efficiency at power back-off decreases, independent of the network size N. In the upper power range the efficiency decreases very fast with about 4 %/dB and has a value of −53 % at −10 dB. In comparison, for a typical class AB amplifier a reduction of 50 % from its maximum efficiency happens at about −6 dB and decreases further at larger back-off. Therefore, the concept of power combining is still more efficient compared to a single amplifier.

Fig. A.4(b) shows the relative mismatch, plotted versus back-off power. Here, the dependence is not unique. For each back-off power level there exist more than one ΔZ. Consequently, N and n can be optimized to find the smallest mismatch values at a certain back-off power without influence on the combining efficiency at that level.

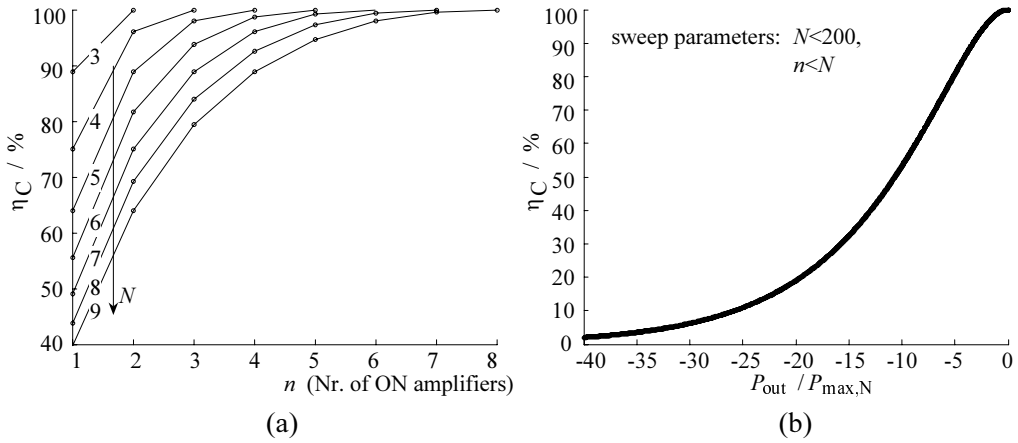


Figure A.3: (a) Decrease of combining efficiency with varying number of ON amplifiers. (b) Upper bound of η_C versus normalized output power back-off.

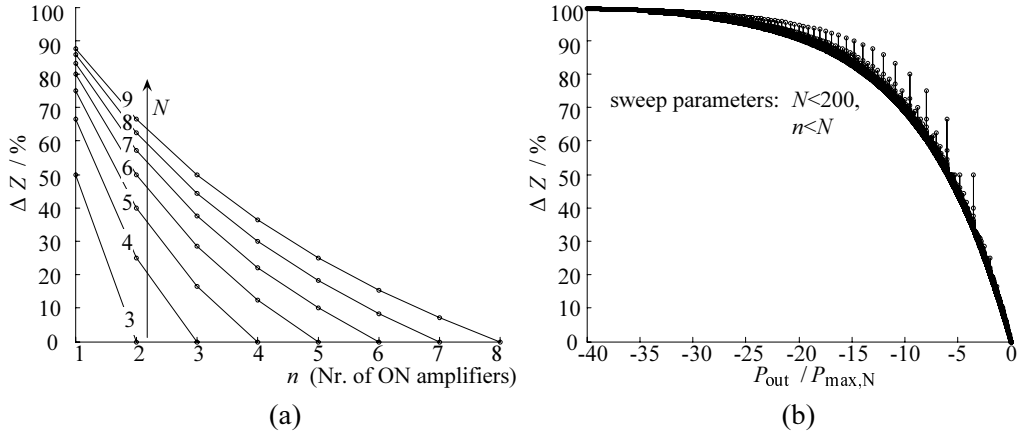


Figure A.4: (a) Increase of mismatch with varying number of ON amplifiers. (b) Upper bound of ΔZ versus normalized output power back-off.

A.3 N-Port Lossless transmission line combiner

The network of the lossless transmission line power combiner together with the scattering matrix for an arbitrary number of ports N are shown in Fig. A.5. The

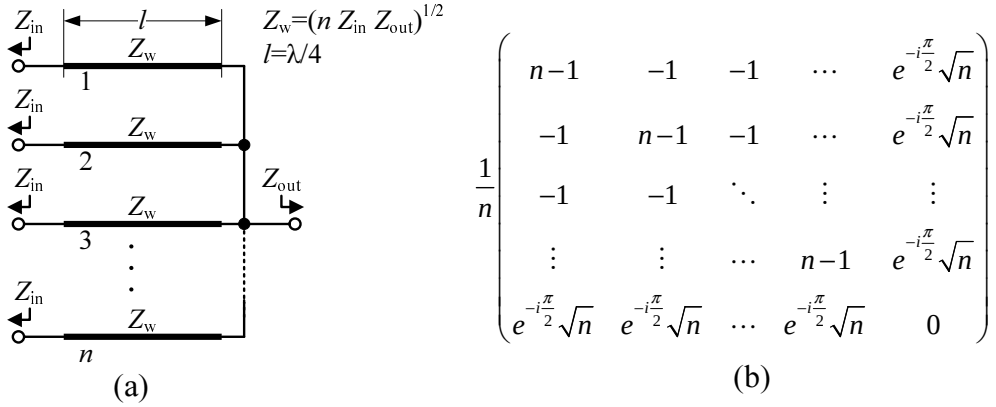


Figure A.5: (a) N-port lossless transmission line power combiner (b) scattering matrix.

differences to the lossless T-junction combiner are an additional 180° phase shift and an impedance transformation at all n input ports. This action is accomplished by transmission lines of certain length $l = \lambda/4$ and impedance $Z_{\mathrm{W}} = (nZ_{\mathrm{in}}Z_{\mathrm{out}})^{1/2}$. Thus, the power combiner can be matched to arbitrary impedances, independent of its size N. This is an advantage over the T-junction combiner where Z_{out} has to be changed for each N and Z_{in}.

The variations of combining efficiency and relative mismatch versus N and n are shown in Fig. A.6 and Fig. A.7. These plots were obtained by successively solving the network equations for various combinations of N and n with Matlab. The procedure

starts with definition of general equations for η_C and ΔZ:

$$\eta_C(N,n) = P_{in}(1-|\Gamma_N|^2)n\left|\frac{b_N}{b_s}\right|^2 \tag{A.10}$$

$$\Delta Z(N,n) = \left|\frac{Z_1(n,N)-Z_1(N-1,N)}{Z_1(N-1,N)}\right| \text{ with } Z_1(n,N) = \frac{1+\left.\frac{b_1}{b_s}\right|_{\Gamma_1=0}}{1-\left.\frac{b_1}{b_s}\right|_{\Gamma_1=0}} \tag{A.11}$$

The ratios b_1/b_s and b_N/b_s are calculated from the network equations and depend on N and n:

$$(b_1,\ldots,b_N)/b_s = (\underline{U}-\underline{S}\cdot\underline{\Gamma})^{-1}\cdot\underline{S}_j = \underline{B}_j, \tag{A.12}$$

with $\underline{U}$ being the NxN unity matrix, $\underline{S}_j$ being the j^{th} column vector of the scattering matrix and $\underline{\Gamma}$ being the diagonal matrix of all port reflection coefficients in dependence on the state, i.e. $\Gamma_{i,\text{OFF}} = -1$ and $\Gamma_{i,\text{ON}} = \Gamma_N = 0$.

These equations are repeatedly solved in a loop for $N > 3$ and $n = 1,\ldots,N-1$ and the results plotted in the contour plot. It is possible to derive general empirical equations from these results - as for the T-junction combiner - but is not done here.

The plots show that the combining efficiency is equal to the one from the T-junction combiner network but the relative mismatch levels are by 50% smaller. The reason for this reduction of ΔZ comes from the additional impedance transforming $\lambda/4$ lines which lead to smaller coupling values $|S_{i,j}| = 1/n < 2/(n+1)$ compared to the T-junction combiner.

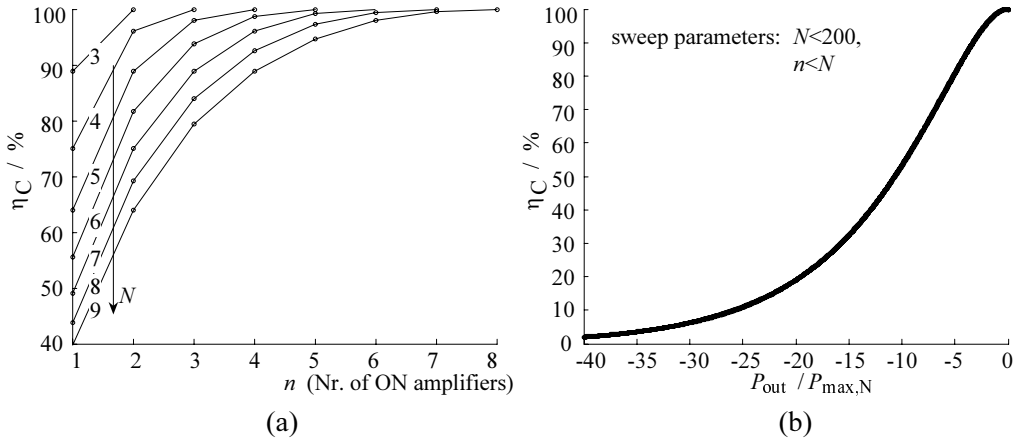


Figure A.6: (a) Decrease of combining efficiency with varying number of ON amplifiers. (b) Upper bound of η_C versus normalized output power back-off.

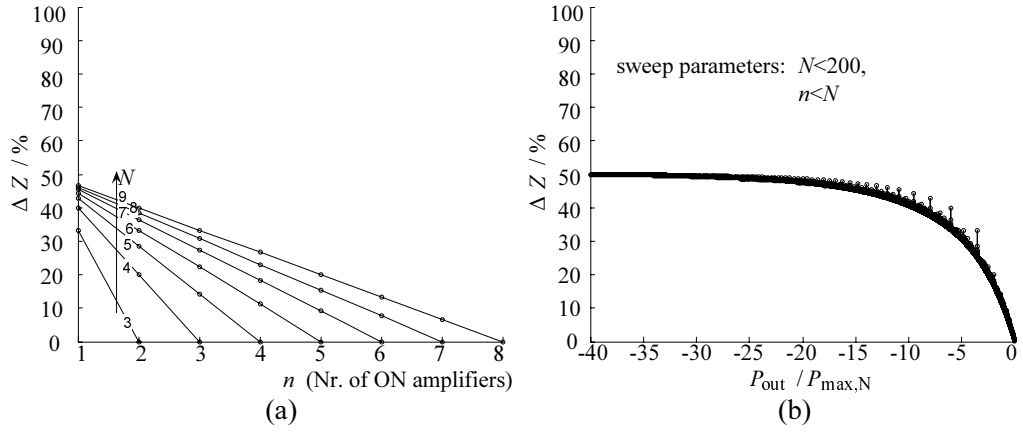


Figure A.7: (a) Increase of mismatch with varying number of ON amplifiers. (b) Upper bound of ΔZ versus normalized output power back-off.

