

Alessandro De Gloria
Editor

Applications in Electronics Pervading Industry, Environment and Society

APPLEPIES 2014

Lecture Notes in Electrical Engineering

Volume 351

Board of Series editors

Leopoldo Angrisani, Napoli, Italy
Marco Arteaga, Coyoacán, México
Samarjit Chakraborty, München, Germany
Jiming Chen, Hangzhou, P.R. China
Tan Kay Chen, Singapore, Singapore
Rüdiger Dillmann, Karlsruhe, Germany
Haibin Duan, Beijing, China
Gianluigi Ferrari, Parma, Italy
Manuel Ferre, Madrid, Spain
Sandra Hirche, München, Germany
Faryar Jabbari, Irvine, USA
Janusz Kacprzyk, Warsaw, Poland
Alaa Khamis, New Cairo City, Egypt
Torsten Kroeger, Stanford, USA
Tan Cher Ming, Singapore, Singapore
Wolfgang Minker, Ulm, Germany
Pradeep Misra, Dayton, USA
Sebastian Möller, Berlin, Germany
Subhas Mukhopadhyay, Palmerston, New Zealand
Cun-Zheng Ning, Tempe, USA
Toyoaki Nishida, Sakyo-ku, Japan
Bijaya Ketan Panigrahi, New Delhi, India
Federica Pascucci, Roma, Italy
Tariq Samad, Minneapolis, USA
Gan Woon Seng, Nanyang Avenue, Singapore
Germano Veiga, Porto, Portugal
Haitao Wu, Beijing, China
Junjie James Zhang, Charlotte, USA

About this Series

“Lecture Notes in Electrical Engineering (LNEE)” is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer’s high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer’s other Lecture Notes series, LNEE will be distributed through Springer’s print and electronic publishing channels.

More information about this series at <http://www.springer.com/series/7818>

Alessandro De Gloria
Editor

Applications in Electronics Pervading Industry, Environment and Society

APPLEPIES 2014

 Springer

Editor

Alessandro De Gloria
Electronic Engineering
University of Genoa
Genoa
Italy

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-3-319-20226-6 ISBN 978-3-319-20227-3 (eBook)
DOI 10.1007/978-3-319-20227-3

Library of Congress Control Number: 2015941875

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Preface

I am proud and honored to introduce this book that collects papers from the Applepies 2014 Conference (Applications in Electronics Pervading Industry, Environment & Society). The papers offer a wide and reasoned overview of electronic applications in several domains, demonstrating how electronics has become pervasive and ever more embedded in everyday objects and processes.

The computational, storage and communication power of current electronic systems is such that we may really say that their applications are limited only by the designer's fantasy. This represents a great challenge for practitioners, managers and academicians in electronic engineering. The challenge also stresses the importance of multidisciplinary knowledge, expertise and collaboration, in order to support a virtuous iterative cycle from user needs to new products and services. The cycle goes through the whole system engineering process, which typically encompasses requirement elicitation, specification management, software and hardware design, lab and user testing and verification, maintenance management.

For either an embedded or cyberphysical system to be successful in the current globalized market competition, at least one of the following features must be provided: innovation, high performance, a good cost/performance ratio. Designing and implementing each one of such features requires deep knowledge of both the system's target application and domain, and of the technologies that are potentially able to fulfill the expected goals.

One of the most important factors for the success of a project consists in the adoption of a suited design flow and related tools. Only seldom are simple top-down or bottom-up methods able to meet the time and cost-related challenges of nowadays market scenarios. Even if every application stems from recognizing one or more key user needs, proper design, implementation, and maintenance requires mastering the most suited technologies and tools in order to support efficient and effective development and life-cycle management of electronics applications. Support tools must also be able to capture and share a team's experience in the design and implementation process, as it allows anticipating possible problems that may not appear on paper.

All these challenging aspects call for the importance of the role of the university as a place where new generation designers can learn and practice with cutting edge technological tools and are stimulated to devise solutions for challenges coming from a variety of application domains, such as healthcare, transportation, education, tourism, entertainment, cultural heritage, energy.

This book reports and discusses several examples of designs. Year after year—and through the fundamental support of several colleagues—I am proud to say that the Applepies Conference is ever more becoming a reference point in the field of electronics systems design, trying to fill at the scientific and technological R&D level a gap that the most farsighted industries have already indicated and are striving to cover.

Alessandro De Gloria

Contents

1	Developments and Applications of Electronic Nose Systems for Gas Mixtures Classification and Concentration Estimation . . .	1
	Calogero Pace, Letizia Fragomeni and Walaa Khalaf	
1.1	Introduction	1
1.2	Electronic Nose System Architecture	2
1.3	Support Vector Machine Approach and Least Square Regression Method	3
1.4	E-Nose for Safety Monitoring Applications in Refinery Environment	4
1.5	E-Nose for Breath Biomarkers Monitoring During Haemodialysis	5
1.6	Conclusions	6
	References	7
2	Machine Learning-Based System for Detecting Unseen Malicious Software	9
	Federica Bisio, Paolo Gastaldo, Claudia Meda, Stefano Nasta and Rodolfo Zunino	
2.1	Introduction	9
2.2	The Malware-Detection Framework	10
2.3	A ML-Based Approach to Malware Detection	11
2.4	Experimental Results	13
2.5	Conclusions	15
	References	15
3	Implementation of a Spread-Spectrum-Based Smart Lighting System on an Embedded Platform	17
	Maurizio Martina, Massimo Ruo Roch and Flavio Ghirardi	
3.1	Introduction	17
3.2	System Overview	18

3.3	System Implementation	21
3.4	Conclusion	22
	References.	23
4	Self-powered Active Cooling System for High Performance Processors	25
	Maurizio Rossi, Luca Rizzon, Matteo Fait, Roberto Passerone and Davide Brunelli	
4.1	Introduction and State of Art	26
4.2	Thermo-Electric Energy Harvesting	27
4.3	Hybrid Cooling	28
4.4	Experimental Results	30
4.5	Conclusions	32
	References.	33
5	High Speed VLSI Architecture for Finding the First W Maximum/Minimum Values	35
	Guoping Xiao, Waqar Ahmad, Syed Azhar Ali Zaidi, Massimo Ruo Roch and Giovanni Causaprano	
5.1	Introduction	35
5.2	Problem Formulation	36
5.3	Architectural Description	38
5.4	Results and Comparisons	40
5.5	Conclusions	40
	References.	41
6	Design and Implementation of a Portable fNIRS Embedded System.	43
	Diego Agrò, Riccardo Canicatti, Maurizio Pinto, Giuseppe Morsellino, Alessandro Tomasino, Gabriele Adamo, Luciano Curcio, Antonino Parisi, Salvatore Stivala, Natale Galioto, Alessandro Busacca and Costantino Giaconia	
6.1	Introduction	44
6.2	Hardware	45
6.3	Experimental Results	48
6.4	Conclusion	49
	References.	49
7	Advancements on Silicon Ultrasound Probes (CMUT) for Medical Imaging Applications	51
	Giosuè Caliano and Alessandro S. Savoia	
7.1	Introduction	51
7.2	Probe Fabrication and Packaging.	52

7.3	Probe Characterization and Ultrasound Imaging	54
7.4	Conclusions	56
	References.	56
8	Open Platforms for the Advancement of Ultrasound Research . . .	59
	Enrico Boni, Luca Bassi, Alessandro Dallai, Gabriele Giannini, Francesco Guidi, Valentino Meacci, Alessandro Ramalli, Stefano Ricci and Piero Tortoli	
8.1	Introduction	59
8.2	ULA-OP Description	60
8.3	ULA-OP 256 Design Specifications.	62
8.4	ULA-OP Applications	63
	References.	64
9	A Robust Tracking Algorithm for Super-Resolution Reconstruction of Vehicle License Plates	65
	Stefano Marsi, Sergio Carrato and Giovanni Ramponi	
9.1	Introduction	65
9.2	Overview of the Proposed Method	66
9.3	License Plate Tracking and Registration.	67
9.4	Observations About the Implementation of the Genetic Algorithm.	69
9.5	Super-Resolution Image Generation.	70
9.6	Conclusions	72
	References.	72
10	c-Walker: A Cyber-Physical System for Ambient Assisted Living	75
	Luca Rizzon, Federico Moro, Roberto Passerone, David Macii, Daniele Fontanelli, Payam Nazemzadeh, Michele Corrà, Luigi Palopoli and Domenico Prattichizzo	
10.1	Introduction	76
10.2	System Description	77
10.3	Hardware/Software Architecture	79
10.4	Preliminary Results	80
10.5	Conclusions	81
	References.	82
11	2D and 3D Palmprint Extraction by an Automated Ultrasound System	83
	Antonio Iula, Gabriel Hine, Alessandro Ramalli, Francesco Guidi and Enrico Boni	
11.1	Introduction	83
11.2	Experimental Setup	85

11.3	Results	86
11.4	Conclusions	88
	References	88
12	AA-Battery Sized Energy Harvesting Power Management Module for Indoor Light Wireless Sensor Applications	91
	Alessandro Vinco, Rashid Siddique, Davide Brunelli and Wensi Wang	
12.1	Introduction	91
12.2	Power Management Module Design and Simulations	93
12.3	System Implementation and Test Results	96
12.4	Conclusions and Future Work	96
	References	97
13	A Framework for Network-On-Chip Comparison Based on OpenSPARC T2 Processor	99
	G. Causaprano, A. Audero, S. Tota and M. Ruo Roch	
13.1	Introduction	99
13.2	OpenSPARC T2 Crossbar	100
13.3	Network-on-Chip Replacement	101
13.4	Results	103
13.5	Conclusions	104
	References	104
14	A GPU 3D Segmentation Framework for Medical Imaging	107
	Francesca Galluzzo, Luca De Marchi, Nicola Testoni and Guido Masetti	
14.1	Introduction	107
14.2	Background	108
14.3	Parallel Implementation	109
14.4	Experimental Results	111
14.5	Conclusions	113
	References	114
15	Augmented Reality Tools for Structural Health Monitoring Applications	115
	L. De Marchi, A. Ceruti, N. Testoni, A. Marzani and A. Liverani	
15.1	Introduction	115
15.2	Impact Localization	116
15.3	Augmented Reality	118
15.4	Experimental Setup and Results	119
15.5	Conclusions	120
	References	120

16 Squeeze the Lemon: Balancing as a Way to Use Every Drop of Energy in a Lithium-Ion Battery 123
 Federico Baronti, Roberto Roncella and Roberto Saletti

16.1 Introduction 123

16.2 Battery Balancing Circuit Topologies. 124

16.3 Balancing Circuit Implementations 127

16.4 Conclusions 129

References. 129

17 Fully Integrated 60 GHz Transceiver for Wireless HD/WiGig Short-Range Multi-Gbit Connections 131
 Sergio Saponara and Bruno Neri

17.1 Reasons for Integrated Transceivers at 60 GHz in Bulk-CMOS 131

17.2 Target CMOS-Bulk Technology 133

17.3 Low Complexity 60 GHz Transceiver Architecture 133

17.4 Implementation Results and Conclusions 136

References. 137

18 Low Cost FMCW Radar Design and Implementation for Harbour Surveillance Applications 139
 Sergio Saponara, Stefano Lischi, Riccardo Massini, Luca Musetti, Daniele Staglianò, Fabrizio Berizzi and Bruno Neri

18.1 Applications for MIC-Based Radar Transceivers 139

18.2 System Specifications for Harbour Surveillance Radar 140

18.3 Low Complexity 10 GHz Transceiver Architecture 142

18.4 Conclusions 144

References. 144

19 Healthcare System for Non-invasive Fall Detection in Indoor Environment 145
 Marco Mercuri, Carmine Garripoli, Peter Karsmakers, Ping Jack Soh, Guy A.E. Vandenbosch, Calogero Pace, Paul Leroux and Dominique Schreurs

19.1 Introduction 146

19.2 Embedded Healthcare System. 147

19.3 Classification Implementation 149

19.4 Experimental Results 150

19.5 Conclusion 151

References. 151

20	Analysis of Spread-Spectrum Clocking Modulations Under Synchronization Timing Constraint	153
	Davide De Caro, Michele De Martino, Nicola Petra and Antonio G.M. Strollo	
20.1	Introduction	153
20.2	Synchronization Constraint for Triangular and Sawtooth Modulations	155
20.3	Simulation Results.	157
	References.	158
21	Towards a Frequency Domain Processor for Real-Time SIFT-based Filtering	161
	Giorgio Lopez, Ettore Napoli and Antonio G.M. Strollo	
21.1	Introduction: SIFT, CVDS, and the TMuC	161
21.2	State of the Art Approaches to SIFT Elaboration.	162
21.3	Frequency Based Approach and Block-Based LoG	163
21.4	Proposed Filtering Processor Architecture.	164
21.5	Architecture Tuning and Experimental Results	165
21.6	Conclusions	166
	References.	166
22	A Real-Time FPGA-based Solution for Binary Image Thinning	169
	Daniele Davalle, Berardino Carnevale, Sergio Saponara, Luca Fanucci and Pierangelo Terreni	
22.1	Introduction	169
22.2	Algorithm Description	170
22.3	Hardware Architecture	171
22.4	Results.	173
22.5	Conclusion	174
	References.	174
23	Low Cost Electrical Current Sensors with Extremely Wide Measurement Range	175
	N. Galioto, F. Lo Bue, L. Mistretta and C.G. Giaconia	
23.1	Introduction	175
23.2	System Overview	176
23.3	System Design	180
23.4	Firmware and Auto-Range Selection	182
23.5	Conclusions	182
	References.	183

24	Pathological Voice Analysis via Digital Signal Processing	185
	Francesco Lo Bue, Natale Galioto and Costantino Giaconia	
24.1	Introduction	185
24.2	Methods.	186
24.3	Formant Analysis	187
24.4	Introducing the Wavelet Transform	187
24.5	Data Analysis	189
24.6	Experimental Test and Results	191
24.7	Conclusions and Future Work	192
	References.	192
25	A Platform-Based Emulator for Mass-Storage Flash Cards Evaluation in Embedded Systems	195
	Francesco Menichelli and Mauro Olivieri	
25.1	Introduction	196
25.2	State of the Art.	196
25.3	Emulator and Platform Architecture.	197
25.4	Simulation Setup and Results	198
25.5	Conclusions	200
	References.	201
26	A Model-Based Methodology to Generate Code for Timer Units	203
	Marco Marazza, Francesco Menichelli, Mauro Olivieri, Orlando Ferrante and Alberto Ferrari	
26.1	Introduction	204
26.2	Reference Example	204
26.3	Methodology	205
26.4	Application to the Reference Example	207
26.5	Conclusion	209
	References.	210

Introduction

Electronics research activity is usually thought as related to new electron devices development, improvement of existing device characteristics, or to the development of innovative circuits.

