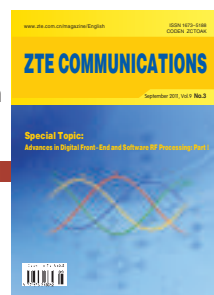


Contents

Http://www.zte.com.cn/magazine/English
Email: magazine@zte.com.cn



Editorial Board

Chairman: Yixin Zhong

Vice Chairmen: Weigui Hou,
Zhengkun Mi

Members (in Alphabetical Order):

Anshi Xu, Bo Ai, Bo Hong,
Changjia Chen, Chengzhong Xu,
Daxiong Xie, Guangxin Yue,
Guo Wei, Heyuan Xu, Hongbin Li,
Houlin Zhao, Hua Jiang,
Huiling Zhao, Jiandong Li,
Jianping Chen, Jie Chen,
Jinkang Zhu, Jinyun Chang,

Lemin Li, Leping Wei, Lintao Jiang,
Lirong Shi, Luoming Meng, Qin Ni,
Shaoqian Li, Shiduan Cheng,
Shixin Cheng, Shiyu He,
Shuangjin Gong, Shumin Cao,
Susu Zhou, Tongxu Zhang,
Wanyi Gu, Weigui Hou, Wen Gao,
Wenguo Tian, Xianming Zhao,

Xiaohu You, Xiaoming Wang,
Xiaoyun Wang, Xing Li, Xiren Xie,
Xisheng Chen, Yimin Yin,
Yixian Yang, Yixin Zhong,
Yongcheng Gu, Yuefeng Ji,
Yumin Wang, Yunfei Guo, Zhen
Yang, Zhengkun Mi, Zhenhui Tan,
Zhenzhou Lei, Zhijiang Zhang

ZTE COMMUNICATIONS
Vol. 9 No.3 (Issue 31)
Quarterly
First Issue Published in 2003

Supervised by:

Anhui Science and Technology
Department

Sponsored by:

ZTE Corporation and Anhui Science
and Technology Information
Research Institute

Staff Members:

Editor-in-chief: Xie Daxiong

Deputy Editor-in-chief: Deng Xin

Executive Deputy

Editor-in-chief: Huang Xinming

Editor in Charge: Zhu Li

Editors: Paul Sleswick, Yang Qinyi, Xu Ye,
Lu Dan

Producer: Yu Gang

Circulation Executive: Wang Pingping

Assistant: Wang Kun

Editorial Correspondence:

Add: 12/F Kaixuan Building
329 Jinzhai Road,
HeFei 230001, P. R. China

Tel: +86-551-5533356

Fax: +86-551-5850139

Email: magazine@zte.com.cn

Published and Circulated

(Home and Abroad) by:

Editorial Office of
ZTE COMMUNICATIONS

Printed by:

Hefei Zhongjian Color Printing Company

Publication Date: September 25, 2011

Publication Licenses:

ISSN 1673-5188

CN 34-1294/TN

Advertising License:

皖合工商广字0058号

Annual Subscription Rate:

USD\$50

Responsibility for content rests
on authors of signed articles and
not on the editorial board of
ZTE COMMUNICATIONS or its sponsors.
All rights reserved.

Special Topic: Advances in Digital Front-End and Software RF Processing: Part I

1 Guest Editorial

3 Adaptation of a Digitally Predistorted RF Amplifier Using Selective Sampling

13 A New Two-Branch Amplification Architecture and its Application with Various Modulated Signals

22 FPGA Implementation of a Power Amplifier Linearizer for an ETSI-SDR OFDM Transmitter

28 Design Technologies for Silicon-Based High-Efficiency RF Power Amplifiers: A Brief Overview

36 Multi-Gbit/s 60 GHz Transceiver Analysis Using FDM Architecture and Six-Port Circuit

42 Millimeter-Wave Heterodyne Six-Port Receiver: New Implementation and Demodulation Results

Research Paper

49 Security Service Technology for Mobile Networks

Development Field

55 Research on LTE Network Coverage Planning

Operational Application

59 Cloud Computing in Mobile Communication Networks

Lecture Series

63 The Internet of Things and Ubiquitous Intelligence (3)

Roundup

12 ZTE, Telkom Kenya-Orange Announce Nation's Fastest Commercial HSPA+ Network

41 ANTEL Selects ZTE for Uruguay National GPON Project

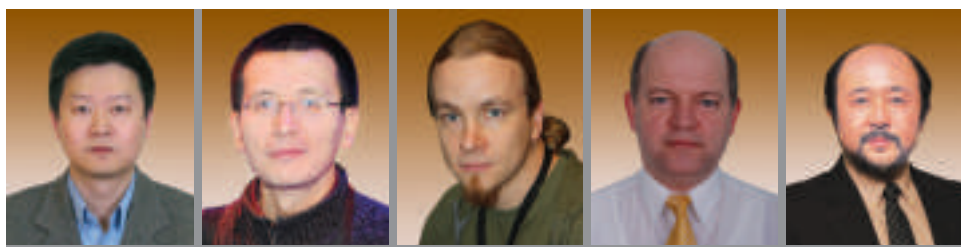
58 ZTE Brings International Mobile Hotspot to Sprint Customers

Departments

48 Ad Index

66 Abbreviation Index

Advances in Digital Front-End and Software RF Processing: Part I



Jun Fang

Fa-Long Luo

Mikko Valkama

Serioja Ovidiu Tatu

Tomohisa Wada

One of the biggest technology trends in wireless broadband, radar, sonar, and broadcasting systems is software radio frequency processing and digital front-end. This trend encompasses a broad range of topics, from circuit design and signal processing to system integration. It includes digital up-conversion (DUC) and down-conversion (DDC), digital predistortion (DPD), DC-offset calibration, peak-to-average power ratio (PAPR) or crest factor reduction (CFR), pulse/noise shaping, delay and gain imbalance compensation, numerical controlled oscillator (NCO), and conversion between analog signals and digital signals. Digital techniques for RF processing have many advantages over traditional techniques in terms of power efficiency, cost and area reduction, flexibility, and programmability. Digital techniques allow software defined radio to be reconfigured to support multiple standards and multimode applications for fast time-to-market solutions. These advantages are becoming increasingly important in future wireless infrastructure development and deployment. Implementing digital processing and circuits in front-end is highly desirable but also highly challenging. Huge efforts are required from industry, research institutes, and regulatory authorities to bring the next generation of wireless communication systems to fruition. Extremely stringent requirements on bandwidth, power consumption, and performance need to be met for future systems. It is also necessary to say a few words about software RF terminology. Software RF might be a pretty new concept but still encompasses digital RF. It belongs in the software defined radio (SDR) category but goes further, extending the SDR frontier. Software RF means almost limitless flexibility and scalability in the way RF signal processing is done, and digital is necessary for software RF. Software RF will ultimately become an operational part of SDR in coming years.

This special issue aims to stimulate and guide the development of new and improved RF systems for wireless communication and digital broadcasting. It aims to be a timely and high-quality forum for scientists, engineers,

technologists, broadcasters, manufacturers, software developers and other professionals to engage in discussion. The call-for-papers for this special issue attracted a good number of excellent submissions. After two-round reviews, fourteen papers have been selected for publication in this special issue, which is organized in two parts and will be published in two consecutive issues in 2011.

The contents of Part I is divided into two categories. The first consists of four papers addressing different aspects of power amplification technologies. The second is devoted to another interesting topic: architecture design and testing of millimeter wave (60 GHz) wideband radio transceivers, which are mainly based on six-port devices and related technology.

The first paper by R. Neil Braithwaite provides a cost-effective method of measuring residual nonlinearities in an adaptive digitally pre-distorted amplifier. This method involves selective sampling of the amplifier output, integrated over the input envelope range, to adapt a fourth-order polynomial predistorter with memory correction. Simulations show that a transmitter using the proposed method can meet the ACLR specification. Inverse modeling of the nonlinearity is proposed as a future extension that will reduce the cost of the system further.

The second paper by W. Hamdane, A. B. Kouki, and F. Gagnon proposes a novel two-branch amplification architecture that combines baseband signal decomposition with RF front-end optimization. The proposed system separates the filtered modulated signals into two components that are amplified independently then combined to regenerate an amplified version of the original signal. A branch using an efficient amplifier transmits a low-varying envelope signal that contains the main part of the information. The other branch is used to amplify the residual portion of the signal. The baseband decomposition and RF part's parameters are optimized, and the optimal configuration is determined for the best power efficiency and linearity.

In the third paper, Suranjana Julius and Anh Dinh deal with implementation of a power amplifier (PA) linearizer for an

ETSI–SDR OFDM transmitter. This paper presents an interesting case study on satellite digital radio application in L–band. An adaptive linearizer is designed and implemented on the same FPGA device. Digital predistortion is used to correct the undesired effects of the PA on the transmitted signal.

This thematic group ends with an excellent paper by Ruili Wu, Jerry Lopez, Yan Li, and Donald Y.C. Lie that provides an updated overview of design technologies for RF PAs. The authors outline several promising design techniques for highly efficient silicon–based RF PAs and their use in mobile broadband wireless communications. Four important aspects in PA design are addressed. The paper is an invaluable source of information on new wireless standard requirements, design methodologies, power amplification architectures, and integrations.

Wireless applications have emerged in the 59–64 GHz ISM band in recent years, and these have attracted increasing interest from the IT and entertainment industries. 60 GHz wireless local area networks (WLAN) are suited to applications with a short range and very high data–rate, such as high–speed home and office wireless networking, high definition television (HDTV), and interactive HD gaming among others. The fifth paper in Part I by Nazih Khaddaj Mallat, Emilia Moldovan, Serioja O. Tatu, and Ke Wu describes how advanced system simulation can be used to analyze and validate compact, low–cost six–port transceivers for future wireless LANs operating at millimeter–wave frequencies. Frequency division multiplexing is used by introducing four QPSK channels in the wireless communication link. A data rate of about 4 Gbit/s in the 60–64 GHz unlicensed band can be reached. Both single carrier and multicarrier architectures are presented, and results are compared. The proposed wireless system is potentially an excellent and efficient candidate for millimeter–wave communication systems operating at quasi–optical data rates.

The last paper in Part I, by D. Hammou, E. Moldovan, and S. O. Tatu, is devoted to a new implementation of a millimeter–wave heterodyne receiver based on six–port technology. The six–port model is implemented in ADS software using S parameter measurements for advanced, realistic simulation systems of a short–range 60 GHz wireless link. The proposed mixer is compared with a conventional balanced millimeter–wave mixer, and some improvements are observed in conversion loss and I/Q phase stability over the LO and RF power range. Furthermore, the bit error rate analysis and error vector magnitude analysis show that the proposed architecture can be successfully used for wireless link transmission up to 10 meters.

We would like to thank all authors for their valuable contributions. We also express our sincere gratitude to all the reviewers for their timely and insightful comments on all submitted papers. It is hoped that the contents of this issue are informative and useful from various technology and implementation aspects. Please stay tuned for the second part of this special issue.

Biographies

Dr. Jun Fang graduated from Shanghai Jiao Tong University in 1982, and received his Ph.D. from Ecole Nationale Supérieure des Télécommunications de Paris (ENST Paris) in 1987. He has been an associate professor at Shanghai Jiao Tong University since 1987. He worked with Alcatel Space Industries from 1990 to 2001 in several R&D and management positions. He was senior vice president of Linkair Communications USA from 2001 to 2005, and then worked as digital design director with TechnoConcepts from 2005 to 2006. During this time he was involved in digital RF–mixed chip projects. Dr. Fang held a senior wireless system position in EDA with Cadence USA during 2006. Since 2009, he has been director of the Electronics & Information Technology Center of the Research Institute of Tsinghua University (RITS) and has led R&D and industrialization activities in wireless, signal processing, EDA, and digital TV multimedia. He has also been involved in wireless activities with CARITS Inc., a subsidiary of RITS in the U.S. Dr. Fang has written one book on information theory and coding, and he has contributed to ITU publications on satellite systems. He has 10 patents and published many technical papers. His current research interests include SDR, digital RF chip design, and MIMO transmission systems.

Dr. Fa–Long Luo is chief scientist at two leading international companies, headquartered in Silicon Valley, CA, that deal with SDR and wireless multimedia. He has been the editor–in–chief of the International Journal of Digital Multimedia Broadcasting since 2007. Dr. Luo is currently chairman of the IEEE Industry DSP Standing Committee and technical board member of the IEEE Signal Processing Society. He has 28 years of research and industrial experience in multimedia, communication and broadcasting with real–time implementation, applications, and standardization. He has received worldwide recognition. Dr. Luo has authored and edited four books, more than 100 technical papers, and 18 patents on these and closely related fields.

Dr. Mikko Valkama is full professor and department head of the Department of Communications Engineering at Tampere University of Technology (TUT), Finland. He has been involved in organizing conferences such as the IEEE SPAWC'07 in Helsinki. He was awarded Best Ph.D. Thesis by the Finnish Academy of Science and Letters. His research interests include communications signal processing, estimation and detection techniques, signal processing algorithms for software defined flexible radios, and signal processing for cognitive radio. He is also interested in digital transmission techniques, such as different variants of multicarrier modulation methods and OFDM, and radio resource management for ad–hoc and mobile networks.

Dr. Serioja Ovidiu Tatu received his M.Sc. and Ph.D. degrees in electrical engineering from the École Polytechnique, Montréal, in 2001 and 2004. From 2004 to 2005, he was a post–doctoral researcher at the Institut National de la Recherche Scientifique–Énergie Matériaux et Télécommunications, Montréal, and is now associate professor at that institute. His current research interests include millimeter–wave circuit design, hardware and software radio receivers, radar, and sensor systems.

Dr. Tomohisa Wada received his B.S. degree in electronic engineering from Osaka University in 1983. He received his M.S.E.E. degree from Stanford University in 1992, and his Ph.D. degree in electronic engineering from Osaka University in 1994. He joined the ULSI Laboratory, Mitsubishi Electric Corp. in 1983. Since 2001, he has been a professor at the Department of Information Engineering, University of the Ryukyus, Okinawa. In 2001, he was the founding member of Magna Design Net Inc., an LSI design company for communication–related digital signal processing such as OFDM. Currently, he is chief scientist at Magna Design Net Inc. and is engaged in the research and development of terrestrial video broadcasting, wireless LAN, and WiMAX.

Adaptation of a Digitally Predistorted RF Amplifier Using Selective Sampling

R. Neil Braithwaite

(Powerwave Technologies, Santa Ana, CA 92705, the U. S.)

Abstract: In this paper, a reduced-cost method of measuring residual nonlinearities in an adaptive digitally predistorted amplifier is proposed. Measurements obtained by selective sampling of the amplifier output are integrated over the input envelope range to adapt a fourth-order polynomial predistorter with memory correction. Results for a WCDMA input with a 101 carrier configuration show that a transmitter using the proposed method can meet the adjacent channel leakage ratio (ACLR) specification. Inverse modeling of the nonlinearity is proposed as a future extension that will reduce the cost of the system further.

Keywords: amplifier distortion; communication system nonlinearities; power amplifier linearization

1 Introduction

A digital transmitter used in wireless communication applications comprises several stages, including a digital baseband, digital-to-analog converter (DAC), modulator, and power amplifier (PA). An efficient PA often has a nonlinear gain that varies as a function of the input signal envelope. As a result, a linearization method is required to compensate for undesired nonlinearities.

Digital predistortion (DPD) involves introducing a nonlinear gain function into the digital transmission path that opposes nonlinearities in the modulator and PA stages. Adaptive DPD measures the residual nonlinearity of the predistorted transmitter and adjusts the DPD coefficients to reduce distortion in the output signal. The measurement circuitry, referred to as the observation path, includes down-conversion and digitization of the PA output signal and typically comprises demodulator and

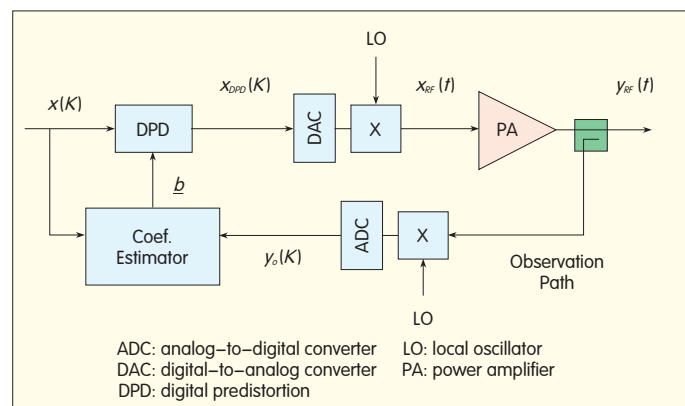
analog-to-digital converter (ADC) stages. The block diagram of a typical transmitter, including adaptive DPD, is shown in Fig. 1.

An inherent problem with adaptive linearization is that the measurement system cannot distinguish distortion generated by nonlinearities in the transmitter from distortion induced by data acquisition components in the observation path. Thus, the observation path must be significantly more linear than the desired linearity of the

transmitter. It must also have sufficient dynamic range to avoid degrading the output spectral mask. As a result, the measurement circuitry used is often expensive.

As an example of the contribution of the observation path to transmitter cost, ADC can be considered. The price of an ADC increases with the sampling rate and resolution. A typical zero-IF observation path for a multicarrier WCDMA [1] signal uses two 14 bit ADCs sampled at 122.88 MHz [2]. If the

Figure 1. ▶ Block diagram of a typical transmitter with adaptive DPD, including an observation path for measuring the PA output signal.



number of ADCs was reduced to one, the resolution reduced to 8 bits, and the sample rate reduced to 32 MHz, ADC-related costs would drop by a factor of 25 [3].

An inexpensive approach for measuring nonlinearities in a digital transmitter is shown in [4]. In this approach, a known calibration signal is transmitted with an amplitude-modulated (AM) component. The AM component excites the nonlinear modes of the PA to generate distortion. In contrast, the measurement circuit contains a cancellation loop that reduces AM variations so that the signal amplitude is nearly constant. Thus, the distortion generated within the measurement circuit is minimal. The dynamic range of the measured signal at the ADC is also reduced. As a result, the requirements of the data acquisition components are relaxed, and the cost of the adaptive system is reduced. The drawback of [4] is that the measurements must be made offline because of the use of a calibration signal.

It is preferable to optimize the system adaptively based on measurements made while transmitting the actual signal. In section 2, a WCDMA signal is sampled selectively to create a probing signal that highlights nonlinear modes of the transmitter. The probing signal allows for online adaptation of the digital predistorter. Section 3 describes the coefficient estimation module, which includes a recursive integration method for a 4th order predistorter. Memory compensation is discussed in section 4, and linearization results for a WCDMA signal are given in section 5. Section 6 describes inverse modeling of the nonlinearity as a future extension.

This paper is an extension of [5], a conference paper titled "Measurement and correction of residual nonlinearities in a digitally predistorted power amplifier," by the author, which appeared in the proceedings of the 2010 75th ARFTG Microwave Measurement Conference © IEEE. The remainder of the introduction includes a brief review of past DPD work done by other

researchers.

1.1 Digital Predistortion Background

The predistorted baseband signal for the digital transmitter in Fig. 1 is

$$\chi_{DPD}(k) = G_{DPD}(|\chi|) \cdot \chi(k), \quad (1)$$

where G_{DPD} is the predistortion gain and a nonlinear function of $|\chi|$. The predistorted baseband signal is up-converted to produce an RF signal

$$\chi_{RF}(t) = h_{DAC}\{\chi_{DPD}(k)\} \cdot \exp(j \cdot \omega_{LO} \cdot t), \quad (2)$$

where $h_{DAC}\{\}$ is a reconstruction filter used in the digital-to-analog conversion (not shown in Fig. 1), and $\omega_{LO}(t)$ is the LO frequency. The output of the PA is

$$y_{RF}(t) = G_{PA}(|\chi_{RF}|) \cdot \chi_{RF}(t), \quad (3)$$

where G_{PA} is the gain of the PA and a nonlinear function of $|\chi_{RF}|$.

Memoryless nonlinearities are often described using AM-AM and AM-phase modulated (PM) curves where the amplitude and phase components of the gain are plotted as a function of the input envelope. The gain curves produced by the DPD module are represented using a polynomial function of order N :

$$G_{DPD}(|\chi|) = \sum_{n=0}^{N-1} b_n \cdot |\chi|^n, \quad (4)$$

where b_n are complex DPD coefficients. The gain of the predistorted transmitter, which is the combination of the DPD and PA nonlinearities, is also represented by a polynomial:

$$\begin{aligned} G_{trans}(|\chi|) &= G_{PA}(|\chi_{RF}|) \cdot G_{DPD}(|\chi|) \\ &= G_0 \cdot \sum_{n=0}^{N-1} a_n \cdot |\chi|^n \end{aligned} \quad (5)$$

where G_0 is the desired (linear) gain of the transmitter, and a_n are complex values referred to as residual memoryless coefficients. Estimates of the residual nonlinearity a_n are used to update the DPD coefficients b_n in an iterative manner, that is,

$$\underline{b}(i+1) = \underline{b}(i) - \alpha \cdot [(\underline{a}_0 - 1) \ a_1 \ a_2 \ a_3]^T, \quad (6)$$

where $(\)^T$ indicates transpose $\underline{b}(i) = [b_0 \ b_1 \ b_2 \ b_3]^T$ at iteration i , and $0 < \alpha < 1$. The iterative sequence (6) has converged when

$$\underline{a} = [a_0 \ a_1 \ a_2 \ a_3]^T = [1 \ 0 \ 0 \ 0]^T.$$

The standard approach to measuring the residual nonlinearities (shown in Fig. 1) is to capture the signal from the PA output, down-convert and digitize it to produce an observation signal $y_o(k)$, then compute the coefficients a_n that minimize [2],[6],[7]

$$J = \sum_k |y_o(k) - G_0 \cdot \sum_{n=0}^{N-1} a_n \cdot BF_n(k) \cdot \chi(k)|^2, \quad (7)$$

where BF_n is a gain basis function and a nonlinear function of $\chi(k)$. For the memoryless case, the gain basis functions are defined by

$$BF_n = |\chi(k)|^n. \quad (8)$$

In general, the observation signal $y_o(k)$ is corrupted by nonlinearities within two paths: the transmitter path from $\chi(k)$ to $y_{RF}(t)$, and the observation path from $y_{RF}(t)$ to $y_o(k)$. The DPD is intended to compensate for the former. Any nonlinearity in the observation path offsets the steady-state DPD coefficients \underline{b} and degrades the adjacent channel leakage ratio (ACLR) measured at $y_{RF}(t)$. Thus, the observation path in Fig. 1 must be significantly more linear than the desired linearity of the transmitter. The observation path should also minimize other impairments, such as demodulator imbalance, LO phase noise, and quantization noise.

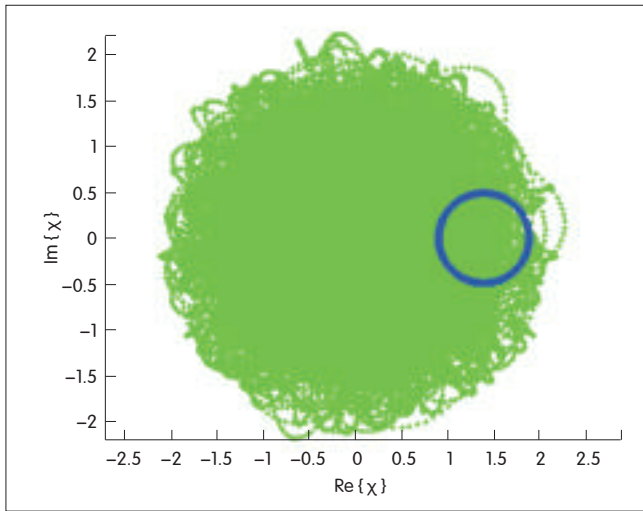
Up to this point, only memoryless PA nonlinearities have been considered. Nonlinear memory is often modeled using delayed digital samples of the input signal. For example, the gain basis functions for a discrete Volterra series would be

$$BF_v = \chi(k - \tau_1) \cdots \chi(k - \tau_l) \cdot \chi^*(k - \tau_{l+1}) \cdots \chi^*(k - \tau_{2l}), \quad (9)$$

where τ_l are integer sample offsets, v is an index, and $(\)^*$ is complex conjugate. In general, the number of basis functions in a Volterra series is too large to be practical. As a result, pruned versions are typically used [8]–[10]. A popular pruned basis function set is

$$BF_v(k) = |\chi(k - \tau)|^n, \quad (10)$$

which is referred to as a memory polynomial [6],[7],[11],[12]. Memory polynomials are also implemented by



◀ Figure 2.
Selective sampling
extracts the desired
probing signal (blue circle)
from the WCDMA signal
(green cloud).

delaying the distortion modes as opposed to delaying the gain component only. This produces predistorted waveforms that are a weighted sum of $|\chi(k-\tau)|^n \chi(k-\tau)$ instead of $|\chi(k-\tau)| \chi(k)$. Other memory models use delayed samples from both the input $\chi(k)$ and output $\chi_{DPD}(k)$ of the predistorter, where the latter forms a feedback loop around a nonlinear kernel. An example using feedback within an artificial neural network can be found in [13].

While both the pruned Volterra series and artificial neural network models discernably improve distortion cancellation, memory models based on delayed or fed-back digital samples are not compatible with the measurement approach proposed in section 2. The compatible memory model described in section 4 is based on derivatives of the input envelope $\delta|\chi|/\delta t$, instead of delayed samples $|\chi(k-\tau)|$. This model is shown in section 5 to improve ACLR performance over the memoryless DPD model presented in [5].

2 Measurement Approach

A new approach to measuring transmitter nonlinearities, suitable for WCDMA signals, is proposed here. The motivation is to replace the standard observation path shown in Fig. 1 with something much cheaper. To achieve this goal, the required linearity, dynamic range, and sampling rate of the

observation path must be reduced. In the standard approach, the WCDMA output signal is captured directly; however, the sampling rate must be several multiples of the Nyquist rate to measure the out-of-band distortion without aliasing [14]. In addition, a large dynamic range is needed in the standard approach to measure distortion below an ACLR2 level of -50 dBc [1] from a signal with a large peak-to-average power ratio (PAPR), on the order of 7.2 dB.

The proposed approach involves creating a probing signal, similar to the one used in [4], that is extracted from a subset of the sampled WCDMA input signal. This selectively sampled probing signal has a lower PAPR and sampling rate than the WCDMA signal. The proposed measurement circuitry reduces the PAPR of the probing signal further, to almost 0 dB, using a cancellation bridge. As a result, the observation path requires far lower linearity, dynamic range, and sampling rate than the standard approach, allowing cost to be reduced.

It is possible to transform the complex input signal $\chi(k)$ into

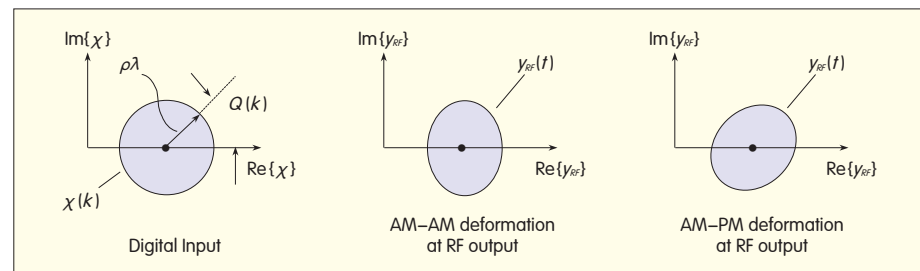
$$\begin{aligned}\chi(k) &= \text{Re}\{\chi\} + j \cdot \text{Im}\{\chi\} \\ &= \rho \cdot \{1 + \lambda(k) \cdot \exp[j \cdot \theta(k)]\},\end{aligned}\quad (11)$$

where both $\lambda(k)$ and $\theta(k)$ vary with time, and ρ is a constant. This transformation $\chi(k) = f\{\rho, \lambda(k), \theta(k)\}$ is similar to a conversion from rectangular to polar coordinates, except that the origin is offset by the constant ρ . Equation (11) is an exact transformation that converts the signal into a form that makes the selective sampling, described below, easier to implement.

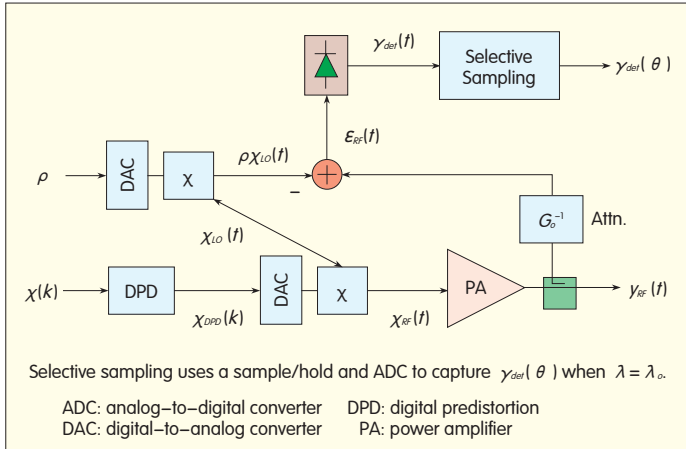
Consider a subset of the input samples $\chi(k) = f\{\rho, \lambda(k), \theta(k)\}$, where λ_0 is a constant. This subset can be viewed as a selective sampling process where a circle within the I-Q space is chosen. An example for a WCDMA signal is shown in Fig. 2. The circular trajectory is specified as a function of the angle θ (Fig. 3). The trajectory is used as a probing signal to highlight nonlinearities within the transmitter, which appear as elliptical deformations in the RF output $y_{RF}(t)$ (Fig. 3). In a typical implementation, several probing signals with different ρ values are tested. These probing signals create circles at various power levels within the I-Q space, although only one circular trajectory is shown in Fig. 2.

Elliptical trajectories at the output are caused by slopes in the gain of the transmitter. Deformations of the circular trajectory due to slopes in the AM-AM (center) and AM-PM (right) curves occur along the horizontal and 45 degree axes, respectively.

The deformation of the circular trajectory provides information about the nonlinear gain of the transmitter around the output operating point $G_o \rho \chi_{LO}(t)$. Any slope in the gain curves



▲ Figure 3. Baseband input and RF output signals, selectively sampled.



$\delta G_{trans}/\delta|\chi|$ creates an elliptical trajectory at the PA output. For example, a downward slope in the AM-AM curve, $\delta G_{trans}/\delta|\chi|$, compresses the circle along the horizontal axis, as shown in Fig. 3 (center). A slope in the AM-PM curve, $\delta_{arg}\{G_{trans}\}/\delta|\chi|$, shears the circle, thereby compressing or expanding the output trajectory along the 45 degree axis, as shown in Fig. 3 (right). Thus, a complex measurement of the nonlinear gain at a specific power level is extracted from elliptical deformations of the circular trajectory.

The deformation in the circular trajectory is measured using a bridge circuit comprising a cancellation loop and a square law detector (Fig. 4). The cancellation loop output is

$$\epsilon_{RF}(t) = G_0^{-1} \cdot y_{RF}(t) - \rho \cdot x_{LO}(t), \quad (12)$$

where

$$x_{LO}(t) = \exp(j \cdot \omega_{LO} \cdot t). \quad (13)$$

The AM component within the measurement system is minimized by the cancellation loop. With the AM component removed, the nonlinear modes of the measurement system are not stimulated, and the dynamic range of the detector output signal, $y_{det} = |\epsilon_{RF}|^2$, is reduced (for the selected samples where $\lambda = \lambda_o$).

The selective sampling module, shown in Fig. 4 following the detector, contains a sample/hold circuit and an ADC. The sample/hold captures detector values y_{det} corresponding to time instants t_s when $\lambda = \lambda_o$. The cancellation loop and selective sampling reduce the resolution and

sampling rate required of the ADC, allowing for the use of lower cost components. Cost is discussed further at the end of this section.

The selectively sampled output of the detector

$$y_{det}(\theta(t_s)) = |\epsilon_{RF}(t_s)|^2 \quad (14)$$

is specified as a function of θ . If the cancellation loop is balanced and the transmitter is linear, the selectively sampled signal $y_{det}(\theta)$ is constant as a function of θ . Misalignment of the cancellation loop creates a first harmonic variation as a function of θ . Nonlinear gain (elliptical deformation) in the transmitter creates second harmonic variations. These three cases are shown in Fig. 5.

Although the use of harmonics of $y_{det}(\theta)$ for measuring nonlinearities is believed to be new (outside of the author's previous work [4],[5]), Cavers used similar first and second harmonics in [15] to measure the offsets and imbalances in modulator circuits for the special case of $\chi(k) = \lambda_o \exp(j\theta(k))$ (no cancellation loop). The author

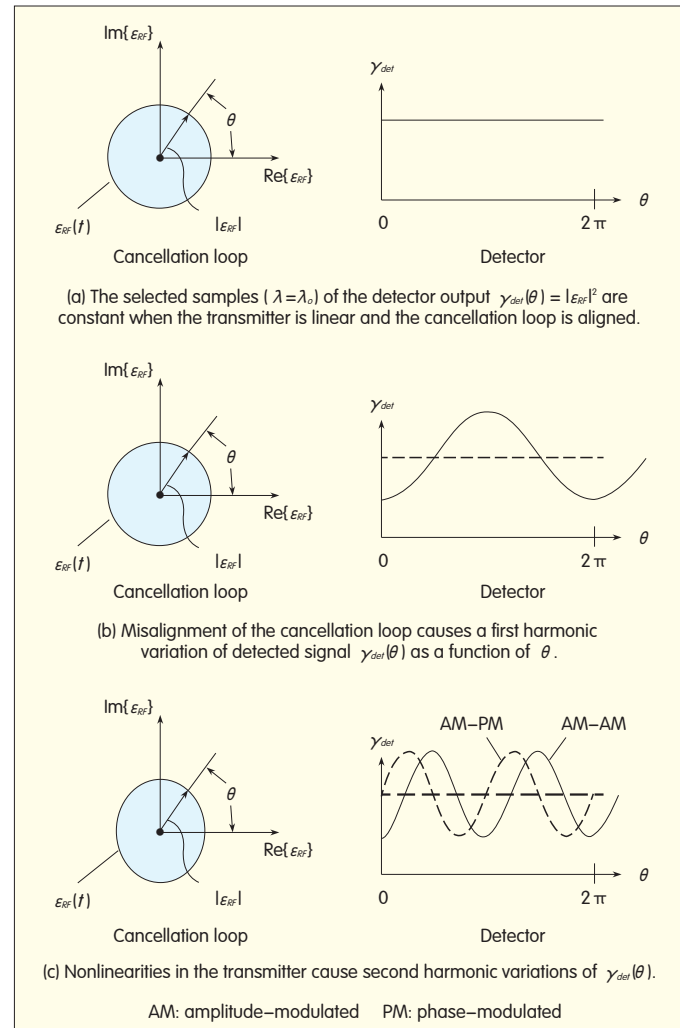


Figure 5. ▶

The selectively sampled outputs of the detector in three cases.

recommends reading [15] to obtain a better understanding of the proposed technique.

A compact method of representing selective samples of the detected signal $\gamma_{det}(\theta(k))$ is to accumulate the measurements within look-up-tables (LUTs). LUTs of the accumulated zero, first, and second-order moments of $\gamma_{det}(\theta(k))$, denoted by L_0 , L_1 , and L_2 , respectively, are

$$L_0(\theta) = \sum_k \beta(k) \quad (15)$$

$$L_1(\theta) = \sum_k \beta(k) \cdot \gamma_{det}(\theta(k)) \quad (16)$$

$$L_2(\theta) = \sum_k \beta(k) \cdot [\gamma_{det}(\theta(k))]^2, \quad (17)$$

where i is the bin index and the quantization of θ is

$$\beta(k) = \begin{cases} 1 & \text{when } |\theta(k) - \theta| < \frac{\pi}{N_{bins}}, \lambda = \lambda_0 \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

The mean and variance of γ_{det} for bin i are

$$E[\gamma_{det}(\theta_i)] = \frac{L_1(\theta_i)}{L_0(\theta_i)} \quad (19)$$

and

$$\sigma^2(\theta) = \frac{L_2(\theta) \cdot L_0(\theta) - [L_1(\theta)]^2}{[L_0(\theta)]^2}, \quad (20)$$

respectively. The first and second harmonics of $\gamma_{det}(\theta)$ are measured by demodulating the mean LUT as a function of θ . The demodulated signal becomes

$$\Gamma_{m\theta} = \frac{1}{N_{bins}} \cdot \sum_{i=1}^{N_{bins}} E[\gamma_{det}(\theta_i)] \cdot \exp(-j \cdot m \cdot \theta), \quad (21)$$

where the first and second harmonics correspond to $m=1$ and $m=2$, respectively. Memoryless measurements (Γ_θ , $\Gamma_{2\theta}$) are obtained for each value of ρ tested.

To show the cost savings of the proposed measurement approach, the approach is compared to the standard observation path. A zero-IF observation path used in [2] is chosen as the bench mark. It uses two 14 bit ADCs sampled at 122.88 MHz to measure a two-carrier WCDMA signal with a 101 carrier configuration and captures the output signal in a 16 K sample buffer. In contrast, only one ADC is needed for the proposed approach, and the measurements are stored in LUTs L_0 , L_1 , and L_2 . Each LUT

has 64 bins for a total storage of 192 bins. The dynamic range required by the proposed approach is negligible because of the circular probing signal and the cancellation loop. However, it will be assumed that an 8-bit ADC is used.

The ADC sampling rate for the proposed approach depends on how the selective sampling module in Fig. 4 is implemented. If sampling asynchronously (when $\lambda = \lambda_0$), then the sample/hold must be placed before the ADC. The required sampling rate for the ADC is determined by the inverse of the minimum time between selective samples. Because samples can be ignored, the sampling period can be made arbitrarily long, limited primarily by the maximum hold time of the sample/hold device. Thus, it is possible to reduce the sampling rate below 1 MHz to allow the use of an ADC that is integrated within a micro-controller (part of the estimation module).

An alternative implementation is to apply the ADC before the sample/hold (which becomes a digital interpolation). The detected output is sampled at about twice the Nyquist rate of the WCDMA signal so that the time instants when $\lambda = \lambda_0$ can be interpolated accurately. Because the Nyquist rate for the 101 WCDMA signal is around 15 MHz, the ADC with 32 MHz sample rate (mentioned in the introduction) would be sufficient. As mentioned in the introduction, reduction in sampling rate, resolution, and number of ADCs reduces the cost by a factor of 25 over the bench mark system.

The drawback of the proposed approach is that the acquisition time needed to measure the nonlinearity and adapt the DPD is increased compared to the standard approach. This is due to the slower accumulation of samples within the selective sampling process and, as discussed in section 3, the fact that measurements are serially obtained from several ρ values in order to estimate the DPD coefficients. However, in the standard approach, 16 K sample blocks are captured infrequently, and large blocks of incoming data between captures are ignored to reduce the DSP

requirements in the estimation module. In addition, it is possible to speed up the proposed approach using parallel measurement circuits with different ρ values, at an increased cost. There is a trade-off between cost and acquisition speed in both the proposed and standard approaches.

It is worth noting that Cavers used both a detector-based observation path (similar to the proposed approach) and the standard over-sampled observation path in [15] so that the errors in the modulator and demodulator could be measured independently. This is an example where it is more important to have several different measurement methods available rather than debate the relative merits of the individual methods used in isolation.

3 Coefficient Estimation

This section describes the estimation of the DPD coefficients \underline{b} from the memoryless measurements (Γ_θ , $\Gamma_{2\theta}$).

The relationship between (Γ_θ , $\Gamma_{2\theta}$) and the residual memoryless coefficients \underline{a} in (5), for the case of $N=4$, is approximated by

$$\mathbf{M} \cdot [(\mathbf{a}_0^* - 1) \ a_1^* \ a_2^* \ a_3^*]^T \approx \begin{bmatrix} \Gamma_\theta \\ \Gamma_{2\theta} \end{bmatrix}, \quad (22)$$

where

$$\mathbf{M} = \begin{bmatrix} \rho^2 \lambda_0 & \rho^3 c_0 \lambda_0 & \rho^4 \lambda_0 & \rho^5 c_0 \lambda_0 \\ 0 & \rho^3 c_0 c_1 \lambda_0 & \rho^4 \lambda_0^2 & \rho^5 c_0 \lambda_0 (c_1 + \lambda_0) \end{bmatrix} \quad (23)$$

$$[c_0 \ c_1] = \left[\sqrt{1 + \lambda_0^2} \ \frac{0.5 \lambda_0}{1 + \lambda_0^2} \right] \quad (24)$$

The residual memoryless coefficients \underline{a} are used to update the DPD coefficients \underline{b} , as shown in (6). However, only two measurements (Γ_θ , $\Gamma_{2\theta}$) are available for a given value of ρ . It is necessary to integrate measurements from several values of ρ to estimate all four DPD coefficients.

A set of recursive equations [16] is used to update the DPD coefficients \underline{b} . Measurements (Γ_θ , $\Gamma_{2\theta}$) from several values of ρ are combined using

$$\mathbf{K}_i = \mathbf{S}_i \cdot \mathbf{M}^H \cdot [\mathbf{M} \cdot \mathbf{S}_i \cdot \mathbf{M}^H + \mathbf{R}_i]^{-1} \quad (25)$$

$$\underline{b}(i+1) = \underline{b}(i) - \mathbf{K}_i^* \cdot [\Gamma_\theta \ \Gamma_{2\theta}]^H \quad (26)$$

$$\mathbf{S}_{i+1} = (\mathbf{I}_{4 \times 4} - \mathbf{K}_i \cdot \mathbf{M}) \cdot \mathbf{S}_i + \mathbf{Q}_i, \quad (27)$$

where $(\cdot)^H$ is conjugate transpose, $\underline{b}(0) = [1 \ 0 \ 0 \ 0]^T$, $S_0 = 2I_{4 \times 4}$, $R_i = 0.0001 I_{2 \times 2}$, and $Q_i = 0.0002 I_{4 \times 4}$. The matrix S_i is the error covariance of \underline{b} . Experiments show that it is beneficial to reset S_i to S_0 after a few cycles of the ρ set while retaining the current estimate of \underline{b} . This is likely due to the approximation used for the matrix M in (23), which assumes that $[a_0 \ a_1 \ a_2 \ a_3] \approx [1 \ 0 \ 0 \ 0]$.

The estimation module would be implemented, typically, in a micro-controller because the amount of computation needed to convert $(\Gamma_\theta, \Gamma_{2\theta})$ into coefficient updates using (23)–(27) is modest. In contrast, the standard approach for estimating coefficients involves auto- and cross-correlations of the gain basis functions (truncated to 16 K samples) and the cross-correlations of the output capture with the gain basis functions. These correlations are often computed using a high performance DSP chip. In general, micro-controllers are less expensive than DSP chips.

4 Memory Compensation

This section discusses memory compensation within the DPD module, including how to measure the PA memory using a cancellation bridge and selective sampling (as was done in section 2 for the memoryless case). In general, it is desirable to use the lowest order DPD model that makes it possible to meet the WCDMA specifications. That is, if the memoryless DPD is adequate, it can be used. However, if additional correction is required, the gain model can be extended to include memory correction.

Although it is possible to model memory using a pruned Volterra series based on delayed digital samples of the input signal, such an approach does not allow for selective sampling. A compatible approach defines the nonlinear gain as a function of $|X|$ and $\partial|X|/\partial t$, both of which are referenced to the time sample t_s when $\lambda = \lambda_0$. One possible model for the DPD gain is

$$G_{DPD}\left(|X|, \frac{\partial|X|}{\partial t}\right) = \sum_{n=0}^{N-1} b_n \cdot |X|^n + h_\omega \left\{ \frac{\partial|X|}{\partial t} \right\} \cdot \sum_{p=0}^{P-1} s_p \cdot |X|^p, \quad (28)$$

where s_p are DPD coefficients associated with the memory correction, P is the polynomial order of the memory, and $h_\omega(\cdot)$ is a bandpass filter used to limit the high frequency noise. The gain model for the predistorted transmitter is

$$G_{trans}\left(|X|, \frac{\partial|X|}{\partial t}\right) = \sum_{n=0}^{N-1} a_n \cdot |X|^n + h_\omega \left\{ \frac{\partial|X|}{\partial t} \right\} \cdot \sum_{p=0}^{P-1} r_p \cdot |X|^p \quad (29)$$

where r_p are coefficients associated with the memory component of the residual nonlinearity.