A.4 N-Port Wilkinson combiner

The network of the Wilkinson power combiner together with the scattering matrix for an arbitrary number of ports N are shown in Fig. A.10.

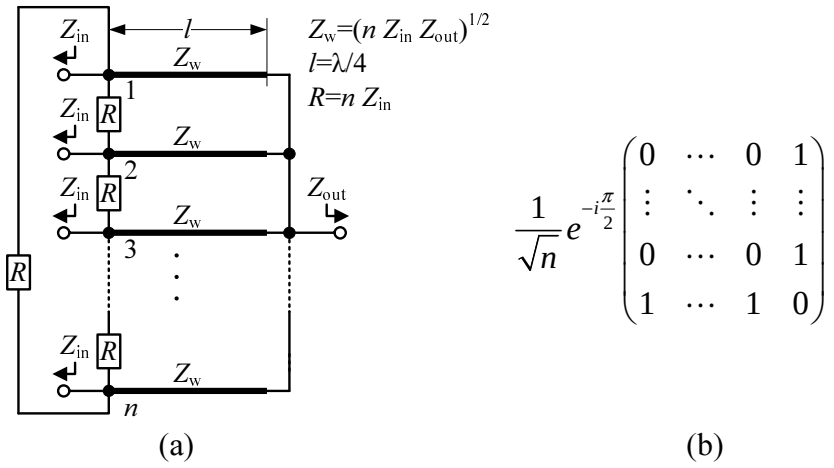


Figure A.8: (a) N-port Wilkinson power combiner (b) scattering matrix.

This particular power combiner circuit offers matching at all ports and perfect isolation between the input ports. This is achieved by the additional resistors $R = nZ_{\text{in}}$ between adjacent inputs. Due to this behavior the combining efficiency versus output power reduces very fast with about 9 %/dB. This is shown in Fig. A.9, which is calculated with the same procedure as described in Appendix A.3. The mismatch level $\Delta Z = 0\%$, on the other hand, is constant for every power back-off state. Due to the simple result no figure is shown here for ΔZ.

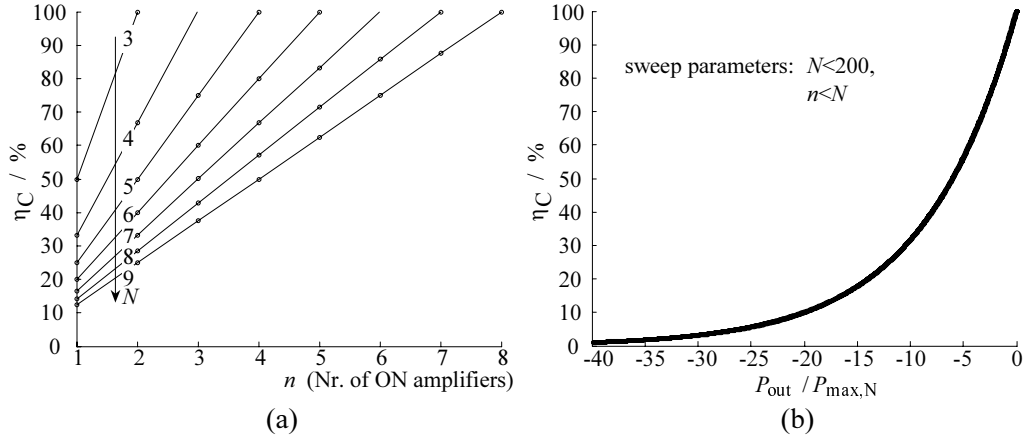
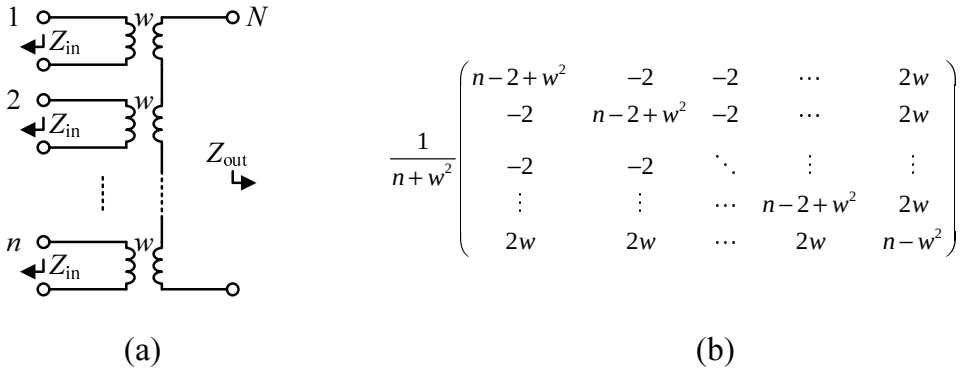


Figure A.9: (a) Decrease of combining efficiency with varying number of ON amplifiers. (b) Upper bound of η_C versus normalized output power back-off.

A.5 N-Port transformer based combiner

The network of the transformer based power combiner together with the scattering matrix for an arbitrary number of ports N are shown in Fig. A.10.



$$\frac{1}{n+w^2}\begin{pmatrix} n-2+w^2 & -2 & -2 & \cdots & 2w \\ -2 & n-2+w^2 & -2 & \cdots & 2w \\ -2 & -2 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & n-2+w^2 & 2w \\ 2w & 2w & \cdots & 2w & n-w^2 \end{pmatrix}$$

Figure A.10: (a) N-port transformer based power combiner (b) scattering matrix.

This type of power combiner is similar to the lossless transmission line combiner, as with the winding ratio w it has an additional degree of freedom to match arbitrary source impedances to arbitrary load impedances. The relation between w, N and $\Gamma_{N,\mathrm{opt}}$ for maximum combining efficiency is:

$$\Gamma_{N,\mathrm{opt}} = S^*_{N,N} = \frac{n - w^2}{n + w^2} \tag{A.13}$$

This equation is evaluated in Fig. A.11 in logarithmic scale for w. For the special case $\Gamma_{1,\ldots,N} = 0$ the winding ratio has to be $w = \sqrt{n}$ for not loosing power when all amplifiers are turned ON.

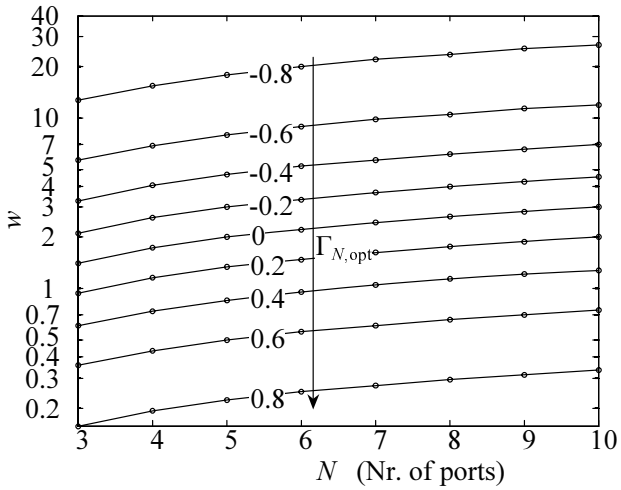


Figure A.11: Contour plot of the relations between the transformer winding ratio w, the number of ports $N = n + 1$ and the optimum load impedance values $\Gamma_{N,\mathrm{opt}}$ for maximum η_{C}

The combining efficiency (η_{C}) and relative mismatch (ΔZ) are calculated as described in Appendix A.3. The results, shown in Fig. A.12 and Fig. A.13, are equal to those of the transmission line combiner. Apparently there are some advantages of the transformer based structure for small size integration on semiconductor chips. The most important are listed below:

- For a given frequency the transformer realization usually uses less area and has a larger operation bandwidth compared to $\lambda/4$ transmission lines.
- The realization works with virtual ground connections, hence a low inductive RF ground connection on chip is not essential. With center tapped transformers these virtual connections can also be used as bias feed.
- The realization has all advantages of a push-pull circuit like immunity to common mode interferers and even order harmonic cancelation.

An efficiency enhanced power amplifier using this kind of power combiner was realized in [Haldi 08, Liu 06, Liu 08].

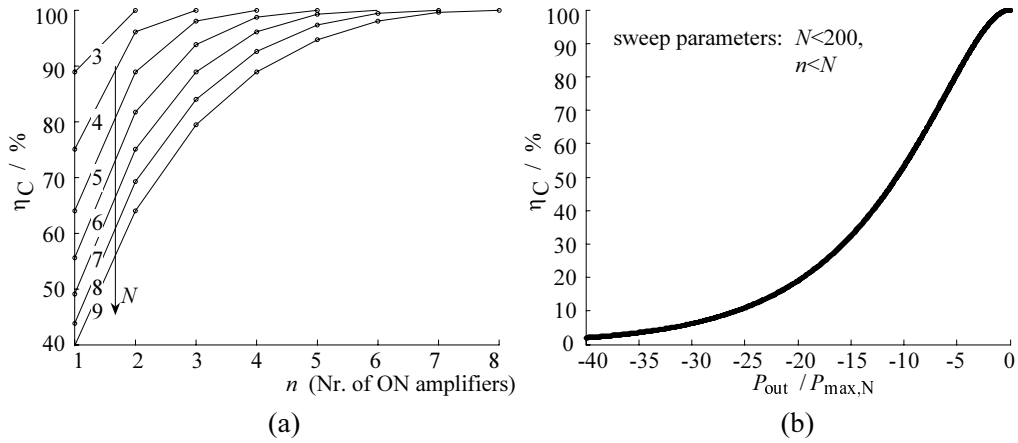


Figure A.12: (a) Decrease of combining efficiency with varying number of ON amplifiers. (b) Upper bound of η_C versus normalized output power back-off.