Technology advances in these three fields led, in the past years, to the integration of billions of components in small integrated circuits, and to a dramatic reduction in power consumption.

Common examples of such results are high performance smartphones, featuring the computing power of a desktop PC, or biomedical implantable devices, able to operate for weeks, without the need to recharge their batteries. Such a huge market of products was born when a strong competence on one or several technologies met an open mind vision of a societal need.

The possibility to have complex devices, including an entire system in a few chips or even in a single-chip (SoC), requires a change in the usual way of thinking. The challenge is now the identification of new applications, accomplishing user needs, able to highlight advantages and features of existing and ready for the market electronics technologies. Moreover, the possibility to have *smart* devices, with increasing computing power inside embedded components, allows for a distributed intelligence pervading objects surrounding human beings. This *smartness* must be thought to improve quality of life, without requiring specific skills in users. We can really speak of natural interfaces, giving us total unconsciousness of electronics existence. The *application* is now the target of current electronics engineer design process, requiring multidisciplinary knowledges. Volendo si può mettere una frase del tipo. Hence, Electronics is at the core of emerging paradigms such as Internet of Things, Cyber Physical Systems and Ambient Intelligence.

To this aim, the *Applications in Electronics Pervading Industry, Environment and Society* (APPLEPIES) conference is a workplace where researchers from both academic institutions and companies can meet, with a variety of knowledge backgrounds, to exchange information on electronic technology advances, challenging applications, and user needs.

The basic thought is that sharing information between people with different fields of interest (i.e. research, industry, society, etc.) will build up new ideas,

innovative solutions to known problems, and a complex vision of the near future of electronics applications.

Aside, an industrial exhibition is hosted, where production level technologies and research development tools are shown.

The APPLEPIES 2014 conference was a 2-day workshop, with 26 original presentations from the academic and industrial world, keynotes from STMicroelectronics and National Instruments, and open discussion sessions.

F. Bellotti
R. Berta
M. Olivieri
C. Pace
M. Ruo Roch
S. Saponara

Chapter 1

Developments and Applications of Electronic Nose Systems for Gas Mixtures Classification and Concentration Estimation

Calogero Pace, Letizia Fragomeni and Walaa Khalaf

Abstract In this paper, electronic nose systems consisting of five low cost gas sensors and three auxiliary sensors are described. The devices are effectively applied to gases mixtures classification in refinery environment and in monitoring of patients' breath on haemodialysis treatment. The systems exploit a classification algorithm based on support vector machine method and a least square regression model for concentration estimation. In particular, in the present work, the systems implementation and the results obtained during data acquisition and post-processing phases are reported.

Keywords Electronic nose · Artificial olfaction · Sensors system · Gas sensors · Clinical diagnosis · Concentration estimation · Least square regression · Support vector machine

1.1 Introduction

Electronic noses hold great promises for many fields of our lives: they may be effectively applied, for example, in agriculture, environmental monitoring, food and beverage manufacturing, biomedical and industrial applications [1–4].

In this paper, the implementation of electronic nose (ENose) systems is described: in particular the present work reports the implementation of two ENose

C. Pace (✉) · L. Fragomeni

Department of Computer Engineering, Modeling, Electronics and Systems Science (DIMES),
University of Calabria, Arcavacata di Rende (CS), 87036, Rende, Italy
e-mail: calogero.pace@unical.it

L. Fragomeni

e-mail: lfragomeni@deis.unical.it

W. Khalaf

Computer and Software Engineering Department, College of Engineering, Almustansiriya
University, Baghdad, Iraq
e-mail: walaakhalaf@yahoo.com

© Springer International Publishing Switzerland 2016

A. De Gloria (ed.), *Applications in Electronics Pervading Industry,
Environment and Society*, Lecture Notes in Electrical Engineering 351,
DOI 10.1007/978-3-319-20227-3_1

devices and their application respectively to gases mixtures classification and concentration estimation in refinery environment and in monitoring human breath during haemodialysis (HD) treatment. The systems exploit a classification algorithm based on Support Vector Machine (SVM) method, which allows for obtaining, during training phase, a model to be used afterwards in the testing phase for gases classification, and Least Square Regression (LSR) algorithm for the concentration estimation of identified gas.

1.1.1 VOCs Detection in Air Monitoring Applications

Gases emissions from new furniture, oil plants, building materials residuals and human activities products often contain components which potentially may harm to public health as pollutants of indoor air quality [5–7]. Specifically, in industrial environments the early and distributed detection of dangerous gases mixtures is a serious task, especially in oil refineries where a great amount of volatile organic compounds (VOCs) and toxic gases are unavoidably emitted and, depending on mixture concentration values, can expose the workers to poisoning and explosion risks. In this context, electronic nose systems allow for real-time monitoring that is essential for simultaneous gases detection.

1.1.2 Volatile Breath Biomarkers in Patients on Haemodialysis Treatment

Breath analysis represents an effective, non-invasive, harmless methodology for detecting disease, monitoring disease progression, or monitoring therapy. In the last decade a number of marker molecules, which are produced endogenously as a result of normal or abnormal physiologies, have been identified in breath thanks to techniques such as gas chromatography (GC) and mass spectroscopy (MS) [8].

As far as renal diseases are concerned, correlations between patients' clinical conditions and the exhalation of VOCs have been demonstrated: in the specific, numerous studies have reported biomarkers variations during haemodialysis (HD) treatments. In [9] isoprene, acetone and pentane were analyzed in patients' exhaled air by means of GC/MS; furthermore, ammonia was analyzed during HD sessions by means of ion mobility spectrometry and cavity ring-down spectroscopy [10].

1.2 Electronic Nose System Architecture

The block diagram of the developed ENose system is reported in Fig. 1.1. The system consists of a sensors array placed in a sealed sensing chamber connected to the PCB where sensors supply circuits and sensors readout circuit are implemented.

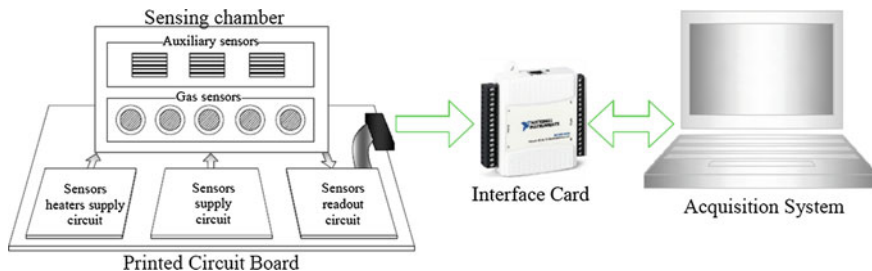


Fig. 1.1 Block diagram of the proposed electronic-nose system

Several low cost, Commercial Off-The-Shelf (COTS) sensors with different selectivity patterns are used according to the specific targeted application. SVM classification algorithm has been used in order to overcome the poor selectivity of each individual sensing device and LSR algorithm has been implemented for concentration estimation. Auxiliary sensors, i.e. temperature, humidity and pressure sensors, are also employed in order to take into account the effects of environment changes, which directly impact on gas sensors responses, and mechanical stress.

All sensors readout circuit outputs are wired to an interface card (NI DAQ) which is connected to a Personal Computer via USB for acquiring data. C-based custom software developed in the LabWindows/CVI environment allows for acquiring data and post-processing them through SVM training and testing algorithms.

LIBSVM-2.82 and 3.17 packages are used for SVM multi-classification training and testing [11].

1.3 Support Vector Machine Approach and Least Square Regression Method

The SVM method strongly relies on statistical learning theory: classification is based on the idea of finding the best separating hyperplane of data sets in the vector space (in our case, eight–dimension vector space). The hyperplane tries to both maximize the margin between different classes and minimize the sum of classification errors at the same time. The error ξ_i of a point (x_i, y_i) (y_i represents the class membership) with respect to a target margin γ and for a hyperplane f (Eq. 1.2) is:

$$\xi_i = \xi((x_i, y_i), f(x_i), \gamma) = \max(0, \gamma - y_i \cdot f(x_i)) \tag{1.1}$$

$$f(x) = \mathbf{w}^T \mathbf{x} + b = 0 \tag{1.2}$$

where ξ_i is called the margin slack variable which measures how much a point fails to have margin. The error ξ_i is greater than zero if the point x_i is correctly classified but with margin smaller than γ

$$\gamma > \xi_i > 0. \quad (1.3)$$

The cost function to be minimized is:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \quad (1.4)$$

where C is a positive, constant regularization parameter that determines the trade-off between accuracy in classification and margin width.

The original patterns can be mapped by means of appropriate kernel functions to a higher dimensional space called *feature space*: a linear separation in the feature space corresponds to a non-linear separation in the original input space. Kernel functions in machine learning range from simple polynomial mappings to sigmoid and radial basis functions: the choice of the best function depends on the specific application.

In conjunction with the classification of volatile compounds, the concentrations of the identified gases must be estimated: to this aim an LSR method has been used. The optimal estimate of the concentration is, in our model, a combination of the outputs of the employed sensors: the concentration c_i of gas i can be evaluated by Eq. 1.5:

$$c_i = \frac{1}{M} \sum_{j=1}^M c_{i-j} \text{ and } c_{i-j} = V_{i-j}^2 \times \alpha_{i-j} + V_{i-j} \times \beta_{i-j} + \delta_{i-j} = \mathbf{V}_{ij}^T \boldsymbol{\theta}_j \quad (1.5)$$

where M is the number of sensors, c_{i-j} represents the concentration estimate of sensor j , V_{i-j} is the voltage response of sensor j to gas i and the weights $\boldsymbol{\theta}$ are obtained by solving the following minimization problem:

$$\min_{\theta_1 \dots \theta_M} \sum_{i=1}^n \left(\bar{c}_i - \sum_{j=1}^M \theta_j c_{i-j} \right)^2 \quad (1.6)$$

with n equal to analyte samples number, \bar{c} actual concentration, c previously calculated concentration.

1.4 E-Nose for Safety Monitoring Applications in Refinery Environment

The ENose system, whose architecture has been reported before, has been firstly applied to the classification and concentration estimation of volatile organic compounds, such as mixtures of methane, hexane, pentane, and hydrogen sulfide [12]. The system was successively employed to address the problem of early and distributed detection of dangerous gases mixtures in refinery environment: in particular the ENose was tested at Midland oil Company refinery of Aldora, Baghdad, Iraq [13].

The developed system consists of five sensors from FIGARO USA INC: two semiconductor sensors (TGS-825, TGS-2611), two catalytic devices (TGS-6810,

Table 1.1 Methane concentration estimation obtained by the LSRM described

Actual concentration (ppm)	Estimated concentration (ppm)	Relative error (%)
2000	1991	0.45
3000	2986	0.47
4000	3905	2.38
5000	5092	1.84
6000	6078	1.30

TGS-6812) and an electrochemical oxygen sensor (KE-50) placed into a 3000 cm³ chamber. In addition two more devices was integrated: a temperature and humidity sensor (HTG-3535 from Measurement Specialties), and a pressure sensor (XFAM from Fujikura Ltd.).