Residual PA memory is measured using a cancellation loop, detector, and selective sampling (Fig. 4). The selective sampling in this case produces time-aligned triples $(\theta, h_\omega\{\partial|X|/\partial t\}, \gamma_{det})$ at each sample instant t_s . This allows the detector output to be expressed as a function of two variables, $\gamma_{det}(\theta, \partial|X|/\partial t)$.

The separable form of (28) allows the DPD coefficients to be estimated using two sets of LUTs: the memoryless LUTs defined in (15) and (16), and memory LUTs defined in (31) and (32). A new measure of the memory is needed, which is

$$\psi(k) = \frac{[\gamma_{det}(k) - \gamma_0] \cdot h_\omega(k)}{|h_\omega(k)|^2 + |\delta|^2} \quad (30)$$

where γ_0 is the expected value of γ_{det} (average radius of the circle formed by the selectively sampled points), $|\delta|$ is a small constant used to prevent a divide by zero, and $h_\omega(k)$ is the filtered derivative $h_\omega\{\partial|X|/\partial t\}$ sampled at time k . The numerator in (30) measures the correlation of the detector output γ_{det} with the filtered derivative $h_\omega\{\partial|X|/\partial t\}$. The subsequent estimation and update of the memory DPD coefficients minimizes the correlation.

The accumulated LUTs used for the memory estimation are

$$L_{0,\partial|X|}(\theta) = \sum_k \beta(k) \cdot |h_\omega(k)|^2 \quad (31)$$

and

$$L_{1,\partial|X|}(\theta) = \sum_k \beta(k) \cdot [\gamma_{det}(k) - \gamma_0] \cdot h_\omega(k). \quad (32)$$

The mean LUT is

$$E[\psi(\theta)] = \frac{L_{1,\partial|X|}(\theta)}{L_{0,\partial|X|}(\theta)}, \quad (33)$$

which provides an estimate of (30) as

a function of θ . Using the mean LUT, the demodulated signal becomes

$$\psi_{me} = \frac{1}{N_{bins}} \cdot \sum_{i=1}^{N_{bins}} E[\psi(\theta)] \cdot \exp(-j \cdot m \cdot \theta). \quad (34)$$

Memory measurements $(\psi_\theta, \psi_{2\theta})$ are obtained for each value of ρ tested.

As in the memoryless case, memory measurements $(\psi_\theta, \psi_{2\theta})$ are integrated over several values of ρ . The relationship between $(\psi_\theta, \psi_{2\theta})$ and the residual memory coefficients \underline{r} , for the case of $P = 4$, is approximated by

$$M \cdot [r_0^* \ r_1^* \ r_2^* \ r_3^*]^T \approx \begin{bmatrix} \psi_\theta \\ \psi_{2\theta} \end{bmatrix}, \quad (35)$$

where M is the same matrix defined in (23). The memory coefficients of the DPD gain in (28) are updated using

$$\underline{s}(i+1) = \underline{s}(i) - \alpha \cdot [r_0 \ r_1 \ r_2 \ r_3]^T, \quad (36)$$

where $\underline{s}(i) = [s_0 \ s_1 \ s_2 \ s_3]^T$ for the iteration i .

The estimations of the memoryless and memory DPD coefficients \underline{b} and \underline{s} are decoupled in this implementation. Decoupling does not impact the convergence when $E[h_\omega\{\partial|X|/\partial t\}] = 0$ for each angle θ of the selectively sampled input signal. This is a reasonable assumption for a WCDMA signal. It is recommended that the memoryless coefficients be adapted first, in isolation, because the uncorrected memoryless distortion tends to be larger than the memory-related distortion. Both components are adapted concurrently once the residual memoryless nonlinearity is reduced to a level comparable to the memory component.

The number of basis functions in the memory model can be adjusted using the polynomial order P . In most cases, the order of the memory is less than the order of the memoryless nonlinearity, that is, $P < N$. In such cases, a subset of the matrix M is used in (35). If needed, the size of the memory model can be increased without increasing the polynomial order P by using a set of filtered derivatives, $h_\omega(n)\{\partial|X|/\partial t\}$, that have different frequency responses. Separate memory measurements $(\psi_\theta, \psi_{2\theta})_{\omega(n)}$ could be obtained for each filter $h_\omega(n)$ and used to estimate additional memory coefficients, $\underline{s}_{\omega(n)}$.

Assuming that the number of filters used is N_n , the total number of memory basis functions available becomes $N_n P$. Incorporating more filters $h_{\omega(r)}$ into the DPD model in (28) is similar to expanding the memory depth of a memory polynomial.

The DPD model in (28) is designed to be compatible with the proposed measurement system and the selective sampling process. It is based on filtered derivatives of $|X|$ instead of delayed values of $|X|$. However, the difference between a derivative and a delay is minor in practice. For example, $h_r \{ \partial |X| / \partial t \}$ can be approximated by $|X(t - \tau T)| - |X(t + \tau T)|$, where T is the sampling period. Thus, the basis functions have the form

$BF_V = |\chi(t - \tau T)|^* |\chi(t)|^{n-1}$, which is similar to the memory polynomial shown in (10). The memory depth of this representation can be made large and expressive, like the memory polynomial, but there is no value in exceeding what is needed.


At this point the question might be asked “Which is better?”, the DPD model in (28) or the pruned Volterra models in (9) and (10)? This question is misguided. When selecting the basis function set, it is important to remember that the goal of linearization is to meet the WCDMA specification, and there is no prize for destroying the ACLR specification by 10 dB. In fact, excess margin is money wasted [17].

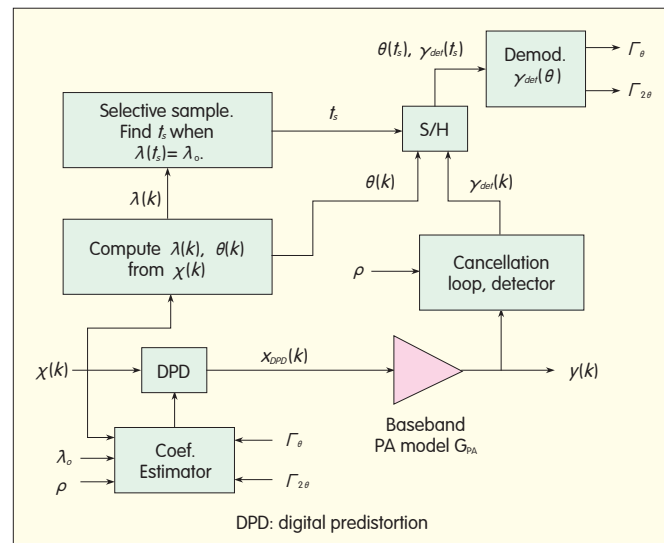
The DPD module is often implemented in an FPGA. The cost of an FPGA, in general, is related to the number of multipliers used. Thus, to minimize the multiplier usage and reduce cost, the smallest basis function set that meets the WCDMA specification should be used.

5 Results

A proof of concept can verify that a polynomial predistorter tuned using selective samples of a WCDMA signal will converge. It can also estimate the ACLR performance of the proposed transmitter.

The author has a library of PA characterizations, without linearization, that comprise digitized data captures of

Figure 6.  A Matlab simulation of the digital transmitter and selective sampling process is used to demonstrate the convergence of the DPD algorithm. Memory correction is included in the simulation but is not shown above.



input and output signals. A class AB-biased amplifier driven by a two-carrier WCDMA signal is chosen for the proof of concept. The WCDMA signal has a 101 carrier configuration and is crest factor reduced to a peak-to-average power ratio (PAPR) of 7.2 dB. The RF output power and center frequency of the PA are 45.3 dBm and 2.14 GHz, respectively. The sample rate used in the input and output data captures is 122.88 MHz. The data captures are 614,400 samples long and time-aligned to remove the delay between the input signal and the linear component of the output signal. The synchronized data captures of the PA input and output signals are denoted by $x(k)$ and $y_o(k)$, respectively. These are used to compute the gain of the PA model, GPA. Note that $y_o(k)$ denotes the digitized data capture of the actual uncorrected PA output signal whereas $y(k)$ is used to denote the simulated predistorted PA output signal computed using $G_{PA}x_{DPD}(k)$.

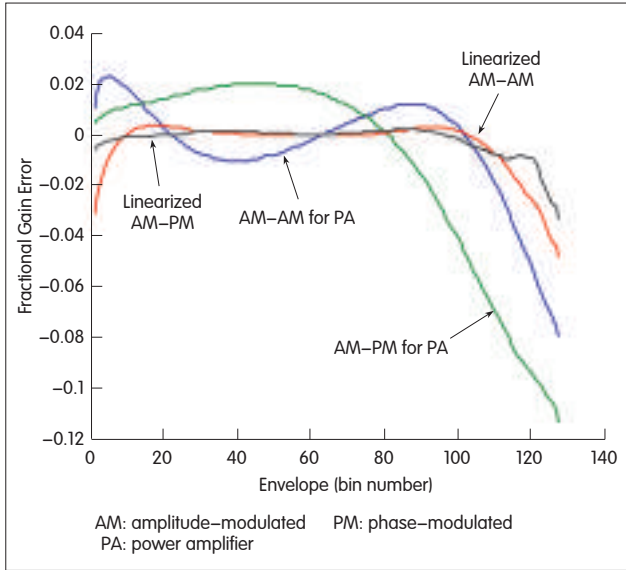
The PA gain GPA and the input data capture $\chi(k)$ are used in a Matlab simulation of the digital transmitter to demonstrate the convergence of the DPD algorithm. The structure of the transmitter is shown in Fig. 6. The memoryless portion of GPA is represented as AM-AM and AM-PM LUTs. The LUT representation was chosen for the PA gain so that any modeling errors would differ from those

associated with the polynomial gain models used in the digital predistortion. GPA is augmented by a memory component estimated from the output capture y_c using a least square technique, similar to (7) (see [2], [6], or [7] for details). Once the algorithm has converged, the steady-state value of the DPD gain function GDPD provides an estimate of the inverse nonlinear gain needed to linearize the PA.

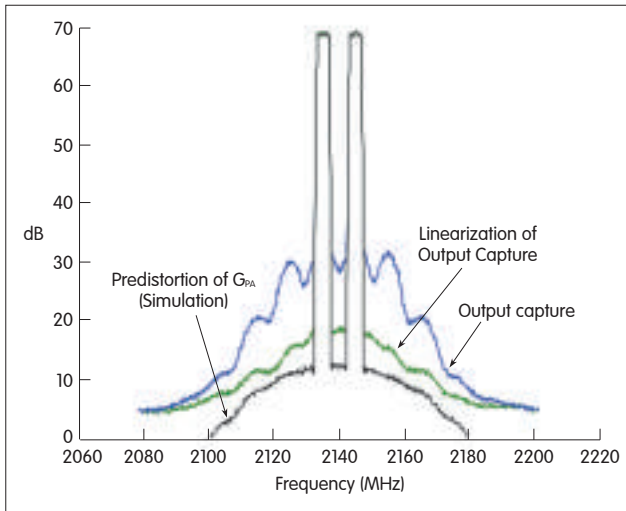
The ACLR results from the Matlab simulation, based on $y(k) = G_{PA}x_{DPD}(k)$, are optimistic because noise and some nonlinear behaviors from the actual PA are not modeled. In contrast, the output capture $y_o(k)$ contains noise and intermodulation (IMD) products generated by the actual PA. As a result, the actual ACLR performance can be estimated by linearizing the output data capture $y_o(k)$ using the inverse PA gain function (G_{DPD}); that is, $y(k) = G_{DPD}y_o(k)$. Although this is not the same as predistortion, the difference in ACLR performance resulting from commuting G_{DPD} and G_{PA} is minimal for mild nonlinearities [9].

The memoryless and memory DPD coefficient estimations are performed independently as decoupled processes. The memoryless adaptation is performed initially in isolation to reduce the residual nonlinearity. After this, both the memoryless and memory adaptations are performed concurrently.

Let us begin with the estimation of the memoryless portion of the DPD gain in



◀ Figure 7. AM-AM and AM-PM curves for the uncorrected and predistorted PA models (From [5] © 2010 IEEE).



◀ Figure 8. WCDMA output spectra of the original output capture y_o (no linearization), the converged predistorted transmitter model $y = G_{PA} \chi_{DPD}$ (Matlab simulation), and the linearized output capture $y = G_{DPD} y_o$.

(28), which has four coefficients ($N=4$). In order to estimate all of the coefficients, several values of ρ are selected, $\rho = [0.5, 0.8, 1.1, 1.4]$, and tested sequentially. For a given value of ρ , the signals $y(k)$ and $\chi_{det}(k)$ are computed for the current setting of the DPD module GDPD. A value for λ_o is chosen ($\lambda_o = 0.35$), and selective sampling is applied to determine the sampling instants t_s where $\lambda = \lambda_o$. Corresponding values of $\theta(t_s)$ and $\chi_{det}(t_s)$ are sampled. The best method for determining t_s is to interpolate $\chi(k)$ to localize the instants when $\chi(k)$ crosses the circle defined by (ρ, λ_o) . However, for ease of implementation using Matlab, the signal $\chi(k)$ is up-sampled by a factor of 4, and

samples are selected if $\chi(k)$ is within $0.025\lambda_o$ of the (ρ, λ_o) circle. $\theta(t_s)$ and $\chi_{det}(t_s)$ are accumulated in LUTs in (15) and (16), from which $E[\chi_{det}(\theta)]$ is computed using (19), then demodulated using (21) for $m=[1, 2]$ to produce the memoryless measurements Γ_θ and $\Gamma_{2\theta}$.

▼ Table 1. ACLR for Various Output Spectra

Output Spectrum	ACLR (dBc)		ACLR ₂ (dBc)	
	Lower	Upper	Lower	Upper
Output Capture, y_o (Uncorrected)	-41.1	-40.3	-39.0	-37.5
Predistorted Model, $y = G_{PA} \chi_{DPD}$	-57.1	-57.2	-57.6	-58.1
Linearized Output Capture, $G_{DPD} y_o$	-52.2	-51.5	-53.4	-53.4
WCDMA Specification [1]	< -45	< -45	< -50	< -50

ACLR: adjacent channel leakage ratio

Fig. 7 shows the AM-AM and AM-PM curves for the uncorrected and predistorted PA models. The measurement system and the recursive (25)–(27) produce steady-state DPD coefficients \underline{b} that flatten the AM-AM and AM-PM curves as desired over the input envelope range where the probability density function (PDF) of $|x|$ has its highest density.

The class AB-biased amplifier requires some memory correction to ensure that the worst case ACLR of the transmitter passes the WCDMA specification with at least 3 dB of margin (for manufacturing tolerances). The memory model chosen comprises one term, s_0 . From (28), where $N = 4$ and $P = 1$, the DPD gain becomes

$$G_{DPD}(|\chi|, \frac{\partial \chi}{\partial t}) = \sum_{n=0}^3 b_n \cdot |\chi|^n + h_\omega \left\{ \frac{\partial \chi}{\partial t} \right\} \cdot s_0. \quad (37)$$

A weighted average of the memory measurements ψ_θ from the four ρ values is used to update the memory coefficient s_0 , that is,

$$s_0(i+1) = s_0(i) - \alpha \cdot \left[\sum_{j=0}^3 w_j \cdot \psi_{\theta_j} \cdot [\rho_j^2 \lambda_o]^{-1} \right]^*, \quad (38)$$

where $\rho_j = [0.5, 0.8, 1.1, 1.4]$, ψ_{θ_j} is the memory measurement ψ_θ obtained for ρ_j , $\alpha < 1$, and w_j is a weight defined by

$$w_j = \frac{\min_i \{L_{0,d|x|}(\theta_i; \rho_j)\}}{\sum_{i=0}^3 \min_i \{L_{0,d|x|}(\theta_i; \rho_i)\}}. \quad (39)$$

That is, the weight used for each ρ_j is determined by the minimum bin value of the $L_{0,d|x|}$ LUT used in the memory estimation (31). Note that $\rho^2 \lambda_o$ in (38) corresponds with $M(1, 1)$ in (23).

Fig. 8 shows the output spectra for three cases: the original output capture $y_o(k)$, the predistorted memory PA model $y(k) = G_{PA} \chi_{DPD}(k)$, and the output

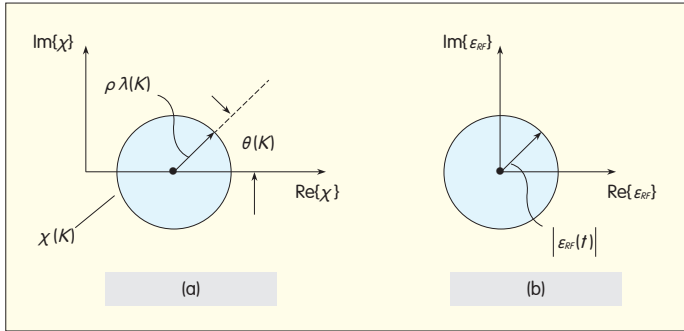


Figure 9. Baseband input and RF output signals for the inverse model.

capture linearized using the converged value of G_{DPD} , that is, $y(k) = G_{DPD}y_o(k)$. Table 1 contains the ACLR measurements.

Comparing the uncorrected capture $y_o(k)$ and predistorted memory results $y(k) = G_{PA}x_{DPD}(k)$, the steady-state DPD coefficients reduce the ACLR2 by 18.6 dB to -57.6 dBc. This is well below the WCDMA specification of -50 dBc. This result validates the memory measurement and shows that the coefficient estimation converges. It also validates the implementation of the memoryless and memory coefficient estimations as decoupled processes. The ACLR measurements are optimistic because they are based on a simulated model of the PA.

The estimate of the actual ACLR performance is obtained from the linearized output capture $G_{DPD}y_o(k)$ and is shown in Table 1. The ACLR2 values pass the WCDMA specification with 3.4 dB of margin, as desired. Thus, the basis function set ($N=4$, $P=1$) used to form the memory DPD (37) provides sufficient linearization of the PA for a 101 WCDMA input signal. The basis function set in (37) is a compact representation (low number of coefficients) allowing the DPD module to be implemented in a smaller FPGA than would be possible if a pruned Volterra series with a large memory depth was chosen for the DPD model.

6 Future Extension

A future extension of the DPD approach includes the use of inverse modeling. In this alternative implementation, the inverse nonlinearity of the transmitter is computed by using the output of the detector y_{det} to control

the selective sampling. The sampling instances t_s are selected when $y_{det}(t_s) = y_o$. The precision required by the measurement system is reduced because the detector output at the selectively sampled instants t_s is constant by definition. For example, a single bit comparator can be used to control the sampling, where the sample times t_s are indicated by changes in the output state of the comparator. The elliptical deformations associated with residual nonlinearities appear in the selected input samples $x(t_s)$ and result in a time-varying value of $\lambda(t_s)$, as shown in Fig. 9(a).

The selected samples of $x(t_s)$ are used to compute (ρ, λ, θ) . The measurement LUTs are created based on (θ, λ) instead of (θ, y_{det}) . The LUT capturing the first-order moment of λ (instead of y_{det}) as a function of quantized values of θ is

$$L_{1\lambda}(\theta) = \sum_k \beta_i(k; y_o) \cdot \lambda(k), \quad (40)$$

where the sampling and quantization are defined by

$$\beta_i(k; y_o) = \begin{cases} 1 & \text{when } |\theta(k) - \theta_i| < \frac{\pi}{N_{bins}} \cdot y_{det} = y_o \\ 0 & \text{otherwise} \end{cases} \quad (41)$$

The mean of λ for bin i is

$$E[\lambda(\theta_i)] = \frac{L_{1\lambda}(\theta_i)}{L_0(\theta_i)}. \quad (42)$$

The demodulated signal for the inverse nonlinearity becomes

$$\Gamma_{mb} = \frac{1}{N_{bins}} \cdot \sum_{i=1}^{N_{bins}} E[\lambda(\theta_i)] \cdot \exp(-j \cdot m \cdot \theta_i). \quad (43)$$

The integration of the measurements uses the same equations as the previous implementation in (25)–(27), except that the sign of the coefficient update is reversed.

The nonlinear gain of the transmitter is indicated by an elliptical trajectory at the input, instead of the output as shown in Fig. 3.

7 Conclusion

A method has been proposed where selective sampling of a WCDMA signal is used to obtain a probing signal that highlights nonlinearities within a transmitter. The measurement circuitry comprises a cancellation loop, detector, sample/hold, and an ADC. The cancellation loop and selective sampling reduce the linearity, resolution, and sampling rate required of the ADC. LUTs defined by (15) and (16) provide a compact representation of the detected signal $y_{det}(\theta)$, reducing the storage requirements compared with an over-sampled capture of the PA output signal. A DPD model with memory correction that is compatible with the proposed measurement system has been presented. Results show that a compact DPD model can linearize a class AB biased PA sufficiently to meet the ACLR specifications for a 101 WCDMA input signal. The approach reduces the cost of the observation path, estimation module, and DPD module compared to a standard transmitter with DPD based on a pruned Volterra series with a large memory depth and adapted using output captures based on over-sampled data.

References

- [1] *Base Station (BS) Radio Transmission and Reception (FDD)*, 3GPP TS 25.104 v8.3.0., 2008.
- [2] R. N. Braithwaite, "Memory correction of a Doherty power amplifier with a WCDMA input using digital predistortion," in *IEEE MTT-S Int. Microw. Symp.*, San Francisco, CA, 2006, pp. 1526–1529.
- [3] Analog Devices [Online]. Available: <http://www.analog.com>
- [4] R. N. Braithwaite, "Method and system for reducing non-linearities," U.S. patent 6566948, May 20, 2003.
- [5] R. N. Braithwaite, "Measurement and correction of residual nonlinearities in a digitally predistorted power amplifier," in *75th ARFTG Microw. Measurement Conf.*, Anaheim, CA, 2010, pp. 14–17.
- [6] R. N. Braithwaite and S. Carichner, "An improved Doherty amplifier using cascaded digital predistortion and digital gate voltage enhancement," *IEEE Trans. Microw. Theory and Tech.*, vol. 57, no. 12, pp. 3118–3126, Dec. 2009.
- [7] R. N. Braithwaite, "Wide bandwidth adaptive digital predistortion of power amplifiers using reduced order memory correction," in *IEEE MTT-S Int.*

- Microw. Symp.*, Atlanta, GA, 2008, pp. 1517–1520.
- [8] A. Zhu, J. C. Pedro, and T. R. Cunha, "Pruning the Volterra series for behavioral modeling of power amplifiers using physical knowledge," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 5, pp. 813–821, May 2007.
- [9] A. Zhu, P. J. Draxler, J. J. Yan, T. J. Brazil, D. F. Kimball, and P. M. Asbeck, "Open-Loop digital predistorter for RF power amplifiers using dynamic deviation reduction-based Volterra series," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 7, pp. 1524–1534, July 2008.
- [10] J. Staudinger, J. Nanan, and J. Wood, "Memory fading Volterra series model for high power infrastructure amplifiers," *IEEE Radio Wireless Symp.*, New Orleans, LA, 2010, pp. 184–187.
- [11] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Comm.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.
- [12] O. Hammi, S. Carichner, B. Vassilakis, and F. M. Ghannouchi, "Synergetic crest factor reduction and baseband digital predistortion for adaptive 3G Doherty power amplifier linearizer design," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 11, pp. 2602–2608, Nov. 2008.
- [13] F. Mkadem, M. B. Ayed, S. Boumaiza, J. Wood, and P. Aen, "Behavioral modeling and digital predistortion of power amplifiers with memory using two hidden layers artificial neural networks," in *IEEE MTT-S Int. Microw. Symp.*, Anaheim, CA, 2010, pp. 656–659.
- [14] R. N. Braithwaite, "Adaptive digital predistortion of nonlinear power amplifiers using reduced order memory correction," presented at workshop on "Highly efficient linear power transmitters for wireless applications based on switching mode amplifiers," *IEEE MTT-S Int. Microw. Symp.*, Atlanta, GA, 2008.
- [15] J. K. Cavers, "New methods for adaptation of quadrature modulators and demodulators in amplifier linearization circuits," *IEEE Trans. Vehicular Tech.*, vol. 46, no. 3, pp. 707–716, August 1997.
- [16] R. G. Brown, *Introduction to Random Signal Analysis and Kalman Filtering*. New York, NY: Wiley & Sons, 1983.
- [17] Ted Heil, vice president, Mini-Circuit Labs, New York, personal communications.
- [18] R. N. Braithwaite, "General principles and design overview of digital predistortion," chapter in *Digital Processing for Front End in Wireless Communication and Broadcasting*, F. Luo (Ed.), Cambridge Univ. Press, 2011.

Biography

R. Neil Braithwaite (nbraithwaite@pwav.com) received his B.Sc. degree in electrical engineering from the University of Calgary in 1985. He received his M.Sc. and Ph.D. degrees from the University of British Columbia in 1989 and 1992. From 1992 to 1995, he conducted postdoctoral research at the University of California, Riverside. From 1985 to 1987 and 1995 to 2002, he worked for Computing Devices Company (Canada), Nortel (Canada), and Agilent Laboratories (USA). Since 2002, he has been working for Powerwave Technologies (USA). He is the author of several papers and patents, as well as a recent book chapter on digital predistortion in RF power amplifiers [18].

Roundup

ZTE, Telkom Kenya–Orange Announce Nations' Fastest Commercial HSPA+ Network

September 8, 2011, Shenzhen—ZTE Corporation, a publicly-listed global provider of telecommunications equipment and network solutions, and its partner, Telkom Kenya–Orange, launched Kenya's fastest HSPA+ network, based on Uni-RAN, in Nairobi. The network allows Kenyans to access high-speed wireless Internet with a download speed of up to 21 Mbps. The launch was attended by Kenyan President Mwai Kibaki and more than 200 guests.

The newly developed network greatly improves user experience, dramatically reduces OPEX, and can evolve to LTE. The joint project reflects the combination of flexible and efficient networking. As part of the agreement, ZTE provided Telkom Kenya–Orange with Uni-RAN-based all-IP solutions. The platform was then used to build a unified 2G/3G mobile network for the operator.

In December 2007, France

Telecom, the world's fourth largest telecom company, acquired a 51 percent stake in Telkom Kenya. The Kenyan government has the remaining stake. France Telecom was granted a license to operate 3G in Kenya in 2010.

At the press conference, Telkom Kenya–Orange CEO, Mickael Ghossein, said, "We are committed to creating a fresh wireless lifestyle for Kenyans. The advanced HSPA+ network built in partnership with China's leading communications firm, ZTE, will provide access service of the highest quality to users and will increase productivity."

"To help Telkom Kenya–Orange fulfill growing market demand for reliable mobile application services, we provided them with world-class telecom technologies and effective solutions with the lowest TCO in the industry," said Xu Chengrong, CEO of ZTE Kenya. "We are committed to the

industry and provide our clients with robust telecom solutions and reliable technologies that can be used to seamlessly upgrade their networks. This enables them to develop value-added services for their customers. We intend to further strengthen our R&D efforts and develop more customized solutions to boost network service quality."

ZTE produces leading Uni-RAN technology, and has made breakthroughs in high-end markets, including markets in Europe. The company has deployed several LTE/HSPA networks in Germany and Belgium by cooperating with KPN; in Hungary and Montenegro by cooperating with Telenor; in Portugal by cooperating with Optimus; and in Sweden by cooperating with H3G. ZTE's Uni-RAN technology is becoming a core solution in the mobile broadband era and a key source of profit for operators. (ZTE Corporation)

A New Two-Branch Amplification Architecture and its Application with Various Modulated Signals

W. Hamdane, A. B. Kouki, and F. Gagnon

(Department of Electrical Engineering, École de Technologie Supérieure, Montréal, Québec H3C 1K3, Canada)

Abstract: This paper proposes a new two-branch amplification architecture that combines baseband signal decomposition with RF front-end optimization. In the proposed architecture, the filtered modulated signals are separated into two components that are then amplified independently and combined to regenerate an amplified version of the original signal. A branch with an efficient amplifier transmits a low-varying envelope signal that contains the main part of the information. Another branch amplifies the residual portion of the signal. The baseband decomposition and parameters of the RF part are optimized to find the configuration that gives the best power efficiency and linearity. For M-ary quadrature amplitude modulation (M-QAM) signals, this technique is limited in terms of power efficiency. However, for filtered continuous phase modulation (CPM) signals, especially for minimum shift keying (MSK) and Gaussian MSK (GMSK) signals, high power efficiency can be achieved with no significant impact on the overall linearity. The results show that this technique gives better performance than the single-ended class-B amplifier.

Keywords: CPM modulation; M-QAM; RF power amplifiers; DC-RF efficiency; linearity; crest factor; shaping filters

1 Introduction

The design of new wireless transmitters is becoming more complex as demand on spectral resources increases and available power for wireless terminals becomes limited. The amount of band occupied by the modulation scheme must be as small as possible in order to accommodate many channels in a given band. For this reason, complex modulation schemes such as M-ary quadrature amplitude modulation (M-QAM) (M up to 256) and M-ary phase shift keying (M-PSK) modulation have been proposed [1], [2]. These modulations are spectrally efficient, especially with baseband filtering such as root raised cosine filtering (RRCF). However, the resulting filtered signals have very high peaks, which are problematic for the power amplifier (PA) stage. To achieve linear amplification

with acceptable power efficiency, many approaches have been proposed [3]–[10]. In one approach, single branch amplification, typically a class-AB or class-B amplifier, is used to reduce power consumption. A linearization stage, such as digital predistortion [3]–[5], is generally needed with such topology to achieve good linearity and meet linearity requirements. Another approach is to use multibranch schemes [6]–[10], which are more complex but offer greater flexibility. The feed-forward technique [6] is very linear, especially for large-bandwidth signaling, but is not power efficient. Linear amplification with nonlinear components (LINC) [7]–[10], based on outphasing, is another solution. LINC is inefficient in terms of high crest factor signals, which measure the ratio of peak to average power.

Modulations with constant

amplitude, such as continuous phase modulation (CPM) [2], [11]–[14] have also been developed. The information is carried in the signal's phase, which is shaped with different types of windows for smooth phase transitions. These modulations have constant amplitude, and efficient non-linear amplifiers can be used, which makes CPM attractive. However, the spectral efficiency of this modulation depends on signal characteristics such as phase shaping and modulation index [2], [11] and is generally quantized in terms of out-of-band radiation, that is, at the side-lobe level. Unfortunately, spectrally efficient CPM signals, particularly those with low modulation index, have degraded bit error rate (BER) performance and require complex receiver architecture [2], [11]. Pulse-shaping CPM signals can therefore be considered. However, the constant envelope property is lost, and

▼ Table 1. Impact of shaping filter on the crest factor

	QPSK			16QAM		
No Filter Used	0	0	0	2.5	2.5	2.5
α_{rc}	0.2	0.35	0.5	0.2	0.35	0.5
Crest Factor (dB)	5	3.7	3.1	7.6	6.2	5.5

QAM: quadrature amplitude modulation

QPSK: quadrature phase-shift keying

the design of a linear, power-efficient transmitter arises as a new issue.

In this paper, we propose a new two-branch amplification architecture that combines digital signal processing and RF front-end optimization. In section 2, we give a detailed description of the transmitter and the motivations for using a two-branch structure. The signal decomposition technique and overall power efficiency are also presented. In section 3, we investigate the use of this system with digitally modulated signals such as M-QAM and filtered CPM. In section 4, we discuss the performance of the proposed architecture and compare it with a single-ended class-B PA. Concluding remarks are given in section 5.

2 Proposed Two-Branch Amplification System

Using a baseband-shaping filter for digital modulations is unavoidable because it limits the amount of band occupied by the transmitted signal and eliminates inter-symbol interference (ISI) according to the Nyquist criteria [2]. However, baseband-shaping filters increase the signal crest factor, which directly impacts efficiency in the amplification stage. For example, a quadrature phase-shift keying (QPSK) signal has a constant amplitude. When filtered with an RRCF filter that has a roll-off (α_{rc}) of 0.2, the crest factor of the signal can rise up to 5 dB. For a 16QAM signal, the crest factor increases from 2.5 dB for a non-filtered constellation to around 7 dB with the same RRCF (Table 1).

In both cases, the amount of spectrum occupied by the transmitted signal is greatly reduced by filtering. But this comes at expense of a dramatic increase in envelope variation.

Using single-ended amplifiers generally requires high back-off levels in order to maintain acceptable linearity, and these high back-off levels reduce power efficiency. Therefore, to improve efficiency while ensuring good linearity, we propose a new two-branch amplification system. This system is based on a special decomposition of the filtered input signal. First, the decomposition is applied in baseband, and then specifications of the RF front-end are determined. This entails specifying the output combining structure and power handling capabilities (P_{1dB} and gain) of the amplifiers that are needed in each branch. Therefore, the amplifiers and output combiner must be designed specifically to meet the established specifications.

2.1 Signal Decomposition

The filtered output of a digitally modulated signal is denoted by $h(t)$. It is decomposed in baseband into two different signals, $m(t)$ and $r(t)$, which have desired power properties and are amplified by optimized amplifiers. $m(t)$ is the main signal, has a constant or low-varying envelope, and carries the main part of the information. $r(t)$ is the remaining part of $h(t)$. Geometrically, this can be represented as a projection of the vector $h(t)$ on the chosen region for $m(t)$ followed by a calculation of $r(t)$. The placement of the projection region is denoted by γ , the ratio of average power of $m(t)$ to average power of $r(t)$. γ impacts the power levels in both branches and the ratio between these power levels. This ratio, in turn, determines the characteristics and performance of the required amplifiers. The optimal choice of γ is a function of the probability density function (PDF) distribution of $h(t)$ and should be such that the difference

between the average powers in both paths is minimal.

2.2 RF Front-End Architecture

The decomposed signal must be amplified and combined using the RF transmitter front-end depicted in Fig 1. The RF transmitter front-end consists of two PAs (main and auxiliary amplifiers) and a directional coupler. The main amplifier generates a high-power output signal and should be a highly efficient class-B, C or F amplifier [15]–[17]. Because the input signal has a low-varying envelope, a non-linear power-efficient amplifier can be used without introducing significant distortion into the main signal component. The auxiliary amplifier amplifies the low-power residual signal. This amplifier must be sufficiently linear to keep the overall system linear. Then, a directional coupler is used to sum the outputs of both amplifiers and generate an amplified version of the filtered signal. A directional coupler is a four-port device typically realized in microstrip or stripline technology by closely spaced transmission lines [6].

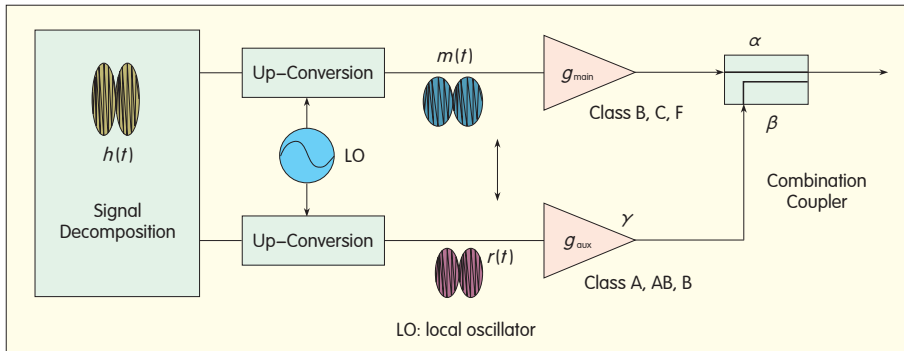
The parameters of this architecture are as follows:

Input parameters

$m(t)$	input signal of the main amplifier
$r(t)$	input signal of the auxiliary amplifier
P_{in_main}	average power of the main signal
P_{in_aux}	average power of the residual signal
γ	ratio of the average power of the main signal to the average power of the residual signal
$\Delta m(t)$	crest factor of the main signal
$\Delta r(t)$	crest factor of the residual signal

Amplification stage parameters

α	insertion loss of the coupler
β	power coupling coefficient
g_{main}	main amplifier's gain
g_{aux}	auxiliary amplifier's gain
P_{DC_main}	DC power of main amplifier
P_{DC_aux}	DC power of auxiliary amplifier
η_{main}	main amplifier power efficiency
η_{aux}	auxiliary amplifier power efficiency
P_{out_main}	average output power of the main amplifier
P_{out_aux}	average output power of the auxiliary amplifier
$P_{1dB}(A_{main})$	P_{1dB} of the main amplifier



▲ Figure 1. Amplification system model.

▼ Table 2. QPSK and 16QAM decomposed signal dynamics

	QPSK			16QAM		
α_{rc}	0.2	0.35	0.5	0.2	0.35	0.5
$\Delta h(t)$: (dB)	5	3.7	3.1	7.6	6.2	5.5
$\Delta m(t)$: (dB)	0	0	0	2.5	2.5	2.5
$\Delta r(t)$: (dB)	10.9	11.3	11.7	15.2	12.7	12.7
γ : (dB)	10.9	11.3	11.7	16.0	16.2	16.2

QAM: quadrature amplitude modulation QPSK: quadrature phase-shift keying

$P_{1dB}(A_{res})$ P_{1dB} of the auxiliary amplifier

The input parameters for the amplification stage are obtained after optimization has been done in the signal decomposition stage. The amplification stage parameters are the parameters for both amplifiers and the directional coupler. Based on the architecture in Fig. 1, we can write the following equations:

$$\gamma = \frac{P_{in_main}}{P_{in_aux}} \quad (1)$$

$$P_{out_main} = P_{in_main} \cdot g_{main} \quad (2)$$

$$\text{and} \quad P_{out_aux} = P_{in_aux} \cdot g_{aux} \quad (3)$$

The total output power of the amplification system, P_{out} , is given by:

$$P_{out} = \alpha P_{out_main} + \beta P_{out_aux} \quad (4)$$

The coupler's parameters satisfy the following equation [6]:

$$\alpha = 1 - \beta \quad (5)$$

For perfect linearity, the two paths must be balanced and must have the same gain. This condition can be mathematically expressed as

$$g_{main} \alpha = \beta g_{aux} \quad (6)$$

With the system parameters established, it is possible to determine

the ratio $P_{1dB}(A_{main})$ to $P_{1dB}(A_{res})$ and the global power efficiency of the proposed amplification architecture. After mathematical simplification (Appendix A), the final expression of the ratio $P_{1dB}(A_{main})$ to $P_{1dB}(A_{res})$, denoted as χ , is

$$\chi = \frac{P_{1dB}(A_{main})}{P_{1dB}(A_{res})} = \frac{\gamma \Delta m(t) \beta}{\Delta r(t) \alpha} \quad (7)$$

Based on χ , it is possible to predict the potential use of this technique with a given signal. As shown in (7), this parameter depends only on the signal envelope statistics and the coupler parameters. The global efficiency is expressed as

$$\eta_{global} = \frac{\alpha \eta_{main} (1 + \frac{1}{\gamma})}{1 + \frac{\alpha \eta_{main}}{(1 - \alpha) \gamma \eta_{aux}}} \quad (8)$$

The power efficiency of the amplification system is then a function of the power efficiency of the two amplifiers, the coupler parameters, and the ratio of the average powers of both signals. For a given set of amplifiers, fixed η_{main} and η_{aux} , and a given signal, fixed γ , the optimal efficiency depends solely on the choice of the coupler. A very low coupling coefficient leads to a highly inefficient auxiliary PA. Too much

coupling leads to considerable loss in the main path.

In the next section, we discuss the use of this amplification technique with different input signals. Based on χ , it is possible to determine whether the system will be power efficient with a given signal or not.

3 Use of The Proposed Amplifier with Various Signals

3.1 The Case of M-QAM Signals

The first step in determining the suitability of this technique for the signals of interest is to decompose the signals and calculate all obtained parameters, particularly γ , $\Delta m(t)$, and $\Delta r(t)$. Then, χ for different coupling coefficients is calculated, which allows us to determine whether there is a potential gain when using this technique. For M-QAM signals, we focus on the QPSK and 16QAM modulations. For the QPSK, the main signal corresponds to its mean value $m(t)$. $r(t)$ is the vectorial difference between $m(t)$ and $h(t)$. For the 16QAM signal, we project $h(t)$ onto the region centered by its mean value and reduce the crest factor of $m(t)$ to around 2.5 dB, the crest factor of the non-filtered 16QAM signal. Again, $r(t)$ is the vectorial difference between $m(t)$ and $h(t)$. This operation is performed for different α_{rc} . The statistics for each signal component are given in Table 2. The crest factor of $r(t)$ is very high, and this increases the power requirements on the residual amplifier, particularly in QPSK modulation.

χ can be plotted using these component statistics for varying coupling factors, and Fig. 2 confirms the above observations. Depending on the constellation and the value of α_{rc} , a high χ is obtained for a coupling factor order of -3 dB or less. This implies that around half the power generated by the main amplifier is used to compensate for the coupler loss. The high crest factor of $r(t)$ imposes a high P_{1dB} on the auxiliary amplifier, in the order of the main amplifier. This technique, accordingly, is not power efficient with

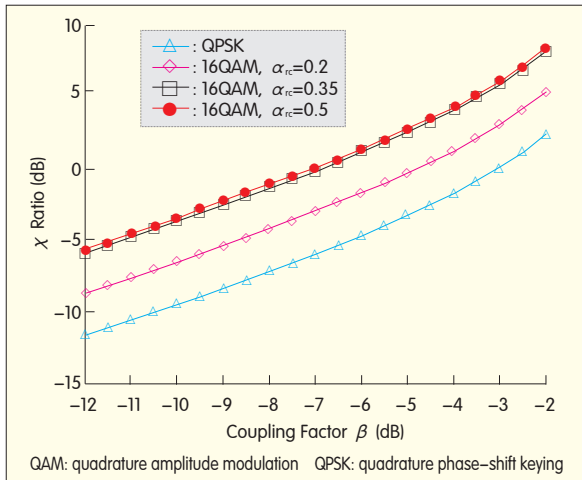


Figure 2. χ vs. coupling factor β (QPSK and 16QAM).

such signals.

3.2 CPM Modulation: The Impact of Filtering CPM Modulations on Spectral Efficiency and BER

CPM modulation is a class of phase modulation that has been widely used in wireless communication systems such as GSM [13], [14]. The expression of the CPM signal is

$$S(t) = \sqrt{\frac{2E}{T}} \cos(2\pi f_c t + \varphi(t, I) + \varphi_0), \quad (9)$$

where f_c is the carrier frequency, φ_0 is the initial phase, E is the symbol energy, T is the symbol duration, and $\varphi(t, I)$ is the time-varying phase of the carrier. The time-varying is defined as

$$\begin{aligned} \varphi(t) &= 2\pi h \sum_{n=-\infty}^n a_n q(t - nT) \\ &= \theta_n + 2\pi h a_n q(t - nT), \end{aligned} \quad (10)$$

where θ_n is the accumulation of all symbols up to time $(n-1)T$, h is the modulation index, and $q(t)$ is the integration of some pulse $g(\tau)$, that is,

$$q(t) = \int_0^t g(\tau) d\tau. \quad (11)$$

The modulation order M , the value of h , and the pulse type directly impact the spectral characteristics of the resulting signal and its detection performance [11]. Small values of h result in CPM signals that occupy a small amount of bandwidth but have poor detection. Compared to rectangular pulses, smoother pulses, such as Gaussian pulses, improve spectral efficiency at the expense of receiver detection. In this paper, we

consider binary CPM where h is 0.5. When $g(\tau)$ is rectangular, the corresponding modulation is called minimum shift keying (MSK). When $g(\tau)$ is Gaussian, the obtained signal corresponds to the Gaussian MSK (GMSK) modulation, which has a more compact spectrum with good error detection.

Both signals have constant amplitude according. This enables the use of highly efficient amplifiers such as class-C or F and makes this type of modulation very attractive. However, the amount of bandwidth occupied is relatively high in terms of fractional out-of-band radiation (Fig. 3). To reduce the amount of bandwidth occupied, we introduce RRC filtering with α_{rc} set to 0.35. For both signals, the side lobes are dramatically attenuated, increasing spectral efficiency (Fig. 3). The resulting change in signal dynamic,

expressed in terms of crest factor, is relatively small:

$\Delta h(t) = 0.62$ for MSK and $\Delta h(t) = 0.77$ for GMSK. The phase transitions in time are too small. Thus, when these signals are filtered, the output signal trajectories stay close to the unity circle, resulting in a small added amplitude modulation. A finite peak-to-peak variation is obtained and does not exceed 2.1 dB.