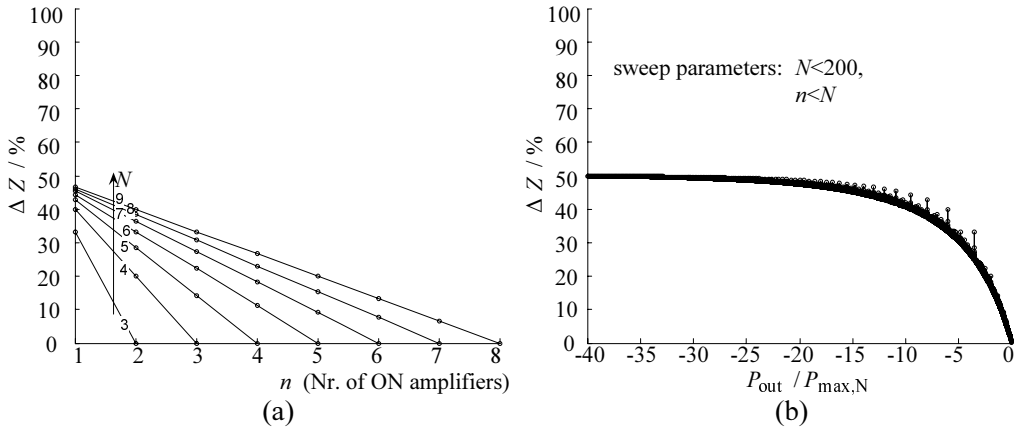
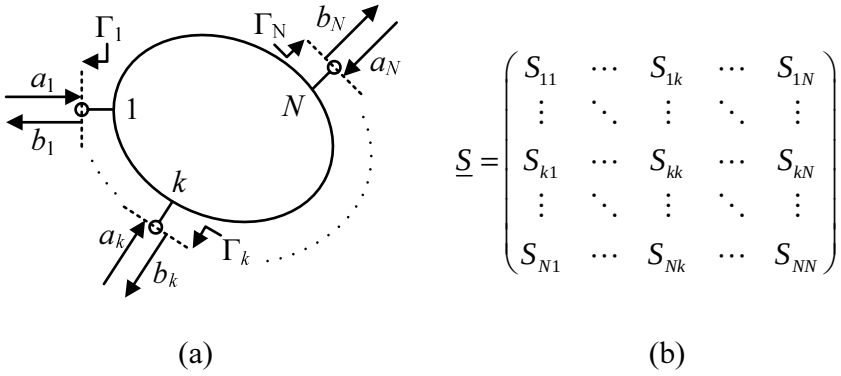


Figure A.13: (a) Increase of mismatch with varying number of ON amplifiers. (b) Upper bound of ΔZ versus normalized output power back-off.

Appendix B

Reflection and Transmission in N-Port Networks with General Port Impedances



$$\underline{S} = \begin{pmatrix} S_{11} & \cdots & S_{1k} & \cdots & S_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ S_{k1} & \cdots & S_{kk} & \cdots & S_{kN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ S_{N1} & \cdots & S_{Nk} & \cdots & S_{NN} \end{pmatrix}$$

Figure B.1: (a) N-port network (b) Scattering matrix.

Given a general N-port network, depicted in Fig. B.1, the relation between an incident wave a_j at port j and the corresponding reflected wave b_i at port i is described by the scattering matrix $\underline{S}$ and all reflection coefficients Γ_k. In case that all the ports are matched with the reference impedance, the fraction b_i/a_j is directly specified as S_{ij}. For the general case of arbitrary port impedances this ratio cannot be directly read off. It is determined by a rational expression which depends on polynomials in $\underline{S}$ and Γ_k. The complexity for deriving this expression increases with the number of ports and depends on whether an analytical/algebraical or numerical analysis is needed. Besides directly *solving the linear system of equations*, algorithms which ease numerical computer-aided evaluations are the *generalized S-matrix* or the *reduced multi-port* approach. *Mason's loop rule* aims at simplifying analytical manual evaluations. Following, these methods are presented shortly.

B.1 Solving the linear system of equations

The scattering parameters define a set of N complex valued equations with $2N$ unknown complex variables, combined in the vectors

$$\underline{b} = (b_1, \ldots, b_N)^{\mathrm{T}} \quad \text{and} \quad \underline{a} = (a_1, \ldots, a_N)^{\mathrm{T}}.$$

Equation system	Matrix form
$b_1 = S_{11}\, a_1 + \ldots + S_{1j}\, a_j + \ldots + S_{1N}\, a_N$	
$\vdots$	
$b_i = S_{i1}\, a_1 + \ldots + S_{ij}\, a_j + \ldots + S_{iN}\, a_N$	$\underline{b} = \underline{S} \cdot \underline{a}$
$\vdots$	
$b_N = S_{N1}\, a_1 + \ldots + S_{Nj}\, a_j + \ldots + S_{NN}\, a_N$	

When a source is applied at port j, the total incident power wave at this port a_j consists of the original source power wave b_s and the incident port power wave due to reflections $\Gamma_j\, b_j$. The remaining ports are loaded with known reflection coefficients, which yields another N complex equations for $\underline{a}$ and $\underline{b}$.

$$a_k = \begin{cases} \Gamma_k b_k & \text{if } k \neq j \\ \Gamma_k b_k + b_s & \text{if } k = j \end{cases}$$

Elimination of a_k gives a set of N equations in $\underline{b}$:

$b_1 = S_{11}\Gamma_1\, b_1 + \ldots + S_{1j}\, (\Gamma_j\, b_j + b_s) + \ldots + S_{1N}\Gamma_N\, b_N$	
$\vdots$	
$b_i = S_{i1}\Gamma_1\, b_1 + \ldots + S_{ij}\, (\Gamma_j\, b_j + b_s) + \ldots + S_{iN}\Gamma_N\, b_N$	$\underline{b} = \underline{S} \cdot \underline{\Gamma} \cdot \underline{b} + \underline{S}_j b_s$
$\vdots$	
$b_N = S_{N1}\Gamma_1\, b_1 + \ldots + S_{Nj}\, (\Gamma_j\, b_j + b_s) + \ldots + S_{NN}\Gamma_N\, b_N$	

with $\underline{\Gamma} = \mathrm{diag}\,(\Gamma_1, \ldots, \Gamma_N)$ and $\underline{S}_j = (S_{1j}, \ldots, S_{Nj})^T$.

Finally, collecting b_k/b_s results in:

$-S_{1j} = (S_{11}\Gamma_1 - 1)\, \frac{b_1}{b_s} + \ldots + S_{1j}\Gamma_j\, \frac{b_j}{b_s} + \ldots +$	$S_{1N}\Gamma_N \quad \frac{b_N}{b_s}$	
$\vdots$		
$-S_{ij} = S_{i1}\Gamma_1 \quad \frac{b_1}{b_s} + \ldots + S_{ij}\Gamma_j\, \frac{b_j}{b_s} + \ldots +$	$S_{iN}\Gamma_N \quad \frac{b_N}{b_s}$	$-\underline{S}_j = (\underline{S} \cdot \underline{\Gamma} - \underline{U}) \cdot \underline{b}/b_s$
$\vdots$		
$-S_{Nj} = S_{N1}\Gamma_1 \quad \frac{b_1}{b_s} + \ldots + S_{Nj}\Gamma_j\, \frac{b_j}{b_s} + \ldots + (S_{NN}\Gamma_N - 1)\, \frac{b_N}{b_s}$		

with $\underline{U} = \mathrm{diag}\,(1, \ldots, 1)$.

A solution vector can be found by either applying the gaussian elimination algorithm to the extended coefficient matrix

$$\left(\underline{U} - \underline{S} \cdot \underline{\Gamma} \mid \underline{S}_j\right) \tag{B.1}$$

or equivalently taking the inverse of the coefficient matrix

$$\underline{b}/b_s = (\underline{U} - \underline{S} \cdot \underline{\Gamma})^{-1} \cdot \underline{S}_j = (B_{1j}, \ldots, B_{Nj})^T = \underline{B}_j. \tag{B.2}$$

The *transmission coefficient* from port j to port i is:

$$\frac{b_i}{b_s} = B_{ij} \tag{B.3}$$

The *reflection coefficient* at port j is:

$$\left.\frac{b_j}{b_s}\right|_{\Gamma_j=0 \Rightarrow b_s=a_j} = \frac{b_j}{a_j} = \left. B_{jj}\right|_{\Gamma_j=0} \tag{B.4}$$

B.2 Signal flow graphs and Mason's loop rule

A signal flow graph is a method of writing a set of equations, such as $\underline{b} = \underline{S} \cdot \underline{a}$. Each variable becomes a node and each constant becomes a directed branch running from an independent to a dependent node [Medley 93]. Fig. B.2 shows the signal flow

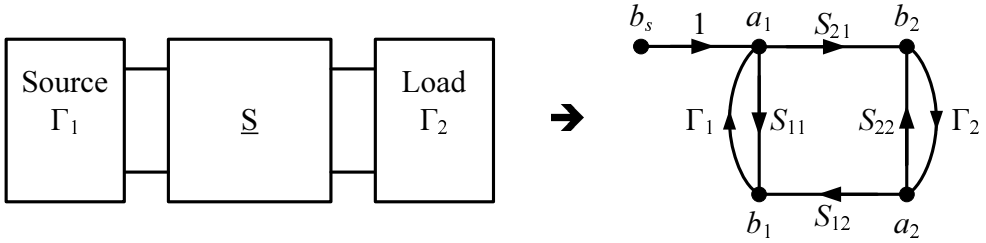


Figure B.2: Signal flow graph representation of generator, 2-port network and load.

graph for a two port network connected to source and load. It has one independent node (b_s), four dependent nodes (a_1, a_2, b_1 and b_2), and seven directed branches running between a_i and b_i, which are related by the constants S_{ij} and Γ_i. Once a microwave network is represented as signal flow graph it can be solved for any combination of independent and dependent variable.

Mason's loop rule or also called *non touching loop rule* provides an algorithm for expressing the ratio of dependent variables (e.g. a_2) to independent variables (e.g. b_s).

$$T_{ij} = \frac{\text{dependent variable at port i}}{\text{independent variable at port j}} = \frac{1}{\Delta G} \sum_{k=1}^{l} P_k \, \Delta^{(k)} \tag{B.5}$$

l	number of different *forward paths* from independent to dependent variable
ΔG	flow graph determinant: $\Delta G = 1 - L_1 + L_2 - L_3 + \ldots$
L_1	sum of all *first order loop gains*
L_i	sum of all i^{th} *order loop gains*.
P_k	k^{th} *forward path gain* from independent to dependent variable
$\Delta^{(k)}$	k^{th} determinant: $\Delta^{(k)} = 1 - L_1^k + L_2^k - L_3^k + \ldots$
L_1^k	sum of all *first order loop gains* not touching path k
L_i^k	sum of all i^{th} *order loop gains* not touching path k

A *forward path* is formed by following a series of directed branches from one node to another such that included branches have the same direction and no node is touched more than once. The *forward path gain* is the product of all branch constants in the path. A *first order loop* is a forward path that begins and ends at the same node and no other nodes may be touched more than once. Its gain is the product of all included branch constants. The *order* i defines the number of non touching first order loops. The i^{th} *order loop gain* is the product of all included first order loop gains.

With help of Mason's loop rule the desired ratios of incident and reflected waves can be calculated on a mixed algebraic and graphical basis. This convenient method has advantages when solving the equations manually, without a computer.