1.4.1 Results

The training data were acquired by feeding the ENose with a constant mixture flux by means of a laboratory gas control system. The SVM training phase performed by using a linear kernel with a leave-one-out cross-validation scheme produced a classification correctness rate of 100 %, with a regulation parameter $C = 1000$.

After completing classification process, the concentration of the classified analyte has been estimated: for the methane case, we obtained the results reported in Table 1.1.

1.5 E-Nose for Breath Biomarkers Monitoring During Haemodialysis

Afterwards an electronic nose for biomedical applications has been developed: the key concept guiding the implementation has been the portability, the easy use and the interface with patients and hospital facilities. In this case, the ENose sensor array faces a small volume chamber (31 cm³): specifically, four metal oxide sensors (TGS-822, TGS-826, TGS-2600, TGS-2602) and an electrolyte sensor (TGS-4161) from FIGARO USA Inc., a temperature sensor (LM35 from National Semiconductor Corporation), a relative humidity sensor (HIH4000 from Honeywell) and a pressure sensor (XFAM115 from Fujikura Ltd.) have been employed.

1.5.1 Results

The system training was performed in laboratory over a wide range of analyte concentrations in air: acetone, ammonia, carbon monoxide and carbon dioxide have

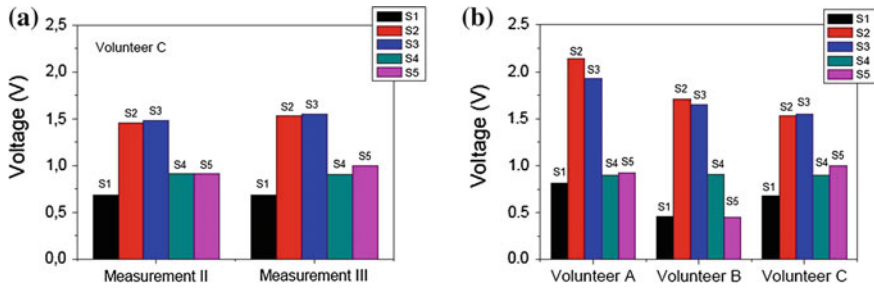
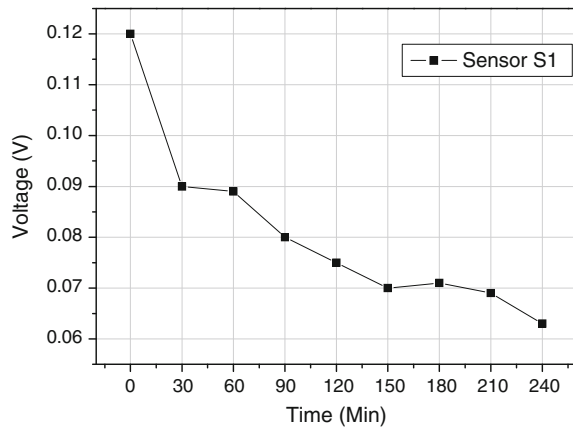


Fig. 1.2 E-Nose measurements of breath samples: **a** repeated measures from the same volunteer, **b** measures from different volunteers

Fig. 1.3 Sensor S1 voltage response variation with respect to dialysis treatment time



been mainly taken into account. The system has been later tested sampling the breath of volunteer subjects: the response patterns of the sensors are reported in Fig. 1.2. Afterwards the apparatus has been tested in hospital, sampling the breath of patients with renal disease during HD treatment. The sensors response patterns changed during the treatment (about 4 h), in particular, sensor S1, specific for ammonia, showed a continuous decrease of its voltage response with dialysis time consistently with data reported in [8, 9] (Fig. 1.3).

1.6 Conclusions

Implemented electronic nose systems have been effectively applied to gases mixtures detection in refinery environment and in monitoring of patients' breath on haemodialysis treatment. A 100 % classification rate have been obtained by employing support vector machine based algorithm. A least square regression

model for concentration estimation has been also used allowing for getting an estimate error of maximum 2.3 %. The real-time breath monitoring of patients under haemodialysis treatment has been also performed to validate the developed analyzer.

Discussed systems can be further improved by adopting some HW/SW solutions, e.g. temperature stabilization, gas sampling and management, advanced classification techniques, innovative reading techniques for sensing.

References

1. Jeong, Y.; Choi—Time horizon selection using feature feedback for the implementation of an E-nose system. *IEEE Sens. J.* **13**(5), 1575–1581 (2013)
2. Saha, P., et al.: Multi-class support vector machine for quality estimation of black tea using electronic nose. In: *Proceedings of the International Conference on Sensing Technology* (2012)
3. Wongchoosuk, C., et al.: WiFi electronic nose for indoor air monitoring. In: *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON* (2012)
4. Güney, S., et al.: Multiclass classification of n-butanol concentrations with k-nearest neighbor algorithm and support vector machine in an electronic nose. *Sens. Actuators B: Chem.* **166–167** (2012)
5. Zhang, L., et al.: Chaos based neural network optimization for concentration estimation of indoor air contaminants by an electronic nose. *S&A, A: Phys.* **189** (2013)
6. Zhanga, L., et al.: Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine. *Sens. Actuators B* **174** (2012)
7. Neri, G., Bonavita, A., Galvagno, S., Pace, C., Donato, N.: Preparation, characterization and CO sensing of Au/Iron oxide thin films. *J. Mater. Sci. Mater. Electron.* **13**, 561–565 (2002)
8. Solga, S.F., et al.: Current status of clinical breath analysis. *Appl. Phys. B* **85** (2006)
9. Goerl, T., et al.: Volatile breath biomarkers for patient monitoring during haemodialysis. *J. Breath Res.* **7** (2013)
10. Neri, G., et al.: Real-time monitoring of breath ammonia during haemodialysis: use of ion mobility spectrometry (IMS) and cavity ring-down spectroscopy (CRDS) techniques. *Nephrol Dial Transp.* **27**, 2945–2952 (2012)
11. Chang, C.C., Lin, C.J.: Libsvm: A Library for Support Vector Machines. Version 3.17 released on April Fools' day, 2013. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
12. Khalaf, W., et al.: Least square regression method for estimating gas concentration in an electronic nose system. *Sensors* **9**, 1678–1691 (2009)
13. Pace, C., Khalaf, W., Latino, M., Donato, N., Neri, G.: E-nose development for safety monitoring applications in refinery environment. *Proc. Eng. Elsevier* **47**, 1267–1270 (2012). ISSN: 1877-7058

Chapter 2

Machine Learning-Based System for Detecting Unseen Malicious Software

Federica Bisio, Paolo Gastaldo, Claudia Meda, Stefano Nasta
and Rodolfo Zunino

Abstract In the Internet age, malicious software (malware) represents a serious threat to the security of information systems. Malware-detection systems to protect computers must perform a real-time analysis of the executable files. The paper shows that machine-learning methods can support the challenging, yet critical, task of unseen malware recognition, i.e., the classification of malware variants that were not included in the training set. The experimental verification involved a publicly available dataset, and confirmed the effectiveness of the overall approach.

Keywords Malware detection · Machine learning · Support vector machine

2.1 Introduction

A large number of computers are everyday infected by viruses, rootkits and trojans, while denial-of-service attacks are carried out by infected networks (“botnets”). The impact of malicious software (*malware*) on both business and Internet security led the scientific community to develop novel protection technologies. It is quite common that new malware variants evolve from old malware. This scenario creates

F. Bisio (✉) · P. Gastaldo · C. Meda · S. Nasta · R. Zunino
DITEN, Polytechnic School, University of Genoa, Via Opera Pia 11a, 16145, Genoa, Italy
e-mail: federica.bisio@edu.unige.it

P. Gastaldo
e-mail: paolo.gastaldo@unige.it

C. Meda
e-mail: claudia.meda@edu.unige.it

S. Nasta
e-mail: stefano.nasta@edu.unige.it

R. Zunino
e-mail: rodolfo.zunino@unige.it

a challenging environment for the reliability of malware detection tools, which cannot rely on conventional, ‘static’ approaches such as the signature method [1, 2].

To address the limitations of signature-based detection techniques, behavior-based antimalware techniques take into account the ‘dynamic’ actions that a program performs at run time [1, 2]. Thus, even when the syntax structure of a malicious software varies, one can exploit the code semantics that the polymorphic variants within the same malware family share. To do so, one executes the tested malware in a controlled environment (‘sandbox’), and monitors the program actions. The resulting execution traces are summarized in reports, and the automated analysis of such reports eventually characterize the type and severity of the possible threat.

The reliable deployment of such a behavior-based technique poses a challenging pattern-recognition task. Firstly, to drive the analysis one has to work out valuable information from execution traces. Secondly, an effective pattern-recognition methodology must be applied. The literature showed that Machine Learning (ML) paradigms can support behavior-based techniques effectively [2, 3]. Those approaches typically characterize suspect executable code by tracking the progression of system calls; different strategies can be applied to achieve this goal [1, 2].

This paper introduces a malware-detection framework based on ML paradigms, which differs from existing ML-based solutions in that it takes into account the non-stationary nature of malware. To measure the performance of a ML-based malware-detection method, one usually chooses a predefined set of malware samples to construct both the training and the test set. Such a procedure, however, does not often match real-world scenarios, in which one must face day-by-day mutations of malicious software.

Therefore, the novelty of the research presented here is a malware-detection framework that attains satisfactory performances even when the tested binary codes do not belong to pre-defined set of malware types. Toward that end, the method adopts a hierarchical approach, suitably designed to classify as malware those samples that do not match the characteristics of benign software.

2.2 The Malware-Detection Framework

Figure 2.1 illustrates the malware-detection system, which includes three steps:

- I. The input binary code is executed in a controlled *sandbox*. The report generated by the sandbox describes the program interactions, such as calls to the Operating System services, including process information (e.g., arguments and return values).
- II. The report undergoes a feature-extraction process. In practice, the report is parsed to assemble a behavioral profile of the executable. The goal is to summarize the code features in a numerical format that is compatible with a ML-based pattern-recognition system.

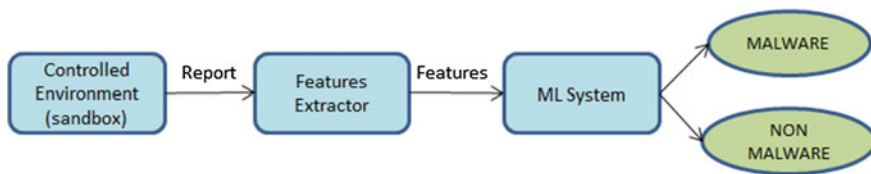


Fig. 2.1 Malware-detection framework

III. A ML-based tool assigns a category (benign/malware) to the tested exec-code. This paper mainly deals with the second and third steps, as they prove to be crucial for the effectiveness of the malware-detection system.

In step 2, the adequacy of the report-generated feature space at rendering the behavior of the executable code heavily affects classification accuracy. Section 2.3 proposes two different feature-extraction approaches that can fruitfully apply in the malware-detection process.

The final step takes advantage of the ability of ML paradigms to deal with multidimensional data characterized by complex relationships, which are learned from examples by using a training algorithm. The ML-based module maps the feature space into the space of target categories (i.e., malware/non-malware).

2.3 A ML-Based Approach to Malware Detection

2.3.1 Feature Extraction

The feature-extraction module processes reports expressed in the *Malware Attribute Enumeration and Characterization* (MAEC) format [2]. MAEC is a standardized language for encoding malware-related information, which includes attributes such as behaviors, artifacts, and attack patterns. In the approach presented here, the feature space characterizes the executable-code activity in terms of API calls. Thus, two descriptors are taken into account:

- Occurrences of API calls (CALLF);
- Occurrences of API categories (CATF).

The former descriptor space considers the number of calls for a set of API's that are deemed relevant; the number, N_{API} , of these API's is a design parameter. Thus feature extraction yields an N_{API} -dimensional vector from a MAEC report; the i -th component of the vector counts the occurrences of the event "the executable code called the i -th API."

The second feature space is designed to aggregate APIs according to a finite set of categories. A category gathers all the APIs that exhibit homogeneous characteristics. For instance, all APIs involving file operations are included in a common

category, whereas those performing registry operations belong to a different category. The design parameter N_{CAT} carries the number of categories of interest. The j -th component of the N_{CAT} -dimensional vector counts the occurrences of the event “the executable code called an API belonging to the j -th category.”

2.3.2 Malware Detection

The basic approach underlying the malware-detection framework stems from two main facts: first, when dealing with malware, one cannot assume that the data distribution is stationary, since new versions of malicious code are released continuously. At the same time, one should consider that the availability of benign software samples is virtually unbounded. To integrate these aspects, the ML-based framework classifies as malware any sample that does not match the characteristics of benign software.

This implies the assumption that the benign-code domain can only be sampled properly. As a consequent requirement, one should limit false negative cases, in which malware is classified as benign software, especially when considering that this kind of errors is critical in a malware-detection domain. Two design approaches are adopted to improve the robustness of the recognition process toward that purpose. Firstly, a two-layer, hierarchical structure allows to benefit from an integrated supervised-unsupervised approach in the set up of the classification tools. Secondly, both feature spaces described in the previous Section are used in the decision-making process. Figure 2.2 outlines the two-layer hierarchical organization.