One issue with the filtered CPM modulations is the impact of RRC filtering on system-level performance. To study this impact, we constructed a complete MATLAB/Simulink model of a wireless link using filtered modulation over an additive white Gaussian noise (AWGN) channel. On the receiver side, we used a matched RRC filter combined with coherent demodulators available in the Simulink library. There is no differential pre-coding, and we use a coherent direct demodulator with maximum likelihood detector. We conducted several simulations of the constructed model to evaluate BER vs. E_b/N_0 of the filtered MSK/GMSK signal over an AWGN channel. The results of these simulations with GMSK modulation are shown in Fig. 4.

There is no significant degradation because a matched RRCF filter is used at the receiver side. ISI is then largely reduced, and the signal is correctly demodulated. Similar results are obtained with the MSK signal. Given the better occupation of spectrum, good BER, and low signal dynamic, the filtered MSK-based modulated signal

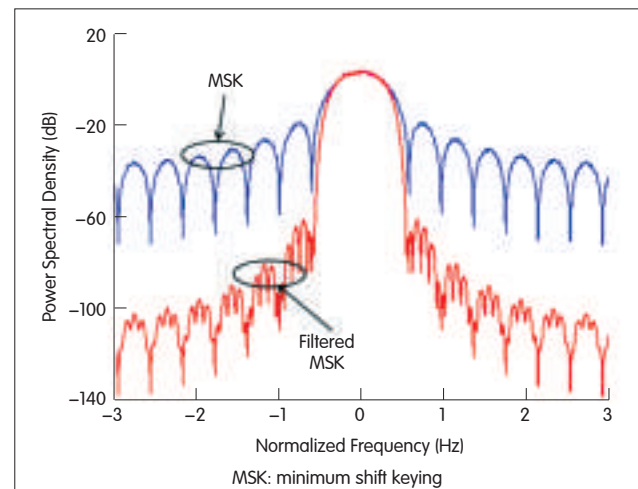


Figure 3. Power spectral density (PSD) of MSK and filtered MSK signals.

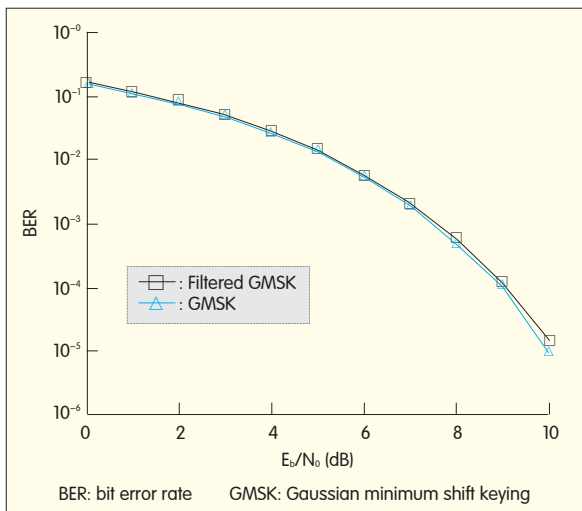


Figure 4.
BER vs. E_b/N_0 for GMSK and filtered GMSK in an AWGN channel.

Table 3. Signal decomposition

	$\Delta h(t)$ (dB)	$\Delta r(t)$ (dB)	γ (dB)	Entire Dynamic (dB)
MSK	0.62	8.7	25.94	1.23
GMSK	0.77	8.4	30.90	2.1
GMSK: Gaussian MSK		MSK: minimum shift keying		

is worth considering.

3.3 Application of the Proposed Amplifier to the Filtered CPM Modulation

To determine the performance of the proposed amplification architecture, we use the MATLAB/Simulink model previously developed and add a baseband model for the RF front-end. We decompose the filtered MSK-based signals. Table 3 shows the dynamics for the residual signal as well as the difference in average power between both branches.

The very high γ can be explained by the level of side lobes for each modulation. As a result, the overall efficiency is expected to be higher in this case. Although the crest factor of the residual signals is considerably high, it is much lower than in QPSK. χ is then plotted against the coupling factor for both considered signals (Fig. 5). For a coupling level of -8 dB, χ is around 15 for GMSK and around 10 dB for MSK. In comparison purpose, for the same coupling level with QPSK signal, χ was -7 . This implies that this architecture is perfectly suitable for this type of signals. In the next section, we study the power efficiency of the

system fed with these signals.

4 Architecture Performance with Filtered CPM Signals

In this section, we consider the amplification of filtered MSK-based signals and its impact on transmitter power efficiency and linearity. We have considered several amplifiers for the residual signal $r(t)$, each corresponding to an amplification class. The main amplifier is assumed to be a 100% efficient class-C or class-F amplifier. For each secondary amplifier,

we compute power efficiency, optimal coupler coefficients, overall efficiency, and the system linearity.

4.1 Single Branch Class-B Amplification

The first intuitive amplification architecture proposed is a single-branch system. Because the input signal has a varying envelope, a class-C amplifier with low conduction angle and high power efficiency causes clipping and is therefore inappropriate. By contrast, a class-B amplifier has a conduction angle of 180° , allowing the entire signal to be amplified. Such an amplifier can be used for the filtered MSK signal. In this paper, we consider a realistic class-B amplifier designed with MRF9060—a 900 MHz Freescale laterally diffused metal oxide semiconductor (LDMOS) transistor. Fig. 6 shows the simulated output power and efficiency of the class-B amplifier versus input power.

This PA has a peak output power of 49 dBm and DC-RF efficiency of 65%. A model of this amplifier is implemented in Simulink for the three MSK filtered signals. For each signal, we compute the average power efficiency, and linearity expressed in terms of adjacent channel power ratio (ACPR) for different back-off levels. The results of these simulations are summarized in Table 4. When the PA is operated at its saturation point, it is highly efficient but has poor linearity. Higher efficiency is obtained for GMSK because the PAR is lower than in MSK. As the input back-off is increased, linearity is

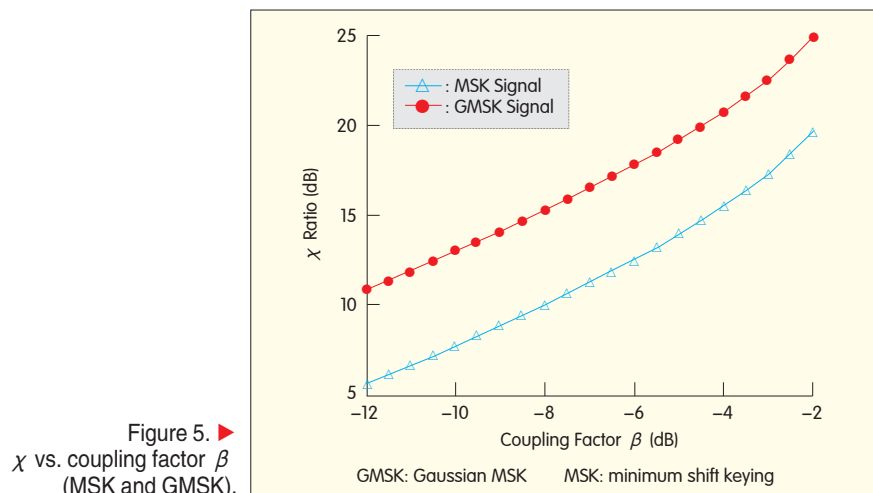
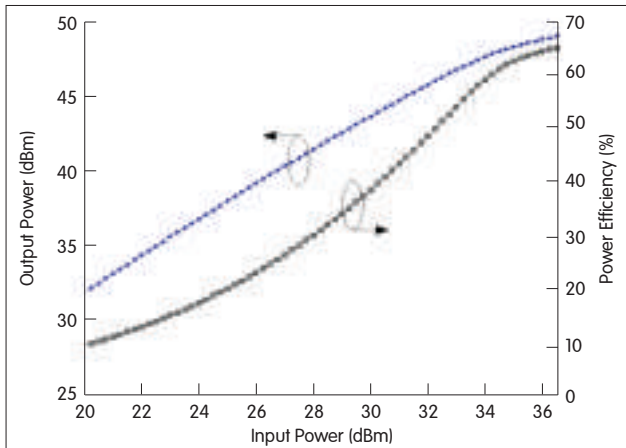


Figure 5. χ vs. coupling factor β (MSK and GMSK).



◀ Figure 6.
Output power and power efficiency vs. input power for a single-ended class-B amplifier using the MRF9060 transistor.

▼ Table 4. Drain efficiency and ACPR for a single-ended class-B amplifier

Back-Off (dB)	Performance	Signal Type	
		GMSK	MSK
0	Efficiency (%)	64.42	64.03
	ACPR (dBc)	-40.9	-34.3
1	Efficiency (%)	62.17	61.48
	ACPR (dBc)	-43.4	-37.9
2	Efficiency (%)	58.34	57.46
	ACPR (dBc)	-50.7	-45.9
3	Efficiency (%)	53.20	52.29
	ACPR (dBc)	-62.7	-50.4

ACPR: adjacent channel power ratio GSMK: Gaussian MSK MSK: minimum shift keying

improved at the cost of power efficiency. Very low ACPR is obtained at a back-off level of 3 dB. These results give a realistic idea of the performance that can be expected from a single-ended PA and will be used as a reference for a subsequent comparison with the proposed two-branch amplification system.

4.2 Amplification of the Residual Signal: Class-AB

Class-AB is an intermediate class between classes A and B. It has better linearity than class-B but lower efficiency. We use a measurements-based model of a 45 W peak Motorola MRF21045 base station amplifier [18]. The amplifier's DC-RF conversion efficiency and power gain are shown in Fig. 7.

Gain varies rapidly around the point of maximum efficiency, and this affects the linearity of the system. Therefore, when assessing the system's

efficiency, linearity also needs to be considered. We use three input power levels (30 dBm, 32 dBm, and 35 dBm) and compute the average efficiency of the class-AB amplifier, assuming that the main amplifier is 100% efficient. We then evaluate the overall efficiency for each signal as a function of the coupling factor β . For a given γ and

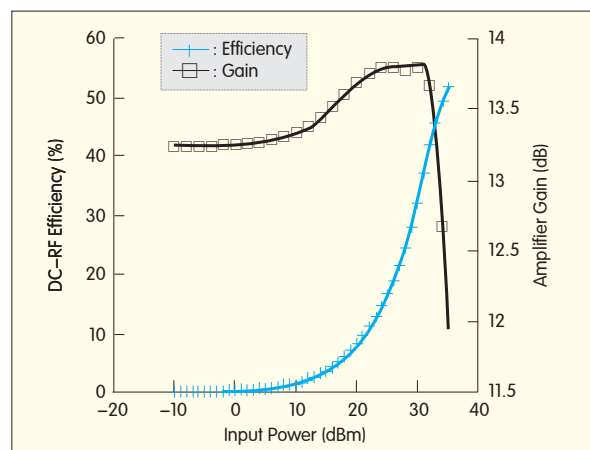
$\eta_{aux}(8)$, optimizing the overall efficiency is equivalent to optimizing the coupler coefficients. For both signals, we calculate the auxiliary amplifier efficiency, the optimal coupler, and the corresponding optimal overall efficiency. To evaluate linearity, we calculate the resulting ACPR. Table 5 summarizes the optimal configuration performance and gives the output signal linearity expressed as ACPR.

For both signals, we have comparable second amplifier efficiency. However, because GSMK has a higher γ , it also has a higher overall efficiency. Efficiency can reach around 87% when the class-AB amplifier is in deep saturation. Overall efficiency of 79.5% can be achieved with the MSK signal. In terms of linearity, this architecture ensures high performance. Distortions are lower than -52 dBc below the carrier level for MSK and -62 dBc below the carrier level for GSMK. Therefore, using a class-AB amplifier substantially increases efficiency without significantly affecting linearity.

4.3 Amplification of the Residual Signal: Class-B

Given the excellent linearity of the class-AB amplifier, further efficiency within an acceptable margin of linearity can be achieved using a more efficient amplification class, such as class-B. Taking the same amplifier used in the single-ended case, we carried out the same simulations as in the previous study. With full power and three back-off levels (1 dB, 2 dB, and 3 dB),

Figure 7. ▶
DC-RF conversion efficiency and power gain of the MRF21045 amplifier.



▼ Table 5. Architecture performance after optimization with a class-AB amplifier in the residual branch

	Signal Type	Class-AB Amplifier Operation Level (dBm)		
		30	32	35
Auxiliary PA Efficiency (%)	GMSK	8.08	10.75	16.0
	MSK	8.42	11.2	16.7
Optimal Coupling Factor (dB)	GMSK	-10.4	-10.9	-11.7
	MSK	-8.3	-8.8	-9.6
Optimal Global Efficiency (%)	GMSK	82.67	84.71	87.2
	MSK	72.7	75.6	79.5
ACPR (dBc)	GMSK	-64.5	-64	-62.1
	MSK	-53.6	-53.4	-52

ACPR: adjacent channel power ratio
GMSK: Gaussian MSK
MSK: minimum shift keying
PA: power amplifier

▼ Table 6. Architecture performance after optimization with a class-B amplifier in the residual branch

	Signal Type	Back-off of the Auxiliary PA (dB)			
		0	1	2	3
Auxiliary PA Efficiency (%)	GMSK	22.73	20.16	17.74	15.52
	MSK	24.09	21.23	18.60	16.22
Optimal Coupling Factor (dB)	GMSK	-12.5	12.2	-12	-11.7
	MSK	-10.4	-10	-9.8	-9.6
Optimal Global Efficiency (%)	GMSK	89.11	88.5	87.8	87
	MSK	82.42	81.44	80.33	79.17
ACPR (dBc)	GMSK	-61.8	-63.1	-63.8	-64.5
	MSK	-51.4	-52.9	-53.0	-54.1

ACPR: adjacent channel power ratio
GMSK: Gaussian MSK
MSK: minimum shift keying
PA: power amplifier

▼ Table 7. ACPR (dBc) vs. phase and gain variation

		ΔG (dB)			
$\Delta \Phi$		0	0.25	-0.25	0.5
0		-52.6	-51.8	-47.4	-51.2
2.5°		-51	-50.7	-46.8	-49.1
-2.5°		-51	-51.7	-46.8	-49.1
5°		-48.1	-47.6	-45.4	-47.7
-5°		-48.1	-47.6	-45.4	-47.7

the power efficiency, optimal coupler, and linearity (expressed as ACPR) are as shown in Table 6. When power efficiency of the class-B amplifier is higher than that of the class-AB, optimal global efficiency is improved. Maximum power efficiency is around 89% for filtered GMSK and 82% for filtered MSK when the class-B amplifier is at full power. An average increase in power efficiency of 2.5% is then

obtained compared to the same architecture with a class-AB amplifier. There is almost no impact on linearity. At worst, ACPR does not exceed -51 dBc, compared to -52 dBc with a class-AB amplifier. These excellent linearity levels can be explained as follows: The nonlinearity of the auxiliary amplifier affects only a small portion of the residual signal, which has a gain variation close to saturation. This

distortion occurs at $-\gamma$ dB ($\gamma > 25$ dB) below the main branch that carries the most important part of the information. Consequently, even when a class-B amplifier (more non-linear than class-AB) is used, linearity is maintained at a high level.

4.4 Considerations for Practical Implementation: Impact of Branch Imbalance on System Linearity

In this paragraph, we discuss the practical implementation issues that arise from non-ideal components in the proposed amplification architecture. In two-branch amplification systems, such as LINC and Feed-forward, non-ideal components cause imbalance between the two branches. This imbalance, in turn, leads to a non-ideal combination of both signal components and results in deteriorated linearity. Imbalances can be caused by transistor aging, temperature, dissimilarity in the up-conversion blocs, and combiner imperfections. To determine the impact of branch imbalance on linearity in the proposed architecture, we use the filtered MSK signal, which is more sensitive to non-linearities. The class-B amplifier without back-off is used in the second branch. We successively introduce phase imbalance, denoted by $\Delta \Phi$ {0°, 2.5°, 5°}, and/or gain imbalance, denoted by ΔG {0 dB, 0.25 dB, 0.5 dB}, and calculate the ACPR for each scenario. A positive $\Delta \Phi$ means that the phase of the main branch is larger than that of the second branch and vice versa. Similarly, positive ΔG means that the gain of the main branch is larger than that of the second branch and vice versa. For negative ΔG , the auxiliary class-B amplifier is used in a deeper compression, and lower overall linearity is expected. The results of these simulations are given in Table 7. An ideal case where $\Delta G = 0$ and $\Delta \Phi = 0$ is the reference. The results show the impact of branch imbalance on system linearity. ACPR is degraded in all cases, and the threshold of -45 dBc is exceeded only when $\Delta G = -0.5$ dB for all $\Delta \Phi$ values (highlighted). The effect of phase variation is the same whether $\Delta \Phi$ is positive or negative, whereas the

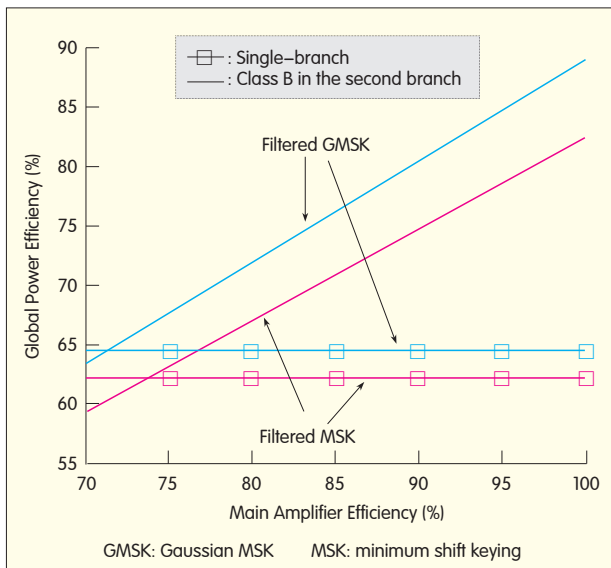


Figure 8. System efficiency vs. main amplifier efficiency.

positivity or negativity of ΔG directly impacts the amount of linearity degradation. This behavior is due to dissymmetry in gain between the branches. For positive ΔG , the signal of the main branch is uniformly amplified by that amount of gain. Uniform distortion is introduced into the system's output signal, leading to low ACPR degradation. However, for negative ΔG , the auxiliary amplifier is excited at its compression region, and an increase in gain in the second branch amplifies both the linearly amplified portion of the residual signal (exciting the PA at its linear region) and the distorted portion of the residual signal (exciting the PA at

its linear region) causing additional distortion. The resulting ACPR for that case is higher. The problem of branch imbalance is well known, and digital predistortion techniques [4], [5] can be used to correct it, as in LINC amplifiers [19], [20].

4.5 Comparison of Architectures

Here, we summarize the results obtained with the single ended class-B amplifier and the new two-branch technique. We previously supposed that the main amplifier was 100% efficient. However, in reality such efficiency cannot be achieved. To determine the efficiency of the proposed architecture with realistic amplifiers in the main branch, we evaluate the sensitivity of the system's global efficiency to the efficiency of the main amplifier. We compute the global efficiency for different efficiencies of the main PA using a class-B auxiliary amplifier. In Fig. 8, we present these results and include the efficiency of the single-branch class-B. Compared to the reference (a single-ended class-B amplifier), the new architecture performs better even when the efficiency of the main amplifier is not 100%. For GMSK, the main amplifier only needs to be about 72% efficient to outperform the single-ended class-B amplifier with excellent linearity. The main amplifier needs to be 77% efficient for the filtered MSK. In recent

literature, drain efficiencies of up to 83% have been reached with class-F amplifiers [16], [17]. With such amplifiers, the proposed architecture outperforms the class-B single-ended amplifier by almost 10% for GMSK signals and 7% for MSK signals. When linearity is taken into account, the proposed architecture becomes even more attractive. Highest efficiency with the single-ended class-B amplifier is reached when the amplifier is operated at full power, but linearity is poor, and ACPR is around -34 dBc for the MSK signal. To meet more stringent ACPR specifications, the PA must operate with sufficient back-off, which lowers power efficiency.

5 Conclusion

This paper presents a new two-branch amplification system combined with appropriate baseband signal decomposition. A main amplifier is fed with a signal that has a crest factor lower than the original filtered signal and an auxiliary amplifier with a low power signal that carries the rest of the information. A signal decomposition technique and RF front-end optimization have been presented. For M-QAM signals, this technique is, at best, similar to a balanced structure. However, with filtered CPM modulations, the newly obtained transmitter offers improved power efficiency of up to 10% compared to a single-ended class-B amplifier using realistic amplifier models. A more compact CPM spectrum is also obtained, with no significant reduction in BER performance. Linearity was also shown to be excellent and largely outperforms the class-B amplifier, and branch imbalance does not significantly impact system linearity. This technique is currently being trialed with filtered CPM signals. A comparison of this technique for M-QAM signals with balanced amplifiers such as LINC has also been given.

Appendix

$$P_{\text{1dB}} \text{ for each amplifier can be written as } P_{\text{1dB}}(A_{\text{main}}) = P_{\text{in,main}} g_{\text{main}} \Delta m(t) \quad (\text{A.1})$$

$$P_{\text{dB}}(A_{\text{res}}) = P_{\text{in res}} g_{\text{res}} \Delta r(t) \quad (\text{A.2})$$

The latter can be expressed using (6)

$$P_{\text{dB}}(A_{\text{res}}) = \frac{P_{\text{in res}} \Delta r(t) g_{\text{main}} \alpha}{\beta} \quad (\text{A.3})$$

Therefore, χ can be expressed simply as

$$\chi = \frac{P_{\text{dB}}(A_{\text{main}})}{P_{\text{dB}}(A_{\text{res}})} = \frac{\gamma \Delta m(t) \beta}{\Delta m(t) \alpha} \quad (\text{A.4})$$

The power efficiency of the system can be written as

$$\eta_{\text{global}} = \frac{P_{\text{out}}}{P_{\text{DC main}} + P_{\text{DC aux}}} \quad (\text{A.5})$$

DC consumption of each amplifier can then be expressed as

$$P_{\text{DC main}} = \frac{P_{\text{out main}}}{\eta_{\text{main}}} = P_{\text{in main}} \frac{g_{\text{main}}}{\eta_{\text{main}}} \quad (\text{A.6})$$

and

$$P_{\text{DC aux}} = \frac{P_{\text{out aux}}}{\eta_{\text{aux}}} = P_{\text{in aux}} \frac{g_{\text{aux}}}{\eta_{\text{aux}}}, \quad (\text{A.7})$$

which can be rewritten as

$$P_{\text{DC aux}} = \frac{P_{\text{in aux}}}{\gamma} \frac{g_{\text{aux}}}{\eta_{\text{aux}}} = \frac{P_{\text{DC main}} \eta_{\text{main}}}{g_{\text{main}}} \frac{g_{\text{aux}}}{\gamma \eta_{\text{aux}}}. \quad (\text{A.8})$$

Since, $\frac{g_{\text{main}}}{g_{\text{aux}}} = \frac{\beta}{\alpha}$ we can write:

$$P_{\text{DC aux}} = \frac{\alpha P_{\text{DC main}} \eta_{\text{main}}}{\gamma \beta \eta_{\text{aux}}}, \quad (\text{A.9})$$

which relates the auxiliary amplifier's DC power consumption to that of the main amplifier. Finally, using (A.9) and (A.6), the total consumed DC power is given by

$$P_{\text{DC total}} = P_{\text{DC main}} \left(1 + \frac{\alpha \eta_{\text{main}}}{\gamma \beta \eta_{\text{aux}}} \right). \quad (\text{A.10})$$

With the DC power known, the total output power of the architecture can be determined. Using (4) and (A.10), we can write

$$P_{\text{out}} = \alpha P_{\text{DC main}} \eta_{\text{main}} + \beta P_{\text{DC aux}} \eta_{\text{aux}}. \quad (\text{A.11})$$

Replacing $P_{\text{DC aux}}$ with its expression in equation (A.10) yields

$$P_{\text{out}} = \alpha P_{\text{DC main}} \eta_{\text{main}} + \frac{\beta \alpha P_{\text{DC main}} \eta_{\text{main}}}{\gamma \beta \eta_{\text{aux}}}, \quad (\text{A.12})$$

which can be put in the following form:

$$P_{\text{out}} = \alpha P_{\text{DC main}} \eta_{\text{main}} \left(1 + \frac{1}{\gamma} \right). \quad (\text{A.13})$$

References

- [1] D. R. Smith, *Digital Transmission Systems*. New-York: Van Nostrand Reinhold, 1993.
- [2] J. G. Proakis, *Digital Communications*. Columbus, OH: McGraw-Hill, 2001.
- [3] J. K. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory

requirements," *IEEE Trans. Veh. Tech.*, vol. 39, no. 4, pp.374–382, Nov. 1990.

- [4] Y. Y. Woo, J. Kim, J. Yi, S. Hong, I. Kim, J. Moon, and B. Kim, "Adaptive digital feedback predistortion technique for linearizing power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 5, pp.932–940, May 2007.
- [5] S. Andreoli, H. G. McClure, P. Banelli, and S. Cacopardi, "Digital linearizer for RF amplifiers," *IEEE Trans. Broadcasting*, vol. 43, no. 1, pp.12–19, Mar. 1997.
- [6] N. Pothecary, *Feedforward Linear Power Amplifier*. Boston: Artech House, 1999.
- [7] S. Cripps, *RF Power Amplifiers for Wireless Communications*. Boston: Artech House, 1999.
- [8] D. C. Cox, "Linear amplification with nonlinear components," *IEEE Trans. Commun.*, vol. 22, no.12, pp. 1942–1945, Dec. 1974.
- [9] H. Chireix, "High power outphasing modulation," *Proc. the IRE*, vol. 23, no. 11, pp. 1370–1392, Nov. 1935.
- [10] M. Helaoui, S. Boumaiza, F. M. Ghannouchi, A. B. Kouki, and A. Ghazel, "A new mode-multiplexing LINC architecture to boost the efficiency of WIMAX up-link transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 2, pp. 248–253, Feb. 2007.
- [11] J. B. Anderson, Tor Aulin, and C.E. Sundberg, *Digital Phase Modulation*. New York: Plenum Press, 1986.
- [12] C.E. Sundberg, "Continuous phase modulation: A class of jointly power and bandwidth efficient digital modulation schemes with constant amplitude," *IEEE Commun. Mag.*, vol. 24, no. 4, Apr. 1986.
- [13] D. M. Balston and R. C. V. Macario, Eds., *Cellular Radio Systems*. London: Artech House, 1993.
- [14] K. Hirade and K. Murota, "GMSK modulation for digital mobile radio telephony," *IEEE Trans. Commun.*, vol. 29, pp. 1044–1050, Jul. 1981.
- [15] F. Ali, A. Gupta, M. Salib, and B. Veasel, "A study of Class C operation of GaAs power HBTs," *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2, pp. 721–724, May 1995.
- [16] D. Schmelzer and S. I. Long, "A GaN HEMT class F amplifier at 2 GHz with > 80% PAE," *IEEE J. Solid-State Circ.*, vol. 42, no. 1, pp. 2130–2136, Oct. 2007.
- [17] M. Helaoui and F. M. Ghannouchi, "Optimizing losses in distributed multiharmonic matching networks applied to the design of an RF GaN power amplifier with higher than 80% power-added efficiency," *IEEE Trans. Microw. Theory Tech.*, vol. 57, no. 2, pp. 314–322, Feb. 2009.
- [18] S. Boumaiza and F. M. Ghannouchi, "Realistic power amplifier characterization with application to baseband digital predistortion for 3G base stations," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 12, pp. 3016–3021, Dec. 2002.
- [19] P. Garcia, J. de Mingo, A. Valdovinos, and A. Ortega, "An adaptive digital method of imbalances cancellation in LINC transmitters," *IEEE Trans. Veh. Tech.*, vol. 54, no. 3, pp. 879–888, May 2005.
- [20] M. Helaoui, S. Boumaiza, A. Ghazel, and F. M. Ghannouchi, "Digital compensation of branches imbalance effects in LINC transmitters," *16th Int. Conf. Microelectronics*, Tunis, Dec. 2004, pp. 688–691.

Biographies

Walid Hamdane (walid.hamdane.1@ens.etsmtl.ca) received his B.Eng. degree in electrical engineering from École Polytechnique de Tunisie and his Ph.D. degree in electrical engineering from École de Technologie Supérieure, Montréal. His research interests include power efficiency enhancement for wireless transmitters, advanced transmitter design for modern radio applications, and frequency synthesizers for radar applications design. From 2008 to 2009, he was a power amplifier designer with Mitec Telecom Inc. Currently, he is an RF hardware engineer at Nanowave Technologies Inc and works with the signal source group. He has 10 publications and 2 patents (pending). Dr. Hamdane was awarded postgraduate scholarships (2004–2008) from the Tunisian Ministry of Higher Education and has been awarded several distinctions.

Ammar B. Kouki (ammkar.kouki@etsmtl.ca) received his B.S. (with honors) and M.S. degrees in engineering science from Pennsylvania State University in 1985 and 1987. He received his Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 1991. Between 1988 and 1991, he was a consultant at the National Center for Supercomputing Applications (NCSA). From 1991 to 1993, he was a postdoctoral fellow at the Microwave Research Laboratory, École Polytechnique de Montréal. From 1994 to 1998, he was a senior microwave engineer at the same laboratory and was involved with power amplifier linearization techniques. In 1998, he co-founded AmpliX Inc., a company that specialized in RF linearizers for wireless and SatCom applications. In 1998, he joined the faculty of the École de Technologie Supérieure, Montréal, where he is currently full professor of electrical engineering and director of the LACIME laboratory. He was also one of the co-founders of ISR Technologies, a software defined radio company.

Professor Kouki has written more than 160 peer-reviewed papers and holds six patents (three are currently under review). He has diverse research interests that cover radio communication and navigation with a focus on devices, intelligent and efficient RF front ends/transceiver architectures, and antenna and propagation. He is currently researching active device modeling and characterization, power-amplifier design, linearization, and efficiency enhancement techniques. He is also involved in computational electromagnetic techniques for modeling and designing passive microwave structures and is active in multiple antenna systems and intelligent antenna research.

François Gagnon (francois.gagnon@etsmtl.ca) received his B.Eng. and Ph.D. degrees in electrical engineering from École Polytechnique de Montréal. Since 1991, he has been a professor in the Department of Electrical Engineering, École de Technologie Supérieure, Montréal. He chaired the department from 1999 to 2001 and is now the holder of the NSERC Ultra Electronics Chair, Wireless Emergency and Tactical Communication, at the same university. His research interests include wireless high-speed communications, modulation, coding, high-speed DSP implementations, and military point-to-point communications. He has been heavily involved in creating the new generation of high-capacity line-of-sight military radios offered by the Canadian Marconi Corporation (now Ultra Electronics Tactical Communication Systems). The company received, a Coin of Excellence from the U.S. Army for the performance and reliability of the radio product. Professor Gagnon is a recognized leader in research management. With an annual budget of \$1.6M, he supervises 18 graduate students, leads a team of seven research professionals, and participates in activities with companies such as Ultra, ISR technology, Sita, Ericsson, Lipso, Nortel, Bell, Octasic Semiconductors, Sierra Wireless, Boomerang, and IREQ. Professor Gagnon was awarded the 2008 NSERC Synergy Award (small and medium-sized companies category) for his fruitful and long-lasting collaboration with Ultra Electronics TCS.

FPGA Implementation of a Power Amplifier Linearizer for an ETSI-SDR OFDM Transmitter

Suranjana Julius and Anh Dinh

(Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, S7N 5A2, Canada)

Abstract: Most satellite digital radio (SDR) systems use orthogonal frequency-division multiplexing (OFDM) transmission, which means that variable envelope signals are distorted by the RF power amplifier (PA). It is customary to back off the input power to the PA to avoid the PA nonlinear region of operation. In this way, linearity can be achieved at the cost of power efficiency. Another attractive option is to use a linearizer, which compensates for the nonlinear effects of the PA. In this paper, an OFDM transmitter conforming to European Telecommunications Standard Institute SDR Technical Specifications 2007–2008 was designed and implemented on a low-cost field-programmable gate array (FPGA) platform. A weakly nonlinear PA, operating in the L-band SDR frequency, was used for signal transmission. An adaptive linearizer was designed and implemented on the same FPGA device using digital predistortion to correct the undesired effects of the PA on the transmitted signal. Test results show that spectral distortion can be suppressed between 6–9 dB using the designed linearizer when the PA is driven close to its saturation region.

Keywords: power amplifier linearization; digital predistortion; ETSI-SDR; OFDM; FPGA

1 ETSI Standard and Power Amplifier Characteristics

In 1992, the US Federal Communications Commission allocated a spectrum in the S-band (2.3GHz) for nationwide broadcasting of satellite-based digital audio radio. Digital radio broadcast via satellite provides a means of delivering high quality audio channels and associated services to fixed and mobile receivers [1]. New standards for satellite radio, such as the European Telecommunications Standard Institute Satellite Digital Radio (ETSI-SDR), are still being developed and deployed [2]–[4]. Authors in [5]–[7] have reported some work on FPGA implementation based on digital audio and video broadcasting standards, but not much has been done for the ETSI-SDR standards. Although

ETSI-SDR standards are fairly new, they are becoming popular.

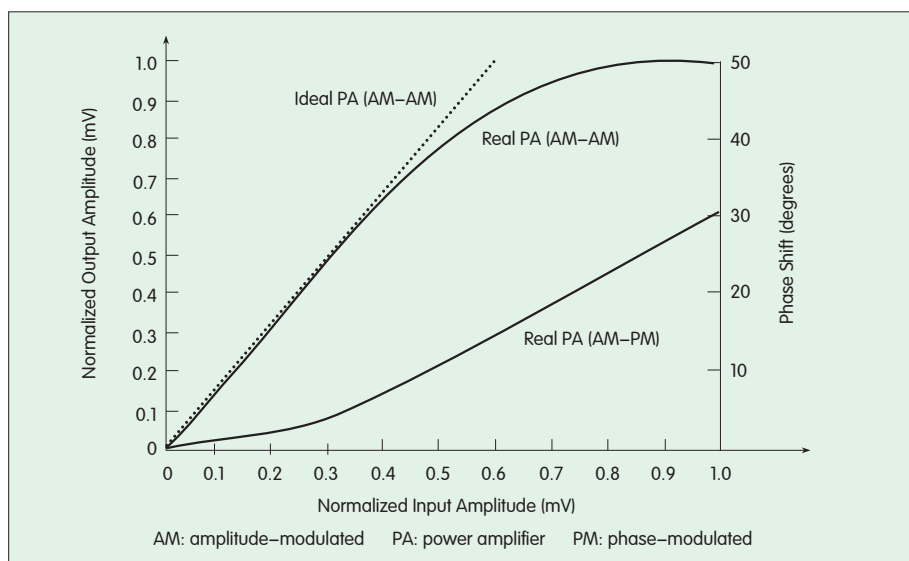
One of the challenges of implementing an ETSI-SDR system is the nonlinear characteristics of the power amplifier (PA). A power-efficient RF amplifier should use a small amount of DC supply power; but unfortunately, there is a trade-off between efficiency and linearity [8]. PAs are most efficient when they are driven close to their saturation, which is also the most nonlinear region of operation. The quadrature amplitude modulation (QAM) or quadrature phase-shift keying (QPSK) signals from the ETSI-SDR are amplified by a nonlinear PA. This not only degrades system performance but also creates out-of-band power leakage that interferes with the systems operating in adjacent channels [9], [10]. The challenges imposed by nonlinearity in the PA are very complex. A back-off of

6–9 dB in input power is common when operating a PA and results in decreased efficiency and increased OPEX. Extending the linear region of operation of a weakly linear PA and decreasing the input power back-off reduces CAPEX and OPEX in ETSI-SDR systems.

An amplifier is linear if its gain is constant throughout the range of the input signal. If this gain is not linear, the output signal is distorted by clipping. The DC bias point is the most important factor in determining the relationship between PA nonlinearity and efficiency. In a real amplifier, the gain $A(s(t))$ and phase-shift $\Phi(s(t))$ are functions of the input signal $s(t)$. The complex transfer function is dependent on the amplifier input power and is given by

$$G(s(t)) = A(st)e^{j\Phi(s(t))}. \quad (1)$$

As a result, the gain decreases and phase-shift changes as the level of the



▲ Figure 1. Power amplifier AM-AM/AM-PM characteristics [11].

input signal drives the amplifier into its saturation region. The PA output amplitude and phase characteristics are known as AM-AM characteristics and AM-PM characteristics, respectively.

Fig. 1 shows typical characteristics of a PA [11]. It is possible to obtain the amplitude-modulated (AM)-AM and AM-phase-modulated (PM) characteristics as a function of complex input signal sample [12], [13]. For an n th complex input sample z_n with a magnitude $|z_n|$ and a phase $\arg(z_n)$, the complex transfer function of the nonlinear PA is given by

$$A(z_n) = AM(|z_n|^2) e^{j(\arg(z_n) + PM(|z_n|^2))}, \quad (2)$$

where $AM(|z_n|^2)$ and $PM(|z_n|^2)$ are the polynomial functions derived from the AM-AM and AM-PM characteristics, and $|z_n|^2$ is the power of the n th complex input sample. For a nonlinear PA without memory, using AM-AM and AM-PM characteristics is sufficient to model the PA's behavior. The complex transfer function takes the form

$$A(z_n) = \sum_{k=0}^{N-1} a_k z_n |z_n|^{2k} \quad (3)$$

where N is the number of polynomial terms, and a_k is the k th polynomial coefficient. The complex transfer function is modeled as a polynomial function of current and previous complex input signal samples. Detailed discussion of this type of PA modeling

with references can be found in [13].

Nonlinearity in a PA is measured by factors such as 1 dB compression point, third order intercept point (IP3), and adjacent channel power ratio and harmonics. One of the undesired products of PA non-linearity is inter-modulation distortion (IMD). In an ETSI-SDR system using QAM modulation, IMD manifests itself in the form of spectral regrowth. This causes in-band distortion and leakage in adjacent channels as a result of spectral spreading. Spectral regrowth degrades the quality of signal and causes adjacent channel interference.

Techniques are required to mitigate the undesired effects of the PA in an ETSI-SDR system. There is a trade-off between efficiency and linearity when using RF PAs. At the base station, linearity in the PA is more important than efficiency [14]. A linearizer is expected to maintain a constant gain at the PA output before the PA reaches saturation. Using linearization, the input signal back-off can be decreased, and a higher level of input signal can be used to drive the PA without degrading output signals. The linearizer type chosen depends on efficiency, complexity, modulation scheme, bandwidth, adjacent channel interference, and dynamic range. Various schemes are classified according to their functionality,

architecture, and application. Details of these schemes and their techniques are described in [8], [10], [14]–[16]. A common linearization technique is predistortion. This technique provides an inexpensive solution in which a nonlinear circuit is inserted between the input signal and the PA. The nonlinear circuit generates IMD products inverse to that produced by the PA and thereby cancels the effect of the PA nonlinearity. This can be viewed as the predistorter having characteristics inverse to the real PA AM-AM curve in Fig. 1. There has been a reasonable amount of research carried out in this area [17]–[21].

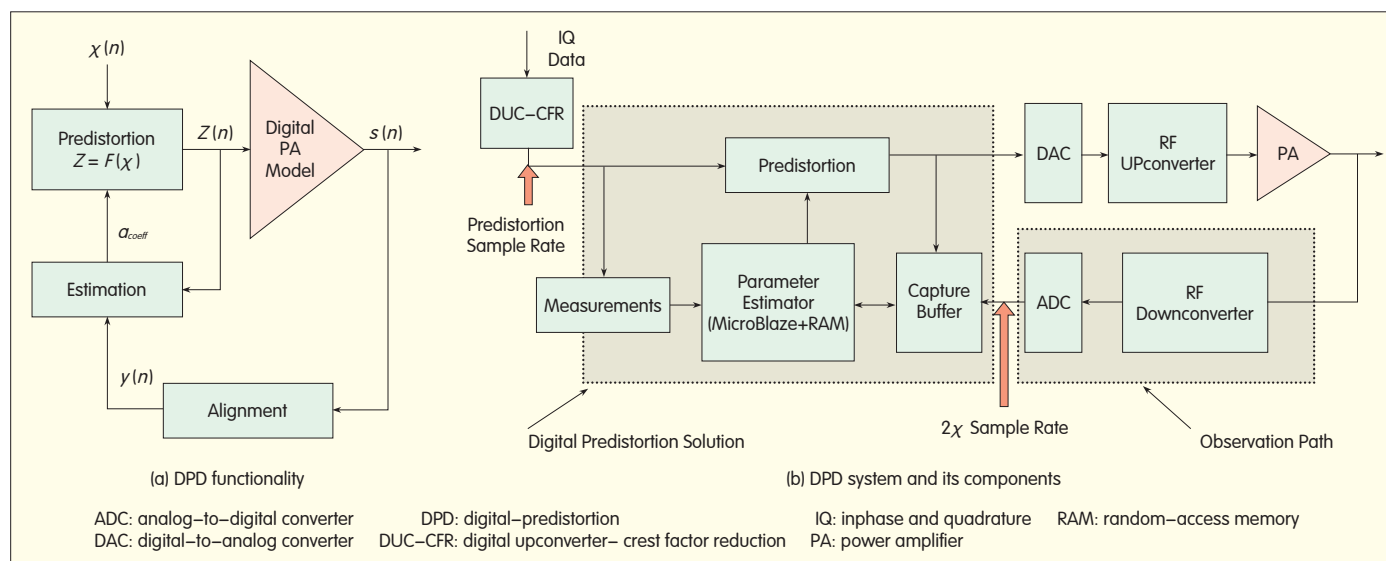
2 Design of a Linearizer for the ETSI Radio Interface Standard

The linearizer design uses a look-up-table (LUT)-based predistortion algorithm. The envelope of the complex input sample is used for LUT-indexing targeting the ETSI radio interface standard. ETSI TS 102 551-2 V2.1.1 (2007-08) [4] uses OFDM for IPL-MC transmission. The mode of interest is Mode-3, with OFDM at 1 k (1024 FFT length) for 1.7 MHz channel spacing. An L-band PA, suitable for use in the transmitter of ETSI-SDR base stations, was chosen (CRF24060, Cree Wireless Devices). The PA nonlinear characteristics causing signal distortion in the OFDM transmission based on ETSI-SDR standard was studied, and nonlinear gain characteristic equations were developed. Using measured data of the PA, the polynomial functions of the $AM(|z_n|^2)$ and $PM(|z_n|^2)$ were obtained:

$$AM(|z_n|^2) = a_3(|z_n|)^3 + a_2(|z_n|)^2 + a_1(|z_n|) + a_0 \quad (4)$$

$$PM(|z_n|^2) = b_4(|z_n|)^4 + b_3(|z_n|)^3 + b_2(|z_n|)^2 + b_1(|z_n|) + b_0 \quad (5)$$

This digital-predistortion (DPD) linearizer is designed to correct 3rd and 5th order IMD products by adaptively developing the inverse characteristics of the PA. This cancels the effects of the PA nonlinearity. The relationship between the input, V_{in} , the PA output



▲ Figure 2. DPD functionality, and DPD system and its components [22].

without DPD, $V_{\text{no-DPD}}$, the DPD linearizer output, $V_{\text{linearizer}}$, and the overall system output with correction, $V_{\text{with-DPD}}$, is given by

$$V_{\text{no-DPD}} = f(nl)kV_{\text{in}} \quad (6)$$

$$V_{\text{linearizer}} = \frac{1}{f(nl)} \quad (7)$$

$$V_{\text{with-DPD}} = V_{\text{no-DPD}} V_{\text{linearizer}} = f(nI)kV_{\text{in}} \frac{1}{f(nI)} = kV_{\text{in}} \quad (8)$$

where k is the linear gain of the PA, and $f(nI)$ is the PA nonlinearity. As a result, the output of the overall system is, theoretically, linear.

Fig. 2(a) shows the function of the DPD. The predistorter models the inverse characteristics of the PA as a function of the magnitude of the input $\chi(n)$. The output of the predistorter $z(n)$ is amplified by the PA generating output $s(n)$. This output is aligned for amplitude, phase, and delay and is used by the estimator. The estimator uses $z(n)$ and the aligned output $y(n)$ to develop the inverse coefficients of the PA given by a_{coeff} in (4). In this way, $s(n)$ is as close as possible to $\chi(n)$. Modifications were made to the VHSIC hardware description language (VHDL)–based DPD linearizer components so that they could be used with the OFDM transmitter generated using MATLAB, SIMULINK, and digital signal processing (DSP) design tools in a Xilinx System Generator environment.

Design parameters were changed to suit the requirements of the ETSI standard.