The *transmission coefficient* from port j to port i is:

$$\frac{b_i}{b_s} = T_{ij} \tag{B.6}$$

The *reflection coefficient* at port j is:

$$\frac{b_j}{a_j} = \frac{b_j/b_s}{a_j/b_s} = T_{jj} \tag{B.7}$$

B.3 Generalized S parameters

An N-port network is commonly represented by a scattering matrix, referenced to a uniform real impedance (e.g. $Z_0 = 50\ \Omega$). This approach requires recalculation of the matrix elements, if the network is connected to impedances different from Z_0. A generalized definition of the S-parameters can be given by using different reference impedances at each port. The equation for calculation of the generalized scattering matrix $\underline{S}'$ out of a scattering matrix $\underline{S}$ normalized to Z_0 is [Kurokawa 65,b]:

$$\underline{S}' = \underline{A}^{-1} \left(\underline{S} - \underline{\Gamma}^{*\mathrm{T}}\right) \left(\underline{U} - \underline{\Gamma}\,\underline{S}\right)^{-1} \underline{A}^{*\mathrm{T}} \tag{B.8}$$

with

$$\underline{A} = \mathrm{diag}\left(A_1, \ldots, A_N\right) , \quad \underline{\Gamma} = \mathrm{diag}\left(\Gamma_1, \ldots, \Gamma_N\right) , \tag{B.9}$$

$$A_k = \frac{1-\Gamma_k^*}{|1-\Gamma_k|}\sqrt{1-|\Gamma_k|^2} \quad \text{and} \quad \Gamma_k = \frac{Z_k - Z_{k,0}}{Z_k + Z_{k,0}^*}. \tag{B.10}$$

For these general formulas it is assumed that the old scattering matrix is referenced to $Z_{k,0}$ at port k and the new one is referenced to Z_k at port k. If the original S-parameters are referenced to a uniform impedance Z_0 then all $Z_{k,0} = Z_0$. With knowledge of $\underline{S}'$ the desired fractions of forward and reflected waves are easily obtained.

The *transmission coefficient* from port j to port i is:

$$\frac{b_i}{b_s} = S'_{ij}\,\frac{A_i}{A_j^*}\,\frac{1}{1-|\Gamma_i|^2} \tag{B.11}$$

The *reflection coefficient* at port j is:

$$\frac{b_j}{a_j} = S'_{jj}\,\frac{A_j}{A_j^*}\,\frac{1}{1-|\Gamma_j|^2}\bigg|_{\Gamma_i=\Gamma_j=0} = S'_{jj}\Big|_{\Gamma_j=0} \tag{B.12}$$

B.4 Reduced multi-port

An N-port network having $(N-n)$ ports terminated in fixed loads is called a *reduced multi-port*. The connected ports are not available any more, so that the resulting quadratic scattering matrix has size n. A derivation of the *reduced n-port scattering matrix* $\underline{S}^{\mathrm{R}}$ is given in [Otoshi 69] and will be outlined in this section.

The original N-port scattering matrix is partitioned after the n^{th} row and n^{th} column, as expressed by

$$\left(\begin{array}{c} b_1 \\ b_2 \\ \vdots \\ b_n \\ \hline b_{n+1} \\ \vdots \\ b_N \end{array}\right) = \left(\begin{array}{cccc|ccc} S_{11} & S_{12} & \dots & S_{1n} & S_{1(n+1)} & \dots & S_{1N} \\ S_{21} & S_{22} & \dots & S_{2n} & S_{2(n+1)} & \dots & S_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \dots \\ S_{n1} & S_{n2} & \dots & S_{nn} & S_{n(n+1)} & \dots & S_{nN} \\ \hline S_{(n+1)1} & S_{(n+1)2} & \dots & S_{(n+1)n} & S_{(n+1)(n+1)} & \dots & S_{(n+1)N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \dots \\ S_{N1} & S_{N2} & \dots & S_{Nn} & S_{N(n+1)} & \dots & S_{NN} \end{array}\right) \cdot \left(\begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_n \\ \hline a_{n+1} \\ \vdots \\ a_N \end{array}\right), \tag{B.13}$$

which can then be rewritten as

$$\left(\begin{array}{c} \underline{B}_1 \\ \hline \underline{B}_2 \end{array}\right) = \left(\begin{array}{c|c} \underline{S}_{11}^{\mathrm{P}} & \underline{S}_{12}^{\mathrm{P}} \\ \hline \underline{S}_{21}^{\mathrm{P}} & \underline{S}_{22}^{\mathrm{P}} \end{array}\right) \cdot \left(\begin{array}{c} \underline{A}_1 \\ \hline \underline{A}_2 \end{array}\right) \tag{B.14}$$

With the load matrix $\underline{\Gamma}_{\mathrm{L}} = \mathrm{diag}\,(\Gamma_{n+1}, \dots, \Gamma_N)$, containing all load reflection coefficients from port $n+1$ to N, the incident power waves at the terminated ports can be expressed by:

$$\underline{A}_2 = \underline{\Gamma}_{\mathrm{L}}\,\underline{B}_2 \tag{B.15}$$

Substituting (B.15) into (B.14) and performing the matrix multiplication yields

$$\left.\begin{array}{rcl} \underline{B}_1 & = & \underline{S}^{\mathrm{P}}_{11}\,\underline{A}_1 + \underline{S}^{\mathrm{P}}_{12}\,\underline{\Gamma}_{\mathrm{L}}\,\underline{B}_2 \\ \underline{B}_2 & = & \underline{S}^{\mathrm{P}}_{21}\,\underline{A}_1 + \underline{S}^{\mathrm{P}}_{22}\,\underline{\Gamma}_{\mathrm{L}}\,\underline{B}_2 \end{array}\right\} \rightarrow \underline{B}_1 = \underline{S}^{\mathrm{R}}\,\underline{A}_1\,, \tag{B.16}$$

where $\underline{S}^{\mathrm{R}}$ is the desired scattering matrix of the reduced multi-port:

$$\underline{S}^{\mathrm{R}} = \underline{S}^{\mathrm{P}}_{11} + \underline{S}^{\mathrm{P}}_{12}\,\underline{\Gamma}_{\mathrm{L}}\,\left(\underline{U} - \underline{S}^{\mathrm{P}}_{22}\,\underline{\Gamma}_{\mathrm{L}}\right)^{-1}\,\underline{S}^{\mathrm{P}}_{21} \tag{B.17}$$

The unity matrix $\underline{U} = \mathrm{diag}(1,\ldots,1)$ has size $(N-n)$.
The single elements of $\underline{S}^{\mathrm{R}}$ can be calculated directly, as well:

$$S^{\mathrm{R}}_{ij} = \frac{\det\begin{pmatrix} S_{ij} & {}^{i}\underline{\mathrm{S}}^{\mathrm{P}} \\ {}^{j}\underline{\mathrm{S}}^{\mathrm{P}} & \underline{U} - \underline{S}^{\mathrm{P}}_{22}\,\underline{\Gamma}_{\mathrm{L}} \end{pmatrix}}{\det\left(\underline{U} - \underline{S}^{\mathrm{P}}_{22}\,\underline{\Gamma}_{\mathrm{L}}\right)}\,, \tag{B.18}$$

where

$$i = 1,2,\ldots,n\,, \quad j = 1,2,\ldots,n\,, \tag{B.19}$$

$${}^{i}\underline{\mathrm{S}}^{\mathrm{P}} = \text{row vector consisting of } i^{\text{th}} \text{ row of matrix } \left(-\underline{S}^{\mathrm{P}}_{12}\underline{\Gamma}_{\mathrm{L}}\right)\,, \tag{B.20}$$

$${}^{j}\underline{\mathrm{S}}^{\mathrm{P}} = \text{column vector consisting of } j^{\text{th}} \text{ column of matrix } \left(\underline{S}^{\mathrm{P}}_{21}\right)\,. \tag{B.21}$$

Now it is assumed, that the original S-matrix is referenced to equal and real impedances at all ports. The terminations of all ports have to be known and the reduced multi port matrix, $\underline{S}^{\mathrm{R},ij}$ between source and load ports j and i, is calculated. The resulting matrix is a two-port scattering matrix, with four entries:

$$\underline{S}^{\mathrm{R},ij} = \begin{pmatrix} S^{\mathrm{R},ij}_{11} & S^{\mathrm{R},ij}_{12} \\ S^{\mathrm{R},ij}_{21} & S^{\mathrm{R},ij}_{22} \end{pmatrix} \tag{B.22}$$

For this reduced two port the *transmission coefficient* from port j to port i is:

$$\frac{b_i}{b_s} = \frac{S^{\mathrm{R},ij}_{21}}{(S^{\mathrm{R},ij}_{11}\,\Gamma_j - 1)(S^{\mathrm{R},ij}_{22}\,\Gamma_i - 1) - S^{\mathrm{R},ij}_{12}\,S^{\mathrm{R},ij}_{21}\,\Gamma_j\Gamma_i} \tag{B.23}$$

If the multi port network is reduced to one port j, the *reflection coefficient* at port j is:

$$\frac{b_j}{a_j} = S^{\mathrm{R}}_{jj} \tag{B.24}$$

Alternatively, the reflection coefficient can also be calculated from the reduced two port scattering matrix:

$$\frac{b_j}{a_j} = S^{\mathrm{R},ij}_{11} + \frac{S^{\mathrm{R},ij}_{12}\,S^{\mathrm{R},ij}_{21}\,\Gamma_i}{1 - S^{\mathrm{R},ij}_{22}\,\Gamma_i} \tag{B.25}$$

Appendix C

Gain Definitions

For analysis of RF networks the gain, which is a ratio of power at the input and power at the output, plays an important role. As this power transfer can be influenced by different load and source impedances of the network, it is necessary to state the gain definition more precisely. For this purpose the meaning of input and output power are specified in case if they are related to source and load impedance or if they are not.

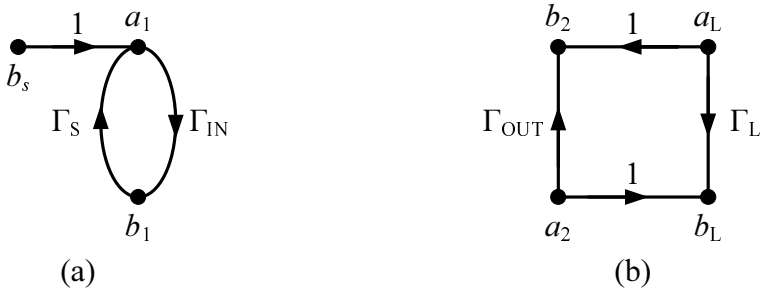


Figure C.1: Signal flow graph for (a) input power (b) output power definition.