In the first layer, the input sample is analyzed by a pair of independent modules, one for each feature space described in Sect. 3.1. Each module, in turn, includes three classifiers (in the present implementation, three binary Support Vector

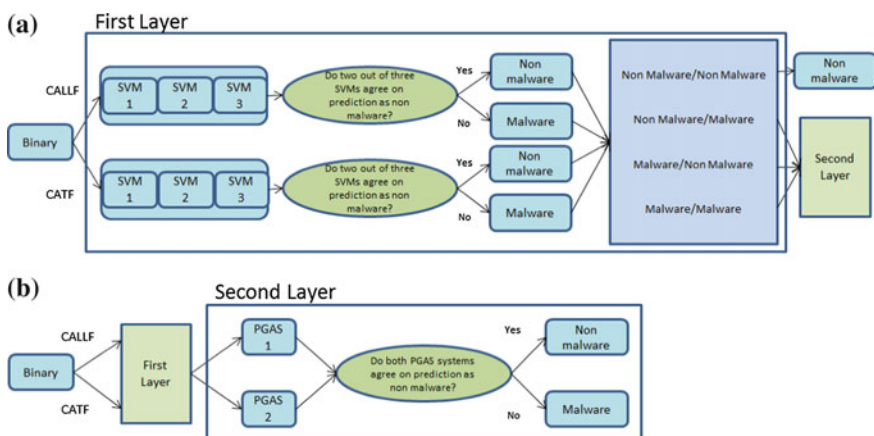


Fig. 2.2 The malware-detection system: **a** first layer; **b** second layer

Machines [4]), and applies a majority-based voting scheme. To set up the two modules, a specific training set is used to tune each SVM classifier. The samples included in each training set are evenly divided between ‘benign’ code and ‘malicious’ code. The same group of malicious code samples is used in all training sets, whereas the three training sets do not share any benign binary.

If both modules agree on categorizing the tested code as benign, that code is classified accordingly. Otherwise, the second layer takes over, and applies clustering paradigms to implement a decision function. This layer still integrates a pair of (unsupervised) learning systems, each operating on a specific feature space. In practice, each classification system is trained with the following procedure:

1. generate a training set evenly divided between benign binaries and malicious binaries;
2. work out clusters of similar code descriptions by applying an unsupervised clustering technique;
3. calibrate each cluster with a classical, majority-based criterion: the class of the cluster is the predominant category within the cluster.

The present implementation adopts the Plastic Neural Gas (PGAS) clustering algorithm [5] to support the second layer training.

2.4 Experimental Results

2.4.1 *Experimental Set up*

The experimental verification of the proposed malware-detection framework involved the monitoring of a computer system supported by a Microsoft Windows OS. The dataset included 17010 binaries, which were worked out from the publicly available database of binaries provided by the CWSandbox web site [6]. The database consisted of two dataset: the ‘reference’ dataset, including 3133 malware samples, and the ‘application’ dataset, covering 33698 binaries (both benign and malware samples). Each malware was labeled by a ‘malware class,’ which identified the type of malicious action performed by the malware. A total of 415 classes were provided.

A set of 5580 binary codes formed the training set, which covered malicious and benign code samples in the same proportion. The malicious binaries were drawn from the ‘reference’ dataset of the CWSandbox database, and spanned a total of 24 different malware classes. The generalization ability of the malware-detection approach was tested by using a set of 11430 binaries, which included 10220 malicious binaries and 1230 benign binaries. To verify the practical impact and the actual consistency of the overall approach, the malicious test samples never belonged to the 24 classes used for training. As a result, the experimental verification addressed a real-world scenario, in which the malware-detection system is expected to analyze binaries that may not belong to known malware classes.

The experiments exploited the open-source Cuckoo Sandbox [7], and reports were formatted according to the MAEC standard. With respect to the discussion in Sect. 3.1, the feature space counting API calls was designed to trace 144 different APIs (i.e., $N_{\text{API}} = 144$); in the following, this feature space will be denoted as ‘CALLF’. Conversely, the feature space grouping API families (‘CATF’) used 9 categories to represent the APIs (i.e., $N_{\text{CAT}} = 9$), namely: file operations, registry operations, window operations, synchronization, process operations, thread operations, network operations, services, other.

2.4.2 Classification Performance

All the SVM classifiers involved in the malware-detection system adopted a Gaussian kernel. A conventional cross-validation procedure supported the adjustment of the machine parameters; to this purpose, a validation set was drawn from the training set. Table 2.1 shows—for each classifier and feature space—the eventual setting for parameters C and σ resulting from the model-selection process.

Table 2.2 reports on the performance attained by the malware-detection system on the test set. The table gives two main descriptors: the detection accuracy (i.e., the percentage of malicious binaries correctly classified as malware) and the false positives rate (i.e., the percentage of benign binaries that have been erroneously classified as malware). Table 2.2 shows that the proposed framework was able to score a detection accuracy greater than 93 %. This in turn means a false negatives rate inferior to 7 %.

A comparison between the proposed framework and other ML-based approaches to malware detection may be complicated by the dissimilarity in the prescribed targets. Rieck et al. [2] tested their malware-detection system by using the CWSandbox database; however, they focused on the detection of malware types that were included in the training set. In [2], the F-measure has been adopted as performance metric; accordingly, the system scored 0.95 when adopting clustering techniques

Table 2.1 Model selection

Feature space		σ	C
CALLF	SVM1	4	1
	SVM2	8	1
	SVM3	8	1
CATF	SVM1	0.5	1
	SVM2	0.5	1
	SVM3	0.5	1

Table 2.2 Performance of the malware-detection system

Descriptor	Value (%)
Detection accuracy	93.37
False positives rate	37.53

and 0.98 when adopting classification techniques. The values reported in Table 2.2 prove that the present approach was able to achieve an F-measure of 0.94 when tested on unseen malware. This outcome indeed confirms the effectiveness of the proposed method. In [3] malware detection is implemented by using SVM classifiers; the paper reported a 93 % of detection rate on unknown malware. In fact, the authors did not exploit a publicly available database; hence, a direct comparison with the framework introduced in the present work cannot be performed.

2.5 Conclusions

Day-by-day mutations of malicious software represent a challenging issue when developing a malware detection system. This paper introduced a ML-based approach to malware detection that can tackle this aspect. A hierarchical structure allows the framework to benefit from an integrated supervised-unsupervised approach in the setup of the classification tools, which tag as malware those samples that do not match the characteristics of benign software. As a future development, the possibility of a hardware implementation is appealing, even if related to the difficult and still unsolved task of the hardware implementation of the SVM.

References

1. Kolbitsch, C., Milani, P., Kruegel, C., Kirda, E., Zhou, X., Wang, X.: Effective and efficient malware detection at the end host. In: Proceedings of the 18th USENIX Security Symposium (Security '09), pp. 351–366, Montreal, Canada, Aug 2009, USENIX (2009)
2. Rieck, K., Trinius, P., Willems, C., Holz, T.: Automatic analysis of malware behavior using machine learning. *J. Comput. Secur.* **19**(4), 639–668 (2011)
3. Kolter, J.Z., Maloof, M.A.: Learning to detect and classify malicious executables in the wild. *The. J. Mach. Learn. Res.* **7**, 2721–2744 (2006)
4. Vapnik, V.N.: *Statistical Learning Theory*. Wiley, New York (1998)
5. Ridella, S., Rovetta, S., Zunino, R.: Plastic neural gas for adaptive vector quantization. *IEEE Trans. Neural Netw.* (2001)
6. Willems, C., Holz, T., Freiling, F.: CWSandbox: towards automated dynamic binary analysis. *IEEE Secur. Privacy* **5**(2) (2007)
7. Cuckoo Sandbox. <http://www.cuckoosandbox.org>

Chapter 3

Implementation of a Spread-Spectrum-Based Smart Lighting System on an Embedded Platform

Maurizio Martina, Massimo Ruo Roch and Flavio Ghirardi

Abstract In the last years smart lighting systems have attracted a lot of attention due to the increasing interest in reducing wasted power consumption. This work describes the implementation on an embedded platform of a smart lighting system, where the lamps communicate together creating a cooperative network, to trim the amount of light a given place. The proposed implementation relies on the spread spectrum technique and on optical orthogonal codes borrowed from optical communication research. Experimental results performed on a Freescale Freedom board, prove the feasibility of the proposed system.

Keywords Smart lighting · Embedded systems · Spread spectrum

3.1 Introduction

Nowadays power consumption is a critical issue in several aspects of people life. Indeed, portable devices are one of the most known examples of power critical applications. Besides, other every-day-applications, such as lighting systems, give

M. Martina (✉) · M.R. Roch

Electronics and Telecommunications Department, Politecnico di Torino, Turin, Italy
e-mail: maurizio.martina@polito.it

M.R. Roch

e-mail: massimo.ruoroch@polito.it

M. Martina

Neodelis, Turin, Italy

M. Martina · M.R. Roch · F. Ghirardi

Xaluxi, Turin, Italy

F. Ghirardi

e-mail: fghirardi@neodelis.com

© Springer International Publishing Switzerland 2016

A. De Gloria (ed.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 351,
DOI 10.1007/978-3-319-20227-3_3

a relevant contribution to power consumption. In the last years several efforts were spent to improve the efficiency of lighting systems moving from incandescence to fluorescence lamps. Recently, Light-Emitting-Diode (LED) technology has become an effective solution to further improve the efficiency of lighting systems. Indeed, LED lighting offers not only power reduction but also several interesting characteristics, including easy change of light color and dimming. Furthermore, LED-based lamps enable the possibility to implement smart lighting systems, where the lamps are part of a cooperative network. Indeed, lamps can communicate together to trim lighting parameters and to create light effects. Moreover, this communication structure can be employed as the substrate for implementing an alternative wireless local area network.

Several works in this field have been proposed in the last years, e.g. [1–3]. In [1] uniform illumination is studied and lamps of different shapes are addressed. A novel illumination model and its implementation on a wireless sensor network are proposed in [2]. Similarly [3] deals with the analysis and optimization of an indoor visible light communication system. An interesting approach is proposed in [4], where spread spectrum theory is adapted to LED based lighting systems. Stemming from [1] we propose to build a network whose Medium-Access-Control (MAC) level relies on spread spectrum. This solution has the great advantage of being compatible with existing radio-frequency and wireless standards for communications. To prove the feasibility of the proposed solution, a prototype based on the Freescale Freedom Systems [5] has been proposed. The prototype relies on a Kinetis-L microcontroller, that is build around the ARM Cortex-M0+ processor and equipped with 128 kB of flash memory and 16 kB of RAM memory.

3.2 System Overview

Implementing a smart lighting system that is organized as a cooperative network requires each node to communicate with the other ones. This can be accomplished by superposing communication to the lighting signal by the means of an optically modulated signal. One of the first tasks of a smart lighting system that works as a cooperative network is to tune the intensity of each lamp. Since we employ LED-based lamps this operation can be accomplished by driving each lamp with a variable duty-cycle, square waveform. Indeed, the intensity of the light produced by each lamp is modulated by the duty-cycle of the square waveform. Moreover, to tune the intensity of each lamp a system able to separate the different contributions to the total light is required. Namely, each lamp should understand its own contribution to the total light and, if necessary, should adjust it. As described in [4], a viable solution to identify the contribution of each lamp to the total light is using code diversity. In particular, stemming from the spread spectrum principles of Code-Division-Multiple-Access (CDMA) systems, that are employed in several telecommunication standards, the light contribution of each lamp is univocally identified by a codeword, that is orthogonal to all the other ones.

3.2.1 CDMA Fundamentals

CDMA is a technique to access a shared channel resorting to orthogonal codes [6]: the main concept is that every user is able to access simultaneously the channel, spreading its information on certain band. The band reserved for the communication (B_{chip}) ought to be wider than the one required by the information itself (B_{bit}), in order to allow the spreading operation. In Fig. 3.1 the spreading operation, both in time and frequency domain is depicted, where $T_{bit} = 1/B_{bit}$ and $T_{chip} = 1/B_{chip}$.

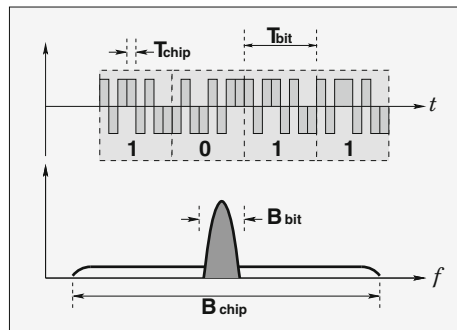
From an implementation point of view the spreading operation can be performed multiplying together the original signal \mathbf{x} = (with band B_{bit} and energy \mathcal{E}_x) and the code \mathbf{a} (with band $B_{chip} \gg B_{bit}$ and energy $\mathcal{E}_w = 1$). The resulting signal \mathbf{t} exhibits the same energy of the original signal (\mathcal{E}_x) spreaded over B_{chip} . If several orthogonal codes are available, more users can access the same channel, provided that each user spreads with a different code. The rate of \mathbf{x} is referred to as *bit-rate* and the rate of \mathbf{a} and \mathbf{t} as *chip-rate*.