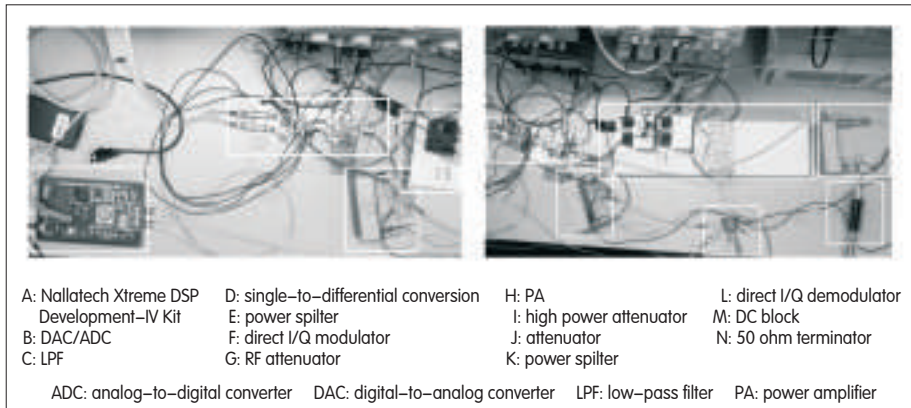
The DPD consists of four components at the highest level: predistorter, capture buffers, measurements system, and parameter estimator, as shown in Fig. 2(b). The predistorter stores the predistortion coefficients for a given signal. These coefficients are obtained by processing a fixed number of samples at the PA input/output signals. The predistortion coefficients are in the form of polynomials of signal magnitude. To save hardware resources, these polynomials are implemented using LUTs addressed by magnitude of the complex sample. The predistorter module receives the 16-bit input I/Q data interpolated by 16 from the OFDM transmitter. The capture buffer collects signal samples from the transmission and observation (feedback) paths, and these samples are used by the estimation function to compute the predistortion coefficients. Data is captured and recaptured according to the standards set by the measurement system.

A simulation test system was developed comprising a simulation model of the ETSI–SDR OFDM transmitter and a simulation model of the linearizer system. The simulation model of the L–band PA was

experimentally obtained from the L-band PA gain and phase-shift characteristics. A commercially available state-of-the-art reference design from Xilinx [22], [23] was used as the basis for this DPD design. Simulations were performed in an integrated environment of MATLAB, SIMULINK, and Xilinx DSP development tools. Spectral regrowth was suppressed by 6–9 dB during testing of the PA in OFDM transmission after linearizer correction. The linearizer system design is suitable for a data rate of up to 92.16 MSPS with a signal bandwidth of 20 MHz. The designed ETSI-SDR OFDM transmitter requires a signal bandwidth of 1.5314 MHz with a data rate of 2.1511 MSPS. The design was implemented in an FPGA to verify its functionality.

3 Implementation, Results, and Analysis

A photograph and block diagram (with signal flow path) of the hardware setup are shown in Figs. 3 and 4. The testing hardware was built using commercially available components and evaluation boards. The designed ETSI–SDR OFDM transmitter and the DPD linearizer were implemented using the Virtex–4 FPGA XC4VSX35 speed grade –10, available on the Nallatech XtremeDSP Development Kit–IV



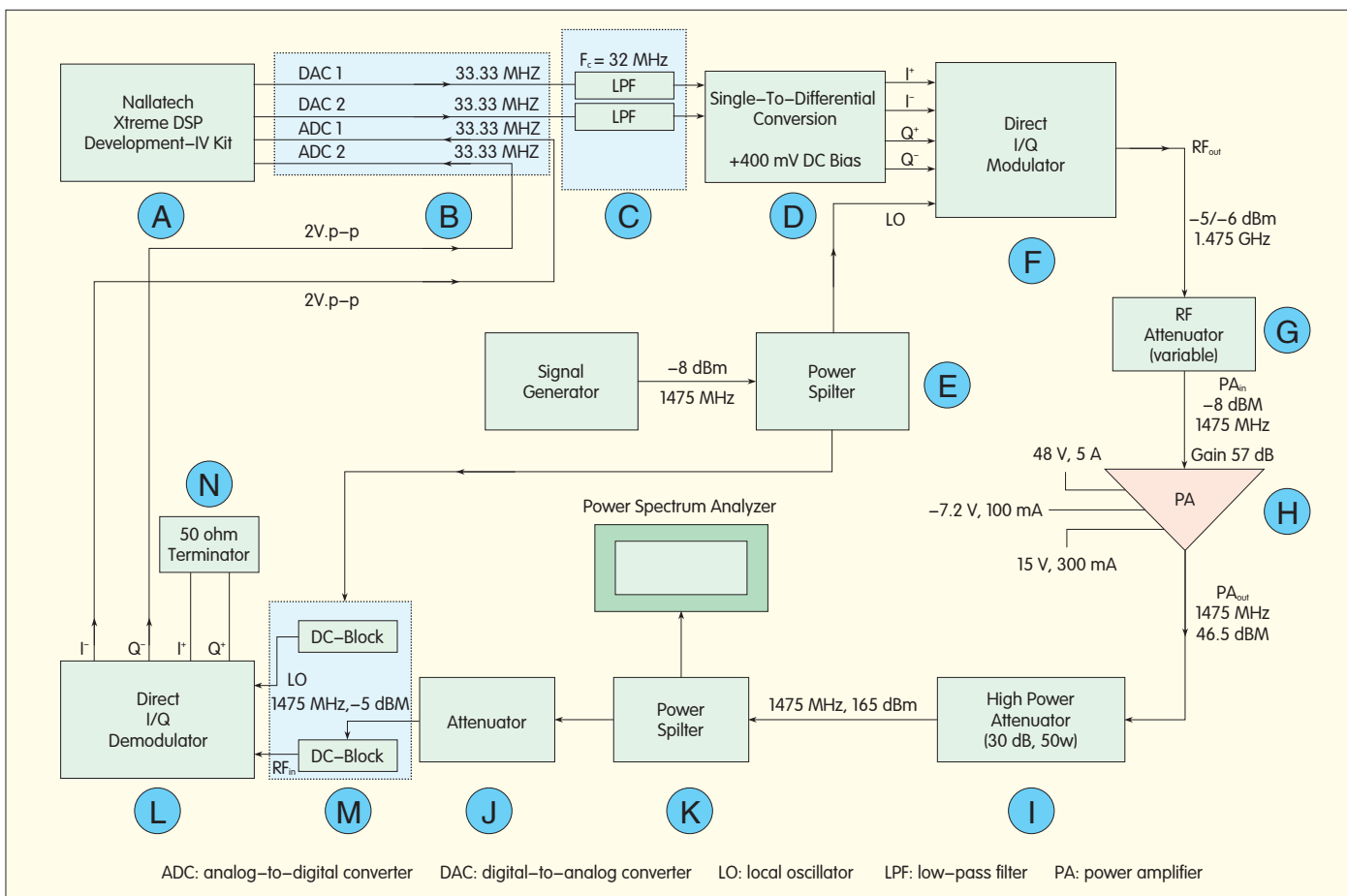
▲ Figure 3. Hardware transmission path and feedback path set-up.

evaluation board. Design synthesis, implementation, and generation of bit files suitable for loading onto the FPGAs were performed in the Project Navigator environment of Xilinx Integrated Software Environment (ISE) Foundation 10.1.3 software. The generated bit files were loaded onto the FPGAs using

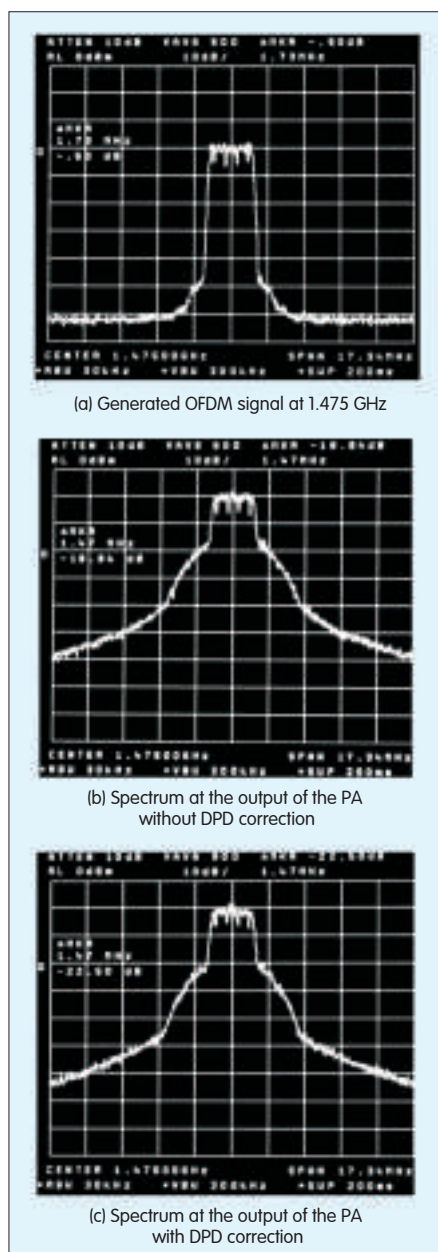
Nallatech Filesystem in Userspace (FUSE). The on-board clocks and resets were also set using FUSE. The main design clock was generated at 33.33 MHz using the onboard programmable oscillator. This clock is used to generate the clocks for the on-chip A/D and D/A converters at

33.33 MHz. A feedback clock signal was brought back to the Virtex-4 FPGA and used as an input to the digital clock manager (DCM). The feedback signal synchronizes the clocks to the digital-to-analog converter (DAC) and analog-to-digital converter (ADC) as well as the rest of the design logic. The DCM is also used to derive clocks that are used in the OFDM transmitter and DPD linearizer. The reset of the signals of the DCM was provided by the FPGA reset.

The logic portion of the design consists of four main modules. The OFDM transmitter module was designed in the Xilinx System Generator for DSP 10.1.3 environment. 16-bit I/Q OFDM transmitter data were applied as input to the second module that up-samples the data to the predistortion sampling rate. The main 33.33 MHz clock is used for this second module. The linearizer system consists



▲ Figure 4. Block diagram of the hardware set-up.



▲ Figure 5. Generated OFDM signal at 1.475 GHz, and spectrum at the output of the PA without/with DPD correction.

of the remaining two modules. One module includes the DPD and measurement system using the 100 MHz DCM output clock. This portion of the design, running at three times the data rate, allows the linearizer to simultaneously process three consecutive data samples. The final module consists of an embedded processor (Microblaze) running at the same 33.33 MHz clock as the

up-sampled data. These DPD linearizer modules are meant for use with pure-VHDL approach in an ISE project environment. The 16-bit I/Q outputs of the DPD linearizer are the predistorted data that have the same sample rate as the on-chip DAC at 33.33 MHz. The DPD system also accepts the received baseband I/Q data from the PA output via the feedback path. These data were received from the output of the 14-bit ADC at 33.33 MHz. The 16-bit I/Q transmitted data from the DPD linearizer are rounded to 14 bits to match the maximum number of bits allowed by the DACs. Conversely, the 14-bit data received from the ADCs are zero-stuffed and used by the 16-bit input ports of the DPD linearizer. The digitally predistorted I/Q data output from the Virtex-4 FPGA are applied as input to the two separate DACs for transmission.

The ETSI-SDR OFDM transmitter was designed and implemented in Virtex-4. This hardware implementation was a fixed-point equivalent of the MATLAB floating point design. The output signals come from the OFDM transmitter at 2.1511 MSPS and have been up-sampled by a factor of 16 to the predistortion sampling rate of 34.4176 MSPS. The channel bandwidth is 1.536 MHz with channel spacing of 1.712 MHz. There is a loss of precision that derives from the floating-to-fixed-point conversion in the design, and round-off noise is introduced into the system during FPGA implementation. The OFDM signal was re-generated at 33.33 MSPS instead of 34.4176 MSPS to match the system clock rate of 33.33 MHz. Therefore, channel bandwidth and spacing were altered by the same ratio of 1:0.9684 (1.487 MHz and 1.657 MHz for bandwidth and spacing, respectively). Altering these parameters to facilitate hardware implementation has negligible effects on the functional characteristics of the system. Analog signal processing, such as mixing, conversion, and attenuation, are self-explained in the block diagram.

Fig. 5(a) is a capture of the OFDM signal spectrum on the Agilent 8564EC

PSA. The power spectrum at the output of the test PA is shown in Fig. 5(b). The PA amplified the up-converted signal from the OFDM transmission at L-band frequency of 1.475 GHz. No DPD compensation was performed at this stage. The data rate of the signal was 33.33 MSPS for comparison with the output signal after DPD compensation. The total power of the input signal was maintained at around -8 dBm to -9 dBm, and care was taken not to drive the PA into saturation. Spectral regrowth at the shoulder was measured at 1.47 MHz from the carrier frequency for comparison with simulation results. Reduction in spectral regrowth at that fixed frequency from the carrier should be sufficient to evaluate correction performed by the DPD linearizer. Fig. 5(b) shows spectral regrowth of about -18 dB relative to the carrier. Figure 5(c) shows the output spectrum of the PA after the DPD linearizer has performed correction. The number of spectral averages, the sweep time, and other parameters required by the PSA for measurement remained unchanged to allow direct comparison. It is clearly shown that the spectral regrowth is reduced to -22.5 dB. Therefore, spectral regrowth was reduced, and the performance of the L-band test PA was improved after compensation by the DPD linearizer.

The main logic of the design was developed and implemented in the Xilinx Virtex-4 XC4VSX35 FPGA device. All design timings were met at 276.48 MHz. Table 1 lists resource use in the design. The design only uses 35% of the available resources on the Virtex-4. This allows future signal processing exploration or the use of smaller devices to reduce the cost of implementation. The ETSI-SDR OFDM transmitter uses only 13% of the available resources, which shows that it is possible to implement radio standards based on software-defined-radio on small FPGA devices. After being interpolated to the predistortion rate, the OFDM transmitter consumes 24% of the total logic resources. A major portion of this use can be attributed to the cascaded interpolating filters that use 23% of the

▼ Table 1. Resource use by logic on the Xilinx Virtex-4 XC4VSX35 FPGA

Resources	Available	OFDM Transmitter	Transmitter with Upsampling	DPD	Total
Slices	15,360	3466 (22%)	6944 (45%)	2826 (18%)	9770 (63%)
FFs	30,720	4702 (15%)	6711 (22%)	3132 (10%)	9843 (32%)
Look-Up-Table	30,720	4844 (16%)	7324 (24%)	3391 (11%)	10,715 (35%)
Block RAMs	192	9 (5%)	9 (5%)	60 (31%)	69 (36%)
DSP logic slices 48s	192	16 (8%)	44 (23%)	19 (10%)	63 (33%)
Digital Clock Manager	8	—	—	1	1 (12%)
DPD: digital–predistortion		FF: Flip Flop		RAM: random–access memory	
DSP: Digital Signal Processing		OFDM: orthogonal frequency–division multiplexing			

available DSP48 logic slices. The DPD linearizer uses 16% of the available resources. The Xilinx power analyzer reported a total dynamic power usage of 1.193 W in which the OFDM transmitter with the up-sampler consumes 204 mW, and the DPD linearizer uses 989 mW.

4 Conclusions

An OFDM transmission system model suitable for L-band satellite radio was designed according to the ETSI-SDR standard specifications and implemented in FPGA. A linearizer system based on state-of-the-art reference design from Xilinx was also designed to improve non-linearity in the PA. A simulation model of the test system was developed in an integrated environment of MATLAB, SIMULINK, and Xilinx System Generator. A simulation equivalent of the PA was modeled as a memoryless polynomial from the experimentally obtained AM-AM and AM-PM characteristics. Baseband filtering, and RF up-conversion and down-conversion (to fit the PA requirements) were performed in analog. The PA outputs with and without DPD compensation were measured and compared. Spectral regrowth was suppressed by 9 dB at 1.5 MHz from the carrier. The OFDM transmitting system and DPD linearizer module were contained in the Xilinx Virtex-4 FPGA and met all timing constraints. It can therefore be concluded that DPD linearizer performance is compromised by the limited capabilities of the built-in ADC and DAC on the FPGA evaluation kit, the combined noise of components in

the analog chain, and signal attenuation in the RF feedback path.

References

- [1] K. Raghunandan, "Satellite digital radio brings new image to broadcasting," in *Proc. IEEE 2nd Int. Conf. on Recent Advances in Space Tech.*, Istanbul, June 2005, pp. 799–804.
- [2] B. Cuesta, I. Moreno, A. Yun, and J. Rodriguez, "Analysis of the convergence between DVB-SH and ETSI-SDR," in *Proc. IEEE 1st Int. Conf. Advances in Satellite and Space Commun.*, Colmar, France, July 2009, pp. 13–18.
- [3] http://www.en.wikipedia.org/wiki/Satellite_radio, "Satellite Radio," accessed March 2010.
- [4] Satellite Earth Stations and Systems; Satellite Digital Radio (SDR) Systems; Inner Physical layer of the Radio Interface; Part 2: Multiple Carrier Transmission, ETSI TS 102 551-2 V2.1.1 (2007-2008), August 2007.
- [5] A. Sghaier, S. Areibi, and R. Dony, "IEEE802.16-2004 OFDM functions implementation on FPGAs with design exploration," in *Proc. FPL 2008 Int. Conf. on Field Programmable Logic and Applic.*, Heidelberg, Germany, Sept. 2008, pp. 519–522.
- [6] A. Sghaier, S. Areibi, and B. Dony, "A pipelined implementation of OFDM transmission on reconfigurable platforms," in *Proc. CCECE 2008, Canadian Conf. Elect. Comput. Eng.*, May 2008, pp. 801–804.
- [7] J. Garcia and R. Cumplido, "On the design of an FPGA-based OFDM modulator for IEEE 802.16-2004," presented at *Int. Conf. Reconfigurable Comput. and FPGAs (RECONFIGURE'05)*, Puebla, Mexico, Sept. 2005.
- [8] M. Hella and M. Ismail, *RF CMOS Power Amplifiers: Theory, Design, Implementation*. Norwell, MA: Kluwer Academic Publishers, 2002.
- [9] D. M. Klymyshyn, "A power amplifier linearizer using adaptive polynomial predistortion," M.S. thesis, Dept. Elect. Eng., Univ. of Saskatchewan, 1995.
- [10] M. Ghadheri, "Adaptive linearization of efficient high power amplifiers using polynomial predistortion with global optimization," Ph.D. thesis, Dept. Elect. Eng., University of Saskatchewan, 1995.
- [11] Hsing-Hung Chen, Ching-Shyang Maa, Yeong-Cheng Wang, and Jiunn-Tsair Chen, "Joint polynomial and look-up-table power amplifier linearization scheme," in *57th IEEE Semiannual Vehicular Tech. Conf. (VTC 2003)*, Jeju, Korea, April, 2003, vol. 2, pp. 1345–1349.
- [12] R. Marsalek, P. Jardin, and G. Baudion, "From post-distortion to pre-distortion for power amplifier linearization," in *IEEE Commun. Lett.*, vol. 7, no. 7 pp. 308–310, Jul. 2003.
- [13] R. Marsalek, P. Jardin, and G. Baudion, "Power amplifier linearization using pre-distortion with memory," in *Proc. 13th Int. Czech-Slovak Scientific Conf. RADIOELEKTRONIKA*, Brno, Czech Republic, 2003, pp. 193–196.
- [14] S. Cripps, *RF Power Amplifiers for Wireless Communications*, Boston, MA: Artech House Inc., Academic Publishers, 1999.
- [15] S. Bruss, "Linearization Methods," Class Lecture, Dep. Elect. Comput. Eng., University of California, Davis, April 2003.
- [16] J. P. Moffatt and J. P. MacEachern, "An adaptive QAM linearizer using postdistortion," in *Proc. IEEE Global Telecom Conf.*, Dallas, TX, Nov 1989, pp. 406–412.
- [17] T. M. Nguyen, J. Yoh, C. H. Lee, H. T. Tran, and D. M. Johnson, "Modeling of HPA and HPA linearization through a predistorter: Global broadcasting service applications," in *IEEE Trans. Broadcasting*, vol. 49, no. 2, pp. 132–141, 2003.
- [18] N. Safari, J. P. Tanem, and T. Roste, "Block based predistortion for amplifier linearization in burst type mobile satellite communications," in *Proc. IEEE 35th European Microw. Conf., Paris*, vol. 3, 2005, pp. 1783–1786.
- [19] N. Ceylan, J. E. Mueller, T. Pittorino, and R. Weigel, "Mobile phone power amplifier linearity and efficiency enhancement using digital predistortion," in *Proc. IEEE 33rd European Microw. Conf.*, Munich, vol. 1, October 2003, pp. 269–272.
- [20] A. Benchahed, A. Ghazel, M. Mabrouk, C. Rebai, and M. F. Ghannouchi, "RF digital predistorter for power amplifiers of 3G base stations," in *Proc. IEEE 13th Int. Conf. Elect., Circ., Syst.*, Nice, December 2006, pp. 999–1002.
- [21] M. Franco, A. Guida, A. Katz, and P. Herczfeld, "Reduction of in-band intermodulation distortion products in radio frequency power amplifiers with digital predistortion linearization," in *Proc. IEEE Int. Microw. Symp. Dig.*, pp. 918–921, June 2006.
- [22] S. Summerfield, H. Parekh, and V. Barnes, "Digital predistortion reference design V2.0, EA2 Release Note", Xilinx Inc. [Online]. Available: http://www.xilinx.com/publications/prod_mktg/pn2061.pdf
- [23] S. Summerfield, H. Parekh, and V. Barnes, "Digital Predistortion Reference Design, Application Note XAPP1044 (v1.0.1)," Xilinx Inc., April, 2008.

Biographies

Suranjana Julius (suj967@mail.usask.ca) received her BSc degree in electronics and communication engineering from Bangalore University in 2002. Between 2002 and 2007, she worked in India and South Africa in various R&D positions and also worked as an applications engineer. Suranjana received her M.Sc. degree from the University of Saskatchewan, Canada, in 2010. Her research interests include digital signal processing, software defined radio, and digital logic design for FPGAs.

Anh Dinh (anh.dinh@usask.ca) received his B.Sc. degree in electrical and computer engineering from Lakehead University, Canada, in 1992. He received his M.Sc. and Ph.D. degrees from the University of Regina, Canada, in 1997 and 2000. He currently works at the Department of Electrical and Computer Engineering at the University of Saskatchewan. His research interests include biosensors and very large scale integration (VLSI) for wireless communications.

Design Technologies for Silicon-Based High-Efficiency RF Power Amplifiers: A Brief Overview

Ruili Wu, Jerry Lopez, Yan Li and Donald Y.C. Lie

(Department of Electrical and Computer Engineering, Texas Tech University, Lubbock, TX 79401, the U. S.)

Abstract: This paper presents a brief overview of several promising design technologies for high efficiency silicon-based radio frequency (RF) power amplifiers (PAs) as well as the use of these technologies in mobile broadband wireless communications. Four important aspects of PA design are addressed in this paper. First, we look at class-E PA design equations and provide an example of a class-E PA that achieves efficiency of 65–70% at 2.4 GHz. Then, we discuss state-of-the-art envelope tracking (ET) design for monolithic wideband RF mobile transmitter applications. A brief overview of Doherty PA design for the next-generation wireless handset applications is then given. Towards the end of the paper, we discuss an inherently broadband and highly efficient class-J PA design targeting future multi-band multi-standard wireless communication protocols.

Keywords: radio frequency power amplifier; silicon-based power amplifier; envelope tracking; class-E amplifier; broadband PA; class-J; Doherty power amplifier

1 Introduction

Silicon-based single-chip transceiver products for GSM, WLANs, Bluetooth, and digital enhanced cordless telecommunications (DECT) have become commercially available in the past decade. Recently, WLAN/Bluetooth integrated circuit (IC) vendors have successfully integrated less demanding silicon-based monolithic PAs into their transceiver ICs. A critical challenge for wireless transmitter design for mobile products is how to maximize the power-added efficiency (PAE) of battery-operated devices. The peak and average PAE of an RF transmitter heavily impacts the size of the battery and heat dissipation as well as reliability, yield, and cost. On the other hand, many modern wireless transmission protocols, such as WCDMA, LTE, and WiMAX use non-constant envelope modulation with

high peak-to-average power ratios (PAPRs) to be spectrally efficient. This causes rapid changes in magnitude/phase of the modulated signals. Therefore, RF PAs for these modern wireless communication systems require very high linearity. The mobile WiMAX (802.16e) standard, for example, has a 1–75 Mb/s gross data rate and up to 20 MHz signal bandwidth using a 64 QAM-1/2 modulation format with a 10–12 dB peak-to-average ratio (PAR). This standard requires an error vector magnitude (EVM) of below –26 dB (5%) with stringent transmission [1]. Future 4G LTE-A systems will require even larger bandwidths, with signal bandwidth extending to 100 MHz or possibly even greater.

Numerous transmitter configurations have been developed for better PAE, high linearity, and broadband operation [1]–[25]. Switching-mode PA typologies (classes D, E, and S) can increase PAE considerably more than

typologies of traditional linear PAs (classes A, AB, B, and C) by operating devices as switches to minimize overlapping of current and voltage waveforms [26]. With relatively easier on-chip integration and high efficiency at RF, silicon-based class-E PAs are very attractive for use in mobile devices [2]. Because of their nonlinear characteristics, class-E PAs are quite suitable for constant envelope modulation schemes; however, they can also be employed in spectrally-efficient modulation systems that require non-constant envelope signals if appropriate linearization techniques are also used.

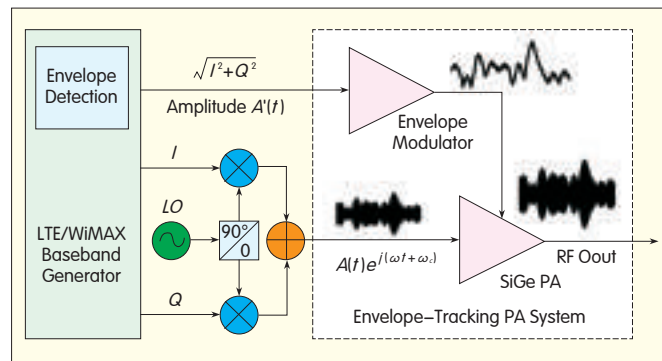
It is desirable to design highly efficient silicon-based RF PAs for mobile broadband applications because they can be cheaper than III–V compound semiconductor PAs and can also achieve higher-level integration with good thermal conductivity. However, III–V compound

semiconductor PAs still dominate the RF handset PA market. SiGe PAs and complementary metal oxide semiconductor (CMOS) PAs are very popular in the WLAN market and are becoming serious contenders for the Watt-level handset market. The ruggedness and reliability of silicon-based PAs needs to be carefully tested under different voltage standing wave ratio (VSWR) mismatched conditions at the output for handset applications (for example, 20:1 VSWR). The latest results suggest that standalone SiGe PAs may be very suitable for Watt-level 2G/2.5G handset PA applications. CMOS PAs have already been successfully demonstrated in GSM PA production.

Another key issue is keeping the RF PA and transmitter efficient not only at the Watt-level peak output power but also at the back-off low-power regions. In this way, average PAE for the transmitter system is excellent. To achieve this goal, slow power supply modulation and/or transistor size switching can be used to increase PAE of an RF PA at low output power levels. However, there is a thornier problem to solve for mobile broadband handset applications. PAR in 3G/4G communications is inherently high, and the PA output can change rapidly (in μs) from 1–3.4 V. This makes the PA mostly inefficient if it is designed to operate with a peak PAE at the high supply level of 3.4 V. Therefore, for a most power-efficient transmitter (TX) system, a fast envelope modulator should be used to track the PA supply voltage in order to somehow modulate its output power and maximize PAE at both high and low output power levels (so that average PAE is maximized). We discuss this envelope tracking (ET) technique later.

The following is a discussion on high-efficiency RF PA and transmitter design with GSM. A TX architecture with attractive PAE and enhanced linearity uses polar modulation of nonlinear PAs. The baseband signal is modulated in the amplitude/phase domain rather than the in-phase/quadrature (I/Q) domain. In the past, polar transmitters were mostly used for

Figure 1. A simplified block diagram of an ET-based RF TX system [1].



high-power base station applications to reduce heat dissipation. However, they have recently become widely used in wireless handset TX design for mass production because they are cheaper and significantly more efficient [3]. Recent research indicates that polar TXs using either envelope tracking (ET) or envelope elimination and restoration (EER) are capable of excellent system efficiency and linearity in several 3G/4G wireless applications. ET-based polar TXs often outperform EER-based polar TXs because of the relaxed requirements on the envelope-tracking amplifier bandwidth. ET-based polar TXs are also not sensitive to timing misalignment between the AM and RF paths [1]. Fig. 1 shows a simplified block diagram of an ET-based RF TX system. The signal bandwidth increases considerably from GSM/EDGE to W-CDMA/LTE, and the bandwidth requirement on the envelope modulator becomes too high and critical to the overall efficiency of the TX system for the EER system to be very attractive. Timing alignment requirements also make it significantly more difficult to meet the linearity specifications. Digital predistortion may be inevitable in this case. We have repeatedly shown that ET-based TX systems outperform EER-based systems in terms of enhanced data rate for GSM evolution (EDGE), WiMAX, WLAN, and LTE applications, especially when SiGe PAs are used [1], [3], [9], [16].

Doherty PA topology [26] is also effective in satisfying the stringent linearity requirement for large output power (Watt level). The topology can be used to achieve high average

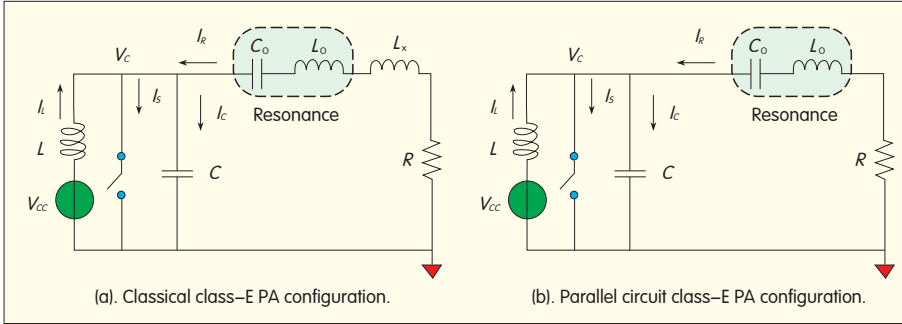
efficiency. Doherty power combining has been frequently used in base station transmitter applications; however, because it has better linearity and efficiency, it has been seriously proposed for handset application as well. Recent research has shown that it can be integrated with lumped element using CMOS technology.

A new mode of operation — class-J PAs [26] — shows the theoretical potential of obtaining linearity with as much broadband efficiency as conventional narrowband class-AB designs. However, class-J PAs do not require a band-limiting harmonic short [4]. Class-J PAs use the 2nd harmonic voltage to realize a phase shift between the output current and voltage waveforms and to provide the reactive termination for the 2nd harmonics. This allows the transmitter to achieve broadband behavior and cover a wide range of frequencies in order to satisfy modern multiband multistandard mobile communication applications.

In section 2, we discuss two considerations when deriving the class-E operation mode, covering the mathematical derivations with a silicon PA design example. In section 3, we discuss envelope tracking (ET) — a very promising technique for enhancing efficiency and linearity in broadband mobile handset applications. In section 4, we discuss the design concept and recent research into Doherty PAs for handset applications. In section 5, we discuss class-J mode PA, and we conclude in section 6.

2 Class-E PA Operation

Among conventional class-A, AB, B,



▲ Figure 2. Two class-E PA configurations.

C, and switching mode PAs, Si-based class-E PAs are quite suitable for portable devices. Its simple topology allows easier integration on-chip and high efficiency at RF. The definition of class-E PA operation proposed in [5] addresses three specific objectives for the collector voltage and current waveforms: 1) the transistor is off when voltage begins to rise across the transistor; 2) the collector voltage goes all the way back to zero when the transistor is turned on; and 3) the slope of the collector voltage should be zero when the transistor is turned on. According to these three conditions, several types of derivations have been reported [6], [7], [26].

2.1 Traditional Class-E Configuration

The classical derivation introduced in [5] assumes that choke inductance is infinite; all the passive elements are ideal (with no parasitics) and the transistor is operated as an ideal switch; and the amplifier is 100% efficient. For a given V_{cc} , P_{out} , operating frequency, and 50% duty cycle, all the component values in Fig. 2(a) can be defined by (1) and (2) from Raab's derivations:

$$R = 0.578 \cdot \frac{V_{cc}^2}{P_{out}}, C = \frac{0.184}{\omega R} \quad (1)$$

$$L_x = \frac{1.1525R}{\omega}, V_{s,max} = 3.56 V_{cc}. \quad (2)$$

C_0 and L_0 can be determined by a desired loaded Q value of the circuit. $V_{s,max}$ is the peak voltage of the collector/drain of the transistor (switch).

2.2 Class-E PA with Parallel Element Configuration

For modern silicon-based IC design,

a very large choke inductor is not suitable for integration on-chip because it occupies a large area and has a very low Q factor. Instead, small inductance or bondwire inductance of 1–2 nH can be used for the choke inductor. In this case, the equations in section 2.1 become inaccurate. Therefore, a subclass-E PA (also called a parallel circuit class-E PA) design analysis and derivation is provided, as in [7]. Fig. 2(b) shows the simplified circuit diagram for this PA configuration. The difference in this configuration is that a finite choke inductance is used, but no L_x (Fig. 2(a)) is needed. For a given V_{cc} , P_{out} , and frequency, the component values in Fig. 2(b) can be defined by (3) and (4):

$$R = 1.365 \cdot \frac{V_{cc}^2}{P_{out}}, C = \frac{0.685}{\omega R} \quad (3)$$

$$L = \frac{0.732 R}{\omega}, V_{s,max} = 3.647 V_{cc}. \quad (4)$$

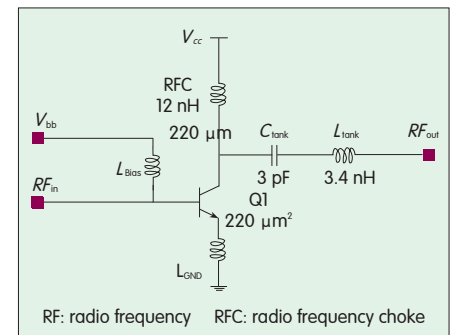
Compared with the classical class-E configuration and derivations, subclass-E PA has the following advantages:

- Small finite choke inductance is used and no additional reactive element is connected in series to the filter. This reduces the number and size of components to enable on-chip integration.
- The load value R in (3) is larger than that in (1) when calculated for this network. This reduces the impedance transformation ratio to the 50 Ω output.
- Theoretically, a slightly higher maximum peak switch voltage can be achieved in the subclass-E circuit, and slightly higher output power is available (provided that device breakdown is not an issue).

Equations (1)–(4) can provide good insights for PA designers, especially for low RF where parasitics are not important. However, even though these theoretical analyses enable engineers to design class-E PAs, the equations ignore device parasitics that may be important at high RF. So the frequencies equations may only be used as rough design guidelines [8].

2.3 Highly Efficient Silicon-Based Monolithic Class-E PA Design

A highly efficient silicon-based monolithic class-E PA design was recently reported in [9]. Fig. 3 shows a simplified schematic for this monolithic one-stage class-E PA designed with IBM 7HP 0.18 μ m BiCMOS SiGe technology. Simulation and measurement using a realistic simulation program with integrated circuit emphasis (SPICE) shows that, to achieve best PAE and P_{out} of –15–22 dBm, a suitable size for the last-stage RF PA (based on a high breakdown device) can be close to 200–280 μ m². This gives a simulated maximum collector current density of <0.5 mA/ μ m², well within the measured HBT safe operation area (SOA). The parasitic effect of bondwire inductance at the emitter node to ground is a little more significant at 2.4 GHz compared to 900 MHz because there is a large increase in effective reactance to ground that acts as strong negative feedback for the bipolar junction transistor (BJT) (by a factor of 2.67). The problem can be avoided by adding more down-bonds to considerably improve the gain and PAE. Our recent study suggests that at least 4–5



▲ Figure 3. Schematic of a one-stage SiGe class-E PA at 2.3–2.4 GHz [10].

down-bonds should be used in a 2.4 GHz SiGe monolithic RF PA design for 20+ dBm P_{out} . PAE is improved by 20% just by adding the down-bonds [10]. Using these design procedures, we achieved 65% peak PAE and 70% collector efficiency (CE) for a one-stage class-E SiGe RF PA. These excellent results were achieved without using any off-chip matching components; however, the radio frequency choke (RFC) is still kept off-chip to reduce loss. For a two-stage SiGe PA, close to 20 dB gain and 50–55% PAE have been achieved [2].

3 Envelope-Tracking-Based RF Polar Transmitter

3.1 Considerations for the System Design of a Polar Transmitter

In practice, RF PAs for mobile applications present many technical challenges. Polar TX systems are known to be sensitive to timing mismatches between the AM and PM paths [12]. The group delay of the two signal paths must be matched to minimize PA distortion, which is difficult to control across all process-voltage-temperature (PVT) corners [26]. The other major obstacle is the larger bandwidth required for the circuits in the polar TX system for broadband wireless. The I/Q to polar transformation at baseband is a nonlinear operation, which inevitably expands the bandwidths of both the AM and PM output signals. Depending on the specific modulation scheme and system specifications, the constant-amplitude phase signal path for the EER-based polar TX may need roughly ten times larger bandwidth than the input signal in order to pass the TX mask requirement and/or the error-vector-magnitude (EVM) specifications [11]. However, these issues can be significantly addressed by using an ET-based polar TX architecture (also known as hybrid-EER or H-EER architecture), as shown in Fig. 1 [1], [12]–[14].

Compared to an EER-based large-signal polar TX system, an

ET-based polar TX system has the following benefits [3]:

- Gain at low output power is higher because the ET-based PA can be operated near saturation but is not always fully saturated, as in EER.
- Sensitivity to timing mismatch (between the RF and amplitude paths) is lower than in EER [14], [15]. The RF path for ET-based TX contains the amplitude modulated signal as well; therefore, the signals in the RF and amplitude paths are very similar. This is not the case in EER.
- For the same linearity performance, ET has lower bandwidth requirement for the envelope amplifier design than in EER [12]. This can be critical because efficiency of the envelope amplifier can limit the overall composite PAE of an ET/EER system, as shown in (8). The higher bandwidth requirement in the amplitude modulator design for EER translates into lower modulator efficiency and lower overall TX system PAE if the PAE of the RF PA alone is kept the same for both ET and EER.
- When the amplitude of the input signal is fixed, it is difficult to keep the gain of an RF PA constant while modulating its collector/drain voltage to adjust the output power, as in EER. Therefore, EER is inherently less linear than ET because its power gain inevitably decreases when the supply voltage is decreased, especially for bipolar devices. ET is superior in this regard because both the DC bias and input signal level can be adjusted to help keep the gain more constant [1]. The high-bandwidth RF limiter required for EER can be power hungry because it needs to operate at high RF.
- ET has less RF feed-forward signal that can appear as distortion in the TX output. Because the drive signal is hard-limited in EER, the RF feed-forward can cause significant distortions (AM-AM, AM-PM) by the large gate-drain or base-collector Miller capacitance in the final RF power device. These distortions couple to the output to cause linearity/EVM issues.

3.2 Design of Envelope Amplifier for ET-Based Polar Transmitter

The overall efficiency of the

ET-based polar TX system is the product of the envelope amplifier efficiency and the PA collector efficiency (CE), which is expressed by

$$\eta_{ET,TX} = \eta_{Envelope_Amp} \cdot \eta_{RF_PA} \quad (5)$$

Therefore, the design of a high-efficiency envelope amplifier is critical to overall system efficiency in a polar TX system using either the EER or ET. Because a wide envelope signal bandwidth above 20 MHz considerably reduces the efficiency of traditional switching direct current–direct current (DC–DC) converters, a linear-assisted switch-mode envelope amplifier (or “split-band” envelope amplifier) proposed in [11] can be used. This system helps attain wideband tracking and maintain high overall system TX efficiency. The simplified discrete circuit schematic is shown in Fig. 4.

The linear-assisted envelope amplifier circuit (often called envelope modulator) has three different modes of operation [3], [11]:

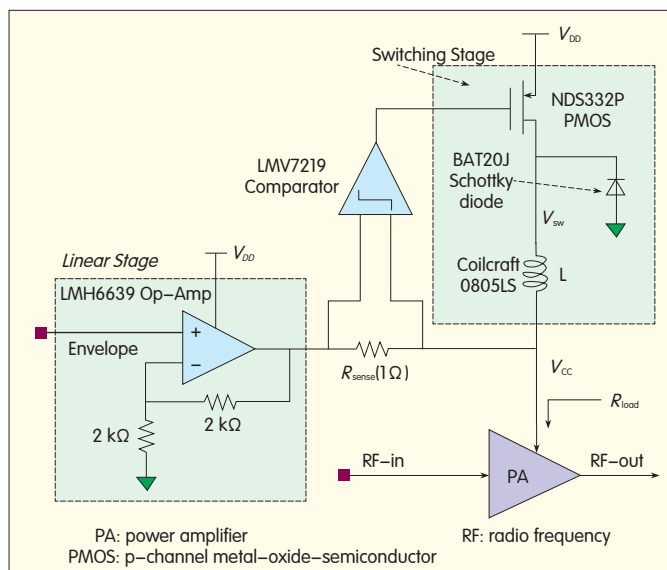
1) linear operation for small-signal envelope (small-signal operation). This occurs when the average slew rate of the switcher current is much larger than the average slew rate of load current. The buck converter can fully support the load current; that is, the switcher stage can provide both DC and AC components of the envelope signal.

2) large-signal operation. This occurs when the average slew rate of the switcher current is much smaller than the average slew rate of the load current. The switcher stage can only provide the DC component of the envelope, and the AC component is provided by the much faster linear stage. The average switching frequency of the buck converter is almost the same as the signal frequency that the current sensing resistor can detect.

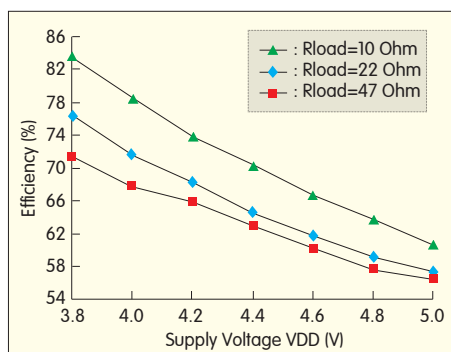
3) matched slew-rate point. In this case, the average slew rate of the switcher current is the same as the average slew rate of the load current.

Fig. 5 shows envelope amplifier efficiency for supply voltage at several load resistances with WiMAX 64 QAM modulated signal of 8.75 MHz.

As the supply voltage decreases, the



◀ Figure 4.
Schematic of a
linear-assisted
switch-mode envelope
amplifier [17].



▲ Figure 5. Efficiency of the envelope amplifier vs. supply voltage with the envelope of a WiMAX 64QAM modulated signals of 8.75 MHz [17].

efficiency always increases. However, lower supply voltage can clip the output envelope waveform and cause more TX distortion. This suggests that if the clipping can be avoided by applying some decreting algorithms to reduce PAR, the supply voltage of the envelope amplifier can be lowered to considerably increase the envelope amplifier's efficiency. The efficiency of the envelope amplifier also increases along with the reduction of its load resistance. For an RF PA output power of 23–28 dBm, the collector impedance seen by the envelope amplifier can be less than 10 Ω , making the envelope amplifier's efficiency 84% at $V_{DD} = 3.8$ V.

3.3 Polar Transmitters Using ET Technique

Recent research has shown

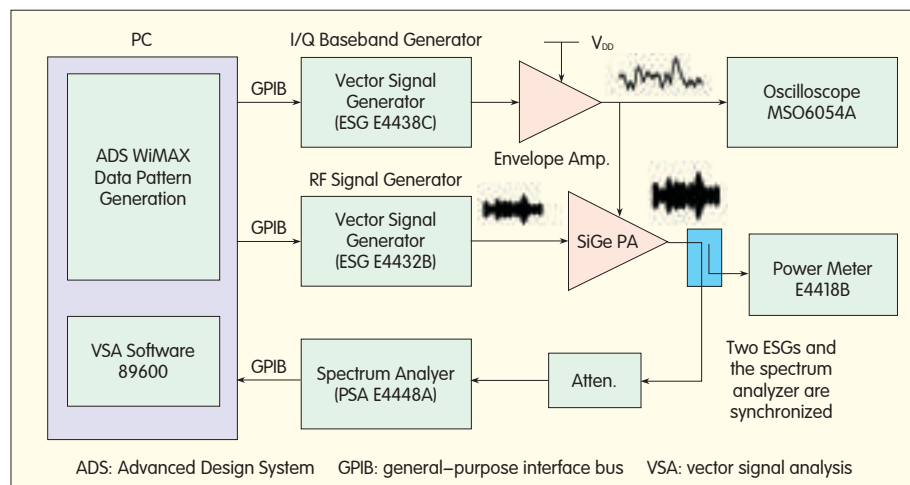
successful ET-based polar TX system designs for modern high-PAR mobile applications such as WiMAX and 3GPP LTE [1], [8], [16], [18].