At the input there exist two power definitions:

- Power into the network (P_{in}): This is the actual power at the input port of the network. It is the source power being reduced by reflections due to mismatch between generator and input. From Fig. C.1(a) this can be written in terms of incident and reflected waves and simplified with Mason's loop rule:

$$P_{\text{in}} = \frac{1}{2}\left(|a_1|^2 - |b_1|^2\right) = \frac{|b_{\text{s}}|^2}{2}\frac{1-|\Gamma_{\text{IN}}|^2}{1-|\Gamma_{\text{S}}\Gamma_{\text{IN}}|^2} \tag{C.1}$$

- Power available from the source (P_{avS}): This is the maximum power that can be delivered by the generator to a connected load ($\max(P_{\text{in}})$). It is obtained when the load is conjugately matched to the generator, i.e. without reflections:

$$P_{\text{avS}} = P_{\text{in}}\Big|_{\Gamma_{\text{IN}}=\Gamma_{\text{S}}^*} = \frac{|b_{\text{s}}|^2}{2}\frac{1}{1-|\Gamma_{\text{IN}}|^2} \tag{C.2}$$

The two power definitions at the output are:

- Power delivered to the load (P_{out}): This is the power actually delivered to a connected load at the output. Reflections between the network output and the load are contained in this definition. In terms of forward and reflected waves from Fig. C.1 (b) it is written as:

$$P_{\text{out}} = \frac{1}{2}\left(|a_{\text{L}}|^2 - |b_{\text{L}}|^2\right) = \frac{|b_{\text{s}}|^2}{2}\left|\frac{b_2}{b_{\text{s}}}\right|^2\left(1 - |\Gamma_{\text{L}}|^2\right) \tag{C.3}$$

  The fraction of outgoing power to incoming power $\frac{b_2}{b_{\text{s}}}$ is calculated with Mason's loop rule:

$$\frac{b_2}{b_{\text{s}}} = \frac{S_{21}}{(1-\Gamma_{\text{S}}S_{11})(1-\Gamma_{\text{L}}\Gamma_{\text{OUT}})} = \frac{S_{21}}{(1-\Gamma_{\text{S}}\Gamma_{\text{IN}})(1-\Gamma_{\text{L}}S_{22})} \tag{C.4}$$

- Power available from the network (P_{avN}): This is the maximum power that can be delivered by the network to a connected load ($\max(P_{\text{out}})$). It is obtained when the load is conjugately matched to the network's output impedance:

$$P_{\text{avN}} = P_{\text{out}}\Big|_{\Gamma_{\text{L}}=\Gamma_{\text{out}}^*} = \frac{|b_{\text{s}}|^2}{2}\left(1 - |\Gamma_{\text{OUT}}|^2\right)\left|\frac{b_2}{b_{\text{s}}}\right|^2_{\Gamma_{\text{L}}=\Gamma_{\text{OUT}}^*} \tag{C.5}$$

The input and output reflection coefficients of the network are in general calculated as:

$$\Gamma_{\text{IN}} = S_{11} + \frac{S_{12}S_{21}\Gamma_{\text{L}}}{1 - S_{22}\Gamma_{\text{L}}} \quad \text{and} \quad \Gamma_{\text{OUT}} = S_{22} + \frac{S_{12}S_{21}\Gamma_{\text{S}}}{1 - S_{11}\Gamma_{\text{S}}} \tag{C.6}$$

Deduced from these power definitions, there are four possible gain definitions: *Transducer gain*, *operating power gain*, *available power gain* and *maximum available gain*. They are summarized in Table C.1.

Gain definition	**Transducer Gain**	**Operating Power Gain**
	$G_{\mathrm{T}} = \dfrac{P_{\mathrm{out}}}{P_{\mathrm{avS}}}$	$G_{\mathrm{P}} = \dfrac{P_{\mathrm{out}}}{P_{\mathrm{in}}}$
	$\dfrac{\text{Power delivered to the load}}{\text{Power available from the source}}$	$\dfrac{\text{Power delivered to the load}}{\text{Power into the network}}$
Load condition	Γ_{S} arbitrary & Γ_{L} arbitrary	$\Gamma_{\mathrm{S}} = \Gamma_{\mathrm{IN}}^{*}$ & Γ_{L} arbitrary
S-parameters	$\dfrac{1-\lvert\Gamma_{\mathrm{S}}\rvert^2}{\lvert 1-\Gamma_{\mathrm{S}}S_{11}\rvert^2}\cdot\lvert S_{21}\rvert^2\cdot\dfrac{1-\lvert\Gamma_{\mathrm{L}}\rvert^2}{\lvert 1-\Gamma_{\mathrm{L}}\Gamma_{\mathrm{OUT}}\rvert^2}$ $\dfrac{1-\lvert\Gamma_{\mathrm{S}}\rvert^2}{\lvert 1-\Gamma_{\mathrm{S}}\Gamma_{\mathrm{IN}}\rvert^2}\cdot\lvert S_{21}\rvert^2\cdot\dfrac{1-\lvert\Gamma_{\mathrm{L}}\rvert^2}{\lvert 1-\Gamma_{\mathrm{L}}S_{22}\rvert^2}$	$\dfrac{1}{1-\lvert\Gamma_{\mathrm{IN}}\rvert^2}\cdot\lvert S_{21}\rvert^2\cdot\dfrac{1-\lvert\Gamma_{\mathrm{L}}\rvert^2}{\lvert 1-\Gamma_{\mathrm{L}}S_{22}\rvert^2}$
Gen. S-parameters	$\lvert S'_{21}\rvert^2$	$\dfrac{\lvert S'_{21}\rvert^2}{1-\lvert S'_{11}\rvert^2}$

Gain definition	**Available Power Gain**	**Maximum Available Gain**
	$G_{\mathrm{A}} = \dfrac{P_{\mathrm{avN}}}{P_{\mathrm{avS}}}$	$G_{\mathrm{MAG}} = \dfrac{P_{\mathrm{avN}}}{P_{\mathrm{in}}}$
	$\dfrac{\text{Power available from the network}}{\text{Power available from the source}}$	$\dfrac{\text{Power available from the network}}{\text{Power into the network}}$
Load condition	Γ_{S} arbitrary & $\Gamma_{\mathrm{L}} = \Gamma_{\mathrm{OUT}}^{*}$	$\Gamma_{\mathrm{S}} = \Gamma_{\mathrm{IN}}^{*} = \Gamma_{\mathrm{mS}}$ & $\Gamma_{\mathrm{L}} = \Gamma_{\mathrm{OUT}}^{*} = \Gamma_{\mathrm{mL}}$
S-parameters	$\dfrac{1-\lvert\Gamma_{\mathrm{S}}\rvert^2}{\lvert 1-\Gamma_{\mathrm{S}}S_{11}\rvert^2}\cdot\lvert S_{21}\rvert^2\cdot\dfrac{1}{1-\lvert\Gamma_{\mathrm{OUT}}\rvert^2}$	$\dfrac{1-\lvert\Gamma_{\mathrm{mS}}\rvert^2}{\lvert 1-S_{11}\Gamma_{\mathrm{mS}}\rvert^2}\cdot\lvert S_{21}\rvert^2\cdot\dfrac{1-\lvert\Gamma_{\mathrm{mL}}\rvert^2}{\lvert 1-S_{22}\Gamma_{\mathrm{mL}}\rvert^2}$
Gen. S-parameters	$\dfrac{\lvert S'_{21}\rvert^2}{1-\lvert S'_{22}\rvert^2}$	$\lvert S'_{21}\rvert^2_{\Gamma_{\mathrm{mS}}=S_{11},\Gamma_{\mathrm{mL}}=S_{22}}$

Table C.1: Gain definitions and their representation with S-parameters and generalized S-parameters (indicated by the inverted comma)

The relations in terms of S-parameters are simply calculated from the power definitions above. These expressions can be used for the relations in terms of generalized S-parameters by first substituting $\Gamma_\mathrm{L} = \Gamma_\mathrm{S} \Rightarrow 0$ and then $S_\mathrm{ij} \Rightarrow S'_\mathrm{ij}$. The reason for this simple substitution can be found in the definition of the generalized scattering parameters. Each port can be referenced to an individual impedance. If the network is actually connected to its reference impedance, then the port is, per definition, terminated without reflection. This means that the reflected power wave at the output $|a'_2|^2$ is zero, which simplifies the expressions of a'_i and b'_i in terms of S'_{ij}:

$$b'_1 = a'_1 S'_{11} + \underbrace{a'_2}_{0} S'_{12} \tag{C.7}$$

$$b'_2 = a'_1 S'_{21} + \underbrace{a'_2}_{0} S'_{12} \tag{C.8}$$

Therefore also the expressions of power are simplified:

$$P_\mathrm{in} = \frac{1}{2}\left(|a'_1|^2 - |b'_1|^2\right) = \frac{|a'_1|^2}{2}\left(1 - |S'_{11}|^2\right) \ , \quad P_\mathrm{avS} = \frac{|a'_1|^2}{2} \ ,$$

$$P_\mathrm{out} = \frac{|b'_2|^2}{2} = \frac{|a'_1|^2}{2}|S'_{21}|^2 \ , \qquad P_\mathrm{avN} = \frac{|a'_1|^2}{2}\frac{|S'_{21}|^2}{\left(1 - |S'_{22}|^2\right)}$$

Regardless of whether generalized or normal scattering parameters are used, the definitions and expressions are only valid for relation of power from one input source port to one output load port. This requires two port scattering matrices. Any linear N-port network can be reduced to a two port network (one input and one output) with the remaining ports being terminated in any load. The above used S_{11}, S_{22}, S_{21} and S_{12} then can be derived from the multi-port as described in Appendix B.4.

Gain definitions with more than one source or load port included - as is the case for power combiners or splitters - are based on the ratio of input or output power sums. The calculation in this case differs from the above but can be done using the presented general power definitions with incident and reflected power waves.