The receiver recovers the information by multiplying the received bit stream \mathbf{r} against \mathbf{a}_l , the same code employed by the transmitter (for the l -th user). If $\hat{\mathbf{y}}$ is the signal after the multiplication, the original information, \mathbf{x} , can be obtained summing $\hat{\mathbf{y}}$ on T_{bit} .

CDMA-based communications exhibit several advantages if compared with other channel access techniques. For instance, given a spreading factor $G = B_{chip}/B_{bit}$, CDMA allows a large number of users to share simultaneously the channel, with a reduced sensitivity to fading effects. Moreover other narrow band access techniques can be employed on the same band: due to the despreading operation the receiver will spread any narrow band contribution over B_{chip} .

The simple concepts behind the CDMA technique hide, as a counterpart, many challenges from a design perspective. Since the signal \mathbf{x} is multiplied by the code \mathbf{a} , the resulting signal rate will become very high. This constrains the receiver to perform the first operations in the decoding chain, in particular the synchronization and the despreading process, working with very high data rates [7, 8].

Fig. 3.1 Spread spectrum transmission principles



Another critical aspect in a CDMA receiver is the synchronization. Indeed, the CDMA receiver can recover the original signal only if it is perfectly synchronized with the transmitter [6]. To achieve synchronization the most common strategy relies on the employment of an header. Namely, the header is a control bitstream, used only to assure the receiver to be synchronized and to start properly the decoding operations. The main task performed by the synchronizer is to evaluate the energy of $\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is the signal obtained integrating $\hat{\mathbf{y}}$ on T_{bit} . The correct operation of the system is strongly based on a characteristics: the code \mathbf{a} satisfies some pseudo-noise properties, in particular its autocorrelation $r_w(\vartheta)$ shows a peak for $\vartheta = 0$ [6]. The pseudo-noise properties of spreading/despreading sequences ensures $\hat{\mathbf{x}}$ energy to show a peak when the synchronization is reached. This operation can be performed comparing the energy measured with a programmable threshold: if the energy is greater than the threshold the synchronization has been accomplished. On the other hand, if the energy is under the threshold, the measure ought to be repeated with a different phase shifting.

3.2.2 Data Shaping

Since the proposed system uses optical signals, the OOK modulation is employed to transmit the signal. Thus, optical orthogonal codes as the ones in [9] have to be employed. A similar idea using Walsh-Hadamard sequences is presented in [4]. In this work the spreaded sequence \mathbf{t} is mapped to pulses of different width as detailed in the following. Terminology and formulation are similar to the ones presented in [4]. Let T_1 be a *slot*, namely the clock period used to drive a LED-based lamp. A *block* is a period $T_2 = N_1 \cdot T_1$ to represent one LED pulse, that is T_{chip} in CDMA terminology. Similarly, a *frame* is the period $T_3 = N_2 \cdot T_2$ to transmit one of the codes

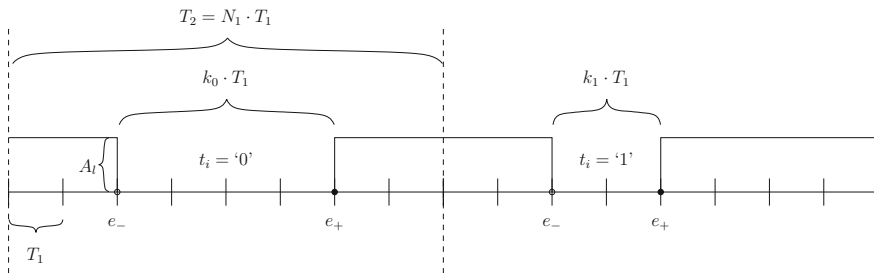


Fig. 3.2 Example of data shaping

that identifies a lamp. As a consequence, T_3 represents T_{bit} in CDMA terminology. Since an optical code of length N_2 is made of several '0' and few '1' [9], direct OOK modulation of \mathbf{t} gives an unacceptable effect on the lighting systems. Thus, we map each *chip* of \mathbf{t} to a pulse as depicted in Fig. 3.2, namely each chip is mapped to a high-to-low (e_-) transition followed by a low-to-high transition (e_+), where A is the pulse amplitude. Let t_i be the i -th chip of \mathbf{t} and p_i the corresponding pulse, then the distance between e_- and e_+ , that is a multiple of T_1 , is $k_0 \cdot T_1$ if the $t_i = \}0'$ and $k_1 \cdot T_1$ if the $t_i = \}1'$. Moreover, as in optical orthogonal codes the occurrence of $t_i = \}0'$ is larger than that of $t_i = \}1'$, we choose $k_0 = N_1/2$, so that the transmitted signal $\mathbf{p} = \{p_0, \dots, p_{N_2-1}\}$ is almost always a 50% duty cycle signal.

3.3 System Implementation

In this application we consider each lamp as made of a stripe of LEDs driven by a supply circuit. In order to apply the technique proposed in Sect. 3.2.2, the supply circuit must have a dimming input. For the sake of simplicity in this work we added an MOS transistor in parallel to the LED stripe (see Fig. 3.3). Thus, even if not optimal, the proposed solution can be used with any lamp and acts as a proof of concept. The MOS is driven by the signal \mathbf{p} obtained with the proposed shaping as described in Sect. 3.2.2. The algorithm to map the data to the signal \mathbf{p} is implemented on the Freedom FRDM-KL25Z platform by Freescale [5]. A periodical interrupt timer has been programmed to read at every *slot* T_1 a buffer containing the pulses corresponding to the values of the spreaded sequence (see Fig. 3.4a). At the receiver side the lighting signal is converted by a photodiode into a digital waveform. Then, a periodical interrupt timer has been programmed to sample the waveform every $T_1/4$ and the sample is put in a buffer. The detection of t_i is implemented comparing the content of the buffer with two masks describing the two possible pulses depicted in Fig. 3.2. Each detected t_i is used to fill a buffer and when the buffer is full the processor toggles a signal to start the synchronization. The synchronization routine of the l -th receiver simply computes the energy of the product between the buffer and the l -th orthogonal code. The position of the maximum indicates where the received sequence and the code are aligned. When the receiver is synchronized a second routine performs the despreaded searching for the header. Once the header is detected, the receiver can recover the data sent by the transmitter (see Fig. 3.4b). In this work the system has been trimmed to work with $T_1 = 40 \mu\text{s}$ and the implemented prototype can support different codes up to $N_2 = 256$. The building blocks described in [10–12] can be used to design a hardware efficient modulation/demodulation system.

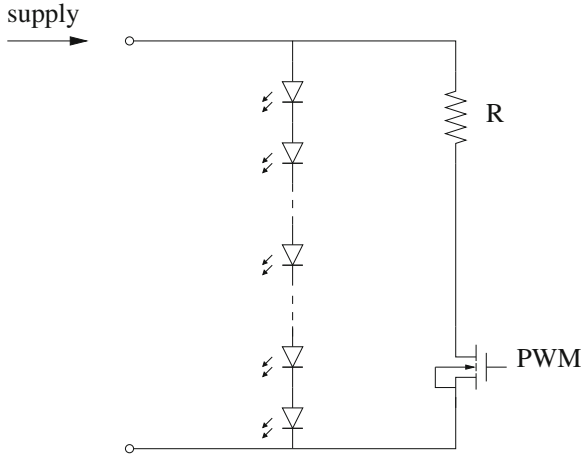


Fig. 3.3 Driving scheme

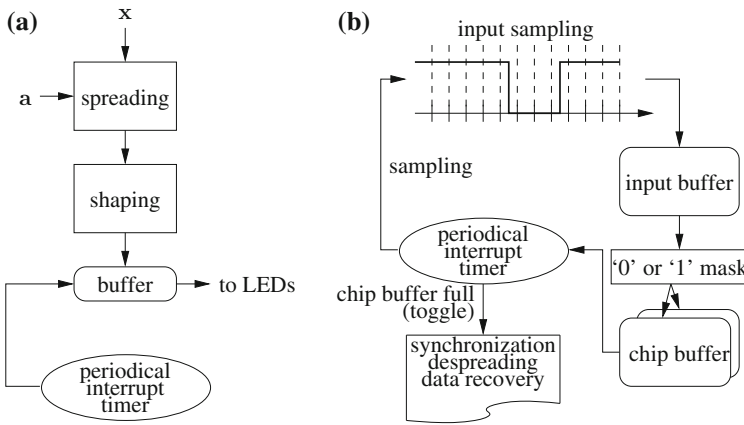


Fig. 3.4 Simplified flow-chart of the proposed system: transmitter (a) and receiver (b)

3.4 Conclusion

In this work a smart lighting system based on the spread spectrum theory and optical orthogonal codes has been described and implemented on an embedded platform, namely a Freescale Freedom board. Experimental results implementing both the transmitter and the receiver show the feasibility of the system. As a future work on field experiments and bit-error-rate measurements have been planned.

Acknowledgments The authors would like to thank Neodelis for supporting this activity.

References

1. Yang, H., Bergmans, J.W.M., Schenk, T.C.W., Linnartz, J.P.M.G., Rietman, R.: Uniform illumination rendering using an array of LEDs: a signal processing perspective. *IEEE Trans. Signal Process.* **57**(3), 1044–1057 (2009)
2. Bhardwaj, S., Ozcelebi, T., Verhoeven, R., Lukkien, J.: Smart indoor solid state lighting based on a novel illumination model and implementation. *IEEE Trans. Consumer Electron.* **57**(4), 1612–1621 (2011)
3. Cui, K., Chen, G., Xu, Z.: Line-of-sight visible light communication system design and demonstration. In: *International Symposium on Communication Systems Networks and Digital Signal Processing*, pp. 621–625 (2010)
4. Linnartz, J.P.M.G., Feri, L., Yang, H., Colak, S.B., Schenk, T.C.W.: Code division-based sensing of illumination contributions in solid-state lighting. *IEEE Trans. Signal Process.* **57**(10), 3984–3998 (2009)
5. freescale: <http://www.freescale.com>
6. Viterbi, A.: *CDMA: Principles of Spread Spectrum Communications*. Addison-Wesley Publishing Company, Redwood City (1995)
7. Louveaux, J., Vandendorpe, L., Cuvelier, L., Pollet, T.: An early-late timing recovery scheme for filter-bank-based multicarrier transmission. *IEEE Trans. Commun.* **48**, 1746–1754 (2000)
8. Rappaport, T.: *Wireless Communications: Principles and Practice*. Prentice Hall, Upper Saddle River (1997)
9. Chung, F.R.K., Salehi, J.A., Wei, V.K.: Optical orthogonal codes: design, analysis and applications. *IEEE Trans. Inf. Theory* **35**(3), 595–604 (1989)
10. Cerato, B., Colazzo, L., Martina, M., Molino, A., Vacca, F.: Parametric FPGA early-late DLL implementation for a UMTS receiver. In: *Asilomar Conference on Signals Systems and Computers*, pp. 1069–1072 (2002)
11. Martina, M., Molino, A., Nicola, M., Vacca, F.: Design of a power conscious, customizable CDMA receiver. *Lect. Notes Comput. Sci.* **2778**, 1028–1031 (2003)
12. Zicari, P., Corsonello, P., Perri, S.: A high flexible early-late gate bit synchronizer in FPGA-based software defined radios. In: *European Conference on Circuits and Systems for Communications*, pp. 252–255 (2008)

Chapter 4

Self-powered Active Cooling System for High Performance Processors

Maurizio Rossi, Luca Rizzon, Matteo Fait, Roberto Passerone
and Davide Brunelli

Abstract Thermal stability in datacenter's computing units is fundamental to ensure reliability, and durability of the equipment, besides, environmental concern and new regulations require a reduction of the power used. For these reasons, a novel energy neutral hybrid cooling system is proposed. We describe the design, and the prototype's performance evaluated both in passive and active cooling modes. During normal operating conditions, the thermo-electric energy harvesting system transforms wasted heat into electric energy, and stores it in super-capacitors while the system is providing passive cooling. Active cooling can be activated when a boost in performance requires CPU overclocking, using free energy from the passive step. After the choice of the most suitable harvesting system we designed and tested the prototype on an ARM based CPU, the future core of low-power server architectures. The proposed governor switches to active cooling mode based on customizable thermal management policies. Experimental results demonstrate good passive cooling performance, and several minutes active cooling exploiting the recovered heat.