Our recent work shows that the test bench setup for the polar TX system characterization is as shown in Fig. 6. Fig. 7(a) shows the measured TX output EVM, gain, and overall PAE of the entire ET-based polar TX system with a discrete envelope modulator and a monolithic PA applying the WiMAX 64 QAM 8.75 MHz signal. The overall system PAE is 30.5% at 17 dBm average output power, and the EVM is 4.4%. The output spectrum of our ET-based polar TX also passed the stringent WiMAX 64QAM mask defined

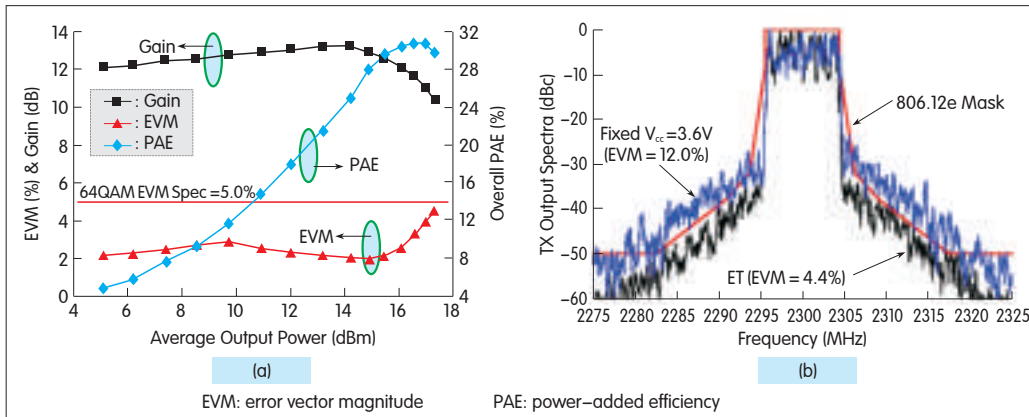
by European Telecommunication Standards Institute (ETSI), as shown in Fig. 7(b). With the ET scheme used, the SiGe PA can operate at its $P_{2\text{dB}}$ compression point (17 dBm) without violating WiMAX linearity specifications. However, the standalone PA (fixed V_{cc} without ET) needs 4–5 dB back-off to satisfy both EVM specifications and spectral mask (not shown here).

The ET-based polar transmitter system designed in [18] performs well for 3GPP LTE application. Fig. 8(a) shows the TX output EVM, gain, and overall PAE of the entire ET-based polar system based on a 16QAM LTE 5 MHz modulation signal. The overall TX system PAE, including the monolithic envelope amplifier and the differential cascode PA, is 33.6% when the average output power is 21 dBm. The EVM is only 7.0%. These results, when compared with those of the standalone PA, show that the output of the ET-based PA can have lower EVM, as shown in Table 1.

Fig. 8(b) compares the TX output spectra of the ET-based PA with the standalone fixed supply PA at an average output power of 21 dBm. The ET technique reduced the out-of-band emission of the TX spectrum (ACPR increased by ~ 10 dB at the 3.5 MHz offset frequency), allowing the polar TX to pass the LTE transmit emission mask. On the other hand, even though we intentionally increased V_{cc} to 4.2 V for the standalone fixed-supply PA, the



▲ Figure 6. Test bench setup for ET-based polar TX systems using discrete linear-assisted switch-mode envelope amplifier and monolithic SiGe PA [17].



▲ Figure 7. (a) Measured TX EVM, gain, and overall PAE vs. average output power of the entire ET-based polar TX system with WiMAX 64 QAM 8.75 MHz input signal; (b) Measured polar TX output spectra before and after applying ET scheme, average output power = 17 dbm [17].

▼ Table 1. EVM and PAE of the SiGe PA before and after using ET with 16QAM LTE signals at 1.42 GHz. No predistortion applied [19]

Supply Voltage	3.3 V	3.8 V	4.2 V	ET
EVM	13%	12%	10%	7%
PAE	45%	38.7%	34.6%	33.6%
Pass Mask	No	No	No	Yes

EVM: error vector magnitude PAE: power-added efficiency

PA still had very high TX out-of-band emission. Furthermore, the standalone fixed-supply PA was also tested under a 6 dB back-off mode from $P_{out} = 21$ dBm (with only 15 dBm output power). As seen in Fig. 8(b), even with $P_{out} = 15$ dBm for the standalone fixed-supply PA, the output spectrum failed once again to pass the LTE transmit spectrum mask. This measurement suggests that ET operation not only keeps the PA efficient but also somehow linearizes it compared to the standalone fixed-supply PA. This is not the first time we have observed this linearization in ET operation; we have also seen this in common-emitter and cascode SiGe PAs. Studies are ongoing to see if similar effects can be observed in CMOS PAs as well.

4 Doherty Power Amplifier

The Doherty amplifier was first proposed by W. H. Doherty in 1936 [19]. The

efficiency and output power of an RF PA can be increased by using the Doherty power combining technique [26]. A Doherty PA consists of a main amplifier and a peaking amplifier. The output load is connected to the main amplifier through an impedance inverter that is usually a quarter-wave transmission line. The Doherty amplifier configuration is shown in Fig. 9(a). Assuming the output current of each PA is linearly proportional to the input

voltage with harmonic short terminations, the efficiency of a Doherty amplifier can be analyzed using the fundamental and DC components only.

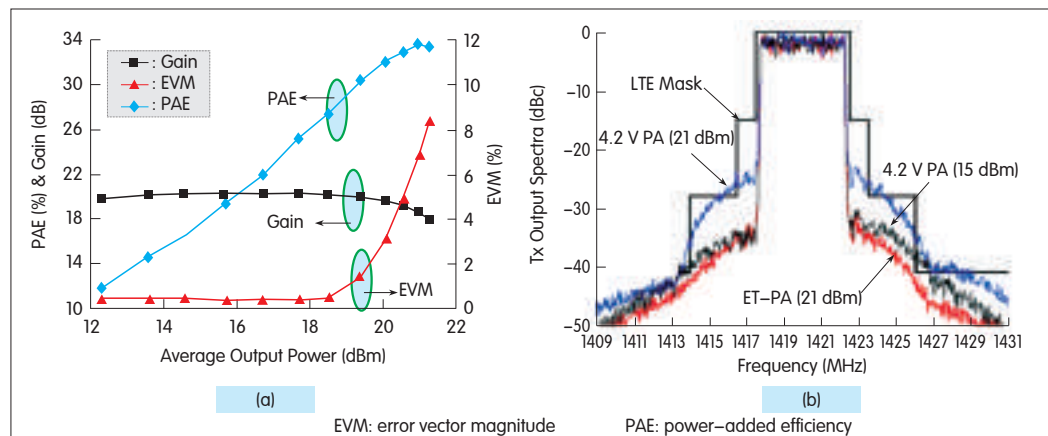
In traditional Doherty operation, the peaking PA is turned on when the main PA, which is at half the maximum input voltage, begins to compress. A load modulation takes place for the PAs according to (6) and (7). Z_{main} and $Z_{peaking}$ are the impedances seen by the output of the main PA and

peaking PA, respectively. I_m and I_p are the currents flowing through the two PAs. In typical operation, at the low power-delivering region, the peaking PA is off, and Z_L is up-converted to the load of the main PA, which enables the linear main PA to be more efficient, as in Fig. 9(b). At the high power-delivering region, the peaking PA is turned on to combine the power with the main PA while also enhancing the linearity, as shown in Fig. 9(c).

$$Z_{main} = \begin{cases} \frac{Z_T^2}{Z_L} & 0 < v_{in} < v_{in,max}/2 \\ \frac{Z_T^2}{Z_L \cdot (1 + \frac{I_p}{I_m})} & v_{in,max}/2 < v_{in} < v_{in,max} \end{cases} \quad (6)$$

$$Z_{peaking} = \begin{cases} \infty & 0 < v_{in} < v_{in,max}/2 \\ Z_L \cdot (1 + \frac{I_m}{I_p}) & v_{in,max}/2 < v_{in} < v_{in,max} \end{cases} \quad (7)$$

Doherty PAs have been used in base



▲ Figure 8. (a) Measured TX EVM, gain, and overall PAE vs. average output power of the entire ET-based polar TX system with 16 QAM LTE 5 MHz signals. Supply voltage = 4.2 V, frequency = 1.42 GHz; (b) Measured TX output spectra before and after the ET technique was applied [19].

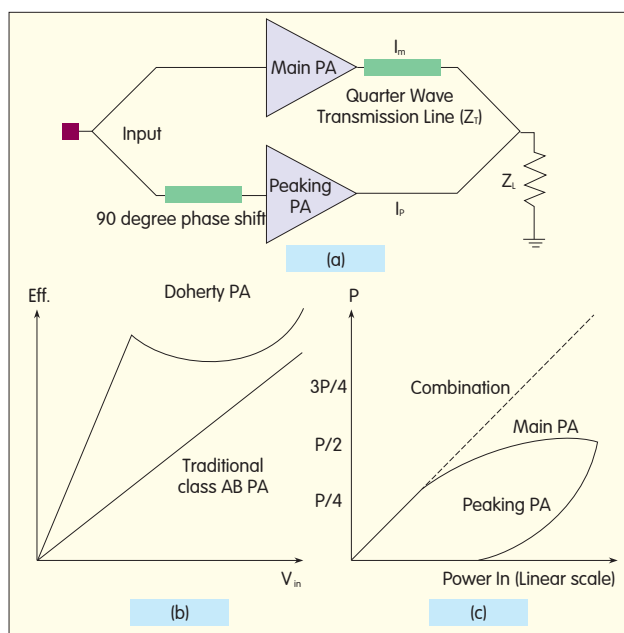


Figure 9. (a) traditional Doherty PA configuration; (b) comparison of efficiency between Doherty PA and traditional class-AB PA with the same load impedance; (c) Doherty power combining.

Table 2. CMOS Doherty PA Performance Summary Chart

Work	Tech. (um CMOS)	Freq. (GHz)	PAE	Power (dBm)	Note
[21]	0.18	2.4	33% at P _{1dB} , 21% at 6 dB back-off	21.4	Off-chip transmission line combine
[22]	0.09	3.65	39% at P _{sat} , 12.5% at 6 dB back-off	28.9	On-chip transmission line combine
[23]	0.18	2.4	14% at P _{1dB} , 10% at 7 dB back-off	21	Lumped element on chip combine
[24]	0.13	1.75	36% at P _{sat} , 18% at 10 dB back-off	31.5	Lumped element on chip combine
[25]	0.13	2.4	60% at P _{1dB} , 35% at 5 dB back-off	22.7	Lumped element off chip combine
[26]	0.09	2.4	26.7% at P _{1dB}	20.5	Transformer on chip combine
CMOS: complementary metal oxide semiconductor				PAE: power-added efficiency	

stations for many years, but it is difficult to integrate the transmission line on chip for handset applications. Voltage standing wave ratio (VSWR) mismatch for handsets, and inherent narrow-band operation of the impedance converter are serious barriers for Doherty PA in handsets. However, recent literature has reported impressive fully integrated CMOS Doherty PAs, as summarized in Table 2. In some of the works, lumped elements are used to realize the impedance inverter for the loading of the main PA.

5 Class-J Mode Operation With Broadband Amplification

The class-J PA topology consists of a slightly inductive load at the fundamental frequency and the only

capacitive load at the 2nd harmonic (Fig. 10). This circuit has no explicit harmonic trapping network, which makes the class-J PA potentially capable of broadband operation [4].

According to [4] and [5], a class-J PA is an unconventional class-AB PA with 2nd harmonic enhancement. The key to a class-J PA design is the 2nd harmonic element contained in the voltage waveform. The class-E-like waveform in [5] makes the PA more efficient, and the non-switching mode makes it linear. Furthermore, the inductive load at the fundamental frequency, and (only) the capacitive load at the 2nd harmonic keeps the drain RF voltage above the knee region [4], which makes class-J operation highly efficient and linear.

The class-J approach uses a phase shift between the output current and

voltage waveforms to render the 2nd harmonic termination purely reactive. This implies significant possibilities for the bandwidth efficiency of the class-J mode of an RF PA. For class-J operation, the load impedance must satisfy the following equation:

$$\begin{cases} Z_{f_0} = R_L \cdot (1 + j) \\ Z_{2f_0} = R_L \cdot (0 - \frac{3\pi}{8}j). \end{cases} \quad (8)$$

When R_L is defined, the harmonic load impedance is formed. Load-pull techniques should be applied to find the value of R_L according to the desired high efficiency or high power. Broadband input matching needs to be applied to realize a broadband PA.

In [5], a class-J PA with a commercially available 10 W GaN high-electron mobility transistor is shown. This PA design has near-rated output power of 39 dBm and very high efficiency of more than 60% across a bandwidth of 1.4–2.6 GHz, centred at 2 GHz. Not many papers have been written about using class-J PAs for handset applications, but we believe class-J is promising for modern multiband mobile TX handset design.

6 Conclusions

In this paper, several highly efficient RF PA design techniques for silicon-based mobile broadband PAs and transmitters have been discussed. Two derivations of highly efficient class-E PAs based on different assumptions are shown in order to provide design insights. However, RFIC still needs to be carefully designed because of parasitics and non-ideal

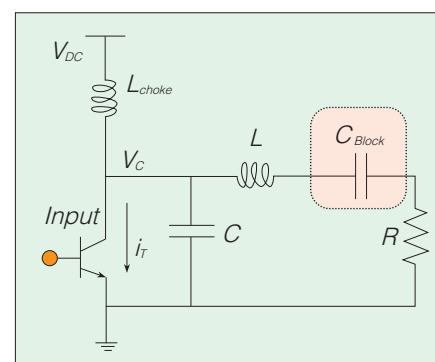


Figure 10. Class-J configuration.

components in practical application. Characteristics and advantages of ET-based polar TX design have been discussed, and a comparison has been made with the EER technique. State-of-the-art ET-based polar TX performance for WiMAX and 3GPP LTE has also been demonstrated. Doherty PA design for high power, linear, and high efficiency operation as well as class-J PA design for broadband handset applications has also been briefly discussed. It is our hope and expectation that some of the promising techniques for high-efficiency silicon-based RF PA design can, someday, revolutionize the cellular RF PA market for future broadband wireless communication.

References

- [1] Yan Li, Po-Hsing Wu, J. Lopez, R. Wu, D.Y.C. Lie, K. Chen, S. Wu, Tzu-Yi Yang, "A highly-efficient RF polar transmitter using SiGe power amplifier and CMOS envelope-tracking amplifier for mobile WiMAX," in *Proc. IEEE VLSI Design, Automation and Test*, Taiwan, Apr. 2011, pp. 1–4.
- [2] D.Y.C. Lie, J. Lopez, J., J.D. Popp, J.F. Rowland, Guogong Wang, Guoxuan Qin, Zhenqiang Ma, "Highly efficient monolithic class-E SiGe power amplifier design at 900 and 2400 MHz," in *IEEE Trans. Circuits Syst. I*, vol. 56, no. 7, pp. 1455–1466, Jul. 2009.
- [3] J. Lopez, Yan Li, J.D. Popp, D.Y.C. Lie, Chia-Chang Chuang, K. Chen, S. Wu, Tzu-Yin Yang, Gin-Kou Ma, "Design of highly efficient wideband RF polar transmitters using the envelope-tracking technique," *IEEE J. Solid-State Circuits*, vol. 44, no. 9, pp. 2276–2294, Sep. 2009.
- [4] P. Wright, J. Lees, J. Benedikt, P.J. Tasker, S.C. Cripps, "A methodology for realizing high efficiency class-J in a linear and broadband PA," *IEEE Trans. Microw. Theory Tech.*, vol. 57, no. 12, pp. 3196–3204, Dec. 2009.
- [5] N.O. Sokal and A.D. Sokal, "Class-E: A new class of high efficiency tuned single-ended switching power amplifiers," *IEEE J. Solid-State Circuits*, vol. 10, no. 3, pp. 168–176, 1975.
- [6] F.H. Raab, "Idealized Operation of the Class E Tuned Power Amplifier," *IEEE Trans. Circuits Syst.*, vol. 24, no. 12, pp. 725–735, 1977.
- [7] A. V. Grebennikov, H. Jaeger, "Class E with parallel circuit—A new challenge for high-efficiency RF and microwave power amplifiers," *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, Aug. 2002, pp. 1627–1630.
- [8] D.Y.C. Lie, P. Lee, J.D. Popp, J.F. Rowland, H.H. Ng, A.H. Yang, "The limitations in applying analytic design equations for optimal class E RF power amplifiers design" in *Proc. IEEE VLSI Design, Automation and Test*, Munich, Aug. 2005, pp. 161–164.
- [9] Yan Li, J. Lopez, D.Y.C. Lie, K. Chen, S. Wu, Tzu-Yi Yang, "SiGe Class-E Power Amplifier With Envelope Tracking For Mobile WiMAX/Wibro Applications," in *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Taipei, Jun. 2009, pp. 2017–2020.
- [10] D.Y.C. Lie, J. Lopez and J.F. Rowland, "Highly efficient class E SiGe power amplifier design for wireless sensor network applications," in *Proc. IEEE Bipolar/BiCMOS Circuits and Tech. Meeting (BCTM)*, Boston, MA, Sep. 2007, pp. 160–163.
- [11] F. Wang, D. F. Kimball, J. D. Popp, A. H. Yang, D. Y. C. Lie, P. M. Asbeck, and L. E. Larson, "An improved power-added efficiency 19-dBm hybrid envelope elimination and restoration power amplifier for 802.11g WLAN applications," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4086–4099, Dec. 2006.
- [12] J. Popp, D. Y. C. Lie, F. Wang, D. Kimball, and L. Larson, "A fully-integrated highly-efficient RF class E SiGe power amplifier with an envelope-tracking technique for EDGE Applications," in *Dig. IEEE Radio and Wireless Symp. (RWS)*, San Diego, CA, Jan. 2006, pp. 231–234.
- [13] D.Y.C. Lie, J.D. Popp, F. Wang, D. Kimball, L.E. Larson, "Linearization of highly-efficient monolithic class E SiGe power amplifiers with envelope-tracking (ET) and envelope-elimination-and-restoration (EER) at 900MHz," in *Proc. IEEE 6th Dallas Circuits and Systems Workshop (DCAS'07)*, Dallas, TX, Nov. 2007, pp. 39–42.
- [14] Feipeng Wang, A.H. Yang, D.F. Kimball, L.E. Larson, P.M. Asbeck, "Design of wide-bandwidth envelope-tracking power amplifiers for OFDM applications," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 4, pp. 1244–1255, Apr. 2005.
- [15] Yan Li, et al., "An SiGe Envelope-Tracking Power Amplifier with an Integrated CMOS Envelope Modulator for Mobile WiMAX/3GPP LTE Transmitters," *IEEE Trans. Microw. Theory Tech.*, submitted in 2011.
- [16] Yan Li, J. Lopez, D.Y.C. Lie, K. Chen, S. Wu, Tzu-Yi Yang, Gin-Kou Ma, "Circuits and system design of RF polar transmitters using envelope-tracking and SiGe power amplifiers for mobile WiMAX," *IEEE Trans. Circuits Syst. I*, vol. 58, no. 5, pp. 893–901, May. 2011.
- [17] Yan Li, J. Lopez, D.Y.C. Lie, K. Chen, S. Wu, Tzu-Yin Yang, "A broadband SiGe power amplifier in an efficient polar transmitter using envelope-tracking for mobile WiMAX," in *Proc. IEEE Silicon Monolithic Integrated Circuits in RF Systems (SiRF)*, Glendale, AZ, Jan. 2011, pp. 137–140.
- [18] Yan Li, J. Lopez, D.Y.C. Lie, K. Chen, S. Wu, Tzu-Yi Yang, "Efficiency enhancement and linearity trade-offs for cascode vs. common-emitter SiGe power amplifiers in WiMAX polar transmitters," in *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Paris, May. 2010, pp. 1915–1918.
- [19] Yan Li, J. Lopez, D.Y.C. Lie, K. Chen, S. Wu, Tzu-Yin Yang, "A highly efficient SiGe differential power amplifier using an envelope-tracking technique for 3GPP LTE applications," in *Proc. IEEE Bipolar/BiCMOS Circuits and Tech. Meeting (BCTM)*, Austin, TX, Oct. 2010, pp. 121–124.
- [20] W. H. Doherty, "A new high efficiency power amplifier for modulated waves," *Proc. IRE*, vol. 24, no. 9, pp. 1163–1182, 1936.
- [21] Y. Chen, et al. "A highly efficient CMOS RF power amplifier with automatic adaptive bias control," *IEEE Microw. Component Letter (MWCL)*, vol. 16, no. 11, pp. 615–617, Nov. 2006.
- [22] M. Elmala, J. Paramesh, K. Sourmyanath, "A 90-nm CMOS Doherty power amplifier with minimum AM-PM distortion," *IEEE J. Solid-State Circ.*, vol. 41, no. 6, pp. 1323–1332, Jun. 2006.
- [23] Li-Yuan Yang, Hsin-Shu Chen, Y-J E Chen, "A 2.4GHz fully integrated cascode-cascade CMOS Doherty power amplifier," *IEEE Microw. Component Letter (MWCL)*, vol. 18, no. 3, pp. 197–199, Mar. 2008.
- [24] N. Wongkomet, L. Tee, P.R. Gray, "A +31.5dBm CMOS RF Doherty power amplifier for wireless communications," *IEEE J. Solid-State Circ.*, vol. 41, no. 12, pp. 2852–2859, Dec. 2006.
- [25] J. Kang, et al. "An Ultra-High PAE Doherty Amplifier Based on 0.13um CMOS Process," *IEEE Microw. Component Letter (MWCL)*, vol. 16, no. 9, pp. 505–507, Sep. 2006.
- [26] S.C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed., Boston, MA: Artech, 2006.

Biographies

Ruili Wu (ruili.wu@ttu.edu) received his B.S. degree in electrical engineering from the Institute of Microelectronics, Tsinghua University, in 2010. He is currently working towards his Ph.D. degree in electrical engineering at Texas Tech University. His current research is focused on the design of integrated PA circuits and systems for highly efficient, linear wideband polar transmitters in LTE-like wireless applications.

Jerry Lopez (jerry.lopez@ttu.edu.) received his B.S. and M.E. degrees at the University of California, San Diego, in 2001 and 2005. He is on track to receive his Ph.D. degree at Texas Tech University in August 2011.

Mr. Lopez has worked for Northrop Grumman Corp. researching and developing the design of highly efficient PAs and polar modulators, including custom RFIC design. He has also worked for the US Navy, where he was a member of the RFIC design group. His research interests include design of analog/RF integrated circuits, such as highly integrated highly efficient CMOS PA/transmitter, and circuits and systems that are highly nonlinear.

Yan Li (yan.li@ttu.edu.) received his B.S. degree in electrical engineering from Southwest Jiao-Tong University, China, in 2007. He received his M.S. degree in electrical engineering from Texas Tech University in 2009. He is currently working towards his Ph.D. degree at TTU. His current research is focused on the design of highly-efficient and linear wideband polar transmitters with integrated PAs for mobile WiMAX/LTE wireless applications.

Donald Y.C. Lie (donald.lie@ttu.edu.) received his M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, in 1990 and 1995. He has held technical and managerial positions at Rockwell International, IBM, and Microtune Inc. He is currently the Keh-Shew Lu Regents Chair Associate Professor (tenured) in the Department of Electrical and Computer Engineering, Texas Tech University. He is also an adjunct associate professor in the Department of Surgery, Texas Tech University Health Sciences Center (TTUHSC). He has authored and coauthored approximately 120 peer-reviewed technical papers and book chapters and holds five U.S. patents. He and his students have won several Best Graduate Student Paper Awards in international conferences in 1994, 1995, 2006, 2008 (twice), 2010 (twice) and in 2011. His students have also won prestigious scholarships. Dr. Lie is the associate editor of IEEE Microwave and Wireless Components Letters (MWCL) and was a guest editor of IEEE Journal of Solid-State Circuits (JSSC) in 2009. His research interests include low-power RF/Analog integrated circuits and System-on-a-Chip (SoC) design and testing, and interdisciplinary research on medical electronics, biosensors, and bio-signal processing.

Multi-Gbit/s 60 GHz Transceiver Analysis Using FDM Architecture and Six-Port Circuit

Nazih Khaddaj Mallat¹, Emilia Moldovan², Serioja O. Tatu², and Ke Wu¹

(1. Poly-Grames Research Center, École Polytechnique de Montréal, Montréal, Québec H3T 1J4, Canada;

2. Institut National de la Recherche Scientifique, INRS-EMT, Montréal, Québec H5A 1K6, Canada)

Abstract: This paper presents an analysis and validation by advanced system simulation of compact and low-cost six-port transceivers for future wireless local area networks (WLANs) operating at millimeter-wave frequencies. To obtain realistic simulation results, a six-port model based on the measurement results of a fabricated V-band hybrid coupler, the core component, is used. A frequency-division multiplexing scheme is used by introducing four quadrature phase-shift keying (QPSK) channels in the wireless communication link. The data rate achieved is about 4 Gbit/s. The operating frequency is in the 60–64 GHz unlicensed band. Bit error rate (BER) results are presented, and a comparison is made between single-carrier and multicarrier architectures. The proposed wireless system can be considered an efficient candidate for millimeter-wave communication systems operating at quasi-optical data rates.

Keywords: millimeter-wave communications; Gbit/s data rates; passive components and circuits; six-port interferometer

1 Introduction

One of the goals of 4G wireless technologies is to simplify wireless systems in homes and enterprises. The coming 60 GHz WLANs are primarily aimed at applications with a short range and very high data rate, such as high-speed home, office, and high definition television (HDTV).

In 2001, the Federal Communications Commission (FCC) allocated a continuous block of 7 GHz of spectrum in the 57–64 GHz band for wireless communications [1] (where oxygen absorption limits long-distance interference). Energy propagation in the 60 GHz band has many unique characteristics and brings advantages such as high security, immunity from interference, and frequency re-use [2], [3]. There are many design challenges for millimeter-wave circuits, including the necessity for low cost, high power efficiency, and accurate computer

aided design models.

Six-port circuits are proposed for low-cost high-performance millimeter-wave transceivers. The six-port is a passive circuit, first developed in the 1970s for accurate automated measuring of the complex reflection coefficient during microwave network analysis [4]. It is a low-cost alternative to a network analyzer or beam direction-finding applications [5]. Various millimeter-wave front-end architectures based on six-port devices have been proposed in recent years. These architectures use various fabrication technologies and modulation schemes [6]–[9].

Section 2 of this paper provides an analysis of a fabricated hybrid coupler that uses miniature hybrid microwave integrated circuit (MHMIC) technology and the six-port model. Section 3 provides an analysis of the proposed 60 GHz transceivers, and simulation results of single-carrier and multicarrier systems are presented [10].

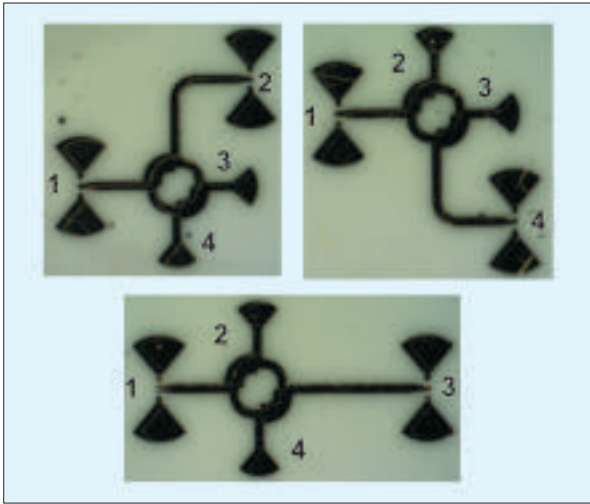
Conclusions are given in section 4.

2 MHMIC Hybrid Coupler and Six-Port Model

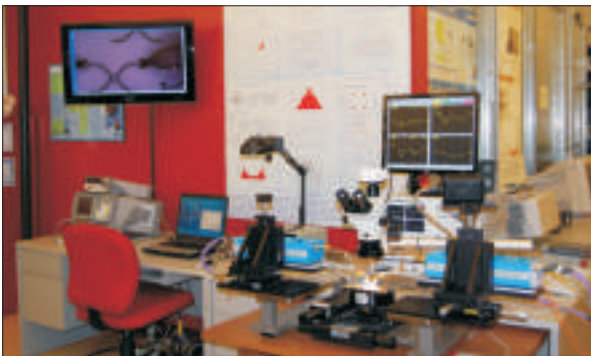
2.1 MHMIC Hybrid Coupler

The four-port 90° hybrid coupler is the core component of the six-port circuit and is designed and fabricated to operate in V-band. Using MHMIC technology, the six-port circuit is integrated on a 125 μm alumina substrate with a relative permittivity of 9.9.

Fig. 1 shows several microphotographs of the MHMIC 90° hybrid coupler. The RF short-circuits (replacing via-holes) for coplanar transitions, and 50 Ω loads are obtained using wideband open-circuited stubs. The diameter of the coupler is around 700 μm , and the 50 Ω line width is nearly equal to the thickness of the alumina substrate. To characterize fabricated MHMIC circuits, on-wafer



▲ Figure 1. Microphotographs of the MHMIC 90° hybrid coupler (with ports 1–4).



▲ Figure 2. S-parameter measurement set-up.

measurements are taken using a Cascade Microtech probe station (equipped with 150 μm pico-probes) connected to an Agilent Technologies E8362B millimeter-wave power network analyzer (PNA) (Fig. 2). Because of measurement limitations, 60–64 GHz is considered for circuit characterization and system simulation.

Fig. 3 shows the phase of transmission-scattering S-parameters (S_{12} and S_{13}). A 90° phase difference is obtained over the 4 GHz band, and the phase imbalance is around 5°. Fig. 4 shows the power coupling, matching, and isolation of the MHMIC-fabricated hybrid coupler. The power splits (S_{12} and S_{13}) over the 4 GHz band are between –3 dB and –4 dB, very close to the theoretical value of –3 dB. The return loss on input port 1 (S_{11}) versus frequency is higher than –20 dB, and the isolation between

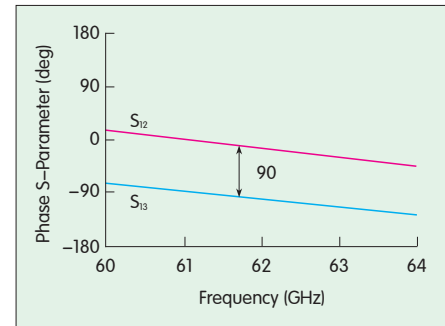
ports 2 and 3 (S_{23}) is higher than –15 dB (Fig. 4). Because of circuit symmetry, measurements of equal isolations between ports 1 and 4, and ports 2 and 3 are obtained. Measurements of return loss at all ports (S_{ii}) are also obtained.

2.2 Six-Port Model

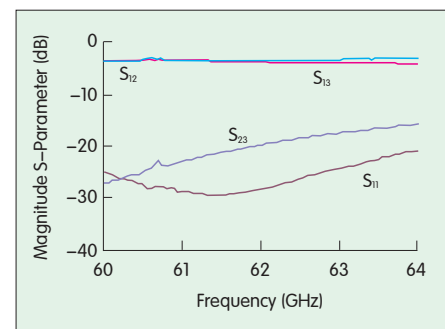
A six-port linear network is represented by a 6×6 characteristic matrix and is generally characterized by a standard two-port vector network analyzer (VNA). The two-port measurements, required for every possible combination of two ports, are taken. The 4 ports not connected to the VNA are terminated with appropriate loads. This multiport circuit is composed of four 90° hybrid couplers and a 90° phase shifter (Fig. 5). The input signals a_s and a_o are normalized waves, from the local oscillator (LO) and radio frequency signal, respectively. The output

detected signals can be calculated based on the multiport block diagram and using the quadratic characteristic of the power detectors, as detailed in [6].

For each two-port measurement, a 2×2 sub matrix of the six-port characteristic matrix is determined. To avoid necessary circuits (fifteen in the



▲ Figure 3. S-parameter phases of an MHMIC hybrid coupler.

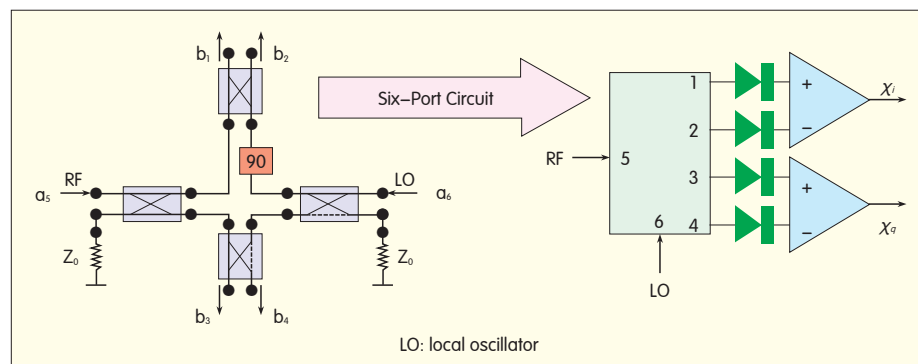


▲ Figure 4. Magnitude S-parameter for transmission, return loss and isolation.

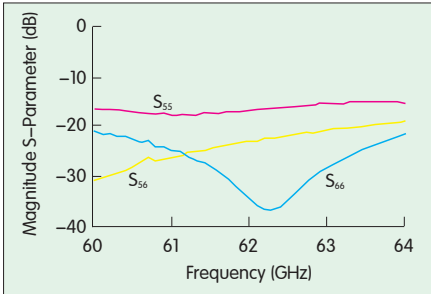
case of a six-port), and to obtain realistic results with minimum fabrication cost, the six-port model is implemented in 2010 Advanced Design System (ADS) software of Agilent Technologies. S-parameter measurements of fabricated 90° hybrid couplers interconnected by transmission lines are also used.

A matching of more than –15 dB and isolation of –20 dB are obtained for the input ports (Fig. 6).

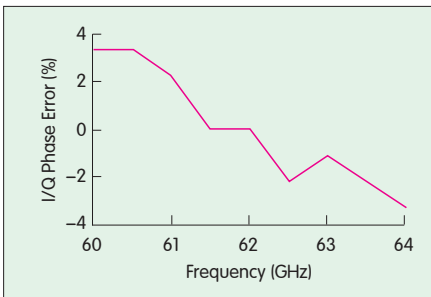
In a direct conversion scheme, the quadrature (I/Q) down-converted signals are obtained using a differential



▲ Figure 5. Six-port circuit.



▲ Figure 6. Simulation of S-parameter magnitude.



▲ Figure 8. Percentage of I/Q error over the 60–64 GHz band.

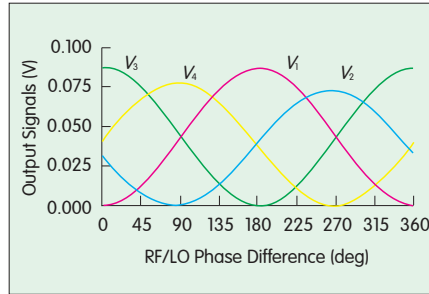
approach [6]:

$$I(t) = V_1(t) - V_2(t) \\ = K \cdot \alpha(t) \cdot |a|^2 \cdot \cos[\Delta\phi(t)] \quad (1)$$

$$Q(t) = V_3(t) - V_4(t) \\ = K \cdot \alpha(t) \cdot |a|^2 \cdot \sin[\Delta\phi(t)], \quad (2)$$

where V_1 , V_2 , V_3 , and V_4 are the six-port output detected signals, K is a constant, a is the amplitude of the LO signal, $\Delta\phi(t) = \phi_6(t) - \phi_5$ is the instantaneous phase difference, and $\alpha(t)$ is the instantaneous amplitude ratio between the RF and LO signals. Harmonic balance simulations are performed for several discrete frequency points over 4 GHz. The RF and LO input powers are set to 0 dBm, and the RF signal phase is swept over 360° . In practice, amplitude and phase imbalances are inherent because of design and fabrication constraints at 60 GHz. S-parameters in simulations can highlight such errors.

Fig. 7 shows the six-port output signals in relation to the phase difference between RF and LO signals. Theoretically, these signals must have equal amplitude and be shifted by 90° and its multipliers. Some imbalance-related errors, both in



▲ Figure 7. Harmonic balance analysis of the six-port at 63 GHz.

amplitude and phase, are observed. A differential approach is proposed in [6] and [7] that reduces these errors. Fig. 8 shows the percentage of six-port down-converter quadrature error.

Because of these inherent errors, I/Q signal phase difference is not exactly 90° and the shapes of demodulated constellations are distorted. However, for simple modulation schemes, that is, amplitude-shift keying (ASK), binary phase-shift keying (BPSK), and quadrature phase-shift keying (QPSK) recommended for low-cost transceivers, phase error of less than 5% is considered acceptable.

3 Proposed 60 GHz Transceiver

3.1 Single-Carrier Architecture

Recent studies have suggested that a V-band receiver based on six-port technology enables the design of compact and low-cost wireless millimeter-wave single-carrier communication receivers for future high-speed wireless communication

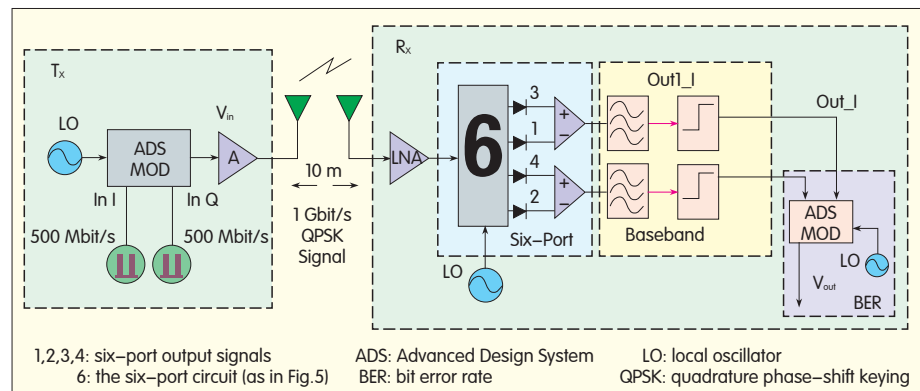
systems [6]–[9]. Millimeter-wave frequency conversion is performed using specific properties of the six-port circuit. This avoids the need for a costly conventional active mixer. A simplified six-port single carrier (SC) homodyne transceiver block diagram is shown in Fig. 9.

The transmission path is simulated by an ADS loss link based on the Friis model. The free-space loss at 62 GHz is around 88 dB and is calculated using the Friis transmission equation

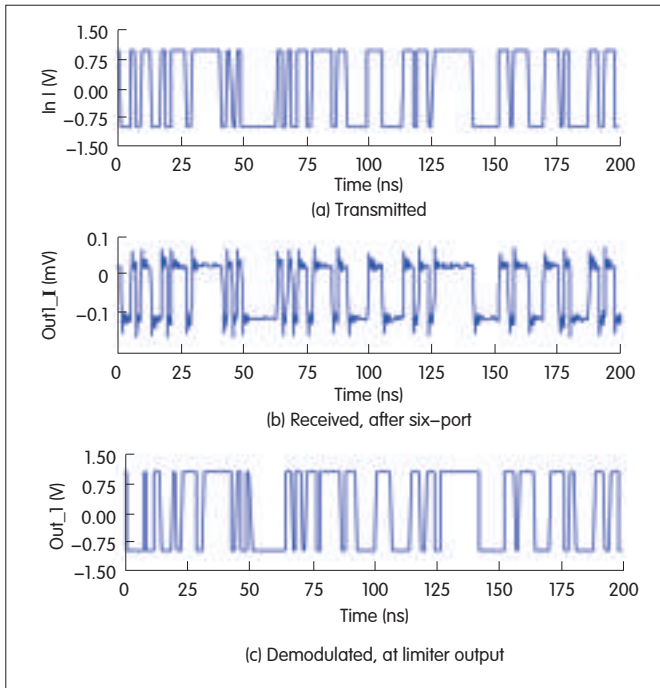
$$\frac{P_R}{P_T} = G_R G_T \left(\frac{\lambda}{4\pi R} \right)^2, \quad (3)$$

where P_R is the ratio of power received by the receiving antenna, P_T is the ratio of power input to the transmitting antenna; G_T and G_R are the antenna gains of the transmitting and receiving antennas, respectively; λ is the wavelength (around 5 mm for 62 GHz), and R is the distance (10 m). In the transmitter part, the parameters are set as follows: LO power = –25 dBm, amplifier gain (A) = 20 dB, and antenna transmitting gain (GT) = 10 dBi. These values are chosen in order to obtain transmitted signal power of 10 dBm (allowed by FCC for a V-band communications system). In the receiver, the antenna receiving gain is 10 dBi, the low-noise amplifier (LNA) gain is 20 dB, and the six-port input signal power is –38 dBm.

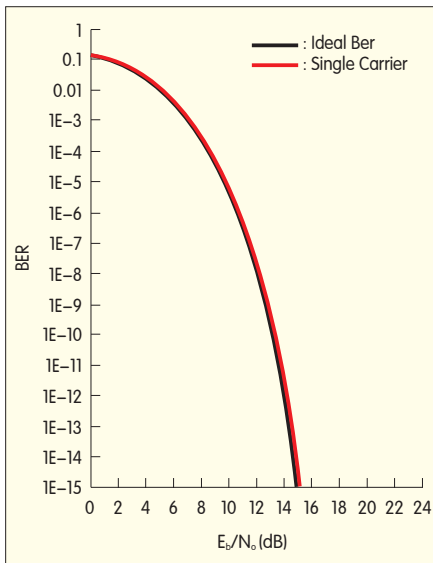
During the simulations, the operating frequency was set at 62 GHz, and the transmitted QPSK-modulated signals were pseudo-randomly generated by ADS with a symbol rate of 500 MS/s (communication data rate = 1 Gbit/s).



▲ Figure 9. Single carrier system block diagram.



◀ Fig. 10. Demodulation of 1 Gbit/s QPSK pseudo-random (I) bit sequence.



▲ Figure 11. BER for 1 Gbit/s QPSK signal.

By using the limiters in the last stage of the receiver, output square waves are generated (Fig. 10). For a bit sequence of 200 ns, the output demodulated (I) signals have the same bit sequence as those transmitted. The same conclusion is obtained for the (Q) signals. Fig. 11 shows the bit error rate (BER) variation in relation to energy per bit over spectral noise density (E_b/N_0) for the same distance of 10 m. The six-port receiver architecture using the single

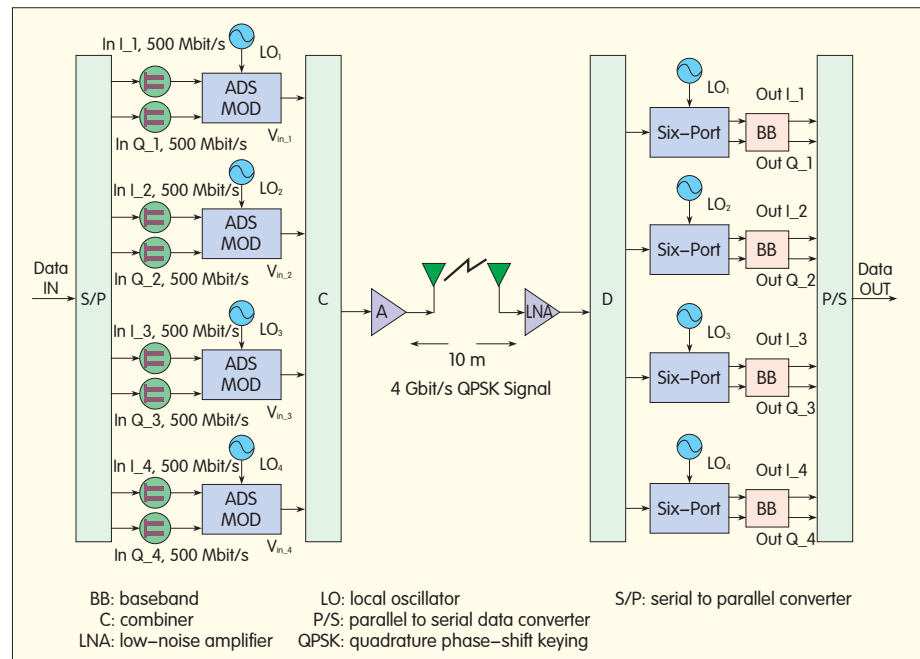
carrier scheme has excellent BER performance, very close to the theoretical ideal.