Appendix D

Multi-Port Network Connection

The complete scattering matrix $\underline{S}_{\text{tot}}$ of two interconnected arbitrary multi-port networks with individual scattering matrices $\underline{S}_1$ and $\underline{S}_2$ is readily obtained using the connection matrix approach, as described in [Medley 93]. For that algorithm the

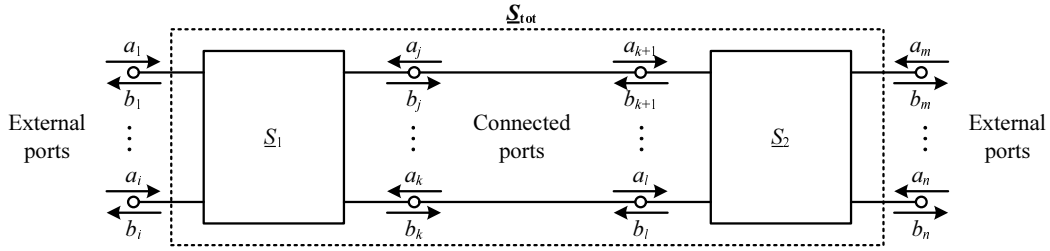


Figure D.1: Multiport interconnection.

ports of the considered networks are first numbered consecutively: $1 \ldots k$ for the first network and $k+1 \ldots n$ for the second network, as shown in Fig. D.1. With this unique index definition the scattering matrices $\underline{S}_1$ and $\underline{S}_2$ are added together along the main diagonal of the auxiliary matrix $\underline{S}$. After that the ports are partitioned into external unconnected ports (subscript e) and internal connected ports (subscript c). According to this partition the matrix $\underline{S}$ is rearranged so that the first rows/columns contain all elements with external port number indices (matrix S_{ee}) and the remaining rows/columns contain the elements with connected port number indices (matrix S_{pp}). The secondary diagonal then contains mixed terms of external and connected ports $\underline{S}_{\text{ec}}$ and $\underline{S}_{\text{ce}}$:

$$\underline{S} = \begin{pmatrix} \underline{S}_1 & \underline{0} \\ \underline{0} & \underline{S}_2 \end{pmatrix} \quad \xrightarrow{\text{rarranged}} \quad \begin{pmatrix} \underline{S}_{\text{ee}} & \underline{S}_{\text{ec}} \\ \underline{S}_{\text{ce}} & \underline{S}_{\text{cc}} \end{pmatrix} \tag{D.1}$$

After that the so called connection matrix $\underline{\Gamma}_{\text{c}}$ is determined, which relates the incident and reflected waves at the connected ports: $\underline{b}_{\text{c}} = \underline{\Gamma}_{\text{c}} \underline{a}_{\text{c}}$. The connection matrix is determined by inspection of the connected ports indices. For that purpose the

indices of the connected ports $j \dots l$ are written besides the rows and columns of a matrix of same dimension and port order as $\underline{S}_{\mathrm{cc}}$:

$$\underline{\Gamma}_{\mathrm{c}} = \begin{array}{c} \\ j \\ \vdots \\ k \\ k+1 \\ \vdots \\ l \end{array} \begin{array}{c} \begin{array}{cccccc} j & \cdots & k & k+1 & \cdots & l \end{array} \\ \begin{pmatrix} 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \\ 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix} \end{array} \tag{D.2}$$

The connection of ports with e.g. index j and $k+1$ is indicated by entering 1 at the intersection of the corresponding row and column number. The remaining elements in the row and column are set equal to 0. This is done for all connected ports, until every row and column is filled.

The complete scattering matrix $\underline{S}_{\mathrm{tot}}$ of the connected multi-ports is then calculated by:

$$\underline{S}_{\mathrm{tot}} = \underline{S}_{\mathrm{ee}} + \underline{S}_{\mathrm{ec}}(\underline{\Gamma}_{\mathrm{c}} - \underline{S}_{\mathrm{cc}})^{-1}\underline{S}_{\mathrm{ce}}, \tag{D.3}$$

Acknowledgements

I would like to express sincere appreciation to my advisor Prof. Dr. Arpad L. Scholtz for his guidance and support throughout my studies and for providing me with the opportunity to work at the wireless communications department of Infineon Munich. My gratitude also goes to Prof. Dr. Robert Weigel for dedicating his time in reviewing my thesis. I am thankful to Prof. Dr. Jan-Erik Müller for careful reading of the manuscript with valuable suggestions for improvement and his support throughout my work and efforts in organizational concerns.

I would like to thank my colleagues Michael Asam, Kevni Bueyuektas, Helmut Herrmann, Boris Kapfelsperger, Wolfgang Thomann and Dr. Markus Zannoth of the former power amplifier development group at Infineon Munich. Their practical experience and participation in many inspiring discussions, which led to new ideas for the realization of my work, were invaluable in my first year as doctoral student. Special thanks go to Dr. Winfried Bakalski not only for being a constant source of help and suggestions over the years but also for offering much motivation and for reviewing this thesis.

I am also grateful to my colleagues from the Technical University of Berlin. I want to thank Norman Wolf for his significant support in programming and automation of the measurement setup and Stephan Leuschner for many inspiring discussions and critical comments on this thesis.

I further like to thank my other colleagues at Infineon Munich, namely, Dr. Ronald Thueringer, Mohit Berry and Sandro Pinarello for useful tips and hints in technical concerns and for keeping things light with varied discussions about finance, politics, life in general and other non-technical topics.

Finally, I would like to thank my mother for her sacrifice and the foundation that she has provided throughout the years and that she always encouraged me in an initiative and sensitive way during my work. I want to extend special thanks to my partner Andrea and her family for their warmth and support and for standing by me in difficult times, never doubting my abilities.

Bibliography

[3GPP 07,a] 3GPP, "3rd Generation Partnership Project; Technical Specifiaction Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception - 3GPP TS 36.101 V9.1.0 (2009-09)", www.3gpp.org, Sept. 2007.

[3GPP 07,b] 3GPP, "3rd Generation Partnership Project; Technical Specifiaction Group Radio Access Network; User Equipment (UE) radio transmission and reception (TDD)- 3GPP TS 25.102 V8.4.0 (2009-05)", www.3gpp.org, Sept. 2007.

[3GPP 09,a] 3GPP, "3rd Generation Partnership Project; Technical Specifiaction Group GSM/EDGE Radio Access Network; Radio transmission and reception - 3GPP TS 45.005 V8.5.0 (2009-05)", www.3gpp.org, Feb. 2009.

[3GPP 09,b] 3GPP, "3rd Generation Partnership Project; Technical Specifiaction Group Radio Access Network; User Equipment (UE) radio transmission and reception (FDD)- 3GPP TS 25.101 V8.7.0 (2009-05)", www.3gpp.org, Mai. 2009.

[Agilent 02] Agilent, "Using Error Vector Magnitude Measurements to Analyze and Troubleshoot Vector-Modulated Signals", Agilent PN 89400-14 product note from http://www.agilent.com, 2002.

[Anadigics 09] Anadigics, "Model:AWT6223R", Internet http://www.anadigics.com/, 2009.

[Andrews 06] David Andrews, *Lumped Element Quadrature Hybrids*, Artech House, Inc., 2006.

[Apel 07] T. Apel, Yu-Lung Tang, and O. Berger, "Switched Doherty Power Amplifiers for CDMA and WCDMA", *Radio Frequency Integrated Circuits (RFIC) Symposium, 2007 IEEE*, vol. , pp. 259–262, June 2007.

[Avago 09] Avago, "Model:ACPM-7371", Internet http://www.avagotech.com/, 2009.

[Bakalski 03] Winfried Bakalski, *Integrated Microwave Power Amplifiers*, PhD thesis, Technical University of Vienna, Austria, 2003.

[Bakalski 04] W. Bakalski, M. Asam, M. Zannoth, and K. Kitlinski, "Optimiertes Anschlußverteilungsnetzwerk für große Transistoren", German patent pending, disclosure no. 2006E50266DE, 2004.

[Bakalski 08,a] W. Bakalski, B. Sogl, M. Zannoth, M. Asam, B. Kapfelsperger, J. Berkner, B. Eisener, W. Thomann, S. Marcon, W. Osterreicher, E. Napieralska, E. Rampf, A.L. Scholtz, and B.-U. Klepser, "A Quad-Band GSM/EDGE-Compliant SiGe-Bipolar Power Amplifier", *Solid-State Circuits, IEEE Journal of*, vol. 43, pp. 1920–1930, Sept. 2008.

[Bakalski 08,b] W. Bakalski, M. Zannoth, M. Asam, W. Thomann, B. Kapfelsperger, P. Pfann, J. Berkner, C. Hepp, A. Steltenpohl, W. Osterreicher, and E. Rampf, "A load-insensitive quad-band GSM/EDGE SiGeC-bipolar power amplifier with a highly efficient low power mode", *in Radio and Wireless Symposium, 2008 IEEE*, pp. 203–206, Jan. 2008.

[Berkner 02] Jörg Berkner, *Kompaktmodelle für Bipolartransistoren*, Expert Verlag, first edition, 2002.

[Bezooijen 05] A. van Bezooijen, R. Mahmoudi, and A.H.M. van Roermund, "Adaptive methods to preserve power amplifier linearity under antenna mismatch conditions", *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 52, pp. 2101–2108, Oct. 2005.

[Birafane 04] A. Birafane and A.B. Kouki, "On the linearity and efficiency of outphasing microwave amplifiers", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 52, pp. 1702–1708, July 2004.

[Blanken 08] P.G. Blanken, R. Karadi, and H.J. Bergveld, "A 50MHz bandwidth multi-mode PA supply modulator for GSM, EDGE and UMTS application", *Radio Frequency Integrated Circuits Symposium, 2008. RFIC 2008. IEEE*, vol. , pp. 401–404, 17 2008-April 17 2008.

[Bode 45] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, D. Van Nostrand, Inc., 1945.

[Campbell 99] C.F. Campbell, "A fully integrated Ku-band Doherty amplifier MMIC", *Microwave and Guided Wave Letters, IEEE*, vol. 9, pp. 114–116, Mar 1999.

[Carlin 56] H. Carlin, "The scattering matrix in network theory", *Circuit Theory, IRE Transactions on*, vol. 3, pp. 88–97, Jun 1956.

[Chireix 35] H. Chireix, "High power outphasing modulation", *Proceedings IRE*, vol. 23, pp. 1370–1392, November 1935.

[Coilcraft 08] Coilcraft, "0403HQ Series (1008)", Internet http://www.coilcraft.com/, 2008.

[Cripps 99] Steve. C. Cripps, *RF Power Amplifiers for Wireless Communications*, Artech House, 1999.

[Cripps 02] Steve. C. Cripps, *Advanced Techniques in RF Power Amplifier Design*, Artech House, 2002.

[Cripps 06] Steve C. Cripps, *RF Power Amplifiers for Wireless Communications*, Artech House, Norwood, MA 02062, second edition, 2006.

[Doherty 36] W.H. Doherty, "A new efficiency power amplifier for modulated waves", *Proceedings IRE*, vol. 24, pp. 1163–1182, September 1936.

[Egolf 07] Steve Egolf, "Intelligent Power Management: A Method to Improve 2G/3G Handset Talk Time", *Microwave Journal*, vol. 50, pp. 94–100, July 2007.

[Eom 94] Soon-Young Eom, Jeong-Ho Kim, , and Choon-Sik Yim, "New N-way Hybrid Power Combiner to Improve the Graceful Degradation Performance", *ETRI Journal*, vol. 16, pp. 58–72, Apr 1994.

[Epcos 08] Epcos, "HQF0402", Internet http://www.epcos.com/, 2008.

[Ernst 77] R.L. Ernst, R.L. Camisa, and A. Presser, "Graceful Degradation Properties of Matched N-Port Power Amplifier Combiners", *Microwave Symposium Digest, MTT-S International*, vol. 77, pp. 174–177, Jun 1977.

[Fano 50] R. M. Fano, "Theoretical Limitation on the Broadband Matching of Arbitrary Impedances", *Journal of Franclin Institute*, vol. , Jan.-Feb. 1950.

[Fidler 92] J.K. Fidler and Y. Sun, "Computer-aided determination of impedance matching domain", pp. 1/1–1/6, Nov 1992.

[Focus microw 08] Focus microwave, "Load Pull Tuner", Internet http://www.focus-microwaves.com/, 2008.