Keywords Thermoelectric generators • Energy harvesting • Hybrid CPU cooling

M. Rossi (✉) · L. Rizzon · M. Fait · R. Passerone · D. Brunelli
University of Trento, Via Sommarive, 9, 38123 Trento, Italy
e-mail: Maurizio.Rossi@unitn.it

L. Rizzon
e-mail: Luca.Rizzon@unitn.it

M. Fait
e-mail: Matteo.Fait@unitn.it

R. Passerone
e-mail: Roberto.Passerone@unitn.it

D. Brunelli
e-mail: Davide.Brunelli@unitn.it

4.1 Introduction and State of Art

Heat is generated by any computing equipment as an unwanted result of processing. It depends on fabrication technology, operating frequency, and voltage. Particularly, CPUs, and GPUs are responsible for the highest heat dissipation in modern general-purpose equipment. To ensure long lasting lifetime and reliability, the thermal stability of such devices must be carefully managed to avoid micro-structural damages and malfunctioning. Besides passive cooling, using metallic elements to ease the heat transfer with the surrounding environment, common active cooling system exploit fans or pumped fluids to keep the temperature under control. Whenever packaging constraints doesn't allow to use such solutions, Dynamic Voltage and Frequency Scaling (DVFS) helps to prevent overheating by actively reducing CPUs' operating frequency (reducing the supply) [1, 2]. The wide spreading of mobile equipment and low-power ARM architectures, which join high performance computing with low voltage and thermal dissipation, pushes this trend forward. Indeed, analysts forecast a increased market share of ARM-architectures even in high performance datacenter environments for the next decade [3]. Datacenters are always on to ensure reliability and Quality-of-Service (QoS) but their energy efficiency is very poor: only the 0.5 % of the total consumed energy is spent to perform useful computations [4], and they are source of a huge thermal dissipation.

In this work we present the design of a hybrid cooling system that mixes passive and active components in a self-powered device, which exploits thermal energy. The passive cooling is provided by the thermo-electric energy harvesting system itself that, in turn, supplies a fan for the forced-convection active cooling. The design of the Thermal Energy Recovery System (TERS) is based on three modules: the thermo-electric harvester, the conditioning circuit, and super-capacitors as storage unit. The prototype's performance has been evaluated both on passive, and active mode of operation. In the latter case we use scavenged energy to supply a fan, and evaluated the boost in cooling to eventually exploit CPU's overclocking.

Thermo-electric energy harvesting has been already presented in literature and similar designs have been proven effective in converting the heat dissipated by computing equipment [5]. However the recovered energy is too small to extend the autonomy, or to replace the primary supply of the cooling system [6]. In [7] a Thermo-Electric Generator (TEG) has been applied on top of a multi-core desktop processor, and the energy scavenged is used to supply a cooling device. However, this solution resulted in up to 25 °C increase of the CPU temperature, causing the computing unit to work near its allowed limit. The proposed system, instead, provides passive cooling during normal workload without impacting the thermal stability; meanwhile the super-capacitors store the recovered energy. Eventually, the energy is used to supply the active cooling whenever overclocking is required. Electrical power generated by TEGs can also be used to supply context-aware computing for monitoring application, and hazard detections [8]. The alternatives offered by those devices generates interest also from the modeling view point, since

having parametric models offers the possibility to designers to determine a priori the amount of power available from the harvesting device when exposed to given thermal gradients [9–11]. Similar designs of thermal energy harvesters are used to supply energy-neutral sensors for Wireless Sensors Networks [12].

In the following, Sect. 4.2 describes thermo-electric energy harvesting devices and performance that are subsequently used to design the hybrid system presented in Sect. 4.3. Section 4.4 presents the experimental results both in terms of passive and active cooling. Finally, Section 4.5 presents conclusive remarks.

4.2 Thermo-Electric Energy Harvesting

TEGs are based on the Seebeck principle: when a temperature difference is applied to a junction of two different metallic materials, a voltage difference is generated between the two faces. The resulting electrical energy is proportional to the thermal gradient (ΔT), but the maximum allowed hot-side temperature must be lower than 80 °C to avoid structural damages, performance degradation and reduced lifetime [13]. This temperature threshold is compliant with ARM CPUs' working conditions. We compared three commercial TEGs to select the most suited to characterize, and build the hybrid cooler: (a) Nextreme eTEG HV56 Thermoelectric Power Generator; (b) Micropelt TE-CORE7 TGP-751 Thermo-Harvesting Power Module with a 33 mm heat sink; (c) Peltier-Cell PE1-12706AC 40 × 40 mm squared cells, that is intended for cooling applications. We realized two custom configurations using the cheap Peltier cells: (i) single layer using one device, and (ii) double layer made by using two stacked cells electrically connected in series. Aluminum sheets were used as heat spreaders between the CPU and the (multiple) layers, while a heat sink on top allows optimal dissipation with the surrounding environment.

We performed the TERS comparison evaluating their electrical power output using a Samsung Arndale Board as high-performance computing equipment; it embeds an Exynos 1.7 GHz dual-core ARM Cortex A15 processor, running Linux kernel v. 3.10. We designed a custom benchmark made of a set of task in sequence to achieve different utilization profiles at the maximum CPU frequency: (a) video encoding using ffmpeg (GPL software) with four threads; (b) multithread application that performs computations with floating point numbers using parallel threads; (c) kernel operations, that are uncompressing, compiling, cleaning and removing the directory. To select the best TERS, we measured the voltage, and the current drained on a matched load with 1 s sampling interval.

Both custom TERSs made of Peltier cells, thanks to the larger size with respect to the others, help to decrease the thermal gradient by more than 20 °C (single cell) and 5 °C (double cells) at the maximum operating point. The maximum hot-side temperature reached during double Peltier test was ~50 °C (~75 °C on the CPU chip) which resulted in the best dissipation, and almost twice the output power of the single cell solution (2.6 vs. 3.6 mW).

The good dissipation performance are reflected also on the total energy scavenged, and benchmark execution time (26 J in 4 h 30 m with double Peltier cell vs. 24 J in 5 h 4 5 m with Nextreme and 20 J in 5 h 20 m with Micropelt), because the commercial devices cannot fully exploit rapid temperature variations that take place in the CPU. The smaller dimensions do not allow dissipating the heat completely and fast enough, hence the DVFS kernel’s module slows down the execution speed resulting in the longer execution time. Based on these observations, we built the energy neutral hybrid cooling system described below starting from the custom double Peltier cells TERS.

4.3 Hybrid Cooling

Thermal stability in datacenters’ computing units depends on their working frequency that is eventually affected by the consolidation policy used. Consolidation means the selection of the number of active servers and corresponding frequency to achieve a specific QoS, and reducing the most the power consumption. Nonetheless, in some applications—for example in high performance computing (HPC) for scientific applications—a boost in performance is sometimes necessary to fulfill specific requirements. The proposed hybrid cooling system is particularly oriented to this scenario where CPU, and cooling can take advantage of the work in symbiosis. In this manner, it is possible to achieve a boost in performance, without compromising thermal stability, and without requiring additional power for cooling.

In this first prototype we realized the conditioning and storage unit on a separate circuit and the governor on a dedicated external PC (Fig. 4.1). The storage uses two super-capacitors to collect the energy from the conditioning circuit while a $40 \times 40 \times 5$ mm fan (5 V, 250 mA rated) realize the active element. To reduce the harvesting phase’ length we selected, by means of further experiments, a

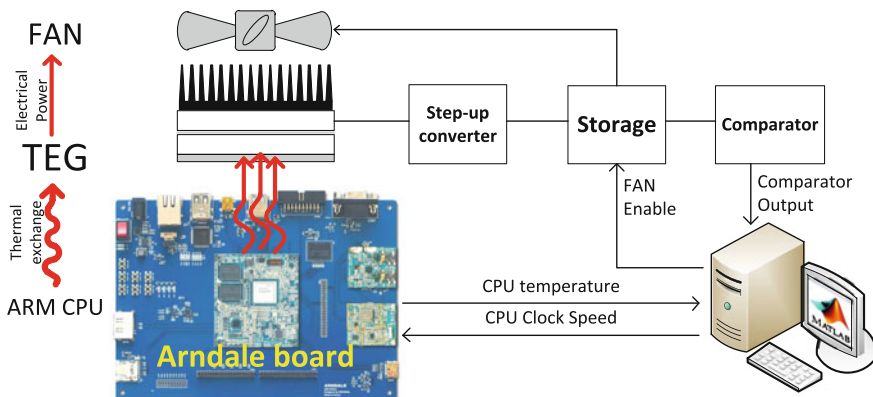


Fig. 4.1 Prototype system block diagram

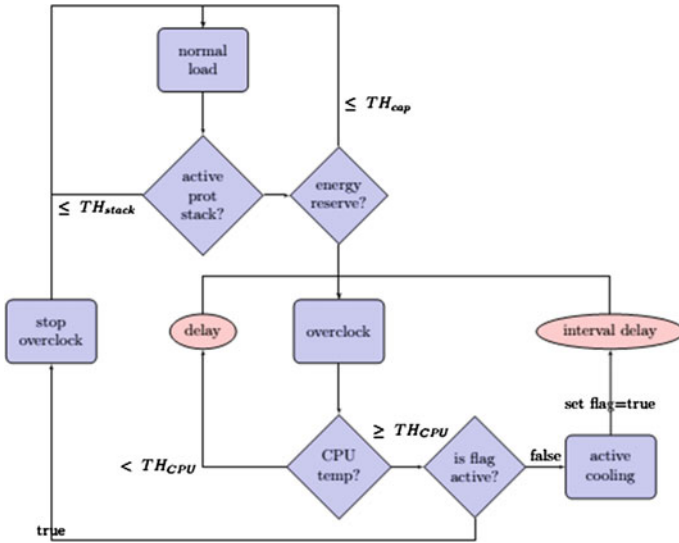


Fig. 4.2 Controller’s flowchart

lower-than-nominal supply voltage (4.4 V) still enough to cool down the heat-sink. The storage comprises two 50 F 2.5 V super-capacitor (due to the high energy required by the fan), and the necessary electric components to switch between charge in parallel configuration, and discharge (supply the fan) in series. This solution allows us to speed up the recharging phase up to 2.2 V (TH_{cap}), and to immediately switch to 4.4 V when it is required.

In our tests the storage is recharged by the TERS until it reaches TH_{cap} , and then switched to series, the resulting 25 F provide enough energy to the fan thanks to the low internal resistance of super-capacitors. A comparator has been used to monitor the storage, and trigger the switching of super-capacitors arrangement. Comparator maximum (V_{max}) threshold is set equal to TH_{cap} , and the minimum is 2.0 V (this last is not used since the active phase’s switch off is managed by the governor).

The flowchart in Fig. 4.2 summarizes the behavior of the governor, which is meant to extend the kernel’s DVFS module. It has been implemented in Matlab, and the data with the Arndale board (frequency, load and temperature) are exchanged by means of the serial communication. Multimeters, and programmable voltage supply were used to monitor the equipment (comparators, storage and temperature probes), and drive active circuitry respectively.

The performance characterization presented in the next section describes the cooling performance in (a) **Standard modality**, when the CPU’s working frequency is intermediate (f_{CLK}), and the hybrid cooling works in passive mode; and (b) **Overclock modality**, when the CPU is pushed at the maximum frequency, the workload approaches 100 % and the hybrid cooling in active mode allows to extend the time with overclock enabled. The governor handles the switching between modalities based on the storage’s state of charge, it periodically polls the output of

the comparator, and only when it is high a new overclocking phase can be triggered. In our tests the overclocking is enabled as soon as the comparator drives its output to the high state, while the fan is kept switched off until the CPU's temperature reaches a predefined temperature threshold TH_{cpu} . Afterwards, when the fan is switched off (and the temperature has been decreased below the threshold) the overclocking can continue until TH_{cpu} is reached again. This policy allows maximizing the time spent in overclocking modality because the temperature increase is slowed down by the presence of the TERS, and the energy scavenged increases with ΔT . The CPU reaches the TH_{cpu} threshold twice, the first time the fan is activated, the second time the system returns to standard modality.

4.4 Experimental Results

In this section, we present and discuss the performance evaluation of the proposed device in terms of energy harvested and amount of heat removed from the processing unit. The processor workload is simulated by running the multithread mathematical application since it allows obtaining an almost flat temperature on the chip, and consequently, constant TERS output power. We are able to modulate the activity of the system by selecting directly the CPU clock speed, and the percentage of time assigned to the application (load). In this manner, we can simulate different workload conditions suitable to evaluate the relation between the workload, and the corresponding cooling device performance. By modeling the datacenter scheduling, and task demand, and using the input-output relation of the proposed system, it is possible to determine the performance of the system in a real deployment.

The lower recharge time can be obtained with the CPU load set at 100 %. We identified two modalities for the CPU speed: (i) the standard condition, during which the clock frequency is set to 1.5 GHz, and the (ii) overclock which corresponds to the CPU maximum designed speed of 1.7 GHz. We do not go beyond the frequency limit imposed by the manufacturer to avoid excessive overheat, and temperature bias outside the operating region of common CPUs. The application controlling the activity of the CPU, and of the harvesting/cooling device monitors the CPU temperature, and it intervenes whenever the temperature goes over $TH_{\text{cpu}} = 68$ °C. This threshold has been set as a precaution to avoid approaching temperature limit suggested by the vendors of the CPU and of the TEG. If the 80 °C limit is overtaken, the DVFS of the CPU will intervene to cool down the chip package, eventually the system may automatically shutoff to prevent micro structural damages. During harvesting, the CPU activity is set in standard mode. The custom harvester ensures a good amount of heat exchange with the air, in fact the CPU without any passive cooling device on top of the CPU package working in standard mode will reach the temperature threshold in 5 min, while with the harvester on top it can run for hours. As shown in Fig. 4.3, the CPU temperature in this phase is almost constant 62 °C and inside the safety region. Under these conditions, the harvesting circuit output is about 300 μA .