3.2 Multicarrier Architecture

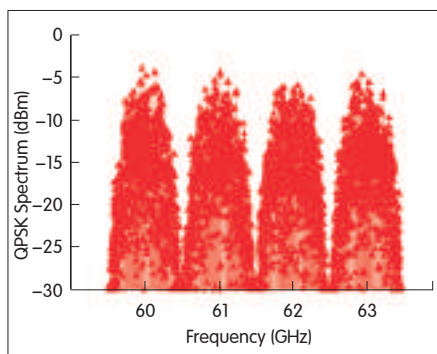
In [11], the same proposed six-port receiver using frequency division multiplexing (FDM) architecture is able

to transmit 2 Gbit/s up to 10 m with BER of 10^{-9} , as required for an uncoded HDTV wireless transmission in a home or office. By spacing the carriers, this data rate can be increased to 4 Gbit/s for a short-range communication of 10 m. The advantage of using FDM scheme is that no central timing synchronization is required for each subchannel. Each subchannel can operate independently. A serial to parallel converter (S/P) with 2×4 parallel outputs, 4 millimeter-wave LO, 4 quadrature modulators, and a millimeter-wave combiner (C) are used to generate the 4 carrier FDM signal. An envelope simulation is then carried out (Fig. 12).

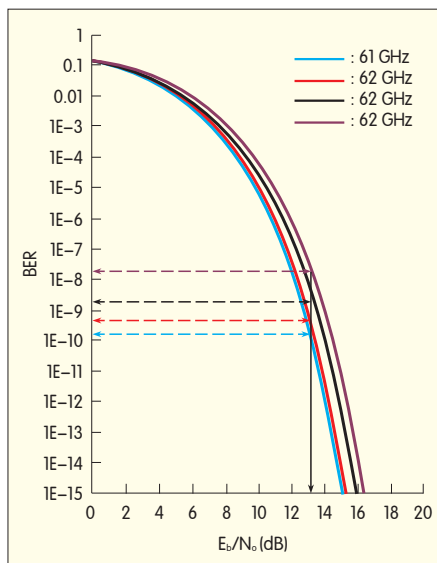
In the simulation, the power of each LO used in the transmitter part is -25 dBm, and their frequencies are 61, 62, 63, and 64 GHz. The received signal is amplified by an LNA (20 dB), split using a millimeter-wave power divider (D), and coherently demodulated using 4 six-port receivers (MP R1–R4) and 4 baseband (BB) circuits. Finally, a parallel to serial data converter (P/S) generates the output data stream. For each frequency, the same power balance as the single carrier system in Fig. 9 is kept. The symbol rate per carrier (SRC) is



▲ Figure 12. Multicarrier system block diagram.



▲ Figure 13. QPSK FDM spectrum at the transmitter.



▲ Figure 14. BER of QPSK signal (different subcarriers).

500 MHz, which provides an FDM signal at 4 Gbit/s. To cover the whole bandwidth of 4 GHz, and considering the ADS convergence properties, the step simulation is fixed at $1/(7 \times \text{SRC})$.

Fig. 13 shows the QPSK multicarrier transmitted signal spectrum. Between the 61–64 GHz frequency channels, there are spectral regions called guard bands that act as buffer zones to prevent interference between frequency subchannels. By taking into consideration the SRC, simulation step, and envelope simulation frequency, the guard bands are minimized as much as possible. Fig. 14 shows the BER values that correspond with each subcarrier. The BER curves are very close to each other, ranging from 10^{-10} to 10^{-7} for an E_b/N_0 of 13 dB. These variations are

considered tolerable for good quality wireless communications, which usually requires a BER of less than or equal to 10^{-9} .

Because of the uncorrelated subchannels and ADS convergence limitation, we calculate the BER of the whole system using an analytical approach. The BER average of the millimeter-wave multicarrier system is the sum of the BER related to each subchannel [12], [13] and is obtained using

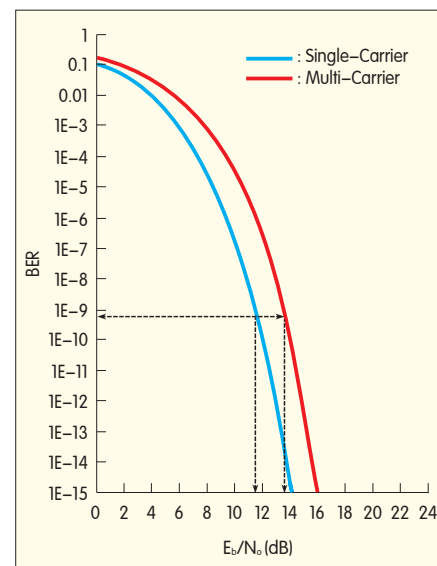
$$P_{sys} = \frac{1}{N \cdot \log_2 M} \cdot \sum_{i=1}^N (BER)_i, \quad (4)$$

where P_{sys} is the global error probability (or BER) of our system, N is the number of carriers used, and M is related to the modulation levels or number of bits per symbol (for QPSK, $M = 2$). $(BER)_i$ is the BER of each subcarrier.

Fig. 15 shows the average BER curve of the system as well as the BER corresponding to the single-carrier communication system. By using multicarrier modulation techniques, and for $BER = 10^{-9}$, E_b/N_0 should be incremented by about 2 dB, which is not a critical disadvantage. In the meanwhile, a high data rate of 4 Gbit/s is attained. This has interesting implications for the new generation of mobile radio communication systems.

4 Conclusion

In this paper, a V-band FDM wireless link based on six-port receiver architecture is proposed. I/Q signal phase error of less than 5% is caused by fabrication errors in the MHMIC coupler. But this error would not degrade some digital modulation schemes, such as QPSK, because of their robustness. The millimeter-wave receiver, based on the six-port junction fabricated on ceramic substrate, is able to transmit at 1 Gbit/s using single-carrier modulation. However, using multicarrier modulation based on frequency division multiplexing (4 subcarriers), 4 Gbit/s is reached. The proposed system allows the design of high-speed, high-performance, and low-cost wireless transceivers for



▲ Figure 15. BER results of QPSK multicarrier system.

future millimeter-wave systems. In the future, the entire six-port circuit will be fabricated and tested, and a full test will be done on the proposed FDM millimeter-wave architecture.

Acknowledgement

The authors gratefully acknowledge the financial support of the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT) and the support of the Centre de Recherche en Électronique Radiofréquence (CREER) of Montreal, funded by the FQRNT, for the MHMIC circuit fabrication.

References

- [1] P. Smulders, "Exploiting the 60 GHz band for wireless multimedia access: prospects and future directions," *IEEE Commun. Mag.*, vol. 40, no. 1, pp. 140–147, 2002.
- [2] D. Cabric, M. Chen, D. Sobel, S. Wang, J. Yang, and R. Brodersen, "Novel radio architectures for UWB, 60 GHz, and cognitive wireless systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2006, no. 2, p. 22, Apr. 2006.
- [3] C. Park and T. Rappaport, "Short-range wireless communications for next-generation networks: UWB, 60 GHz millimeter-wave WPAN, and ZigBee," *IEEE Wireless Commun.*, vol. 14, no. 4, pp. 70–78, 2007.
- [4] G. F. Engen, "The six-port reflectometer: An alternative network analyzer," *IEEE Trans. Microw. Theory Tech.*, vol. 25, no. 2, pp. 1077–1079, Dec. 1977.
- [5] T. Yacabe, F. Xiao, K. Iwamoto, F. Ghannouchi, K. Fujii, and H. Yabe, "Six-port based wave correlator with application to beam direction finding," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 2, pp. 377–380, Apr. 2001.
- [6] S.O. Tatu and E. Moldovan, "V-band multi-port

heterodyne receiver for high-speed communication systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2007, no. 1, p. 45, Jan 2007.

- [7] S.O. Tatu, E. Moldovan, Ke Wu, R.G. Bosisio, and T. A. Denidni, "Ka-band analog front-end for software-defined direct conversion receiver," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 9, pp. 2768–2776, Sept. 2005.
- [8] N. Khaddaj-Mallat, E. Moldovan, and S.O. Tatu, "Comparative demodulation results for six-port and conventional 60 GHz direct conversion receivers," *Progress In Electromagnetics Research*, vol. 84, pp. 437–449, 2008.
- [9] X. Xu, R. G. Bosisio, and Ke Wu, "Analysis and implementation of six-port software-defined radio receiver platform," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 7, pp. 2937–2943, Jul. 2006.
- [10] N. Khaddaj Mallat, E. Moldovan, S.O. Tatu, and Ke Wu, "Ultra-high-speed six-port frequency-division-multiplexing V-band transceiver," in *IEEE Radio and Wireless Symp. (RWS 2011)*, Phoenix, AZ, 2011.
- [11] S. O. Tatu, N. Khaddaj Mallat, and E. Moldovan, "Ultra wideband frequency division multiplexing millimeter-wave multi-port receiver analysis," in *2009 European Microw. Week, IEEE Conf.*, Rome, pp. 108–111.
- [12] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [13] Jun Lu, Tjeng Thieng Tjhong, Adachi, F., and Cheng Li Huang, "BER performance of OFDM-MDPSK system in frequency-selective Rician fading with diversity reception," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1216–1225, Jul. 2000.

Biographies

Nazih Khaddaj Mallat (nazih@ieee.org) received his bachelor's degree in electrical and computer engineering from the Lebanese University in 2000. He received his master's degree from Ecole Nationale Supérieure des Télécommunications de Bretagne (ENSTB), France, in 2003. He received his Ph.D. degree in telecommunications from the University of Quebec Institut National de la Recherche Scientifique (INRS) in 2010. He is a postdoctoral fellow at Ecole Polytechnique de Montreal. Dr. Khaddaj Mallat's main research interests are passive microwave/millimeter-wave circuit design, and telecommunication systems. He is the IEEE Montreal section chair in 2011. He has served at many IEEE conferences: EPC2007, SMC2007, EPEC2009, CNSR2010, MWP2010, FBW2011, CCECE2012 and IMS2012.

Emilia Moldovan (moldovan@emt.inrs.ca) received her B.Sc. degree in electrical engineering from the Polytechnic University of Cluj-Napoca, Romania, in 1980. She received her M.Sc.A. and Ph.D. degrees in electrical engineering from the École Polytechnique of Montréal, Canada, in 2001 and 2006. From 1982 to 1997, she was a telecommunication engineer with the Quality of Service Department, National Company of Telecommunications, Rom-Telecom, Romania. She is currently a research associate at the Institut National de la Recherche Scientifique (INRS), Canada. Her research interests include passive microwave/millimeter-wave circuit design, and telecommunication and radar systems.

Serioja O. Tatu (tatu@emt.inrs.ca) received his B.Sc. degree in radio engineering from the Polytechnic University, Bucharest, 1989. He received his M.Sc.A. and Ph.D. degrees in electrical engineering from the École Polytechnique de Montréal in 2001 and 2004. From 1989 to 1993, he was an RF engineer and head of the Telecommunications Laboratory at the National Company of Telecommunications, Rom-Telecom, Romania. From 1993 to 1997, he was a technical manager at the Telecommunication Laboratory. He is currently associate professor at the Institut National de la Recherche Scientifique (INRS). His current research interests are the millimeter-wave circuit design, hardware and software radio receivers, and radar systems.

Ke Wu (ke.wu@ieee.org) is professor of electrical engineering and Tier-I Canada Research Chair in RF and millimeter-wave engineering at the École Polytechnique de Montréal. He was director of the Poly-Grames Research Center and founding director of the Center for Radiofrequency Electronics Research of Quebec (funded by FRQNT). He has authored or coauthored more than 700 refereed papers and a number of books and book chapters. He holds numerous patents. Professor Wu's current research interests include substrate integrated circuits (SICs), antenna arrays, and development of low-cost RF and millimeter-wave transceivers and sensors for wireless systems and biomedical applications.

Roundup

ANTEL Selects ZTE for Uruguay National GPON Project

On September 19, 2011, ZTE Corporation, a publicly-listed global provider of telecommunications equipment and network solutions, announced it has been selected by the Administración Nacional de Telecomunicaciones (ANTEL) to assist in providing 300,000 subscribers in the country with a gigabit passive optical network (GPON). The tender was one of the largest ever put out in South America.

ANTEL is Uruguay's largest fixed-line operator and mobile carrier, with a 95.5 percent share of the nation's broadband market as of June 2010. Under the contract, ZTE will

help the operator deliver broadband speeds of 100 Mb/s.

The GPON will offer broadband under a range of FTTx scenarios, including FTTH, FTTB, backhaul and enterprise networks. ZTE will provide a next-generation, converged, full-service optical access platform called ZXA10 C300. The platform meets ANTEL's requirements and is upgradable to NG PON and WDM PON.

"ZTE has cooperated with ANTEL in a number of areas, including home gateways and cellphones," said ZTE Vice President Xu Ming. "With our understanding of the Uruguayan

telecom market and our experience deploying more than 100 million lines of xPON across the globe, our goal is to build this project into a model deployment for South America that is scalable worldwide."

ZTE was one of the first vendors to develop and commercialize optical access network products.

According to Ovum's "4Q10 Market Share—FTTx, DSL, and CMTS ports", ZTE ranks first among major GPON vendors in revenue growth. To date, 108 million lines of ZTE xPON products, over 55 million of which are broadband equipment, have been deployed worldwide. (ZTE Corporation)

Millimeter-Wave Heterodyne Six-Port Receiver: New Implementation and Demodulation Results

D. Hammou, E. Moldovan, and S.O. Tatu

(Institut National de la Recherche Scientifique-Énergie Matériaux et Télécommunications, Montréal, Québec H5A 1K6, Canada)

Abstract: This paper presents a new implementation of a millimeter-wave heterodyne receiver based on six-port technology. The six-port model is implemented in Advanced Design System (ADS) using S-parameter measurements for realistic advanced simulation of a short-range 60 GHz wireless link. Millimeter-wave frequency conversion is performed using a six-port down-converter. The second frequency conversion is performed using conventional means because of low IF. A comparison between the proposed receiver and a conventional balanced millimeter-wave mixer shows that the proposed receiver improves conversion loss and I/Q phase stability over the local oscillator (LO) and RF power ranges. The results of demodulating a V-band quadrature phase-shift keying (QPSK) signal at a high data rate of 100 Mb/s–1 Gb/s are discussed. The results of a bit error rate (BER) and error vector magnitude (EVM) analysis prove that the proposed architecture can be successfully used for wireless link transmission up to 10 m.

Keywords: millimeter wave; six-port; frequency conversion; heterodyne; front-end; wireless LAN

1 Introduction

Recent dramatic growth in wireless communication systems has caused microwave frequency bands to become overcrowded.

The millimeter-wave spectrum provides multi-Gb/s data transmission [1], [2] and is a possible solution to this problem.

Millimeter-wave frequencies enable the design of compact, low-cost wireless millimeter-wave communication front-ends. These can offer convenient terminal mobility and high-capacity channels.

This paper describes a millimeter-wave heterodyne receiver based on a new six-port architecture dedicated to high-speed wireless communication systems. Six-port technology is widely used in microwave applications for low-cost circuit characterizations [3] and has also been used as an unconventional technique

for performing frequency down-conversion [4], [8].

2 Millimeter-Wave Six-Port Circuit

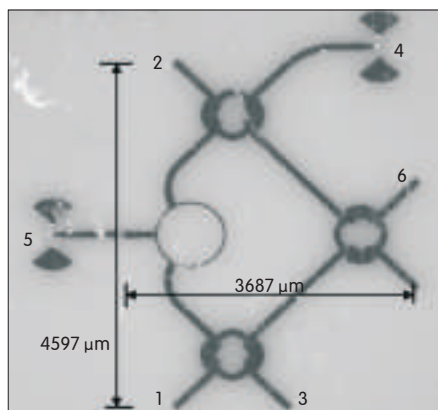
A number of six-port architectures have been described in literature. All of these architectures are functions of fabrication technologies and specific applications, such as down-converter, direct modulator, and reflection coefficient measurement.

The best-known six-port architecture for communication purposes comprises a 3 dB Wilkinson power divider and three 90° hybrid couplers. At high frequencies, the two branches of the Wilkinson power divider must be placed very close to each other to be connected to the 100 Ω resistor. A strong, undesirable mutual coupling between the output lines occurs when using a conventional Wilkinson power divider [9], [10].

Therefore, a second six-port is built with four 90° hybrid couplers to avoid the Wilkinson power divider. The power divider is implemented using one of the 90° hybrid couplers and a transmission line of 90°. This deteriorates the desired amplitude and phase-split ratio over a wide band and increases phase and amplitude imbalance. The S-parameter measurements for this architecture show that, for the coupled ports being considered, phase imbalance is less than 10% over the entire bandwidth [11]. However, the amplitude imbalance reaches 1.5 dB at higher frequencies.

In addition, S-parameter measurements for the ring power divider developed in [12] show that the circuit has a quasi-null amplitude imbalance of 0.25 dB and additional insertion loss of <0.5 dB from 60 GHz to 70 GHz.

In light of recent progress in power divider design, a new six-port



▲ Figure 1. Microphotograph of the proposed six-port circuit.

architecture is proposed. This circuit is constructed using three hybrid couplers ($H-90^\circ$) and a new ring power divider [13].

For full S-parameter characterization of a six-port circuit, 15 circuits prepared for two port measurements are needed. To reduce the cost of circuit fabrication, and regarding the symmetry of the six-port architecture, our approach is to evaluate the most important S-parameters from port 5. These parameters include transmission, output phase balance, and RF input isolation.

Five six-port circuits (5-1, 5-2, 5-3, 5-4, and 5-6), prepared for two port measurements, are integrated on a TPS superstrate with relative permittivity of 9.9 and thickness of 127 μm . Fig. 1 shows this circuit in S_{54} configuration.

S-parameter measurements are taken in the 60–65 GHz band because of the characteristics of the measuring equipment. Measurements show that the RF inputs at ports 5 and 6 are well matched; that is, results are better than -15 dB across the whole frequency band. Isolation between the two RF input ports is at least 15 dB and reaches 30 dB at the central frequency. High isolation is needed to avoid local oscillator (LO) leakage to the antenna.

The S-parameter measurements for transmission show <0.5 dB of additional loss, and the relative amplitude imbalance between requested ports does not exceed 8.33% across 60–65 GHz.

The S-parameter measurements for

phase performance between two requested ports show a relative phase imbalance of only 7.6%, related to the expected quadratic reference of 90° across 60–65 GHz.

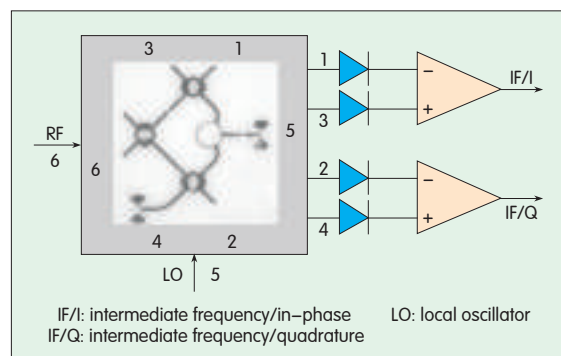
3 Mixer Topologies

Here, we describe a six-port millimeter-wave mixer (SPMM) that performs down-conversion and can improve mixing signals in terms of conversion loss versus RF signal and LO power sweep.

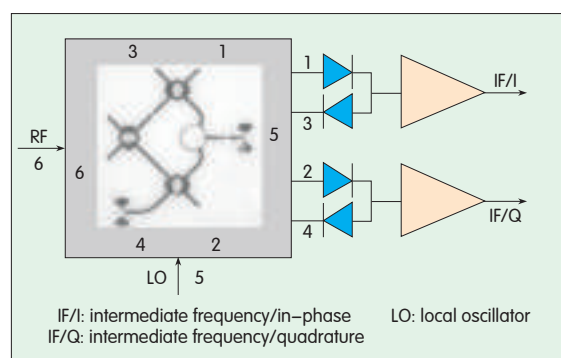
The proposed down-converter comprises the new six-port circuit as well as two pairs of parallel Schottky diodes that are connected to two differential amplifiers and act as power detectors (Fig. 2).

A conventional I/Q mixer uses couplers and two anti-parallel pairs of Schottky diodes acting as LO-driven switches (Fig. 3). For comparison with the proposed architecture, a six-port circuit is used instead of conventional mixer couplers. The intermediate frequency (IF) amplifiers have the same gain as the IF differential amplifiers of the proposed mixer. The block diagram for this architecture is shown in Fig. 3.

To estimate the conversion loss for both mixers and the stability of the I/Q phase, harmonic balance simulations are performed using Advanced Design System (ADS). Zero-bias Schottky diode (Spice model HSCH-9161) is



▲ Figure 2. SPMM using parallel diodes.



▲ Figure 3. Mixer using anti-parallel diodes.

used as a non-linear element in both topologies. First, we analyze the behaviour of the two topologies in relation to the LO drive in the range -15 – 15 dBm. RF power is fixed at -25 dBm, and the amplifier gain is fixed at 10 dB for both mixers.

Over the entire LO power range, the output I/Q power of the proposed mixer is better than that of a conventional mixer by 42 dBm (Fig. 4). The conversion loss of the proposed mixer is improved by 42 dBm. In addition, the LO driving power of the

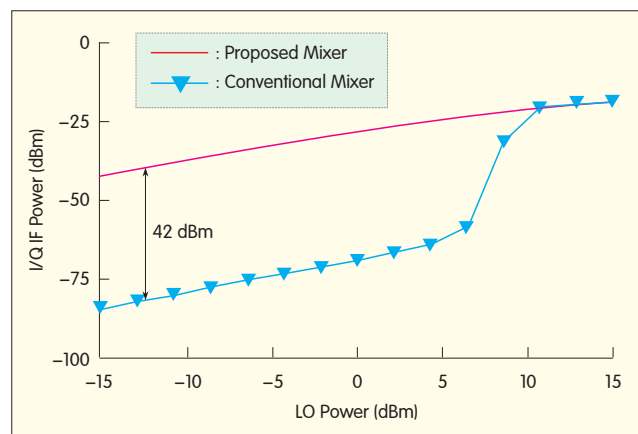


Figure 4. ▶
Simulation results:
conversion loss vs. LO.

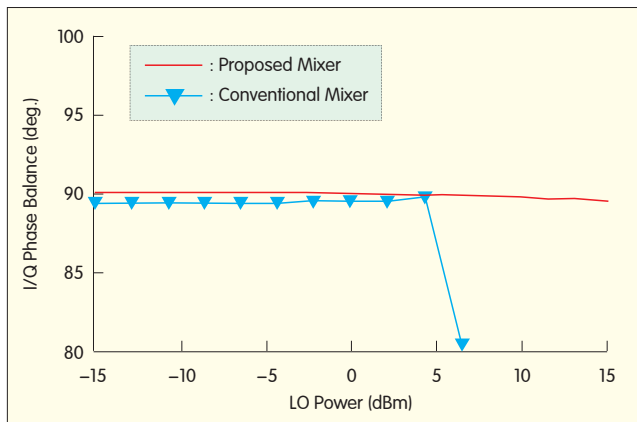


Figure 5. Simulation results: I/Q IF phase balance vs. LO.

quasi-conventional mixer should be at least 10 dBm to move it into saturation area.

The phase performances of both mixers are quasi-similar and close to the quadratic reference of 90° until LO power of 5 dBm, at which point I/Q phase performance of the conventional mixer deteriorates dramatically (Fig. 5). At an LO power of 6 dBm, phase performance is 10° . The proposed mixer has a phase disparity of only 0.5° over the entire LO drive range.

Second, we analyze the two topologies in relation to RF drive in the range -50 – 0 dBm. LO power is -10 dBm for the proposed mixer, and the sweep is -10 – 5 dBm for the quasi-conventional mixer. The amplifier gain is 10 dB for both mixers.

Fig. 6 shows the results of HB analysis in terms of I/Q IF power versus RF drive. The proposed mixer gives excellent results. When RF power is -50 dBm, I/Q IF power is approximately -60 dBm for the proposed mixer and -80 dBm for the quasi-conventional mixer with 5 dBm of LO power.

This represents a 20 dBm improvement in conversion loss with 30 times less LO power.

Keeping the same parameters for the mixers, we analyze I/Q phase (Fig. 7). The I/Q phase of the proposed mixer is almost the same as the quadratic reference of 90° from -50 – 0 dBm. The I/Q phase of the quasi-conventional mixer has quasi-similar performance for RF power over a limited range of -36 – 10 dBm.

Low-cost receivers need reduced

LO power and high I/Q phase stability. So the proposed mixer is an excellent candidate for the applications dealt with in the next section.

4 Receiver Architecture and Operating Principle

Heterodyne architecture is widely used in wireless communications

because it has many well-known advantages. However, the main problem in the millimeter-wave domain is the design of a high-quality low-cost mixer. To overcome this problem, the receiver uses the new SPMM described in the previous section instead of a conventional mixer to perform millimeter-wave down-conversion. Fig. 8 shows the proposed heterodyne receiver architecture.

In this section, we show that the first down-conversion from RF to IF, shown in Fig. 8(a), can be performed using specific properties of the SPMM.

The SPMM inputs are connected to a low-noise amplifier (LNA) and to the first (millimeter-wave) local oscillator (LO1), respectively. The intermediary frequency module (IFM) allows quadrature IF signals to be obtained using detected output six-port signals and differential amplifiers. The second local oscillator (LO2), two mixers, and two low-pass filters (LPF) are used to

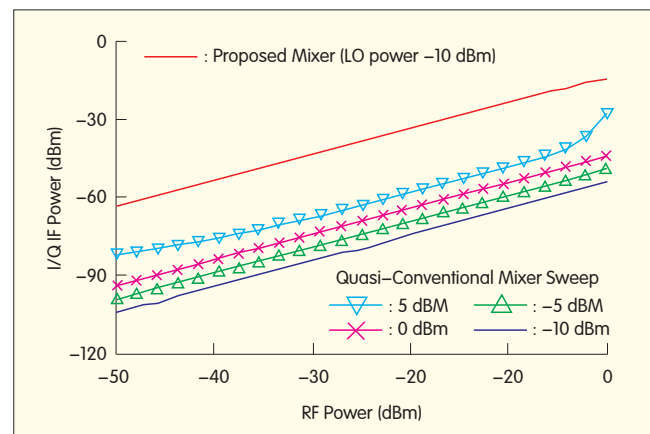


Figure 6. Simulation results: conversion loss vs. RF.

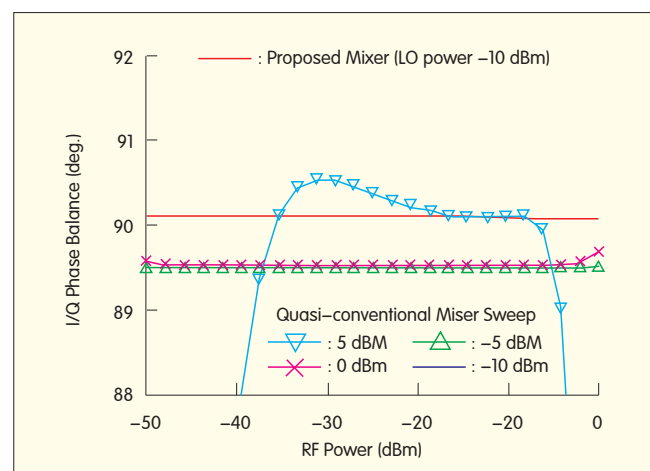


Figure 7. Simulation results: I/Q IF phase balance vs. RF.

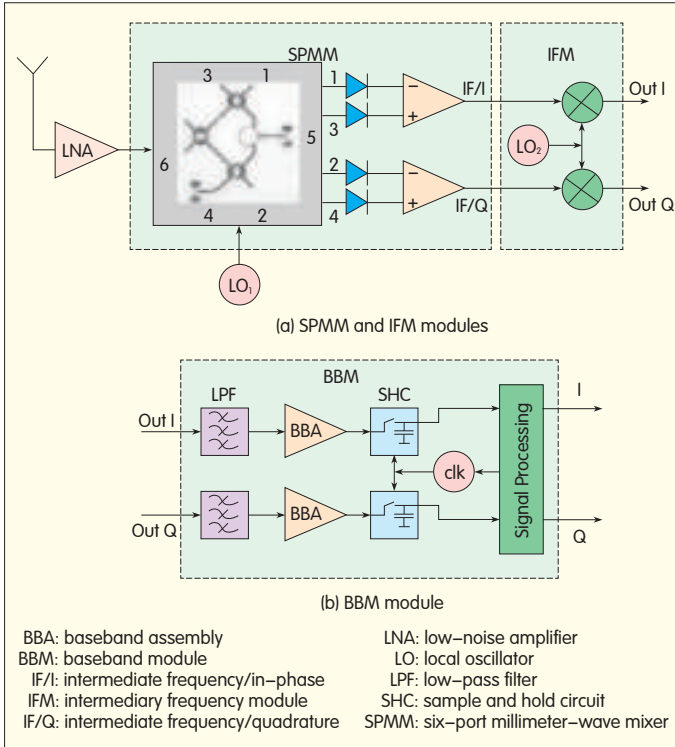


Figure 8. Block diagram of the heterodyne six-port receiver.

generate the primary baseband signals. The baseband module (BBM) amplifies these signals, and sample and hold circuits (SHCs) operating at the bit-rate frequency generate improved demodulated output signals. Signal processing techniques compensate for rotation of the demodulated constellation caused by instability of the LOs, especially LO1.

The operating principle of the six-port direct-conversion receiver is described in [14]. In this paper, we demonstrate that a six-port circuit can down-convert millimeter-wave modulated signals to intermediary frequency IF. The second frequency conversion, from IF to baseband, can be easily obtained by conventional means related to relative low frequency operation.

Generally, the output signals (b_i) of a multiport can be expressed with the dispersion parameters S_{ij} as

$$b_i = \sum_{j=1}^6 S_{ij} a_j, \quad i = 1 \text{ to } 6. \quad (1)$$

The scattering matrix of the proposed six-port circuit can be obtained using the scattering matrix of a 90° hybrid coupler and the power divider in Fig. 1.

$$[s] = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 & 0 & -j & j \\ 0 & 0 & 0 & 0 & 1 & j \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & -j & -1 \\ -j & 1 & 1 & -j & 0 & 0 \\ j & j & 1 & -1 & 0 & 0 \end{bmatrix} \quad (2)$$

It can be assumed there are two normalized wave inputs: a_5 from the LO and a_6 from the RF input (Fig. 8). These two normalized input waves have an amplitude ratio α , phase difference $\Delta\phi(t) = \phi_6(t) - \phi_5$, and frequency shift $\Delta\omega = \omega - \omega_0$. These are expressed as

$$a_5 = a \cdot e^{j(\omega_0 t + \phi_5)} \quad (3)$$

$$a_6 = \alpha \cdot a \cdot e^{j(\omega t + \phi_6(t))} = \alpha \cdot a_5 \cdot e^{j(\Delta\omega t + \Delta\phi(t))} \quad (4)$$

Assuming a perfect match $a_1 = a_2 = a_3 = a_4 = 0$, the four normalized output waves can be calculated using (1) and the six-port scattering matrix (2):

$$b_i = a_5 \cdot S_{5i} + a_6 \cdot S_{6i}, \quad i = 1 \text{ to } 4 \quad (5)$$

$$|b_i| = \frac{a}{2} \cdot \left| 1 + \alpha \cdot e^{j[\Delta\omega t + \Delta\phi(t) + (3-i)\frac{\pi}{2}]} \right|, \quad i = 1 \text{ to } 4. \quad (6)$$

To obtain the IF output signals, four power detectors are connected to the

multiport outputs. The output voltage of an ideal power detector is proportional to the square magnitude of the RF input signal:

$$v_i = K_i \cdot |b_i|^2 = K_i \cdot b_i \cdot b_i^*, \quad i = 1 \text{ to } 4 \quad (7)$$

Supposing identical power detectors are used, that is, $K_i = K$, for $i = 1$ to 4, the output voltages are

$$v_{1,3}(t) = K \frac{a^2}{4} \cdot \{1 + \alpha^2 \mp 2 \cdot \alpha \cdot \cos[\Delta\omega \cdot t + \Delta\phi(t)]\} \quad (8)$$

$$v_{2,4}(t) = K \frac{a^2}{4} \cdot \{1 + \alpha^2 \mp 2 \cdot \alpha \cdot \sin[\Delta\omega \cdot t + \Delta\phi(t)]\}. \quad (9)$$

The output voltages at ports 1 and 3, and 2 and 4 are phase opposites. Therefore, the quadrature output signals can be obtained using two differential amplifiers at the outputs of the SPMM stage:

$$v_{IF}^I(t) = A_{IF} \cdot [v_3(t) - v_1(t)] = \alpha \cdot K \cdot a^2 \cdot A_{IF} \cdot \cos[\Delta\omega \cdot t + \Delta\phi(t)] \quad (10)$$

$$v_{IF}^Q(t) = A_{IF} \cdot [v_4(t) - v_2(t)] = \alpha \cdot K \cdot a^2 \cdot A_{IF} \cdot \sin[\Delta\omega \cdot t + \Delta\phi(t)]. \quad (11)$$

After the second frequency conversion and low-pass filtering, the output I/Q signals are obtained:

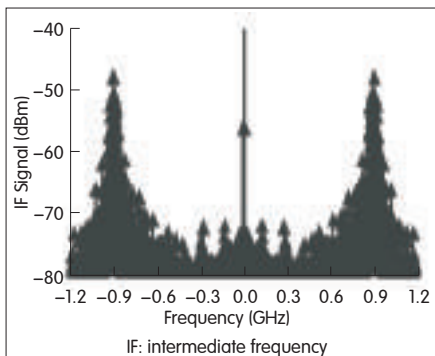
$$I(t) = \frac{1}{2} \cdot \alpha \cdot a^2 \cdot K \cdot A_{IF} \cdot A_{BB} \cdot \cos(\Delta\phi(t)) \quad (12)$$

$$Q(t) = \frac{1}{2} \cdot \alpha \cdot a^2 \cdot K \cdot A_{IF} \cdot A_{BB} \cdot \sin(\Delta\phi(t)). \quad (13)$$

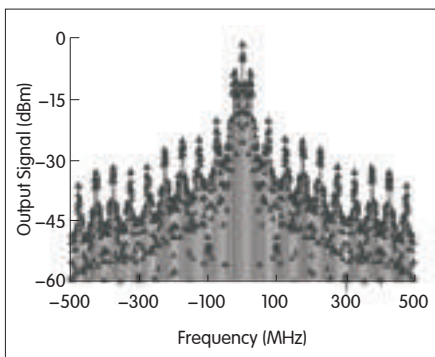
Equations (3)–(13) depend on the amplitude ratio α , phase difference $\Delta\phi(t)$, and frequency shift $\Delta\omega$. The proposed heterodyne receiver based on the new six-port architecture can demodulate arbitrary phase-shift keying (PSK), QPSK, M-ary phase-shift keying (MPSK), and M-ary quadrature amplitude modulation (M-QAM) schemes.

5 Demodulation Results for the Heterodyne Six-Port Receiver

To perform realistic simulations using ADS, and to evaluate demodulation performances, we used our V-band six-port circuit model, validated by S-parameter measurements in [13]. Envelope simulations were performed



▲ Figure 9. Typical spectrum of a quadrature IF signal (I or Q) for six-port receiver.



▲ Figure 10. Typical spectrum of a baseband quadrature signal (I or Q) for six-port receiver.

using pseudorandom QPSK-modulated signals. The transmitter uses a vector modulator and two pseudorandom sources (I and Q input signals) with two different voltage levels of ± 1 V. The carrier consists of a 61 GHz V-band signal that is modulated in the range 100 Mb/s–1 Gb/s. The propagation path simulation is performed using the Friis model, and the receiver incorporates the configuration in Fig. 1.

According to the Friis equation [15], the free space line-of-sight (LOS) attenuation is 88 dB for $d = 10$ m. In this system analysis, the antenna gains are set at 10 dBi. The LNA gain and noise figure (NF) are 21 dB and 3.8 dB, respectively. These are common values for today's 60 GHz integrated amplifiers. Limiters are used in the baseband stage of the receiver to obtain data output squared waves, and consequently, perfect demodulated constellations. In keeping with future high-speed requirements of the IEEE 802.15.3.c wireless standard, IF of

900 MHz was chosen. The LO is perfectly synchronized, an important advantage of this architecture compared to direct conversion. Perfect synchronization can be obtained by controlling the lower frequency LO from the digital processing block to IFM block (LO2).

Fig. 9 and Fig. 10 show typical spectrums of a quadrature IF signal (I or Q) centered at 900 MHz and a baseband quadrature signal (I or Q) obtained after the second down-conversion. The results show the performance of a six-port receiver designed to replace a conventional millimeter-wave mixer. The signals are pseudorandom QPSK-modulated at 100 Mb/s.

Fig. 11 shows the demodulation results of a pseudorandom QPSK bit sequence of 700 ns. The demodulated signal shapes are the same as the I/Q signals generated by the transmitter. The gray line represents the baseband signal before it reaches the SHC, which improves the demodulated signal shape.

To evaluate the wireless link quality for Gb/s pseudorandom

QPSK-modulated signals, a BER analysis is performed for the pseudorandom QPSK modulation data rate 100 Mb/s–1 Gb/s. BER results are presented as a function of E_b/N_0 , where E_b is the average energy of a modulated bit, and N_0 is the noise power spectral density (Fig. 11).

Simulations show that BER performance is less than 10^{-6} for a typical E_b/N_0 ratio of 14 dB in all cases. This is an excellent result if we consider that it corresponds to a millimeter-wave LO stability of 10^{-3} , compared to $10^{-6}/^\circ\text{C}$ of commercial oscillators.

Fig. 12 shows only 2.2 dB additional shift in E_b/N_0 . This is an excellent result if we consider that it corresponds to the tenfold increase in data rate from 100 Mb/s–1 Gb/s. To reach a BER of 10^{-9} (required for uncoded HDTV transmission), E_b/N_0 must increase a further 2.8 dB to reach 16 dB.

To calculate error vector, we consider the demodulation results for a pseudorandom QPSK bit sequence of 700 ns in Fig. 11. A comparison is made for data rate of 100 Mb/s–1 Gb/s.

Fig. 13 shows the ideal and QPSK

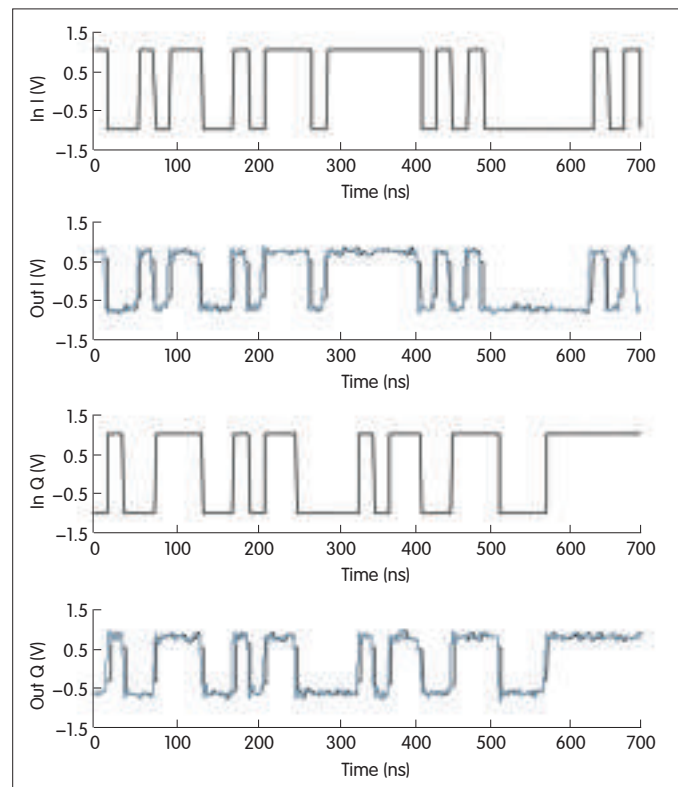


Figure 11. ▶ Demodulation results of 100 Mb/s pseudorandom bit sequence.

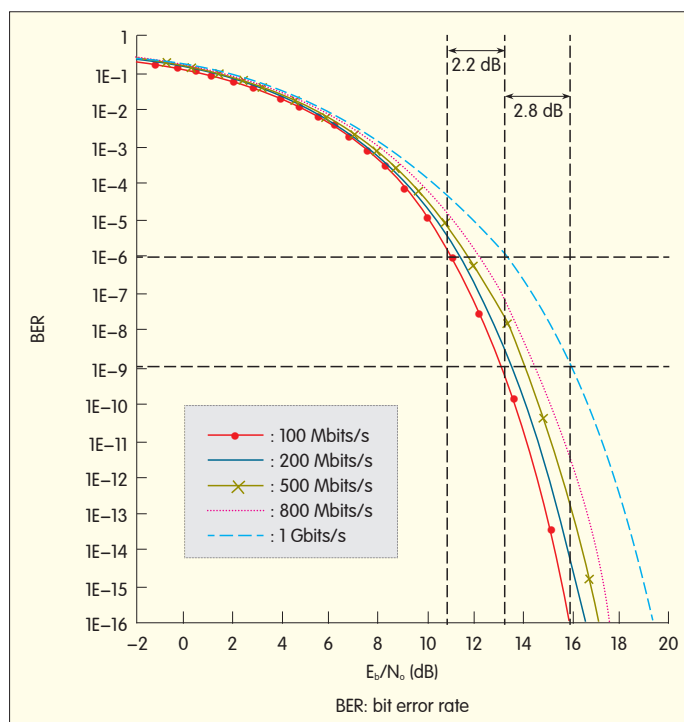


Figure 12. BER performance for a QPSK modulation data rate of 100 Mb/s–1 Gb/s.

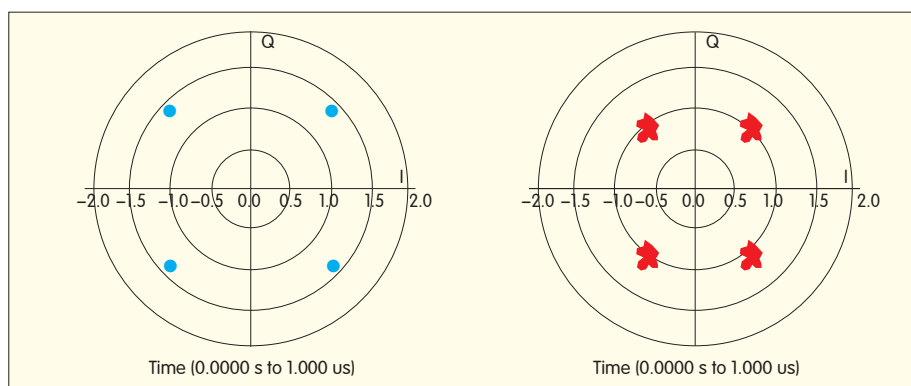


Figure 13. Pseudorandom QPSK demodulation constellation diagrams.

demodulation constellation diagrams at 100 Mb/s. The four clusters are very individualized, the output signals are square waves, and the output constellations are situated at the corners of a square that is centered at the origin.

However, constellation normalization must be enabled in order to effectively calculate the error vector magnitude (EVM) [16].