[Freescale Se 08] Freescale Semiconductor, "RFX300-30 Datasheet", Internet http://www.freescale.com/, 2008.

[Frlan 89] E. Frlan, S. Meszaros, M. Chuaci, and J.S. Wight, "Computer aided design of square spiral transformers and inductors", *in IEEE J. MTT-S Dig*, pp. 661–664, 1989.

[Gonzales 84] G. Gonzales, *Microwave Transistor Amplifiers Analysis and Design*, Prentice Hall, Englewood Cliffs, NJ 07632, 1984.

[Gotch 07] Damian Gotch, "A Review of Technological Advances in Solid-state Switches", *Microwave Journal*, vol. 50, pp. 24–34, November 2007.

[Grebennikov 07] A. Grebennikov, B. Sogl, H. Herrmann, C. Roth, and W. Thomann, "High-efficiency balanced switched-path monolithic SiGe HBT power amplifiers for wireless applications", *Microwave Conference, 2007. European*, vol. , pp. 1189–1192, Oct. 2007.

[Greenhouse 74] H. M. Greenhouse, "Design of planar rectangular microelectronic inductors", *in IEEE Trans. Parts, Hybrids, Pack.*, pp. 101–109, June 1974.

[Hakala 05] I. Hakala, D.K. Choi, L. Gharavi, N. Kajakine, J. Koskela, and R. Kaunisto, "A 2.14-GHz Chireix outphasing transmitter", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 53, pp. 2129–2138, June 2005.

[Haldi 08] P. Haldi, D. Chowdhury, P. Reynaert, Gang Liu, and A.M. Niknejad, "A 5.8 GHz 1 V Linear Power Amplifier Using a Novel On-Chip Transformer Power Combiner in Standard 90 nm CMOS", *Solid-State Circuits, IEEE Journal of*, vol. 43, pp. 1054–1063, May 2008.

[Hammi 06] O. Hammi, J. Sirois, S. Boumaiza, and F.M. Ghannouchi, "Design and performance analysis of mismatched Doherty amplifiers using an accurate load-pull-based model", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 54, pp. 3246–3254, Aug. 2006.

[Hammi 07] Oualid Hammi, Jerome Sirois, Slim Boumaiza, and Fadhel M. Ghannouchi, "Study of the Output Load Mismatch Effects on the Load Modulation of Doherty Power Amplifiers", *Radio and Wireless Symposium, 2007 IEEE*, vol. , pp. 393–394a, Jan. 2007.

[Han 08] Kichon Han and Bumman Kim, "A Parallel Power Amplifier With a Novel Mode Switching Control", *Microwave and Wireless Components Letters, IEEE*, vol. 18, pp. 200–202, March 2008.

[Hau 01] G. Hau, T.B. Nishimura, and N. Iwata, "A highly efficient linearized wide-band CDMA handset power amplifier based on predistortion under various bias conditions", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 49, pp. 1194–1201, Jun 2001.

[Infineon 05] Infineon and Lucent, "Abschlussbericht zum BMBF-Verbundprojekt -HitSpeed- "Höchstperformante SiGe-Technologien fr High-Speed-Kommunikationsnetze" (01M3139A)", Bundesministerium für Bildung und Forschung (BMBF), 2005.

[Infineon 08] Infineon, "BAR65 Datasheet", Internet http://www.infineon.com/, 2008.

[Infineon Tec 08] Infineon Technologies, "Smarti UE PMB 5703 Datasheet", Internet http://www.infineon.com/, 2008.

[Iwamoto 01] M. Iwamoto, A. Williams, Pin-Fan Chen, A.G. Metzger, L.E. Larson, and P.M. Asbeck, "An extended Doherty amplifier with high efficiency over a wide power range", *in Transactions on Microwave Theory and Techniques*, Phoenix, AZ, USA, December 2001. IEEE.

[Kee 03] S.D. Kee, I. Aoki, A. Hajimiri, and D.B. Rutledge, "The class-E/F family of ZVS switching amplifiers", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 51, pp. 1677–1690, June 2003.

[Keerti 05] A. Keerti and A. Pham, "Dynamic output phase to adaptively improve the linearity of power amplifier under antenna mismatch", pp. 675–678, June 2005.

[Kennington 00] Peter B. Kennington, *High-Linearity RF Amplifier Design*, Artech House, 2000.

[King 00] Joel King and Hillsboro R., "High VSWR mismatch output stage", October 2000.

[Klein 99] W. Klein and Klepser, B.U.H., "75GHz Bipolar Production Technology for the 21th Century", *in Proceedings of the ESSDERC 1999*, pp. 88–94. IEEE, 1999.

[Kobayashi 00] K.W. Kobayashi, A.K. Oki, A. Gutierrez-Aitken, P. Chin, Li Yang, E. Kaneshiro, P.C. Grossman, K. Sato, T.R. Block, H.C. Yen, and D.C. Streit, "An 18-21 GHz InP DHBT linear microwave Doherty amplifier", *Radio Frequency Integrated Circuits (RFIC) Symposium, 2000. Digest of Papers. 2000 IEEE*, vol. , pp. 179–182, 2000.

[Krauss 80] H. L. Krauss, C.W. Bostian, and F.H. Raab, *Solid State Radio Engineering*, New York Wiley, 1980.

[Kroemer 57] H. Kroemer, "Theory of a wide-gap emitter for transistors", *in Proceedings of the IRE 45*, pp. 1535–1537, 1957.

[Kurokawa 65,a] K. Kurokawa, "Design Theory of Balanced Transistor Amplifiers", *Bell System Technical Journal*, vol. , pp. 1675–1689, Oct. 1965.

[Kurokawa 65,b] K. Kurokawa, "Power Waves and the Scattering Matrix", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 13, pp. 194–202, Mar 1965.

[Lawrence 72] J. D. Lawrence, *A Catalog of Special Plane Curves*, New York: Dover, 1972.

[Lee 98] Thomas H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998.

[Lee 05] Jaewoo Lee, Chang-Han Je, Seongweon Kang, and Chang-Auck Choi, "A single-pole 6-throw (SP6T) antenna switch using metal-contact RF MEMS switches for multi-band applications", *Microwave Symposium Digest, 2005 IEEE MTT-S International*, vol. , pp. 4 pp.–, June 2005.

[Lee 06] Jongsoo Lee, J. Potts, and E. Spears, "DC/DC converter controlled power amplifier module for WCDMA applications", *Radio Frequency Integrated Circuits (RFIC) Symposium, 2006 IEEE*, vol. , pp. 4 pp.–, June 2006.

[Lee 08] Yong-Sub Lee, Mun-Woo Lee, and Yoon-Ha Jeong, "Highly linear and efficient Doherty amplifier employing power tracking bias supply scheme for WCDMA applications", *Radio and Wireless Symposium, 2008 IEEE*, vol. , pp. 73–76, Jan. 2008.

[Liu 06] Gang Liu, Tsu-Jae King Liu, and A.M. Niknejad, "A 1.2V, 2.4GHz Fully Integrated Linear CMOS Power Amplifier with Efficiency Enhancement", *Custom Integrated Circuits Conference, 2006. CICC '06. IEEE*, vol. , pp. 141–144, Sept. 2006.

[Liu 08] Gang Liu, P. Haldi, Tsu-Jae King Liu, and A.M. Niknejad, "Fully Integrated CMOS Power Amplifier With Efficiency Enhancement at Power Back-Off", *Solid-State Circuits, IEEE Journal of*, vol. 43, pp. 600–609, March 2008.

[Masuda 98] M. Masuda, N. Ohbata, H. Ishiuchi, K. Onda, and R. Yamamoto, "High power heterojunction GaAs switch IC with P-1 dB of more than 38 dBm for GSM application", *Gallium Arsenide Integrated*

Circuit (GaAs IC) Symposium, 1998. Technical Digest 1998., 20th Annual, vol. , pp. 229–232, Nov 1998.

[Mauri microw 08] Mauri microwave incorporation, "Load Pull Tuner", Internet http://www.marymw.com/, 2008.

[Maury 09] Maury, "Model:8045P", Internet http://www.maurymw.com/, 2009.

[McCarroll 00] C.P. McCarroll, G.D. Alley, S. Yates, and R. Matreci, "A 20 GHz Doherty power amplifier MMIC with high efficiency and low distortion designed for broad band digital communication systems", *Microwave Symposium Digest., 2000 IEEE MTT-S International*, vol. 1, pp. 537–540 vol.1, 2000.

[Medley 93] Max W. Medley, *Network Analysis and Feedback Amplifier Design*, Artech House, Inc., 1993.

[Meng 06] Fanfan Meng, A. van Bezooijen, and R. Mahmoudi, "A Mismatch Detector for Adaptive Antenna Impedance Matching", pp. 1457–1460, Sept. 2006.

[Messaoudi 07] N. Messaoudi, O. Hammi, J. Sirois, S. Boumaiza, and F.M. Ghannouchi, "A Comparative Study of Power Amplifiers Sensitivity to Load Mismatch: Single Branch vs. Doherty Architectures", *Electrical and Computer Engineering, 2007. CCECE 2007. Canadian Conference on*, vol. , pp. 1543–1546, April 2007.

[Mileusnic 97] M. Mileusnic, P. Petrovic, and J. Todorovic, "Design and implementation of fast antenna tuners for HF radio systems", volume 3, pp. 1722–1726 vol.3, Sep 1997.

[Mingo 04] J. de Mingo, A. Valdovinos, A. Crespo, D. Navarro, and P. Garcia, "An RF electronically controlled impedance tuning network design and its application to an antenna input impedance automatic matching system", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 52, pp. 489–497, Feb. 2004.

[Mohan 99] S. S. Mohan, M. M. Hershenson, S. P. Boyd, and T. H. Lee, "Simple inductance formulas for radio coils", *in IEEE J. Solid-State Cric.*, pp. 1419–1424, 1999.

[Moré 80] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, "User Guide for MINPACK 1", Technical report, Argonne National Laboratory, 1980, ANL-80-74.

[Murata 08] Murata, "LDC43850M03B-401", Internet http://www.murata.com/, 2008.

[Niknejad 00] Ali M. Niknejad and Robert G. Meyer, *Design, Simulation and Applications of Inductors and Transformers for Si RF ICs*, Kluwer Academic Publishers, Sept. 2000.

[Ogawa 99] K. Ogawa, T. Matsuyoshi, and K. Monma, "An analysis of the performance of a handset diversity antenna influenced by head, hand and shoulder effects at 900 MHz", volume 2, pp. 1122–1125 vol.2, Aug 1999.

[Otoshi 69] T.Y. Otoshi, "On the scattering parameters of a reduced multiport (correspondence)", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 17, pp. 722–724, Sep 1969.

[Palankovski 99] V. Palankovski and Sigfried Selberherr, "Thermal Models for Semiconductor Device Simulation", *in Proceedings of 3rd European Conference on High Temperature Electronis HITEN*, pp. 25–29, Berlin, Germany, 1999.