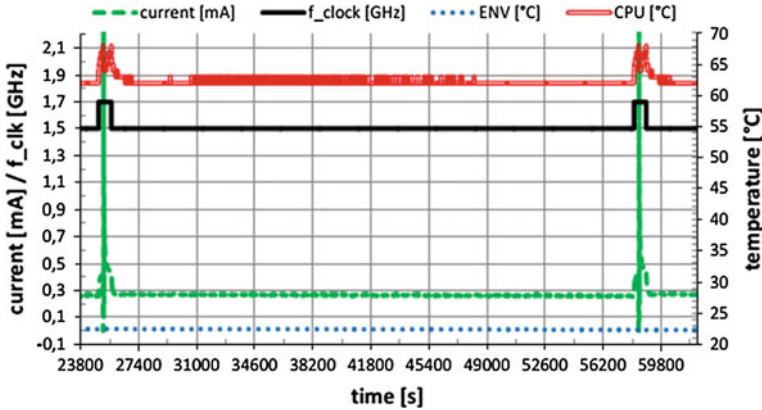


Fig. 4.3 Long-term temperature stability during passive mode

Active cooling can be performed by activating the mechanical fan for an amount of time selected according to the ambient temperature. We analyze the CPU temperature behavior when running the fan for a variable amount of time, ranging from 2 up to 89 s. Figure 4.4 shows the temperature gradient obtained while running the CPU at full clock speed and 100 % load. Providing airflow for 34 s allow the CPU temperature to decrease by 10 °C. In our prototype, the maximum amount of time the system supplies the fan corresponds to 30 s and has been imposed because of the recharging time to recover the amount of required energy. We evaluated the system by running successive experiments for several days, and we obtained an average overclocking time of 13 min (as in Fig. 4.5), while the time required to scavenge energy (corresponding to the time last between two overclocking phase, as in Fig. 4.4) is about 10 h. However, system performance depends on environmental

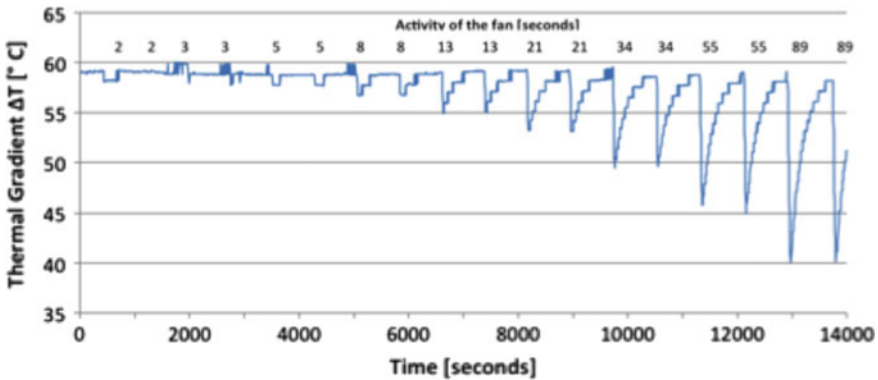


Fig. 4.4 Temperature on the CPU while running at full clock speed and 100 % load. The fan spins for a variable amount of time ranging from 2 up to 89 s

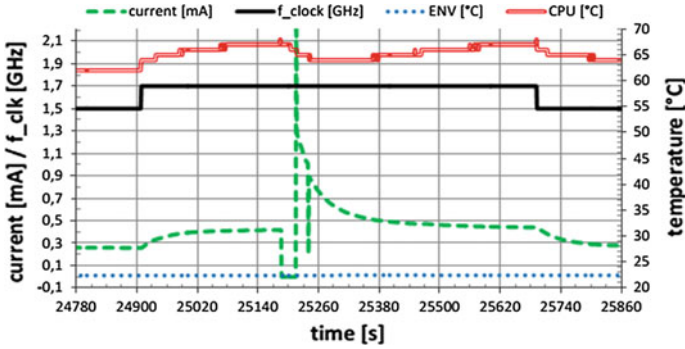


Fig. 4.5 Active cooling in overclock mode. Supply the fan for 30 s makes it possible to control the chip temperature for several minutes during which the CPU is overclocked

temperature. Consequently, in a real deployment with many harvesting/cooling devices mounted within a server, they will perform differently depending on where they are placed given the great variability of temperatures in the cabinets. The devices places near to a fresh air inlet will have faster recharge time with respect to harvesters that receive a flow of relatively hot air.

4.5 Conclusions

We present a device that combines passive chip cooling with self-powered active cooling performed with mechanical fan. Self-powering is achieved thank to thermoelectric harvesting performed while working as a passive cooler. Passive cooling performance have been characterized under different conditions and the pack consisting of aluminum spreaders, Peltier cells and heat sink ensure to always operate under safety temperature threshold. Duty cycled forced convection allow the CPU temperature to drop by 5 °C during a phase in which the CPU works at its full potential, sustaining a period of computational boost that can last for 13 min. The prototype we realized requires about 10 h to collect the energy needed to perform its activity. Adopting novel start-up circuit that reduces initial recharge time can further enhance the performance of the proposed system. Currently, the authors are focused on the design and development of an embedded prototype that embeds harvester, conditioning circuit, storage circuitry and logic in a single unit, together with the development of an optimized scheduler that will run on the embedded boards target of the cooling.

Acknowledgments This work is supported by the European FP7 Project Green Data Net, grant n. 609000. <http://www.greendatnet-project.eu/>

References

1. Choi, K., Soma, R., Pedram, M.: Dynamic voltage and frequency scaling based on workload decomposition. In: Proceedings of the 2004 International Symposium on Low Power Electronics and Design, 2004. ISLPED '04, pp. 174–179 (2004)
2. Murali, S. et al.: Temperature control of high-performance multi-core platforms using convex optimization. In: Proceedings of the Conference on Design, Automation and Test in Europe, DATE '08, ACM, pp. 110–115 (2008)
3. Merritt, R.: Dell, IBM give thumbs up to ARM Servers. In: EE Times (2010)
4. Data Centre Specialist Group: Meeting the energy efficiency and financial challenges in it. Technical report, BCS (2007)
5. Suski, E.D.: Method and apparatus for recovering power from semiconductor circuit using thermoelectric device, 30 May 1995. US Patent 5,419,780
6. Solbrekken, G. et al.: Experimental demonstration of thermal management using thermoelectric generation. In: IThERM '04. The Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, pp. 284–290 (2004)
7. Wu, C.-J.: Architectural thermal energy harvesting opportunities for sustainable computing. Comput. Architect. Lett. (2013)
8. Rizzon, L., Rossi, M., Passerone, R., Brunelli, D.: Wireless sensor networks for environmental monitoring powered by microprocessors heat dissipation. In: Proceedings of the 1st International Workshop on Energy Neutral Sensing Systems, ENSSys '13. ACM (2013)
9. Rossi, M., Rizzon, L., Fait, M., Passerone, R., Brunelli, D.: Energy neutral wireless sensing for server farms monitoring. IEEE J. Emerg. Sel. Top. Circuits Syst. **4**(3), 324–335 (2014)
10. Prijic, A., Vracar, L., Vuckovic, D., Milic, D., Prijic, Z.: Thermal energy harvesting wireless sensor node in aluminum core PCB technology. IEEE Sens. J. **15**(1), 337–345 (2015)
11. Kiflemariam, R., Lin, C.-X.: Numerical simulation and parametric study of heat-driven self-cooling of electronic devices. J. Therm. Sci. Eng. Appl. ASME **7** (2015)
12. Rizzon, L., Rossi, M., Passerone, R., Brunelli, D.: Self-powered heat-sink SoC as temperature sensors with wireless interface: design and validation. In Proceedings of IEEE Sensors 2014, (2014)
13. Laird Technologies: Thermoelectric Handbook

Chapter 5

High Speed VLSI Architecture for Finding the First W Maximum/Minimum Values

Guoping Xiao, Waqar Ahmad, Syed Azhar Ali Zaidi,
Massimo Ruo Roch and Giovanni Causaprano

Abstract VLSI architectures for finding the first W maximum/minimum values are highly demanded in the fields of K -best MIMO detector, non-binary LDPC decoder and product-code decoder. In this paper, a VLSI architecture based on parallel comparing scheme is explored for finding the first W maximum/minimum values from M inputs. The place and route results using a TSMC 90-nm CMOS technology show that, despite some hardware cost, it achieves on average a 3.6x faster speed performance compared to the existing partial sorting architectures.

Keywords Max/min values generator · Non-binary LDPC decoder · MIMO decoder · Product-codes decoder

5.1 Introduction

In the telecommunication field, VLSI architectures for partial sorting are of great importance in many applications including Turbo codes [1, 2], Low-Density-Parity-Check (LDPC) decoder [3–5] and multiple types of MIMO detectors [6–8]. In the

G. Xiao (✉) · W. Ahmad · S.A.A. Zaidi · M.R. Roch · G. Causaprano
Department of Electronics and Telecommunications, Politecnico di Torino,
C.so Duca Degli Abruzzi 24, 10129 Turin, Italy
e-mail: guoping.xiao@polito.it

W. Ahmad
e-mail: waqar.ahmad@polito.it

S.A.A. Zaidi
e-mail: syed.zaidi@polito.it

M.R. Roch
e-mail: massimo.ruoroch@polito.it

G. Causaprano
e-mail: giovanni.causaprano@polito.it

binary LDPC decoder of [3], an architecture for searching the first two minimum values is routinely applied. As we can find in the published literature, only [9, 10] propose algorithms focusing on the general problem of finding the first two maximum/minimum values. While VLSI architectures to find the first $W > 2$ maximum/minimum values from M inputs ($M \leq W/2$) are required in K-Best MIMO detectors [6], non-binary LDPC decoders [4, 5] and Turbo product code decoders [2]. Unfortunately no published works put effort on the investigation for the general issue of $W > 2$.

In this brief, grounded on [10] we deeply analyze the tree-based algorithm with N level comparing stages, and finally obtain an optimized derivation to minimize the comparators in which the architecture only consists of one level of comparing stages with radix of M . In order to achieve a good performance in terms of latency, the implementation of the architecture is highly parallized.

The rest of the paper is as follows. In the next section a problem formulation and analysis is presented. Section 5.3 describes the hardware implementation of the architecture in details. In Sect. 5.4 performance data and comparison results are shown. Finally, a conclusion is drawn in Sect. 5.5.

5.2 Problem Formulation

According with the notation introduced in [10], we can write the problem of finding the first W maximum/minimum values as follows. Given a set $\mathcal{X}^{(M)} = \{x_0, \dots, x_{M-1}\}$ made of M elements we want to find the first W maximum/minimum values, namely $\mathbf{y}^{(M)} = \{y_0^{(M)}, y_1^{(M)}, \dots, y_z^{(M)}, \dots, y_{W-1}^{(M)}\}$ where $y_0^{(M)} = \max(\mathcal{X}^{(M)})$, $y_1^{(M)} = \max(\mathcal{X}^{(M)} \setminus \{y_0^{(M)}\})$, \dots , $y_z^{(M)} = \max(\mathcal{X}^{(M)} \setminus \bigcup_{k=0}^{z-1} \{y_k^{(M)}\})$, \dots , $y_{W-1}^{(M)} = \max(\mathcal{X}^{(M)} \setminus \bigcup_{k=0}^{W-2} \{y_k^{(M)}\})$ (similarly substituting max with min). For the sake of simplicity in the following we will discuss only the max case as the min equivalent solution can be straightforwardly derived.