In the complex plane, we define the baseband signal $\Gamma_{in,n}$ and $\Gamma_{out,n}$ as

$$\Gamma_{in,n} = I_{in,n} + jQ_{in,n}, \text{ and } \Gamma_{out,n} = I_{out,n} + jQ_{out,n}. \quad (14)$$

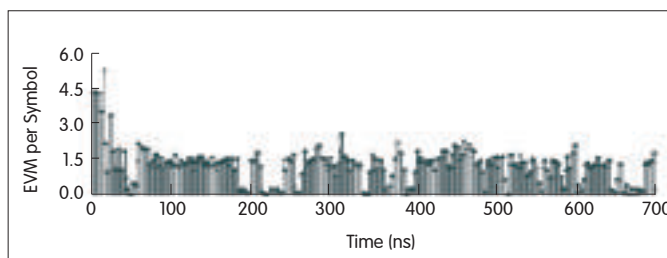
These vectors represent the transmitted and ideal baseband

signals. Constellation normalization is achieved by affecting the output constellation $\Gamma_{out,n}$ with a K coefficient. The symbols in the two diagrams in Fig. 13 become correlated [16]

$$\Gamma_{out,n} = K I_{out,n} + K j Q_{out,n}. \quad (15)$$

Thus, when symbols have been

Figure 14. EVM-per-symbol for a transmitted sequence at 1 Gb/s.



normalized, EVM is defined as the root-mean-square (RMS) value of the difference between a simulated symbols and ideal symbols:

$$EVM(\%) = \sqrt{\frac{1}{N} \sum_{n=1}^N |\Gamma_{in,n} - K \Gamma_{out,n}|^2} \cdot 100 \quad (16)$$

where N is the number of symbols.

Fig. 14 shows the EVM-per-symbol calculations for a pseudorandom QPSK bit sequence of 700 ns at 1 Gb/s. The instantaneous error vector magnitude is <1.5 over 700 ns.

The effective EVM for each QPSK data rate from 100 Mb/s–1 Gb/s corresponds to the mean of the EVM-per-symbol with the bit sequence length (number of symbols).

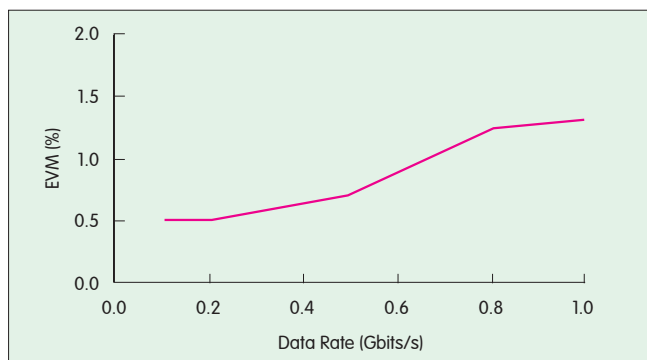
These results are shown in Fig. 15. As expected from the EVM-per-symbol calculations, the effective EVM is less than 2% for 100, 200, 500, 800 Mb/s and 1 Gb/s QPSK data rates.

These calculations confirm the BER results in the previous paragraph and give an indication of the capability of the six-port heterodyne receiver for the millimeter-wave wireless high data-rate communications systems.

6 Conclusion

A new millimeter-wave heterodyne six-port receiver is presented in this paper. To perform millimeter-wave frequency conversion, this receiver uses the specific properties of the new six-port circuit in Fig. 1. This avoids the need for a conventional I/Q mixer with anti-parallel diodes and conventional couplers or a costly active mixer. Realistic system simulations for a short-range 60 GHz wireless link were performed using a measurement-based six-port model and a QPSK-modulated signal of

D. Hammou, E. Moldovan, and S.O. Tatu



◀ Figure 15. Effective EVM vs. QPSK data rates at 0.1–1 Gb/s.

1 Gb/s. The demodulation results are validated by BER and EVM calculations. The proposed six-port heterodyne architecture enables the design of compact, low-cost wireless millimeter-wave receivers for future high-speed wireless communication systems.

Acknowledgement

The financial support of the National Science Engineering Research Council (NSERC) of Canada is gratefully accepted. The authors would like to express their gratitude to chief technologist, Jules Gauthier, Poly-Grames Research Center, École Polytechnique de Montréal, for his technical assistance.

References

- [1] Federal Communication Commission, *Amendment of Parts 2, 15 and 97 of the Commission's Rules to Permit Use of Radio Frequency above 40 GHz for New Radio Applications*, FCC 95–499, ET Docket No. 94–124, RM–8308, Dec. 15, 1995.
- [2] P. Smulders, "Exploiting the 60 GHz band for local wireless multimedia access: Prospects and future directions," *IEEE Commun. Mag.*, vol. 40, no. 1, pp. 140–147, Jan. 2002.
- [3] G. F. Engen, "The six-port reflectometer: An alternative network analyzer," *IEEE Trans. Microw. Theory Tech.*, vol. 25, no. 12, pp. 1075–1080, Dec. 1977.
- [4] J. Li, R. G. Bosisio, and Ke Wu, "Computer and measurement simulation of a new digital receiver operating directly at millimeter-wave frequencies," *IEEE Trans. Microw. Theory Tech.*, vol. 43, no. 12, pp. 2766–2772, Dec. 1995.
- [5] J. Li, R. G. Bosisio, and Ke Wu, "Dual-ton calibration of six-port junction and its application to the six-port direct digital millimetric receiver," *IEEE Trans. Microw. Theory Tech.*, vol. 44, no. 1, pp. 93–99, 1996.
- [6] S. O. Tatu, E. Moldovan, Ke Wu, and R. G. Bosisio, "A new direct millimeter-wave six-port receiver," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 12, pp. 2517–2522, Dec. 2001.
- [7] S. O. Tatu, E. Moldovan, G. Brehm, Ke Wu, and R. G. Bosisio, "Ka band direct digital receiver," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 11, pp. 2436–2442, Nov. 2002.
- [8] S. O. Tatu, E. Moldovan, Ke Wu, R. G. Bosisio, and T. Denidni, "Ka-band analog front-end for software defined direct conversion receiver," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 9, pp. 2768–2776, Sep. 2005.
- [9] D. Antsos, R. Crist, and L. Sukamto, "A novel wilkinson power divider with predictable performance at K and Ka-band," 1994 *IEEE MTT-S Digest*, pp. 907–910, Jun. 1994.
- [10] F. D. L. Peters, D. Hammou, S. O. Tatu, and T. A. Denidni, "Modified millimeter-wave wilkinson power divider for antenna feeding network," *Progress In Electromagnetics Research Letters*, vol. 17, pp. 11–18, 2010.
- [11] D. Hammou, E. Moldovan, Ke Wu, and S. O. Tatu, "60 GHz MHMIC six-port model analysis," *Microw. and Optical Tech. Lett.*, vol. 52, no. 9, Sep. 2010.
- [12] D. Hammou, E. Moldovan, and S. O. Tatu, "Novel ring millimeter wave power divider/combiner," in *Proc. 2011 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2011)*, Ontario, pp. 280–283.
- [13] D. Hammou, N. Hakem, E. Moldovan, N. Kandil, and S. O. Tatu, "New V-band MHMIC six-port architecture," accepted to *9th IEEE International NEWCAS Conference*, Bordeaux, Jun. 2011.
- [14] S. O. Tatu, E. Moldovan, Ke Wu, R. G. Bosisio, and T. Denidni, "Ka-band analog front-end for software defined direct conversion receiver," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 9, pp. 2768–2776, Sept. 2005.
- [15] D. M. Pozar, *Microwave and RF Design of Wireless Systems*. New York: John Wiley and Sons, 2001.
- [16] M. D. McKinley, K. A. Remley, M. Myslinski, J. S. Kenney, D. Schreurs, and B. Nauwelaers, "EVM Calculation for Broadband Modulate Signals," NIST PML, Gaithersburg, MD, Tech. Rep. 2005.

Biographies

Djilali Hammou (hammou@emt.inrs.ca) received his B.Sc. degree and M.Sc. degree in electrical engineering from the University of Technology, Oran, Algeria, in 1983 and 2002. He received his Ph.D. degree in telecommunications from the Institut National de la Recherche Scientifique–Énergie Matériaux et Télécommunications, Montréal in 2011. Dr. Hammou is currently a post-doctoral researcher at the Research Laboratory Télébec in Underground Communications, Quebec University, Val d'Or. His research interests include passive microwave/millimeter-wave circuit design, hardware and software radio receivers, radio propagation, and telecommunication systems.

Emilia Moldovan (moldovan@emt.inrs.ca) received her B.Sc. degree in electrical engineering from the Polytechnic University of Cluj–Napoca in 1980. She received her M.Sc.A. and Ph.D. degrees in electrical engineering from École Polytechnique de Montréal in 2001 and 2006. From 2006 to 2008, she was a post-doctoral researcher at the Institut National de la Recherche Scientifique–Énergie Matériaux et Télécommunications, Montréal. Dr. Moldovan is currently a research associate at that institute. Her research interests include passive microwave/millimeter-wave circuit design, and telecommunication, radar, and sensor systems.

Serioja Ovidiu Tatu (tatu@emt.inrs.ca) received his B.Sc. degree in radio engineering from Polytechnic University, Bucharest, in 1989. He received his M.Sc.A. and Ph.D. degrees in electrical engineering from the École Polytechnique de Montréal in 2001 and 2004. From 2004 to 2005, he was a post-doctoral researcher at the Institut National de la Recherche Scientifique–Énergie Matériaux et Télécommunications, Montréal. Dr. Tatu is now an associate professor at that institute. His current research interests include millimeter-wave circuit design, hardware and software radio receivers, and radar and sensor systems.

AD Index

A1 and Back Cover: ZTE Corporation

Security Service Technology for Mobile Networks

Abstract: As mobile networks become high speed and attain an all-IP structure, more services are possible. This brings about many new security requirements that traditional security programs cannot handle. This paper analyzes security threats and the needs of 3G/4G mobile networks, and then proposes a novel protection scheme for them based on their whole structure. In this scheme, a trusted computing environment is constructed on the mobile terminal side by combining software validity verification with access control. At the security management center, security services such as validity verification and integrity check are provided to mobile terminals. In this way, terminals and the network as a whole are secured to a much greater extent. This paper also highlights problems to be addressed in future research and development.

Keywords: mobile network security; security service; trusted computing; access control

Aiqun Hu
Tao Li
Mingfu Xue

(Information Security Research Center of
Southeast University, Nanjing 210096,
P. R. China)

1 Security Architecture and Threats in Mobile Telecommunication Systems

1.1 Security Mechanism Proposed by 3GPP

To secure 3G mobile telecommunication systems, 3GPP has developed a security architecture with five feature groups [1] in three strata, as shown in Fig. 1.

These five feature groups and their functions are:

Group 1: network access security. This provides users with secure access to 3G services and protects against attacks on the radio access link.

Group 2: network domain security. This enables nodes in the provider domain to securely exchange signaling

data and protects against attacks on the wireline network.

Group 3: user domain security. This secures access to mobile stations.

Group 4: application domain security. This enables applications in the user domain and provider domain to securely exchange messages.

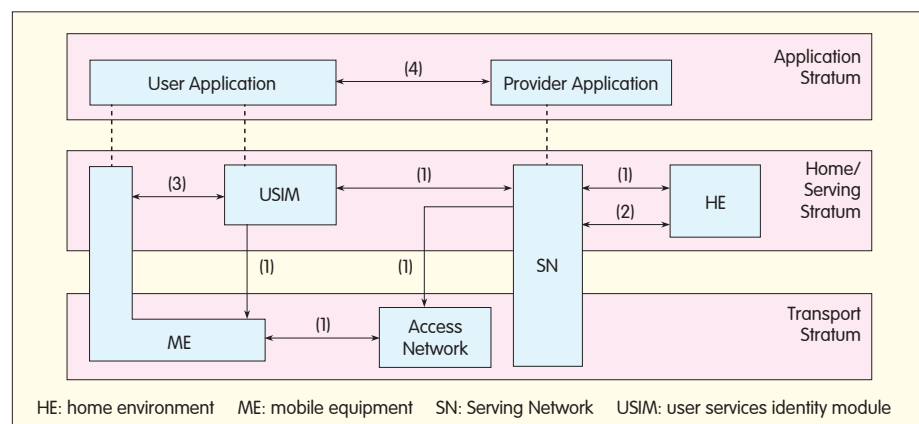
Group 5: visibility and configurability of security. This enables users to determine whether a security feature is in operation or not and whether the use and provision of services should

depend on the security feature.

In 3GPP's security architecture, the emphasis is on the network access security mechanism, including mutual authentication, universal terrestrial radio access network (UTRAN) ciphering, and integrity protection of signaling data. Network access security mechanisms mainly fall into three categories:

- identification by temporary identities such as Temporary Mobile Subscriber Identity (TMSI)
- identification by a permanent identity such as International Mobile Subscriber Identity (IMSI)
- authentication and key agreement (AKA).

Among these, AKA is important and



▲ Figure 1. 3G security architecture by 3GPP.

This work is funded by the National High-Technology Research and Development Program of China ("863" Program) under Grant No. 2009AA01Z427.

is a hot topic in research on 3G network security mechanisms. AKA mutually authenticates the mobile station and network, creating a new cipher key and integrity key. Other security mechanisms of 3G networks include data ciphering and integrity mechanisms. The data ciphering mechanism uses F8 ciphering algorithm to encrypt information between mobile equipment (ME) and the radio network controller (RNC). The data integrity mechanism applies F9 ciphering algorithm to authenticate data integrity and validity of a signaling message.

Protection measures such as mutual authentication, long secret keys, high-density ciphering, integrity algorithms, and signaling integrity protection mechanism are already used in 3G systems. A mechanism to protect core nodes in the telecommunication network is also used [2]. However, new services, open IP networks, and enhanced attack technologies still threaten the security of mobile networks.

1.2 Security Mechanism for 4G Networks

The security requirements of 4G networks fall roughly into four categories [3]:

- network access security. This provides access to 4G services that is secured against attacks on the radio access link.
- network area security. This enables nodes in the provider domain to securely exchange data and protects against attacks on the wired network and network entities.
- user area security. This enables secure access to ME/user services identity module (USIM) and provides a security environment in the ME/USIM.
- application security. This enables applications in the user and provider domains to securely exchange messages.

The security architecture of 4G networks therefore includes these four feature groups, and functions of these groups are similar to those in 3G networks.

In terms of security requirements, the difference between 3G and 4G networks is that in 4G networks, the integrity of the hardware, software, and operating system in the mobile platform on ME/USIM needs to be protected. A trusted computing environment for mobile entities is created. Only with a secure mobile platform can the security of user information be guaranteed. Research shows that the security of a mobile platform depends on mobile terminals themselves and also security management and services of the security servers in the mobile network.

1.3 Security Threats Faced by Mobile Telecommunication Networks

3G mobile telecommunication networks are developing into all-IP networks. As a result, they are facing more security threats from viruses, trojans, junk e-mails, spam SMS, and eavesdropping. These will also affect future 4G networks.

Unlike legacy mobile telecommunications, 3G mobile telecommunications is oriented toward network application services. The mobile terminal is both a platform for users to enjoy services and an Internet terminal. Security problems in mobile terminals have a great impact on the entire 3G network. This is reflected in two ways. First, high-speed 3G mobile phones access to the Internet significantly increases the amount of sensitive data being transmitted over the air interface. Second, viruses and baleful information are transferred from mobile phones to other terminals or nodes in the network. Currently, 3G specifications on terminal data security are incomplete. Except for user authentication, they do not provide effective protection for data coming from the user side. The openness of the Internet is also a great threat to mobile terminal security. Easy access to the Internet increases the possibility that mobile phones will pick up viruses or be hacked. Security problems arising from network openness are a great challenge to 3G networks. With a large number of protocols being adopted, 3G systems are prone to certain security

risks [4].

2 Research on the Security of Mobile Telecommunication Networks

To address threats to 3G and 4G mobile telecommunication networks, much research has been conducted on the security of mobile telecommunication networks.

2.1 Technologies for Architecture Security

Researchers from Anglia Ruskin University in the U.K. have proposed a new hybrid approach using symmetric/asymmetric authentication protocol for future mobile networks [5]. To overcome the defects of existing authentication schemes of 3G mobile systems—including leaking of mobile terminal identities and high update overhead of temporary identities—this approach uses a secure authentication mechanism. With this mechanism, the number of messages between authentication entities is reduced from five to four in the initial authentication procedure. The subsequent authentication procedure only contains two message exchanges. The number of messages between the mobile terminal and authentication center is reduced, and information congestion and slow processing at the authentication center is avoided. Moreover, authentication time delay, call setup time, and signaling traffic are minimized. This mechanism can also be used to counter network attacks such as replay attacks and guessing attacks. The mechanism meets the security requirements of 3G communication systems.

Researchers at the University of Illinois at Urbana-Champaign have proposed a lightweight, component-based, reconfigurable security mechanism [6]. This mechanism applies Tiny SESAME architecture to mobile networks in order to improve authentication and IP-based multimedia security services. This enhances the security functions of mobile devices. SESAME architecture,

first introduced in Europe, is designed for distributed systems. The mechanism adopts asymmetric encryption and a special attribute certificate to verify security attributes such as identity and privilege.

Researchers at the University of Florida have proposed an enhanced AKA protocol (called AP-AKA) so that 3GPP AKA protocol is not vulnerable to false base station attacks [7]. In a false base station attack, user traffic is redirected from one network to another, and authentication vectors are used to impersonate other networks. AP-AKA protocol resists such attacks.

In [8], an IPSec-based virtual private network (VPN) is used to secure multimedia services over 3G networks. The IP multimedia subsystem (IMS) defined by 3GPP is an open network and vulnerable to attack. So security and quality of service (QoS) have become major issues. The IPSec-based VPN solution provides end-to-end security for real-time multimedia transmission while guaranteeing QoS.

AKA protocol-related problems in 3G mobile networks—including authentication vector attacks—were analyzed in [9]. For emerging 4G telecommunication systems, the AKA protocol is enhanced with Secure Sockets Layer (SSL)/Transport Layer Security (TLS) protocol, which has been proven secure and effective.

Security challenges in 4G systems were studied in [10], and the X.805 standard was used to analyze the Y-Comm framework of 4G systems. Y-Comm is a 4G security architecture with an integrated security module and has a specific security model for protecting data, servers, and users.

The AKA protocol used by 3GPP System Architecture Evolution (SAE) release 8 was analyzed in [11], and outstanding security problems were detailed. Security defects in the AKA protocol include user identity disclosure, compromised authentication vectors, and shared key leakage. In [11], a novel 3GPP SAE AKA protocol is proposed. A public key cryptosystem is used to encrypt user identities and authentication vectors in

the network domain, and random numbers are used for public keys and local authentication.

Researchers at Beijing University of Posts and Telecommunications (BUPT) studied the security of 3GPP AKA and analyzed four attacks that affect the protocol [12]. They proposed a public cryptology-based AKA protocol that can be applied to two cases: location update and location immovability. Results showed that this protocol is secure. Compared with existing protocols, this protocol improves security to a certain degree.

Researchers at Tongji University of China proposed a terminal system scheme for establishing an IPSec VPN connection in a mobile network [13]. This system uses the Network Driver Interface Specification (NDIS) intermediate driver to traverse the firewall and ensure normal transmission of IPSec data packets. The system also uses secure intelligent cards to store X.509 certificates for identity verification. This prevents unauthorized users.

2.2 Technologies for Terminal Security

The terminal is the source of data and storage and is where most attacks are initiated. If each terminal in a mobile system is authenticated and authorized and its operations comply with security policies, security of the entire network system can be guaranteed [14]. Therefore, research on terminal security is attracting greater attention. Terminal security can be divided into two types: protection by antivirus and trusted terminal-based security.

A trusted computing based security architecture for 4G mobile terminals was proposed in [3]. Based on Trusted Mobile Platform (TMP) and Public Key Infrastructure (PKI), this architecture provides a robust platform for users to access sensitive services and data in 4G systems. A hybrid AKA scheme is also introduced in [3].

It was suggested in [15] that antivirus software be installed on the terminal; and on the network side, bypass detectors and filters be added to scan and kill viruses that threaten intelligent terminals.

In [16], multilayer security control means were suggested to be taken from network access layer to application service layer. On the terminal side, security applications were developed; and on the network side, strategies for controlling secure network access and application service access were applied. These security control means work with security servers to guarantee security in mobile networks.

Both [14] and [17] advocate a trusted computing-based terminal security architecture so that there is consistency between trusted terminals and trusted network as well as security in the system. In [14], trusted computing based architectures and their effectiveness in terminal security protection was summarized. Future research directions were also given. However, [14] does not take into account system efficiency in actual situations nor does it show how to set up a unified security architecture within a network.

In short, the two approaches to terminal security are important for security in mobile networks. The first approach focuses on securing the air interface, and the second approach prevents viruses from invading the terminal. The first approach is relatively mature, while the second has attracted great attention in the past two years because of the emergence of mobile viruses. This paper focuses on the second approach.

3 Service-Based Security Architecture for Mobile Networks

As previously discussed, there are two main methods of protecting mobile terminals from viruses: scanning terminals and killing viruses, and creating a trusted computing environment for terminals. Neither is independent of security service architecture. In the first method, an updatable virus library and virus-killing service center is needed in the mobile network so that mobile terminals are provided with periodic and online antivirus services. In the second

method, a trust check is needed for setting up and maintaining the trusted environment for terminals. Unlike computer users, mobile users often lack technical knowledge. So security in mobile networks is hard to guarantee if there is no security service.

Supposing identity authentication, data integrity, and confidentiality are guaranteed between the mobile terminals and the access network, security threats to mobile terminals will come mainly from the Internet, similar to a computer accessing the Internet. People expect that a trusted terminal can resist attacks from the Internet. If a terminal is trusted, its performance should satisfy expectations, and there should be no unlicensed software on the terminal. In other words, all software installed or executed on the terminal must be licensed by the security server. In this way, a virus on the Internet will not be transplanted to the terminal.

3.1 Trusted Service-Based Security Architecture

We propose a trusted service-based security architecture in which a mobile trusted module (MTM) [18] is added to the mobile terminal. This module is independent and secure, has computing capability, and can communicate securely with the security service provider (SSP). It calculates the integrity of all software installed on a mobile terminal and reports it to the SSP. It also checks the validity of software that is to be installed and executed on a mobile terminal. If the software is not authorized by the SSP, the module forbids its installation or execution. SSP is also added to the mobile network. The main function of the SSP is to provide software validity to the mobile terminal. Before software is installed or run on a mobile terminal, the software provider (SWP) must provide a validation certificate for the software issued by the SSP. That is, the software must hold the digital certificate issued by the SSP. Fig. 2 illustrates the trusted service-based mobile network security architecture.

In Fig. 2, the SSP server is connected to the access network (AN) server. To establish this architecture, only small

changes need to be made to the existing system. Once a mobile terminal is authenticated and granted access to the network, the SSP checks its integrity. If the integrity of the terminal's software is found to be compromised, the terminal may have a virus. In this case, the terminal is forbidden to access the network in order to avoid the viruses spreading to other terminals. If the integrity has not been compromised, the SSP monitors the installation and running processes of the software and provides the mobile terminal with dynamic security services.

3.2 Trusted Computing Environment for Mobile Terminals

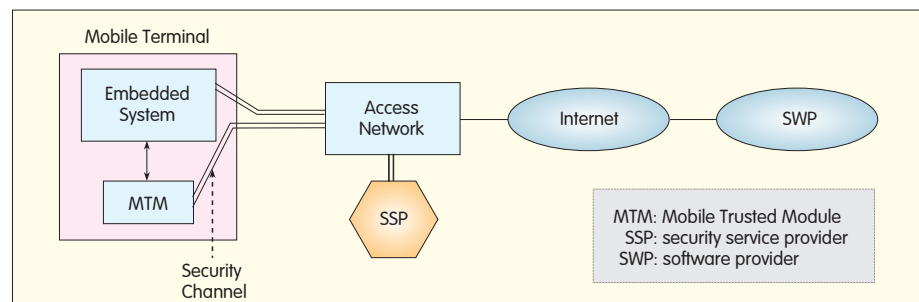
The key to an efficient service-based security architecture is a trusted computing environment for mobile terminals. For the purpose of this paper, the mobile terminal has two states: startup and started. A trusted computing environment established in the startup state is called static trusted environment, and that established in the started state is called dynamic trusted environment.

The method of creating a static trusted environment has already been widely discussed, and here we only intend to give a brief description. Once the power to a mobile terminal has been switched on, the trusted codes fixed in the trusted module are run. There are hardware features to ensure the codes cannot be modified. First, the trusted codes check the integrity of the system's loader codes. If the loader codes are complete, the trusted codes hand control over to the loader codes. The loader codes then check the integrity of the kernel of the operating system. If the kernel is complete, it is

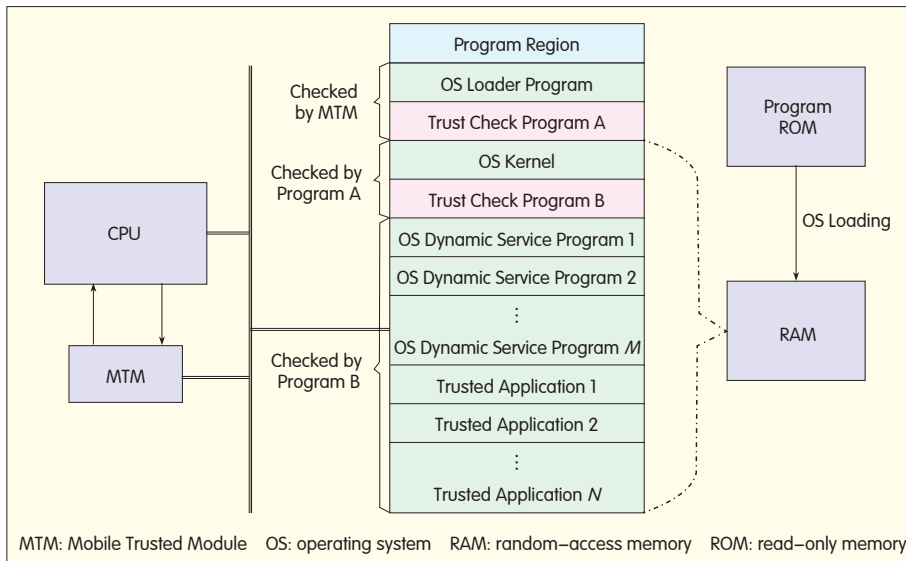
loaded. Next, the OS kernel checks the integrity of other parts of the OS. If all parts pass the integrity check, the OS is enabled. Finally, the integrity of upper-layer applications is checked. The terminal cannot be used until the integrity of all applications has been verified. If any integrity check fails in any of the above steps, the old configuration must be restored or the system must be re-installed. If the chain of trust is transferred away from trusted codes; that is, from the OS to applications by means of integrity check, a static trusted environment is established for the entire system. Fig. 3 illustrates the resource structure when the system starts. The trust check program is attached to the end of each program and is responsible for checking the integrity of the program that follows.

As shown in Fig. 3, to ensure system security, the OS kernel and trust check program B cannot be damaged after the static trust chain has been established. These two programs need to be protected by the MTM because after the system has started, the kernel program runs and trust check program B dynamically checks all applications that are to be run (or gathers integrity information of mobile terminals for SSP).

Once a system is started, protection relies on the dynamic trusted mechanism because the system's integrity changes dynamically with applications that are enabled by users. For example, when a browser is opened, the browser software is executed. The system has to run browser-related programs, so its integrity changes. If integrity information of the system cannot be updated correctly and in a timely



▲ Figure 2. Trusted service-based security architecture for mobile networks.



▲ Figure 3. System resource structure in startup state.

manner, malicious software may have an opportunity to damage the system. Dynamic integrity computing consumes computing resources. This is unavoidable. Therefore, implementing security over a running system is still a difficult problem for trusted computing. Trusted computing should

- periodically determine the integrity of a program or resources in a specified area and report it to the SSP
- check the validity of the program to be installed, and if the program is found invalid, forbid its installation
- check the validity of the program to be executed, and if the program is found to be invalid, forbid its execution
- ensure the security mechanism itself is secure
- ensure the security mechanism does not significantly decrease system efficiency.

It must first be determined whether the running system software or software resources have been modified. A common method for determining the integrity of a segment of codes is to compute the abstract value of the codes and then compare it with the existing abstract value (provided by the SSP). Another method is to select a random number of code bits and compute the abstract value. If the codes are long, the second method is advantageous.

When a program is to be executed on

a mobile terminal, its integrity and validity must be checked. Verifying the validity of a program is often done by checking whether the program is licensed by the SSP. A validity certificate is provided by the SSP. The SSP signs contracts with SWPs and saves version and abstract information of legal software in the SSP server so that mobile terminals can query it. This process is illustrated in Fig. 4.

To ensure the security mechanism itself is secure, the key is to protect the trust check program B (Fig. 3) from being damaged. This often requires special hardware. An effective protection method is to use a hardware circuit to ensure the address space of trust check program B is not re-written, unless the writing is authorized by the

MTM. Such a hardware circuit can be integrated into the MTM.

3.3 Security Server

The two main functions of a security server are:

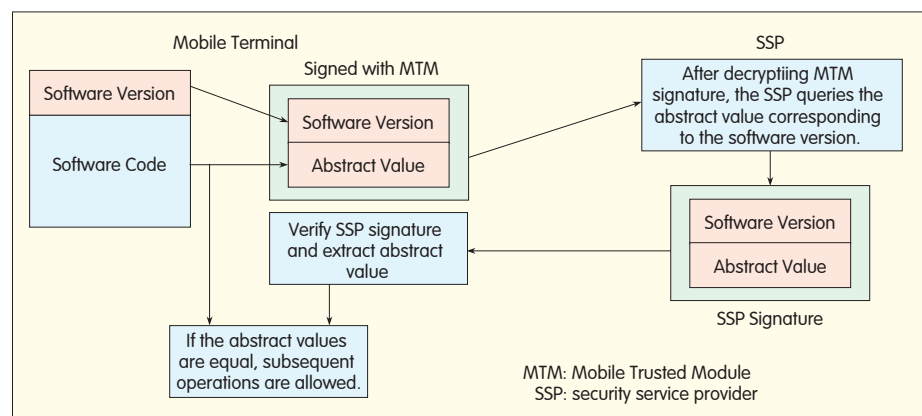
- to check the integrity of software on the mobile terminal. If integrity has been compromised, the terminal may have a virus and it is not allowed to access the network. If the mobile terminal were to access the network, it may spread a virus to other terminals or network devices.

- to provide validity query service for the mobile terminal. When a mobile terminal wants to install or execute software, it first queries the validity of the software in the local MTM. If no result is returned, it sends a query request to the security server. Upon receiving the request, the security server verifies the identity of the terminal and returns a result.

Other functions of the security server include interacting with SWPs (to audit the security of software and gather validity information); interacting with the authentication, authorization and accounting (AAA) server of the mobile network operator (to authenticate identity and to bill); and interacting with the access network server of the operation network (to control access based on the integrity of mobile terminals).

4 Future Research Direction

With network architecture and transmission protocols being mature,



▲ Figure 4. Integrity check and validity verification process of software.

the focus of future mobile security will be on terminals. Existing embedded platforms and operating systems on mobile terminals were not originally designed for mobile telecommunication, so their security problems are more serious than others. Integrated program and data space in the embedded OS makes it difficult to protect critical programs. An area of interest in mobile terminal security is the development of an embedded system architecture that protects critical programs.

Domain separation technology such as TrustZone [19], is a good way to implement read/write protection in critical locations in the memory. Once the OS and applications are loaded in the memory, the technology monitors the memory address accessed by the CPU and controls read/write operations for those critical locations. If the read/write operation is controlled with hardware, the impact on the system's running efficiency is almost negligible. Domain separation is an effective approach to real-time protection of the system. However, it requires the assistance of the embedded OS to provide, for instance, complete address information managed by the memory. For the sake of security, the architecture of the embedded OS must be studied and improved.

The key to software platform security lies in the OS. Access control is an effective means; but so far, it lacks theoretical support and security measurement methods. This is also a problem that trusted computing technologies have to handle. In research on OS security, the focus is often on how to construct a security model using theoretical knowledge.

Determining the integrity of system software and applications is critical in constructing a secure architecture. But this requires a great deal of computing. It is most pragmatic to design an effective integrity measurement algorithm with small computing burden. For example, the computing burden dramatically decreases if only bits of information are extracted from software codes and an abstract value is computed. How to extract such

information and how to synchronize with the verification entity requires further study.

5 Conclusions

It is expected that mobile phones will become genuine Internet terminals in the near future. A mobile phone can be used in the same way as a PC for e-commerce, e-mails, and mobile wallet. However, if mobile phone security problems cannot be solved, mobile phones will be a bottleneck for these applications.

Security in mobile networks should be solved from the perspective of the entire network. The traditional approach, where users assume responsibility for antivirus functions, should be replaced by security services provided by the network operator. Network operators have advantages over individual users in terms of technology, facilities, and management and can more effectively guarantee mobile network security.

Before a good solution to mobile network security is worked out, security events will occur from time to time. Whether an operator can provide high quality security service will be a critical factor for their commercial success. Service-based security solutions will become mainstream in the development of information security technologies for future networks.

References

- [1] Min Lei, Hai Bi, and Zhengjin Feng, "Security architecture and mechanism of third generation mobile communication," in *2002 IEEE Region 10 Conf. Comput., Commun. Control and Power Eng. (TENCON'02)*, Beijing, vol. 2, pp. 813–816.
- [2] H. Yang, F. Ricciato, S. Lu, and L. Zhang, "Securing a wireless world," in *Proc. IEEE*, vol. 94, no. 2, pp. 442–454, 2006.
- [3] Y. Zheng, D. He, W. Yu, and X. Tang, "Trusted computing-based security architecture for 4G mobile networks," in *6th Int. Conf. Parallel and Distr. Comput. Appl. and Tech. (PDCAT'05)*, Dalian, 2005, pp. 251–255.
- [4] Q. Dai and Y. Wang, "Network security ushering in the era of 3G mobile communication," *China Inform. Security*, no. 5, pp. 34–39, 2010.
- [5] M. Al-Fayoumi, S. Nashwan, S. Yousef, and A. R. Alzoubaidi, "A new hybrid approach of symmetric/asymmetric authentication protocol for future mobile networks," in *3rd IEEE Int. Conf. Wireless and Mobile Comput., Networking and Commun. (WIMOB'07)*, White Plains, 2007, p. 29.
- [6] J. Al-Muhtadi, D. Mickunas, and R. Campbell, "A lightweight reconfigurable security mechanism for 3G/4G mobile devices," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 60–65, 2002.
- [7] M. Zhang and Y. Fang, "Security analysis and enhancements of 3GPP authentication and key agreement protocol," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 734–743, 2005.
- [8] W. B. Diab and S. Prism, "VPN solution for securing voice over third generation networks," in *2nd Int. Conf. Internet Multimedia Services, Architecture and Appl. (IMSAA'08)*, Bangalore, pp. 6.
- [9] G. Kambourakis, A. Rouskas, and S. Gritzalis, "Using SSL/TLS in authentication and key agreement procedures of future mobile networks," in *4th Int. Workshop on Mobile and Wireless Commun. Networks (MWCN'02)*, Stockholm, 2002, pp. 152–156.
- [10] M. Aiash, G. Mapp, A. Lasebae, and R. Phan, "Providing security in 4G systems: unveiling the challenges," in *6th Adv. Int. Conf. on Telecommun. (AICT'10)*, Barcelona, 2010, pp. 439–445.
- [11] Y. Deng, H. Fu, X. Xie, J. Zhou, Y. Zhang, and J. Shi, "A novel 3GPP SAE authentication and key agreement protocol," in *2009 IEEE Int. Conf. on Network Infrastructure and Digital Content (IC-NIDC'09)*, Beijing, pp. 557–561.
- [12] F. Lu, K. Zheng, X. Niu, Y. Yang, and Zhongxian Li, "Security analysis of 3GPP authentication and key agreement protocol," *Journal of Software*, vol. 21, no. 7, pp. 1768–1782, 2010.
- [13] X. Wang, "Research on cellphone security system based on IPsec VPN," *Network and Computer Security*, no. 1, pp. 52–54, 2009.
- [14] W. Liu, J. Hu, Y. Fang, and Changxiang Chen, "Research and Development on the secure architecture of terminal based on trusted computing," *Computer Science*, vol. 34, no. 10, pp. 257–269, 2007.
- [15] H. Lv, Q. Chen, and X. Wu, "Research on development and countermeasures of virus in smart phone," *Information Security and Communications Privacy*, vol. 30, no. 1, pp. 80–82, 2008.
- [16] D. Huang and Z. Sang, "Analysis of anti-virus method in mobile communication networks," *Study on Optical Communications*, no. 2, pp. 39–43, 2008.
- [17] Zhenqiang Wu and J. Ma, "Research on TPM-based mobile Internet trusted architecture," *Network Security Technology and Application*, no. 11, pp. 18–20, 2007.
- [18] NTT, DoCoMo, IBM, and Intel Corp. (2005, Jun). "Trusted mobile platform: hardware architecture description," [Online]. Available: <http://xml.coverpages.org/TMP-HWAdv10.pdf>
- [19] T. R. Halfhill, "ARM Dons Armor – TrustZone security extensions strengthen ARMv6 architecture," Microprocessor Report. [Online]. Available: <http://www.mdronline.com/mpr/p/2003/0825/173401.pdf>

Biographies

Aiqun Hu (aqhu@seu.edu.cn) is director of the Information Security Research Centre of Southeast University, China. He is also a professor and doctoral advisor. His research interests include wireless networks and their security technologies.

Tao Li (leotao1984@hotmail.com) is studying for his doctoral degree at the School of Information Science and Engineering of Southeast University, China. His research direction is trusted computing technologies for mobile networks.

Mingfu Xue (57233806@163.com) is studying for his doctoral degree at the School of Information Science and Engineering of Southeast University, China. His research direction is trusted computing technologies for mobile networks.

Research on LTE Network Coverage Planning

Jun Gu and Ren Sheng

(ZTE Corporation, Shenzhen 518004, P. R. China)

Abstract: When deploying an LTE network, coverage planning is critical to reduce construction costs and ensure network quality. This paper considers actual network planning requirements and combines theory with simulation analysis to study LTE wireless access link and network characteristics. A theory for LTE cellular coverage planning and application methods is proposed that lays the basic foundation for LTE cellular networks.

Keywords: coverage design; geometry factor; power control; interference margin

factors that affect its coverage performance, including resource sharing mode and system interference characteristics. Traditional coverage planning and link budget schemes are far from adequate for LTE [4]. When planning the coverage of an LTE system, it is first necessary to analyze the potential requirements of LTE network construction and dissect the technical and network features. Then, a coverage architecture can be worked out that is feasible for LTE network construction and that can be improved constantly. Improving spectral efficiency is the key to creating a competitive edge with LTE. This paper focuses on spectral efficiency as well as LTE network design in the context of same-frequency networking.

1 LTE Network Planning Process and Coverage Planning Strategy

For frequency-division duplexing (FDD) LTE, network planning is similar to that of 2G and 3G networks. The planning process involves five stages: requirement collection and analysis, coverage and capacity design, site selection, simulation, and report preparation. Of these, coverage and capacity design is the core of the entire process. For this design, network characteristics need to be thoroughly analyzed, and network size needs to be evaluated based on the actual demands of customers. This paper analyzes coverage planning methods for LTE system.

For FDD LTE systems, the main goal of coverage planning is to estimate the coverage distance of an eNodeB with parameter settings based on actual cell edge coverage requirements in order to meet network size requirements. Planning strategies for FDD LTE system

With increasing demand for wireless services, the capacity of 2G and 3G networks is reaching saturation point. To deal with the shortage of mobile network resources, 3GPP has proposed Long Term Evolution (LTE), and this is now developing from theory into practice. With powerful service bearing capability, efficient resource utilization, low network construction and operation costs, and flexible network deployment, LTE has won favor from leading operators around the world [1].

Mature LTE standards are necessary for promoting the industry. With 3GPP freezing Release 8 in 2008, LTE standards are now in a mature state. Telecom equipment manufacturers are now conducting R&D on LTE products, and deployment of pilot offices and related tests are progressing well. Despite substantial progress in LTE research, complexity and flexibility in LTE systems brings uncertainty. The telecom industry is still exploring system features, network construction approaches, and optimization strategies. In terms of LTE network

planning, the lack of systematic theory and application solutions is still a problem.

LTE systems adopt advanced radio link technologies in the air interface, including Orthogonal frequency-division Multiplexing (OFDM), multiple-input multiple-output (MIMO), adaptive modulation and coding (AMC), and hybrid automatic repeat reQuest (HARQ). It also applies radio resource configuration algorithms—such as dynamic scheduling and inter-cell interference coordination (ICIC)—as well as power control to increase efficiency and flexibility in air interface resource configuration [2],[3]. From the perspective of LTE network design, these technologies and algorithms improve network performance but increase system complexity as well. To work out an efficient, reliable network coverage planning scheme for LTE, it is necessary to study the system's technical features comprehensively using system theories, simulations, and tests. Compared with 2G and 3G networks, LTE has quite different core

coverage can be divided into three kinds according to application scenario and requirements:

- network size estimation based on uplink edge rate requirement. This strategy is mainly applied in coverage scenarios where only the uplink edge rate is limited. The uplink coverage radius is calculated using the uplink rate and related link budget parameters. Then the downlink edge rate is predicted from the uplink coverage radius.

- network size estimation based on downlink edge rate requirement. This strategy is mainly applied in coverage scenarios where only the downlink edge rate is limited. The downlink coverage radius is first calculated using the downlink rate and related link budget parameters. Then, the uplink edge rate is predicted from the downlink coverage radius.

- network size estimation based on uplink and downlink edge rate requirements. This strategy is mainly applied in coverage scenarios where both uplink and downlink edge rates are limited. The uplink and downlink coverage radiuses are calculated using the uplink and downlink rates and related link budget parameters. Then, the limited coverage radius is obtained by comparing the uplink coverage radius with downlink coverage radius.

In network planning, a proper coverage planning strategy is selected according to actual requirements of the application scenario, and it must be able to flexibly handle problems that occur.

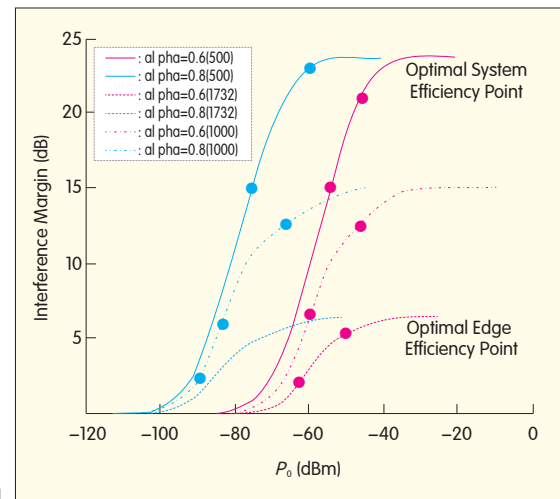
2 Key Technologies for LTE Uplink Coverage Planning

The key issue for coverage planning in LTE is how to work out the coverage based on the edge service rate requirements in the uplink or downlink. In some special scenarios or services, the coverage performance of the control channel must also be taken into account. Here we mainly discuss coverage planning in scenarios where the service channel is limited. Suppose the service rate is given, network coverage planning requires analysis of

technical characteristics of the links and system.

Research on uplink coverage planning technologies is mainly focused on links and system. In the LTE uplink, single carrier frequency division multiple access (SC-FDMA) is introduced as the multiple access scheme, and users within a cell are orthogonal. Most interference comes from active users of neighboring cells. Consequently, the selection of uplink power control policies directly impacts the inter-cell interference mode and interference density [5],[6]. In LTE uplink coverage design, interference margin is critical for network planning. Its value is mainly dependent on power control technique and application scenario. To obtain a feasible reference setting for the interference margin in uplink coverage planning, it is necessary to study the interference in different contexts by using system-level simulation.

Depending on the focus of network construction, uplink interference features may be affected by power control strategies or other factors. For LTE systems, the direct measurement for uplink interference features is average interference over thermal (IoT) level. The distribution of uplink IoT depends on actual application scenarios and uplink power control parameters. The power control scheme in LTE uplink includes open loop power control and closed loop power control. Normally, open loop power control determines the interference mode of the system, and closed loop power control is used to adjust the system parameters in response to changes in service or interference during network operation. Open loop power control meets specific network design requirements by setting suitable power control parameters P_0 and α . Different combinations of these parameter settings bring about different network coverages and capacities. To work out a network planning scheme for a specific scenario, it is necessary to analyze the settings of the parameters in various application scenarios in order



▲ Figure 1. Relationship between mean IoT and P_0 .

to find out which settings meet the requirements. It is also necessary to study the system's IoT as well as the uplink interference margins in different parameter settings. Analyses show that different parameter settings lead to differentiated system performance indexes and interference features. In actual planning, it is necessary to select proper parameter settings according to system coverage and capacity requirements. The relation between mean IoT and the parameter P_0 is shown in Fig.1.

In the early stage of LTE network construction, design is focused on coverage. Power control parameters are used to reduce interference and achieve maximum coverage. Because the target load varies from one network to another, the network interference level under different loads, as shown in Fig. 2, should be considered as a reference for coverage planning.