[Peregrine 08] Peregrine, "PE4246 Datasheet", Internet http://www.peregrinesemi.com/, 2008.

[Powell 70] M. J. D. Powell, *Numerical Methods for Nonlinear Algebraic Equations*, chapter A Fortran Subroutine for Solving Systems of Nonlinear Algebraic Equations, P. Rabinowitz, 1970.

[Pozar 93] David M. Pozar, *Microwave Engineering*, Addison-Wesley Publishing Company, 1993.

[Pusl 01] Joe Pusl, Srikanth Sridharan, David Helms, Philip Antognetti, and Mark Doherty, "A Silicon Germanium, High Efficiency Power Amplifier Chipset for GSM/DCS-PCS/WCDMA Handset Applications", *in European Microwave Conference, 2001. 31st*, pp. 1–4, Oct. 2001.

[Qiao 05] Dongjiang Qiao, Yu Zhao, Tsaipi Hung, D. Kimball, Mingyuan Li, P. Asbeck, D. Choi, and D. Kelly, "Antenna impedance mismatch measurement and correction for adaptive CDMA transceivers", pp. 4 pp.–, June 2005.

[Raab 77,a] F. H. Raab, "Class-F power amplifiers with maximally flat waveforms", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 45, pp. 2007–2012, November 1977.

[Raab 77,b] F. H. Raab, "Idealized operation of the Class E tuned power amplifier", *Circuits and Systems, IEEE Transactions on*, vol. 24, pp. 725–735, December 1977.

[Raab 02] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power Amplifiers and Transmitters for RF and Microwave", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 50, 2002.

[Reisch 03] Michael Reisch, *High-Frequency Bipolar Transistors*, Springer, first edition, 2003.

[RFMD 09] RFMD, "Model:RF3158", Internet http://www.rfmd.com/, 2009.

[Rudolph 04] M. Rudolph, F. Schneider, and W. Heinrich, "Investigation of Thermal Crunching Effects in Fishbone-Type Layout Power GaAs-HBTs", *in Gallium Arsenide applications symposium*, Amsterdam, Oct. 2004.

[Saleh 80] A.A.M. Saleh, "Improving the Graceful-Degradation Performance of Combined Power Amplifiers", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 28, pp. 1068–1070, Oct 1980.

[Schröter 05] M. Schröter, "HICUM - A Geometry Scalable Physics-Based Compact Bipolar Transistor model", Internet http://www.iee.et.tu-dresden.de/iee/eb/forsch/Hicum_PD/Hicum22/HICUM-L2V2p2.pdf, 2005.

[Scuderi 05] A. Scuderi, L. La Paglia, F. Carrara, and G. Palmisano, "A VSWR-protected silicon bipolar RF power amplifier with soft-slope power control", *Solid-State Circuits, IEEE Journal of*, vol. 40, pp. 611–621, March 2005.

[Scuderi 08] A. Scuderi, C. Santagati, M. Vaiana, F. Pidala, and M. Paparo, "Balanced SiGe PA Module for Multi-Band and Multi-Mode Cellular-Phone Applications", *in Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, pp. 572–637, Feb. 2008.

[Sevic 97] John F. Sevic, "Statistical Characterization of RF Power Amplifier Efficiency for CDMA Wireless Communication Systems", *Wireless Communications Conference*, vol. , pp. 110–113, 1997.

[Shirvani 03] Alireza Shirvani and Bruce A. Wooley, *Design and Control of RF Power Amplifiers*, Kluwer Academic Publishers, 2003.

[Shumaker 02] J. Shumaker, Basset, R., and Skuratov, A., "High-Power GaAs FET Amplifiers: Push-Pull versus Balanced Configurations", *Applied Microwave & Wireless*, vol. may, pp. 26–32, 2002.

[Sim 07] Jaewoo Sim, Jaeyoung Choic, Kwon Kim, Myoungkyu Park, Wonwoo Kang, and SeungGon Kim, "Analysis and Design of Doherty Power Amplifiers for Digital Pre-distortion Linearizer", *Microwave Conference, 2007. APMC 2007. Asia-Pacific*, vol. , pp. 1–4, Dec. 2007.

[Skyworks 09] Skyworks, "Model:SKY77162", Internet http://www.skyworksinc.com/, 2009.

[Sogl 07,a] B. Sogl, W. Bakalski, M. Zannoth, M. Asam, B. Kapfelsperger, J. Berkner, B. Eisener, W. Osterreicher, E. Rampf, A.L. Scholtz, and B.-U. Klepser, "A Quad-Band GSM/EDGE-Compliant SiGe-Bipolar Power Amplifier with 35.9/32.3 dBm Output Power at 56 %/44 % PAE in Low/High-Band", *Bipolar/BiCMOS Circuits and Technology Meeting, 2007. BCTM '07. IEEE*, vol. , pp. 98–101, 30 2007-Oct. 2 2007.

[Sogl 07,b] B. Sogl, J.E. Mueller, and W. Thomann, "Power Efficient Transmitter with High Dynamic Range", Patent pending: INFAP119US, 2007.

[Sogl 09] B. Sogl, W. Thomann, J.E. Mueller, W. Bakalski, and A. L. Scholtz, "A multi-mode power amplifier for enhanced PAE in back-off operation with load insensitive architecture", *Radio and Wireless Symposium, 2009. RWS '09. IEEE*, vol. , pp. 542–545, Jan. 2009.

[Sokal 75] N. O. Sokal and A. D. Sokal, "Class E, a new class of high-efficiency tuned single-ended power amplifiers", *Solid-State Circuits, IEEE Journal of*, vol. , pp. 168–176, June 1975.

[Srirattana 05] N. Srirattana, A. Raghavan, D. Heo, P.E. Allen, and J. Laskar, "Analysis and design of a high-efficiency multistage Doherty power amplifier for wireless communications", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 53, pp. 852–860, March 2005.

[Sze 81] S. M. Sze, *Physics of Semiconductor Devices*, John Wiley and Sons, Inc., second edition, 1981.

[Tholl 76] Herbert Tholl, *Bauelemente der Halbleiterelektronik*, B.G.Teubner Stuttgart, first edition, 1976.

[Tinella 06] C. Tinella, O. Richard, A. Cathelin, F. Reaute, S. Majcherczak, F. Blanchet, and D. Belot, "0.13um CMOS SOI SP6T Antenna Switch for Multi-Standard Handsets", *Silicon Monolithic Integrated Circuits in RF Systems, 2006. Digest of Papers. 2006 Topical Meeting on*, vol. , pp. 44pp.–, Jan. 2006.

[Toftgard 93] J. Toftgard, S.N. Hornsleth, and J.B. Andersen, "Effects on portable antennas of the presence of a person", *Antennas and Propagation, IEEE Transactions on*, vol. 41, pp. 739–746, Jun 1993.

[Tombak 08] A. Tombak, R.J. Baeten, J.D. Jorgenson, and D.C. Dening, "Integration of a cellular handset power amplifier and a DC/DC converter in a Silicon-On-Insulator (SOI) technology", *Radio Frequency Integrated Circuits Symposium, 2008. RFIC 2008. IEEE*, vol. , pp. 413–416, 17 2008-April 17 2008.

[Tosaka 03] H. Tosaka, T. Fujii, K. Miyakoshi, K. Ikenaka, and M. Takahashi, "An antenna switch MMIC using E/D mode p-HEMT for GSM/DCS/PCS/WCDMA bands application", *Microwave Symposium Digest, 2003 IEEE MTT-S International*, vol. 1, pp. A5–A8 vol.1, June 2003.

[Vendelin 82] George D. Vendelin, *Design of Amplifiers and Oscillators by the S-Parameter Method*, John Wiley and Sons, Inc., 1982.

[Vidkjaer 78] J. Vidkjaer, "A computerized study of the class-C-biased RF-power amplifier", *Solid-State Circuits, IEEE Journal of*, vol. 13, pp. 247–258, Apr 1978.

[Vuolevi 01] Joel Vuolevi, *Analysis, measurement and cancellation of the bandwidth and amplitude dependence of intermodulation distortion in RF power amplifiers*, PhD thesis, Department of Electrical Engineering, University of Oulu, Finland, 2001.

[Vuolevi 03] Joel Vuolevi and Timo Rahkonen, *Distortion in RF Power Amplifiers*, Artech House, Norwood,MA 02062, first edition, 2003.

[Wheeler 28] H. A. Wheeler, "Simple inductance formulas for radio coils", *in Proceedings of the IRE*, pp. 29–35, 1928.

[Yamamoto 02] K. Yamamoto, T. Asada, S. Suzuki, T. Miura, A. Inoue, S. Miyakuni, J. Otsuji, R. Hattori, Y. Miyazaki, and T. Shimura, "A GSM/EDGE dual-mode, 900/1800/1900-MHz triple-band HBT MMIC power amplifier module", *in Radio Frequency Integrated Circuits (RFIC) Symposium, 2002 IEEE*, pp. 245–248, 2002.

[Yamauchi 96] K. Yamauchi, K. Mori, M. Nakayama, Y. Itoh, Y. Mitsui, and O. Ishida, "A novel series diode linearizer for mobile radio power amplifiers", *MTT-S International Microwave Symposium Digest*, vol. , pp. 831–834, June 1996.

[Zinke 87] O. Zinke and H. L. Hartnagel, *Lehrbuch der Hochfrequenztechnik, Band 2*, Springer Verlag New York Heidelberg Berlin Tokyo, third edition, 1987.

Curriculum Vitae

Personal Data

Name	Bernhard Sogl
Address	Dreyhausenstr. 5/5, 1140 Vienna, Austria
Email	bsogl@gmx.net
Date/Place of birth	27 March 1980, Vienna, Austria
Nationality	Austria

Education

since 12/2005	Doctorate study, Vienna University of Technology, Thesis: "Multi-Mode Power Amplifiers for Mobile Handsets", in cooperation with Infineon Technologies AG.
11/2005	Dipl.-Ing. Degree passed with distinction, Thesis: "Modular RF transceiver Circuits AM and IQ De/modulation, Mixers and Amplifiers".
10/1998 - 11/2005	Electrical Engineering study, Vienna University of Technology Specialization: Communications and RF Engineering.
06/1998	"Matura" (High Shool Diploma) passed with distinction, Bundesrealgymnasium, Linzerstr. 1140 Vienna

Experience

12/2005 - 11/2009	Research and Product Development Engineer, Doctorate student Infineon Technologies AG, Munich, Germany
10/2004 - 06/2005	Teaching Assistant, Vienna University of Technology Department: Electrical Measurements and Circuit Design
03/2004 - 06/2004	Teaching Assistant, Vienna University of Technology Department: Communications and RF Engineering

Trainings

07/2006	TARGET Summer School on PA design, Castelldefels Spain
02/2006	Cadence Design Training for Spectre RF Tools, Schematic and Layout Editor, Munich Germany

Languages German (mother tongue), English (fluently)