At the link level, uplink coverage design is primarily concerned with optimal configuration of user bandwidths at given data rates. Special edge data rate requirements can be met by assigning users different bandwidths. However, this may result in different coverage performance. By studying channel capacity and analyzing link-level simulation results it can be seen that optimizing bandwidth configuration at a given data rate enhances coverage. In uplink coverage design, the following should be taken

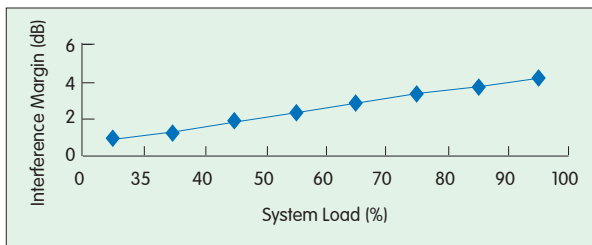


Figure 2.
Uplink interference margin
versus system load.

into consideration and analyzed: link performance in link simulations and actual system tests, power efficiency of the terminals, and system performance at given data rates and with different user bandwidth allocations. From these analyses, optimal uplink bandwidth configuration can be obtained for different service rates, and good coverage can be achieved. Fig.3 shows the relationship between data rate and optimal uplink bandwidth configuration.

With the above system- and link-level analyses, the two core elements for uplink budget can be determined: interference margin and uplink transmission bandwidth (the corresponding modulation and coding scheme (MCS) and signal to interference plus noise ratio (SINR) should be worked out with link simulation). Then, the maximum allowed path loss (MAPL) of the uplink at a given edge rate can be calculated based on traditional link budget and calculation methods.

3 Key Technologies for LTE Downlink Coverage Planning

As in the uplink, coverage planning in the downlink also involves the links and system. In researching link technologies, simulations with different link settings, such as MCS and bandwidth, should be performed to determine the link quality (carrier to interference ratio (CIR)) requirements in different channel environments and at different service rates. The simulation results are then used as the basis for coverage planning. In system-level research, system simulations are performed to determine how received signal strength, interference density, interference margin, and CIR change with the application scenario and

coverage range. The purpose of these analyses is to understand the limitations of traditional link budget schemes (based on interference margin) and to find out the relationship between signals and interference. In this way, existing coverage planning methods can be improved.

Interference in the downlink is affected by networking mode (interference coordination method) and system load, and it varies with the cell radius. The critical parameter of link budget, the interference margin, also changes with the cell radius, as shown in Fig. 4. Because of the strong correlation between interference margin and cell radius, the traditional method whereby the cell radius is calculated using a given interference margin is no longer applicable. Hence, new, relatively stable intermediate parameters have to be worked out for cell coverage analysis.

Analysis of a large number of systems shows the geometry factor (GF), with its unique characteristics, is an ideal bridge for downlink budget. GF is defined as the ratio of the signal to interference plus noise. Under full load

and with same frequency networking—that is, frequency reuse factor is 1—the GF's cumulative distribution functions (CDFs) in the case of different cell radiuses show that the distribution of GFs almost overlap under different coverage radiuses, as shown in Fig. 5. This characteristic of GF makes it a stable intermediate parameter for LTE downlink budget.

In network design, GF corresponding to a coverage area of 95% is often selected as the reference for coverage planning. Simulation under conditions of same frequency networking and full load shows the GF is 3 dB, as shown in Fig. 6. In an actual network, the design objectives are different and GF varies with the system load. It is therefore necessary to work out the GF reference values for different loads. Once the GF reference value is obtained, a definite mathematical relationship between the GF and SINR can be established for downlink coverage analysis. Specifically, in actual link budget, the SINR required by the downlink edge is determined by specific requirements. Then the minimum received signal strength required by the edge is calculated. Finally, the MAPL is worked out from the transmit power of the eNodeB. The formulae for these calculations are as follows:

$$SINR_{\text{require}} = \frac{S_{\text{require}}}{1+N} = \frac{S_{\text{require}}}{GF \times S_{\text{require}} + N} \quad (1)$$

$$S_{\text{require}} = \frac{SINR_{\text{require}} \times N}{1 - GF \times SINR_{\text{require}}} \quad (2)$$

Figure 3. ▶
Optimal uplink bandwidth
configuration at given data
rates.

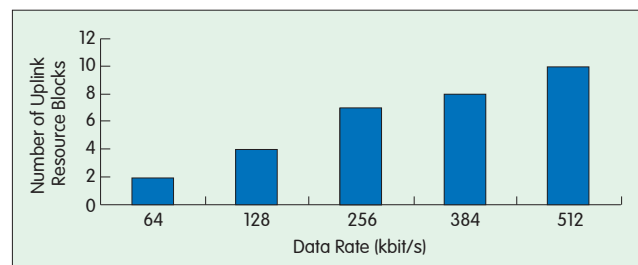
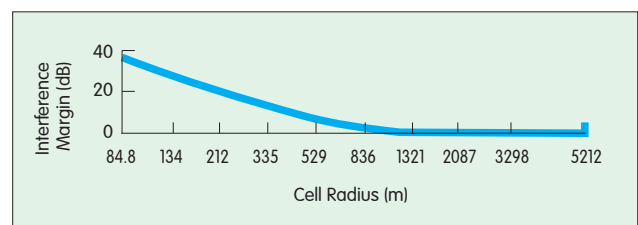


Figure 4. ▶
Downlink interference
margin versus cell radius.



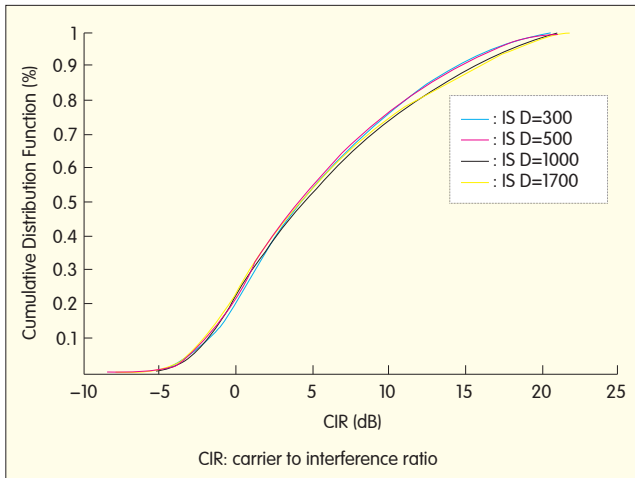


Figure 5.
Cumulative distribution
function of downlink CIR.

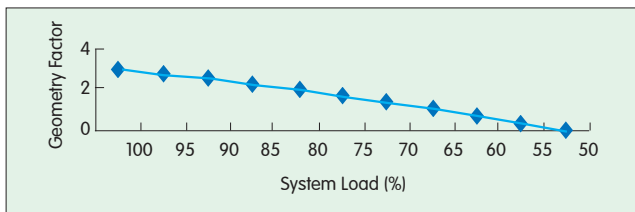


Figure 6.
Relation between geometry
factor and system load.

$$MAPL = P - S_{\text{require}} - sh_margin - Loss + Gain \quad (3)$$

where $SINR_{\text{require}}$ is the required SINR, S_{require} is the minimum receiving power required by the receiving end, P is eNodeB's transmit power, sh_margin is shadow margin, $Loss$ is the loss of all devices including feeders, and $Gain$ is the gain of all devices including antennas.

4 Conclusions

The openness and flexibility of LTE

networks brings great challenges to network design. On the whole, research on LTE networking is still exploratory. This paper introduces the LTE network planning and requirements and analyzes key technologies for coverage planning and system features. Based on these analyses, it gives approaches to LTE network planning for different coverage requirements. These approaches provide guidance for converting commercial demands into technical demands. Moreover, this paper presents overall technical

approaches to uplink and downlink budget of LTE systems and analyzes the settings of critical parameters and their applications. Thus, it lays the basic foundations for LTE network design.

References

- [1] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*, John Wiley, New York: 2009.
- [2] *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Overall Description*, 3GPP TS36.300, 2008.
- [3] *Evolved Universal Terrestrial Radio Access (E-UTRA): Physical Layer Procedures*, 3GPP TS 36.213, 2008.
- [4] Syed Abdul-Basit, "Dimensioning of LTE network: description of models, tools, coverage and capacity estimation of 3GPP LTE radio interface," M.S. thesis, Dept. Elect. Commun. Eng., Helsinki Uni. Tech., Finland, 2009.
- [5] N.J. Quintero, "Advanced Power Control for UTRAN LTE Uplink," M.S. thesis, Dept. Elect. Syst., Aalborg University, June 2008.
- [6] B. Muhammad, "Closed Loop Power Control for LTE Uplink," M.S. thesis, School Eng., Blekinge Inst. Tech., Sweden, 2008.

Biographies

Jun Gu (gu. jun@zte.com.cn) graduated from Beijing University of Posts and Telecommunications and is now an engineer at ZTE Corporation. He has long been engaged in researching key technologies for 2G and B3G systems. He is currently focused on network planning and optimization research of LTE FDD systems. He has published two academic papers, has been granted one international patent, and has been granted five patents in China.

Ren Sheng (shengren@zte.com.cn) graduated from Hefei University of Technology and is now an engineer and LTE FDD product director at ZTE Corporation. He has long been engaged in mobile network planning and optimization. At present, he is working on service planning for LTE FDD products.

Roundup

ZTE Brings International Mobile Hotspot to Sprint Customers

On September 14, 2011, ZTE Corporation announced the availability of the Sprint International Mobile Hotspot by ZTE. The device is the first international hotspot available from Sprint.

The Sprint International Mobile Hotspot allows users to access the Internet through as many as five Wi-Fi enabled devices, including laptops and music players. Casual customers

can increase their access to information and entertainment, and business customers can improve working efficiency while away from the office. The solution is particularly beneficial for customers who travel internationally because the hotspot provides Internet access via both GSM and CDMA networks and includes multiple country adapters.

"ZTE continues to expand its

presence in the U.S. telecommunications market by introducing functional and accessible mobile devices," said Lixin Cheng, CEO of ZTE USA and president of ZTE North American Region. "We are privileged to partner with Sprint to develop their first international hotspot and enhance their customers' mobile experience."

(ZTE Corporation)

Cloud Computing in Mobile Communication Networks

Xinzhi Ouyang

(ZTE Corporation, Shenzhen 518004, P. R. China)

Abstract: Cloud computing makes computing power universally available and provides flexibility in resource acquisition. It allows for scalable provision of services and more reasonable use of resources. This article considers cloud service deployment and virtualization from the perspective of mobile operators. A solution is proposed that allows mobile operators to maximize profits with minimal investment.

Keywords: resource sharing; service deploy; virtualization

1 New Requirements in the Communication Industry

With the development of 3G networks, more service platforms are being deployed by carriers. In addition to the widely-used WAP/WEB gateway and Short Message Center, Multimedia Messaging Center, new service application platforms will emerge as services are developed. Regardless of service functions and office size, an independent construction mode is used for current service platforms.

For service products of different companies, modules such as billing management module, user management module, configuration management module, maintenance management module, and log/report module are similar. Only the core service processing modules are different. Modules can be integrated

using specific methods. For the core processing part of each service, the service logics are different, and the process is complicated. Therefore, capabilities cannot be shared in the service layer. The dispersive construction mode of multiple services creates barriers for the rapid development of the mobile communication industry. The causes are as follows [1]–[4]:

- The types of purchased software and hardware used by each service platform are different. The service interruptions and system faults caused by the purchased parts are hard to control or prevent.
- The maintenance methods provided by different service platforms are different. Carriers have to train many technicians for system maintenance. This training increases maintenance costs.
- The service platforms are

constructed independently. The loads in different regions and service processing capabilities are not equal. Use of the constructed hardware resources is low.

Only virtualization technology does not provide service-level flexibility to invoke and control. The following describes a total solution to service dispatching and virtualization. In this solution:

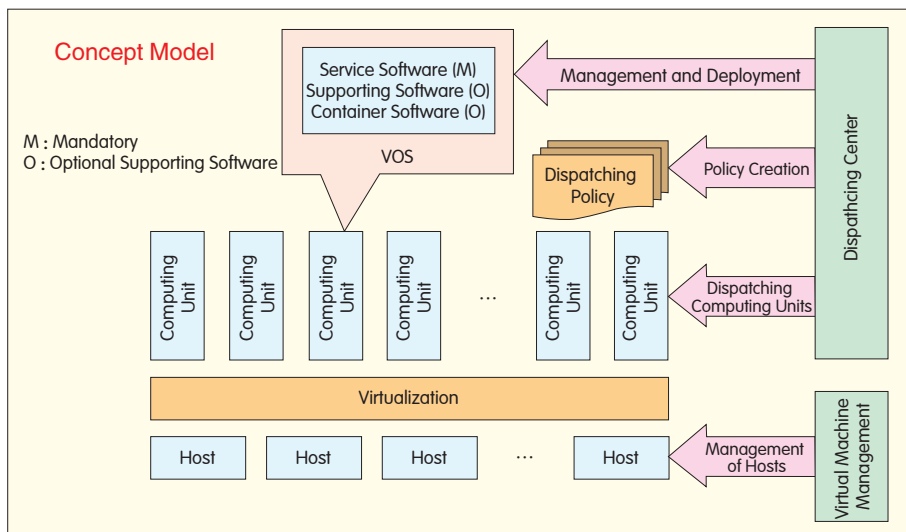
- The service can obtain computing resources according to the actual processing requirements. The carrier need not estimate computing resources before providing a service. This eliminates the risks of pre-investment, and a service can be started on a small scale. With an increase in demand, hardware resources occupied by services can be expanded through service dispatching and virtualization technology.
- The problem of service imbalance in different regions and at different times is solved. Hardware power consumption is reduced when the service volume is low. By expanding hardware resource occupation during holidays or for unexpected service peaks, operational risk is avoided.
- Many mobile network resources can also be rented externally. After the computing resource is virtualized, charges are paid quickly. The tenant can use the computing resource on demand.

2 Service Dispatching and Virtualization

Based on the urgent requirements of mobile carriers, this article provides a total solution that integrates virtualization and service dispatching. The model architecture is shown in Fig. 1.

The core management components include a virtual machine management system and a service dispatching center. In the schematic design, bottom-layer physical device virtualization and service-layer processing capability are separate.

An application requires a computing mode, a storage mode, and a communication mode. For flexibility



▲ Figure 1. Model of service dispatching and virtualization.

and unlimited mirroring of the computing resource, the most practical method is virtualization of resources. The multiplexing and sharing mechanisms are hidden for applications. Different public computing is distinguished according to the abstractness and management layer.

In this scheme, the mobile communication service computing cloud is managed in two layers. In the first layer, physical hardware is virtualized to an abstract computing unit. The process is not affected by upper layer services. The attributes of all computing units are consistent. The second layer is the dynamic dispatching system for the differential service. Flexibility control for the service system can be implemented according to different service processing logics, service performance requirements, and expected resource occupation. With the cooperation of the service dispatching center and virtual machine management center, the requirements of multiple-service real-time dynamic resource adjustment are met.

At present, virtual machine technology is sophisticated. Most mainstream virtual machine factories construct hardware CPU and memory virtual units through XEM and KVM technologies. Virtual machine technology has the following four features:

- It runs multiple virtual units on a single physical server
- Virtual units are mutually separated on a virtual machine of the same physical hardware device
- The complete virtual unit is saved in the file. The virtual unit is moved and copied by moving and copying these files
- The relationship between the virtual unit and bottom layer physical hardware is shielded. Smooth migration can occur on any server without any modification.

Virtualization technology converts physical resources into a resource pool that is easy to divide. In its design, virtual technology complies with the conditions of cloud computing and is capable of universal resource dispatching. The resource to be dispatched is not limited to the virtual unit only. Mobile carriers require a solution that provides centralized capability control for different services. In this solution, dynamic service traffic of multiple services can be monitored, and service load can be determined intelligently. Then, hardware and virtual unit resource dispatching can be balanced.

By constructing service dispatching module based on virtual machine technology, shortage of the control over the communication service in the technology can be covered. The dispatching center works with the

virtual machine management system to complete dispatching. The model is shown in Fig. 2.

Four functional modules are contained in the dispatching center. They are:

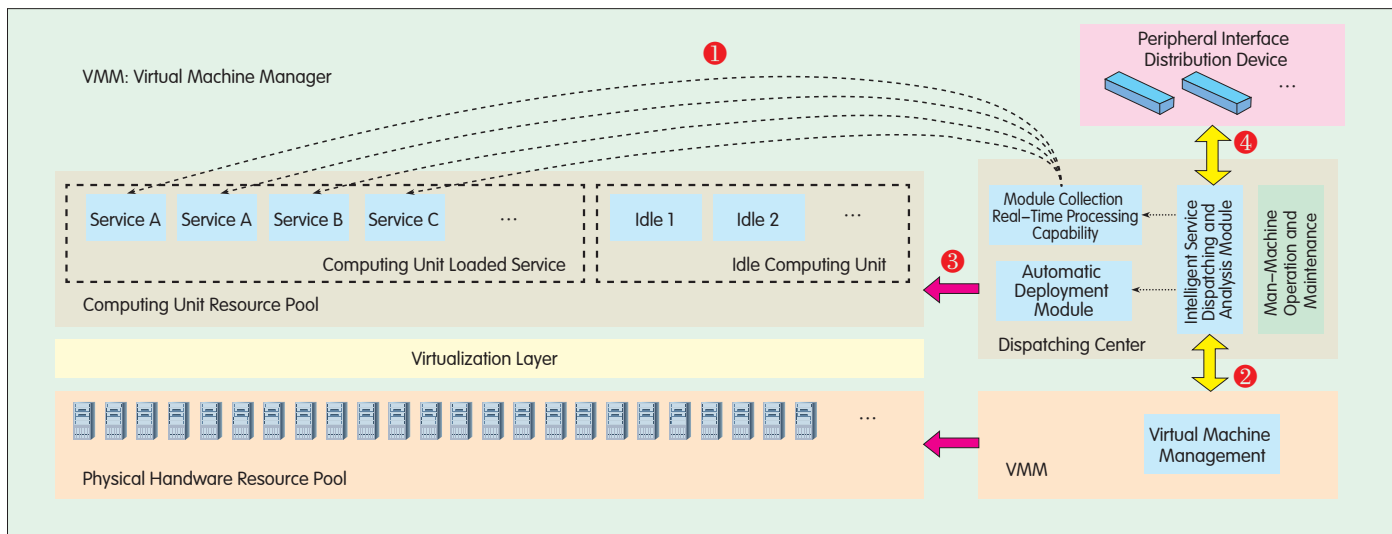
- service intelligent dispatching and analysis module. As the core processing module of the dispatching center, the module analyzes the processing capability of the service layer and adjusts the service licenses according to real-time monitored and collected service running data. When necessary, it applies for an idle computing unit or releases an occupied redundant computing unit by interacting with a virtual machine management system. Services are quickly loaded and unloaded through automatic deployment of modules to dynamically adjust the service license processing capability. At the same time, the module is also responsible for notifying the peripheral interface distribution devices of the service node flexibility.

- real-time processing capability collection module. By interacting with the service processing, this module collects information such as real-time message processing traffic, database resource occupation requirements, and processing capability. Two collection modes are supported: Scheduled service process reporting mode and dispatching sub-system sent message initiative driving mode. The collected data is written into the dispatching analysis database for intelligent dispatching policy analysis.

- automatic deployment module. This module loads the service packets into the specified computing unit according to the deployment message of the service-intelligent dispatching and analysis module. Or it stops the service to clear the service packets from the computing unit.

- man-machine maintenance. This provides man-machine operation interface. The service module running status can be monitored, and manual interference with the dispatching is provided.

The dispatching center implements dynamic adjustment and resource



▲ Figure 2. Dispatching management model.

multiplexing for processing the mobile communication service. It does this through cooperation between the modular structure and virtualization management platform. The detailed procedure is shown in Fig. 3.

Service license flexibility is dynamically controlled by tracing and monitoring the actual service volume of the service processing unit. It is also controlled by tracing and monitoring the dispatching policy and threshold configured in the intelligent dispatching and analysis center.

The intelligent dispatching and analysis policy includes [5]–[7]:

- a disaster-tolerant dispatching policy that checks whether other service processing units can share the workload of the service node if the service processing unit is abnormal. It applies for a new virtual computing unit to take over the original service processing to ensure stable running of the system.

- a periodic hibernation policy that determines periodic adjustments according to the service traffic. It releases and applies for a computing unit according to regularity. For quick start, stop, and switch, the released computing unit can retain the original service program. Hibernation and activation are implemented in status to save energy.

- a service development adjustment policy that determines whether

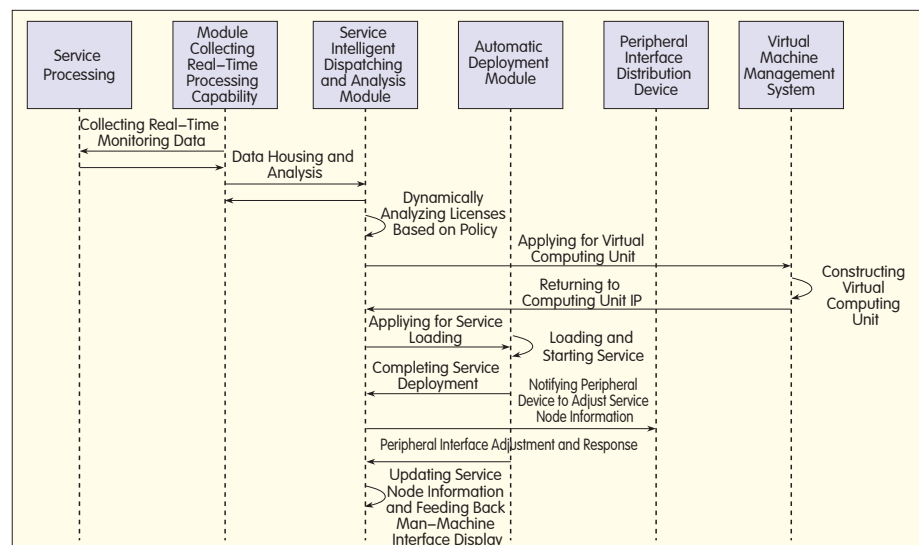
occupation of computing resources should be increased or decreased. It makes this determination based on the actual condition of the service development and complete automatic loading and unloading of services.

These three policies are implemented through the core component—the intelligent dispatching and analysis module of the dispatching center. This module is responsible for controlling virtual resources according to the monitored data. No-interference dynamic adjustment requires a number of comparison algorithms to evaluate multiple indexes. A resource dispatching command is sent

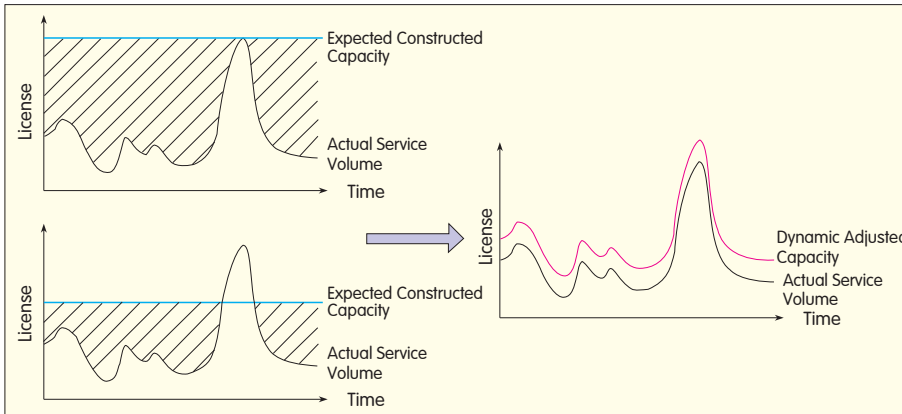
according to the evaluation results [8]. A universal computing model is as follows:

- 1) Sampling condition
 - Sampling interval: 1 s
- 2) Sampling data
 - The processing license of the sampling point virtual unit bearing "service type 1" is: Llic
 - The sampling point virtual unit occupied CPU is: Lcpu
 - The sampling point virtual unit occupied memory is: Lmemory
 - The sampling point virtual unit occupied Input/Output (I/O) resource is: Lio

The weights of the parameters in the



▲ Figure 3. Flow chart of the dispatching process.



▲ Figure 4. Comparison between the dispatched resource distribution and actual service volume.

computing are R_1-R_4 . The weight indicates the occupied resource offset in the computing unit of different service applications [9]. If the coefficient $R_i (R_1-R_4)$ cannot reflect the load of applications, it is modified until more accurate values are found [10]–[12].

3) Calculation formula of the sampling value

$$\text{LOAD}(N) = R_1 \times \text{Llic}(N) + R_2 \times \text{Lcpu}(N) + R_3 \times \text{Lmemory}(N) + R_4 \times \text{Lio}(N)$$

4) Determination period and method

Comprehensive determination of the weighted loading values can be made using consecutive sampling. When setting the cycle of the collected weight values, short cycle more accurately reflects the instant loading of each computing unit. Frequent collection increases the burden on the tested computing unit or increases unnecessary network load [13]. To solve this problem, the cycle for collecting load information can be adjusted (10–15 seconds is recommended). At the same time, the slide window can be used to prevent jittering of the sampling data.

5) Dispatching policy:

A comparison is made based on the collected periodic sampling data and the loading section of the virtual unit to determine the load of the computing unit and then take relevant dispatching policy. The requirements for constructing flexible service platforms are met using the dispatching center and a virtual machine scheme. As a result, CAPEX and OPEX are reduced, investment is saved, and greater profit

is obtained.

3 Conclusions

At present, service platforms of mobile communication networks are constructed independently. In the early stages of investment and construction, carriers usually evaluate the scale of their service platforms based on the predicted maximum service peak in holidays. Even though these predictions are usually accurate, a certain amount of investment is also wasted. If the peak value is under-estimated, the configuration may be insufficient. Service requests of excess users may be rejected, and these rejected users will not bring in any revenue. Users may also lose confidence and refuse to use the service again because of the poor service experience.

As shown in Fig. 4, dynamic dispatching of the service layer and the application of virtualization technology enable resource distribution to be consistent with the service volume curve.

Service dispatching and virtualization discussed in this article is an optional scheme for mobile communication network clouding. The scheme has the following features:

- Unlimited cloud computing resources are obtained. Cloud computing users do not need to plan for computing resources long before the service provision.
- Prior investment of cloud users is not necessary. A service can be started

on a small scale, and hardware resources can be expanded with development requirements.

- Computing resources can be used on demand in short time. Resources can be released when they are not required. Idle machines and storage units are released and costs are saved.

Service dispatching and virtualization technology provides an idea and method for bringing cloud computing to mobile computing. The author believes that service dispatching and virtualization technology will gradually become the main construction mode of the mobile communication industry.

References

- [1] M. Armbrust, A. Fox, R. Griffith, et al., "Above the clouds: A Berkeley view of cloud computing," Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Tech. Rep. UCB/EECS-2009-28, 2009.
- [2] Amazon.com CEO Jeff Bezos on Animoto. [Online]. Available: <http://blog.animoto.com>.
- [3] M.A. Vouk, "Cloud Computing—Issues, research and implementations," *Proc. 30th Int. Conf. on Inform. Techn. Interfaces (ITI'08)*, Dubrovnik, Croatia, 2008, pp. 31–40.
- [4] L. Barroso, U. La Holzle, "The case for energy-proportional computing," *IEEE Computer*, vol. 40, no. 12, pp. 33–37, Dec. 2007.
- [5] Dean J. Ghemawat S., "MapReduce: Simplified data processing on large clusters," *Proc. 6th USENIX Symp. Operation Syst. Design and Implementation (OSDI'04)*, San Francisco, CA, 2004, p. 10.
- [6] W. M. Bulkeley W. M. (2007, Oct. 8). "Universities Combine 'Cloud' Forces." *Wall Street Journal* [Online]. Available: http://www.washington.edu/ni/apps/dailyclips/scraped/WSJ_2007-10-08.html
- [7] A. Demers, K. Petersen, M. Spreitzer et al., "The Bayou architecture: Support for data sharing among mobile users," *Proc. 1st IEEE Workshop on Mobile Comput. Syst. Applic. (WMCSA'94)*, Santa Cruz, CA, 1994, pp. 2–7.
- [8] S. Garfinkel, "An evaluation of Amazon's grid computing services: EC2, S3 and SQS," *Comp. Sci. Group, Harvard Uni., Rep. TR-08-07*.
- [9] S. Ghemawat, H. Gobioff, S. Leung et al., "The Google file system," *Proc. 19th ACM SIGOPS Symp. Operating Syst. Principles (SOSP'03)*, New York, 2003, pp. 29–43.
- [10] J. Gray, *Distributed Computing Economics*, New York: ACM Press, 2008, pp. 63–68.
- [11] J. Gray, D. Patterson, *Conversation with Jim Gray*, New York: ACM Press, 2003, pp. 8–17.
- [12] J. Hamilton. (2008, Nov. 28). "The cost of power in large-scale data centers," [Online]. Available: <http://perspectives.mvdirona.com/2008/11/28/CostOfPowerInLargeScaleDataCenters.aspx>
- [13] J. Hamilton, "Internet-scale service efficiency," *Proc. 2nd Workshop on Large-scale Distributed Systems and Middleware (LADIS'08)*, New York, 2008.

Biography

Xinzhi Ouyang (ouyang.xinzhi@zte.com.cn) graduated from Nanjing University of Science and Technology with a degree in computer science and application. He works as the chief engineer of ZTE Service Research Institute and is an expert in VoIP, SMS, WAP, VAS, and Internet service.

The Internet of Things and Ubiquitous Intelligence (3)

*Dongliang Xie
Yan Shi*

(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, P. R. China)

Editor's Desk:

The traditional Internet is oriented towards person-to-person connection, whereas the Internet of things (IoT) is oriented towards connections between inanimate objects. IoT covers a larger range of connections and involves more semantics than traditional Internet. Traditional Internet and telecom networks focus on information transfer, but IoT focuses on information services. By combining sensor networks, Internet, telecom networks, and cloud computing platform, IoT can sense, recognize, affect, and control the physical world. The physical world can be unified with the virtual world and human perception. In this part, we discuss ubiquitous network convergence and cooperation technologies in terms of their application scenarios, technical advantages and research directions. We also briefly introduce IoT services.

increases network construction cost greatly. Research shows that even advanced signal processing technologies, such as MIMO and adaptive antennas, cannot solve the coverage problem. An effective approach is to develop a new network architecture that combines cellular and multihop wireless technologies.

Self-organizing relay between mobile terminals (mobile relay stations) is often applied in group communication, emergency communication, security mechanisms, and military communication. Mobile terminals are often regarded as valuable supplements to fixed relay stations in public communications, that is, fixed relay stations deployed by the operator. These have fewer functions and cost less than base stations and access points (APs).

Scenario 3: Self-organizing among terminals in a hybrid network comprising cellular networks and wireless sensor networks

In traditional wireless sensor networks, a large number of nodes are required to complete a user-specified task by means of peer-to-peer, multihop and self-organizing wireless communication. Networks based on static data access nodes have inherent problems, including unbalanced energy consumption, low transmission efficiency, inflexible deployment, and single network architecture. They are also prone to routing voids, coverage holes, and bottlenecks and overall network performance is decreased.

A hierarchical network architecture that combines wireless sensor networks and cellular networks can effectively solve coverage and transmission problems. This type of network architecture is more suitable than others for future ubiquitous, heterogeneous and cooperative

5 Application Scenarios and Technical Advantages of Terminal Cooperative Technologies

Terminal cooperative technologies are mainly applied in the following scenarios:

Scenario 1: Self-organized networking and cooperation among multiple terminals of one user

With terminal cooperative technologies, multiple terminal devices of one user can be networked via radio interfaces in a self-organized way. A self-organizing network can manage and maintain multiple computing devices in a distributed environment. These devices are frequently changed and loosely coupled, and they can dynamically discover, select,

aggregate and match terminal capabilities according to user preferences, service characteristics, locations and environments. Thus, terminals can cooperatively support the same service.

Scenario 2: Self-organizing among terminals in network architecture comprising a cellular network and wireless multihop technologies

The greatest challenge for future cellular mobile networks is coverage. Most cellular networks operate at relatively high frequency bands and are likely to have poor penetration and poor non-line-of-sight (NLOS) capability in future high-transmission 4G networks. Coverage and transmission will be unstable, and coverage holes may even arise. One solution to these problems is to increase the number and power output of base stations. But this

networks. Mobile terminals (mobile data access points) such as mobile phones, PDAs, and notebooks have distinct advantages in energy efficiency, bandwidth, reliability, transmission range, and mobility. Mobile terminals with multiple interfaces can use their relatively high-rate wireless interfaces to send and receive network data. They can also use the positioning functions of cellular networks and sensor networks to provide users with diverse location-based services. Cooperation among mobile terminals can also significantly enhance the performance of wireless sensor networks. By handing complex work such as data processing, access processing, data forwarding, and routing maintenance over to mobile terminals, wireless sensor networks can minimize data errors (or packet loss) arising from multihop wireless transmission. Moreover, they can use the powerful computing capability of mobile terminals to process information within the networks, thus lightening their own loads. Mobile terminals in a heterogeneous network can use their own wireless interfaces to communicate with the fixed cellular base stations in the upper layer or access points of other wireless networks or even to form an ad hoc network with other mobile data access points. In this way, mobile terminals can effectively exchange data with each other by means of cooperation and can also greatly expand the applications of wireless sensor networks.

Terminal cooperative technologies can bring many advantages.

First, cooperative technologies can exploit available terminal capability and user network resources for ubiquitous, heterogeneous convergence. By organically integrating different wireless technologies, cooperative technologies can take advantage of the capabilities of several terminals of one user. They can also take advantage of self-organizing relays among terminals and the rich capabilities of sensor networks to converge ubiquitous and heterogeneous networks and services.

Second, cooperative technologies can increase the flexibility of

communication. They offer users and terminals diverse access modes and service provision methods. They can use the advantages of both infrastructure and non-infrastructure networks to attain good network performance, flexibility and scalability. Convergence of infrastructure and non-infrastructure networks is likely to be important in future mobile communication systems.

Third, cooperative technologies solve the coverage hole problem and expand the coverage of wireless networks. When cellular networks are combined with multihop wireless communication technologies, single-hop, poor-performance radio links are replaced with high-performance multihop ones. This eliminates coverage holes and increases the coverage of each base station or AP.

Fourth, cooperative technologies increase system capacity. The multihop forwarding mechanism can reduce the transmit power, which enhances the spatial multiplexing of channels. When local services are set up in a self-organizing way, the bottleneck in access points is mitigated. By transferring the traffic of hotspot cells, resources are dynamically adjusted, and resource use is improved. All these help increase system capacity.

Fifth, cooperative technologies balance load. The traffic of wireless coverage areas is often unbalanced and changeable. Dynamic self-organizing relays can distribute the traffic of one cell to surrounding cells, reducing the probability of service congestion and handover failure that arises from insufficient capacity.

Sixth, cooperative technologies save radio resources. Mobile nodes that are close to each other can directly communicate with each other in a self-organizing way. In this way, the system load is decreased and system capacity is increased.

6 Research Direction of Terminal Cooperative Technologies

Terminal cooperative technologies

support heterogeneous network convergence and are regarded as one of the core components in future mobile communications. At present, they are a focus of research. However, in an environment where infrastructure networks and non-infrastructure networks are converged on the basis of terminal cooperation, there are many problems that arise from heterogeneous wireless technologies and many problems that arise from dynamic, multihop self-organizing non-infrastructure networks and cooperation among terminals. These problems involve routing protocol, network control, mobility management, service provision, and evaluation of system performance.

Among other issues, the following are in urgent need of attention:

Issue 1: Cooperation-based self-organizing adaptive network control mechanism

The infrastructure network and non-infrastructure network have different network control mechanisms. Most infrastructure networks have a centralized network control mechanism, whereas decentralized non-infrastructure networks have a distributed control mechanism. The differences between the two control mechanisms is reflected in their control and management technologies, that is, routing protocol, QoS guarantee, and resource management. It is necessary to develop a self-organizing and adaptive network control mechanism that combines the advantages of centralized and distributed mechanisms. Such a mechanism should effectively control dynamic, multihop, multipath wireless communication that comes with cooperative technologies.

Issue 2: Mobility management technology

Most traditional mobility management technologies focus on location management and horizontal handover control within a specific network. Mobility management technologies for heterogeneous networks focus on hierarchical location management and vertical handover control over different access technologies. In infrastructure

networks, mobility management technologies are relatively mature; but in non-infrastructure networks, there is still no systematic mobility management technology. The introduction of cooperative technologies enables infrastructure and non-infrastructure networks to be organically integrated. A mobile terminal may have several network interfaces and support several communication modes, including cellular, self-organizing and hybrid. It needs information about the relative and absolute location of other mobiles for effective, self-organizing cooperation. Therefore, cooperation-based mobility management technologies need to integrate location management, positioning technologies, and location-based services in order to implement handover control in self-organizing adaptive communication.

Issue 3: Service cooperation technology

In ubiquitous, heterogeneous networks, there are a large number of intelligent devices that provide a range of information and services. The peer-to-peer feature of nodes in non-infrastructure networks enables each mobile terminal to act as a service provider. Hence, self-organizing among terminals requires efficient service cooperation technologies, including cooperative service provision modes as well as broadcast, discovery and request mechanisms for dynamic, self-organizing, and lightweight services. Efficient use of network resources and services is key for automatic configuration of application services.

Issue 4: Evaluating the impact of terminal cooperation on system performance

Existing cooperation schemes are mainly designed to address a specific problem in a specific scenario with the goal of optimizing local performance. In these schemes, only intuition and qualitative analysis are used to evaluate the impact of self-organizing multihop communication and multihop access on system performance. There is still no systematic quantitative analysis, and

the advantages of cooperative technologies to performance are still in dispute. It is necessary to quantitatively analyze the impact of cooperative technologies on network coverage, system capacity, service availability, and reliability. This will provide a useful reference for network planning, cost accounting, and determining network performance and optimization.

7 Services of the Internet of Things

IoT has applications in daily life as well as in industry automation. Typical applications include intelligent power grids, intelligent banking, intelligent transportation, intelligent home, and intelligent medical systems. The services of IoT can be classified according to different criteria. They can be roughly divided into four types according to technical features and development trend: identity-related, information aggregation, cooperative sensing, and ubiquitous intelligence [1].

Identity-related services are delivered using identification technologies such as RFID and two-dimensional code. In information aggregation services, the IoT platform manages the terminals, data, applications, and services in a unified way. Terminals collect and report data and do not need to communicate with each other. Because terminal technologies tend to be diverse, intelligent, and multimode in IoT, communication between the terminals and between terminals and individuals is more frequent and complicated than ever. As a result, one trend in the development of IoT is that terminals cooperate to complete a service. The vision of IoT is to provide ubiquitous intelligent services anytime, anywhere and in any way.

Ambient Intelligence (Aml) was proposed by the Information Society and Technology Advisory Group (ISTAG) of the European Commission. It was used for the launch of the sixth framework programme (FP6) in Information, Society and Technology (IST). Aml builds on pervasive computing, ubiquitous computing, and

human-centered computer interaction design. It is an adaptive digital environment in which many sensing and computing devices are embedded. It can judge a person's intention and reaction by identifying the person's posture, physiological condition, gestures, and voice in different situations. Aml has the following features:

- context-awareness. Aml can recognize the situational context of a user. Sensors embedded in the context change the physical information into valid statuses or operations, such as identity authentication and acquisition of the user's current location.

- personalization and adaptability. Aml can be tailored to a user's specific needs and can meet changes in user needs.

- anticipation. Aml can anticipate the desires of users without conscious mediation. This enables users to unconsciously interact with the environment, process information in a simple and natural way.

In brief, ubiquitous and intelligent services are new implicit services that are designed to be user-centered in human living and work environments. Delivering these services requires IoT in the ubiquitous sensing layer as well as the powerful information processing of cloud computing. (To be continued)

Reference

- [1] Xiaojiang Xing, Jianli Wang, Mingdong Li, "The Internet of Things and its Key Technologies," *ZTE Communications*, vol.2, no. 16, pp.27-30.

Biographies

Dongliang Xie (xiedl@bupt.edu.cn) is a director and associate professor at the Broadband Network Center of State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. He researches wireless and mobile network technologies, wireless sensor networks, mobile Internet QoS, network cooperation, and ubiquitous intelligence. He has published more than 40 papers.

Yan Shi (shiyan@bupt.edu.cn) has a Ph.D in engineering. She is a lecturer in the Broadband Network Research Center of the State Key Laboratory of Networking and Switching Technology of Beijing University of Posts and Telecommunications. He researches wireless ubiquitous heterogeneous network technologies, with a particular interest in mobility management and heterogeneous network convergence and cooperation. He has published over 20 papers to date.

Abbreviation Index

A

AAA: authentication, authorization and accounting
 ACLR: adjacent channel leakage ratio
 ADC: analog-to-digital converter
 ADS: advanced design system
 AKA: authentication and key agreement
 AM: amplitude-modulated
 AMC: adaptive modulation and coding
 AN: access network
 ASK: amplitude-shift keying
 AWGN: additive white Gaussian noise

B

BBM: baseband module
 BER: bit error rate
 BJT: bipolar junction transistor
 BPSK: binary phase-shift keying

C

CDF: cumulative distribution functions
 CE: collector efficiency
 CFR: crest factor reduction
 CIR: carrier to interference ratio
 CMOS: complementary metal oxide semiconductor
 CPM: continuous phase modulation

D

DAC: digital-to-analog converter
 DECT: digital enhanced cordless telecommunications
 DPD: digital-predistortion
 DSP: digital signal processing
 DUC: digital up-conversion
 CFR: crest factor reduction

E

ET: envelope tracking
 EVM: error-vector-magnitude

F

FCC: federal communications commission
 FDD: frequency-division duplexing
 FDM: frequency division multiplexing
 FF: flip flop
 FPGA: field-programmable gate array
 FUSE: filesystem in userspace

G

GF: geometry factor
 GMSK: Gaussian MSK

GPB: general-purpose interface bus

H

HARQ: hybrid automatic repeat request
 HE: home environment

I

I/O: Input/Output
 ICIC: inter-cell interference coordination
 IF: intermediate frequency
 IFM: intermediary frequency module
 IMD: inter-modulation distortion
 IMS: IP multimedia subsystem
 IoT: interference over thermal
 IQ: inphase and quadrature
 ISE: integrated software environment
 ISI: inter-symbol interference

L

LDMOS: laterally diffused metal oxide semiconductor
 LINC: linear amplification with nonlinear component
 LNA: low-noise amplifier
 LO: local oscillator
 LPF: low-pass filter
 LUT: look-up-table

M

MAPL: maximum allowed path loss
 MCS: modulation and coding scheme
 MHMIC: miniature hybrid microwave integrated circuit
 MIMO: Multiple-Input Multiple-Output
 M-PSK: M-ary phase shift keying
 M-QAM: M-ary quadrature amplitude modulation
 MSK: minimum shift keying
 MTM: Mobile Trusted Module
 NCO: numerical controlled oscillator
 NDIS: Network Driver Interface Specification

N

NF: noise figure

P

PA: power amplifier
 PAE: power-added efficiency
 PAPR: peak-to-average power ratio
 PAR: peak-to-average ratio

PDF: probability density function
 PKI: public key infrastructure
 PM: phase-modulated
 PVT: process-voltage-temperature

Q

QAM: quadrature amplitude modulation
 QPSK: quadrature phase-shift keying

R

RAM: random-access memory
 RF: radio frequency
 RFC: radio frequency choke
 RMS: root-mean-square
 RNC: radio network controller
 RRC: root raised cosine filtering

S

SAE: system architecture evolution
 SC-FDMA: single carrier frequency division multiple access
 SDR: satellite digital radio
 SHC: sample and hold circuit
 SMSC: short message service center
 SN: serving network
 SOA: safe operation area
 SP: service provider
 SPICE: simulation program with integrated circuit emphasis
 SPMM: six-port millimeter-wave mixer
 SRC: symbol rate per carrier
 SSL: Secure Sockets Layer
 SSP: security service provider
 SWP: software provider

T

TLS: transport layer security
 TMP: trusted mobile platform
 TPM: trusted platform module
 TUT: Tampere University of Technology

U

USIM: user services identity module

V

VHDL: VHSIC hardware description language
 VNA: vector network analyzer
 VSA: vector signal analysis
 VSWR: voltage standing wave ratio

W

WLAN: wireless local area networks