Tree-based architectures proposed in [9, 10] for $W = 2$ rely on N levels of comparing stages, where each level l is made of O_l comparing stages. These structures can be extended to $W > 2$ as depicted in Fig. 5.1 and detailed in the following. The i -th comparing stage of level $l = 1$ compares $K_{1,i}$ elements taken from $\mathcal{X}^{(M)}$ and produces W results, leading to $K_{1,i} \geq W$. In the following we will refer to $K_{l,i}$ as the radix of the comparing stage, where

$$M = \prod_{l=1}^N K_l \quad K_l = \frac{1}{O_l} \sum_{i=0}^{O_l-1} K_{l,i}. \quad (5.1)$$

Let $\mathcal{X}^{(K_{1,i})} \subseteq \mathcal{X}^{(M)}$ be the set of elements compared by the i -th stage of level $l = 1$, with $\bigcup_{i=0}^{O_1-1} \mathcal{X}^{(K_{1,i})} = \mathcal{X}^{(M)}$, $\mathcal{X}^{(K_{1,i})} \cap \mathcal{X}^{(K_{1,j})} = \emptyset$ and $i, j = 0, \dots, O_1 - 1$, $i \neq j$. Similarly, we define $y_0^{(K_{1,i})} = \max(\mathcal{X}^{(K_{1,i})})$ and $y_z^{(K_{1,i})} = \max(\mathcal{X}^{(K_{1,i})} \setminus \bigcup_{k=0}^{z-1} \{y_k^{(M)}\})$

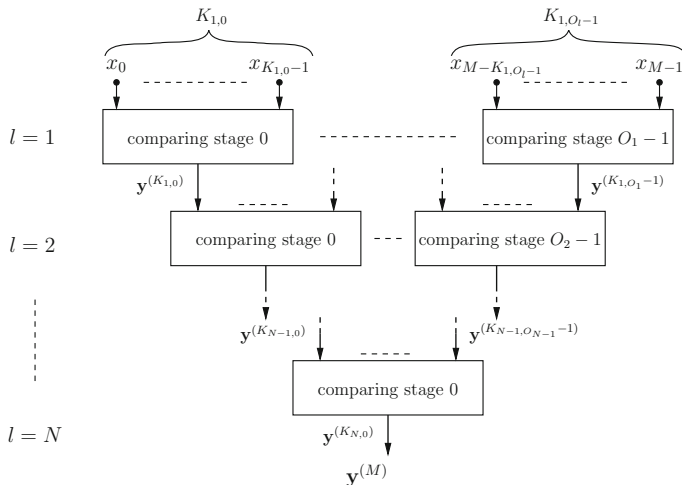


Fig. 5.1 Tree structure for finding the first W maximum values

with $z = 1, \dots, W - 1$ as the output of the i -th stage of level $l = 1$. The number of comparators required by a comparing stage depends on its radix and the total number of comparators at $l = 1$ is

$$C_{l=1} = \sum_{i=0}^{O_1-1} \binom{K_{1,i}}{2} = \sum_{i=0}^{O_1-1} \frac{K_{1,i} \cdot (K_{1,i} - 1)}{2}. \quad (5.2)$$

If all the comparing stages have the same radix, (5.2) simplifies to $M \cdot (K_1 - 1)/2$, where K_1 is the radix at $l = 1$. Since a similar approach can be followed for $l > 1$, then the total number of comparators can be obtained through

$$C = C_{l=1} + C_{l>1} = \sum_{l=1}^N \sum_{i=0}^{O_l-1} \gamma_l \cdot \frac{K_{l,i} \cdot (K_{l,i} - 1)}{2}, \quad (5.3)$$

where $\gamma_l = 1$ when $l = 1$, otherwise $\gamma_l = W \cdot (W + 1)/2$. If all the comparing stages within a level have the same radix ($K_{l,i} = K_l$), meanwhile all the levels have the same radix ($K_l = K$), thus (5.3) simplifies to

$$C = \frac{M \cdot (K - 1)}{2} + \frac{W \cdot (W + 1)}{2} \cdot \frac{M - K}{2}. \quad (5.4)$$

An interesting result can be obtained from (5.4) by computing the partial derivative of C with respect to K . Result in (5.4) can be minimized in two cases: $K = M$ and $K = W$. Since the second term in (5.4) tends to grow as M^3 when W approaches $M/2$, we choose $K = M$ to minimize C .

5.3 Architectural Description

As highlighted in Sect. 5.2, the architecture with a maximum radix of $K = M$ worth investigation. Since this architecture requires only one level of comparing stages, it computes the first W maximum values among its M inputs by the means of $M \cdot (M - 1)/2$ comparators working totally in parallel. Let $x_p, x_q \in \mathcal{X}^M$ be two inputs of a certain comparator. Let $s_{p,q}$ be the result of the comparison between x_p and x_q , namely $s_{p,q}$ is the sign of $x_p - x_q$, which is equal to 0 if $x_p > x_q$, otherwise $s_{p,q} = 1$. If x_n is the first maximum, then $s_{q,n} = 1$ for every q such that $0 \leq q \leq M - 1$ and $q \neq n$. Let \mathcal{N} be the array whose p -th element is $\mathcal{N}_p = \bigwedge_{q=0, q \neq p}^{M-1} s_{q,p}$, where \bigwedge represents the logic-and operation. As it can be observed, \mathcal{N} is the one-hot representation of n and can be used as the selection signal of a mux-like structure. The mux-like structure is the same one proposed in [10], namely for each $x_u \in \mathcal{X}^{(M)}$ with $0 \leq u \leq M - 1$ let $x_{u,v}$ be the v -th bit of x_u and d the number of bits to represent x_u . Then, $y_{0,v}^{(M)}$, the v -th bit of $y_0^{(M)}$ is

$$y_{0,v}^{(M)} = \bigvee_{u=0}^{M-1} x_{u,v} \wedge \mathcal{N}_u \quad (5.5)$$

where \bigvee is the logic-or operation.

The approach suggested in [10] to find $y_1^{(M)}$ by the means of a masking circuit can be extended in this work to find $y_n^{(M)}$ with $n = 1, \dots, W - 1$ (see Fig. 5.2). As argued in [10] the following formulation should ease understanding the underlying idea, even if, from a formal point of view, it has some redundancy. Let $\hat{t}_{p,q} = s_{p,q} \wedge \overline{\mathcal{N}_q}$ where $\overline{(\cdot)}$ is the logic-not operation and $\hat{t}_{p,q} = 1$ if $x_q > x_p$ and $x_q \neq y_0^{(M)}$. Then, let $\mathcal{M}^{(i)}$ be the array whose p -th element is $\mathcal{M}_p = \bigwedge_{q=0, q \neq p}^{M-1} t_{q,p}$, where $t_{q,p} = \hat{t}_{q,p} \vee \mathcal{N}_q$. If x_m is the second maximum value, then \mathcal{M} is the one-hot representation of m and it can be used as the selection signal of a mux-like structure to obtain $y_1^{(M)}$: $y_{1,v}^{(M)} = \bigvee_{u=0}^{M-1} x_{u,v} \wedge \mathcal{M}_u$, as shown in Figs. 5.2 and 5.3. Thus, $y_z^{(M)}$ can be obtained as

$$y_{z,v}^{(M)} = \bigvee_{u=0}^{M-1} x_{u,v} \wedge \mathcal{M}_u^{(z)} \quad (5.6)$$

where $\mathcal{M}_u^{(z)} = \bigwedge_{q=0, q \neq u}^{M-1} t_{q,u}^{(z)}$, and $t_{q,u}^{(z)} = \left(t_{q,u}^{(z-1)} \wedge \overline{\mathcal{M}_u^{(z-1)}} \right) \vee \mathcal{M}_q^{(z-1)}$ with $t_{q,u}^{(0)} = s_{q,u}$ and $\mathcal{M}_q^{(0)} = \mathcal{N}_q$.

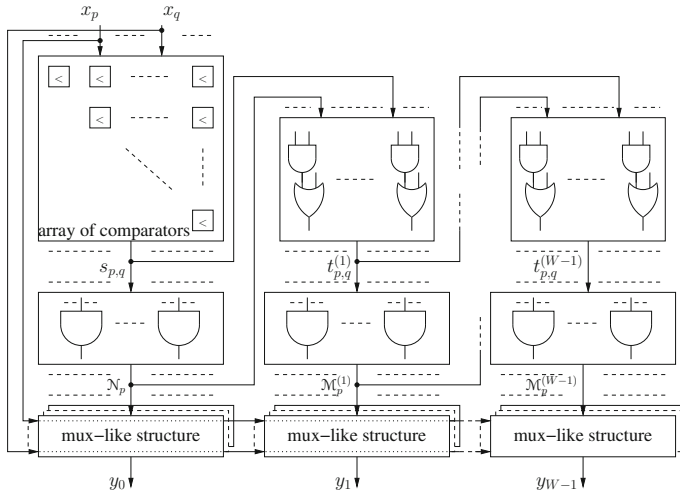


Fig. 5.2 The structure of the proposed architecture

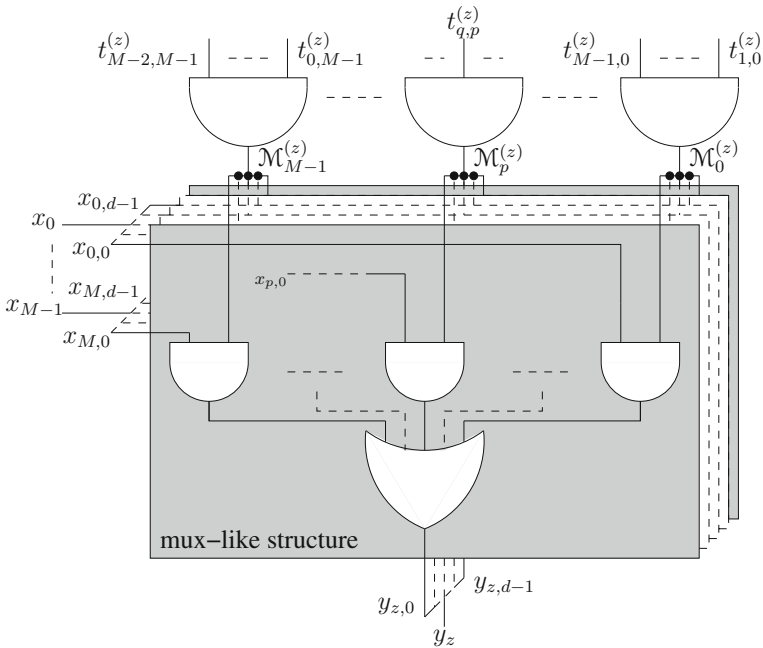


Fig. 5.3 one-hot signals and mux-like structure

Table 5.1 Performance results

(a)							
M, W, bit-width	20, 5, 16	80, 10, 16	32, 16, 5	64, 16, 5	32, 3, 5	64, 4, 5	128, 5, 5
Shabany [6]	19.68	40.98	40.95	44.17	13.75	18.97	27.58
Boutillon [4]	17.73	45.38	37.96	40.86	11.13	15.94	27.45
Tsai [7]	15.13	45.35	31.56	34.25	8.02	10.27	20.20
Proposed	2.89	10.35	7.50	12.27	2.21	3.60	7.30
(b)							
Shabany [6]	33484	93171	25721	35538	18583	24858	43756
Boutillon [4]	48185	113538	29735	47157	27817	46071	68593
Tsai [7]	62221	406108	20649	42905	19177	41584	170769
Proposed	67913	1532464	234952	943145	45815	349082	1415706

(a) latency [ns]. (b) area [μm^2]

5.4 Results and Comparisons

The proposed solutions have been implemented for some cases of study taken from K-best MIMO detectors [6, 7], Non-binary LDPC decoder [4, 5] and product code [2] architectures. In order to get a clear performance comparison, the architecture proposed in this paper, as well as partial sorting architectures in [4, 6, 7] are described in VHDL and synthesized and placed and routed in the TSMC 90-nm digital CMOS process, using Synopsys and Cadence design flows. The performance results are listed in Table 5.1 in terms of area and latency. As we can see from Table 5.1(a), although the data of occupied area in Table 5.1(b) shows the hardware cost of the architecture proposed in this paper is not as good as others, it has an overwhelming advantage on the speed which achieves 3.6 times faster on average and even 5.2 times faster at best compared to the best reference under the 7 configurations.

5.5 Conclusions

In this paper, VLSI architecture for finding the first W maximum/minimum values with high speed is developed. Experimental results show that, owing to the parallel comparing scheme, it achieves an excellent performance on the speed but with some cost on the area.

References

1. Papaharalabos, S., Mathiopoulos, P.T., Masera, G., Martina, M.: Novel non-recursive max* operator with reduced implementation complexity for turbo decoding. *IET Commun.* **6**(7), 702–707 (2012)
2. Leroux, C., Jego, C., Adder, P., Jezequel, M., Gupta, D.: A highly parallel turbo product code decoder without interleaving resource, In: *IEEE Workshop on Signal Processing Systems*, pp. 1–6 (2008)
3. Condo, C., Martina, M., Masera, G.: VLSI implementation of a multi-mode turbo/LDPC decoder architecture. *IEEE Trans. Circuits Syst. I* **60**(6), 1441–1454 (2013)
4. Boutillon, E., Conde-Canencia, L.: Bubble check: a simplified algorithm for elementary check node processing in extended min-sum non-binary LDPC decoders. *IET Electron. Lett.* **46**(9), 633–634 (2010)
5. Zhang, X., Cai, F.: Reduced-complexity decoder architecture for non-binary LDPC codes. *IEEE Trans. VLSI* **19**(7), 1229–1238 (2011)
6. Shabany, M., Glenn Gulak, P.: A 675 mbps, 4x4 64-QAM K-Best MIMO detector in 0.13 um CMOS. *IEEE Trans. VLSI* **20**(1), 1063–8210 (2012)
7. Tsai, P., Chen, W., Lin, X., Huang, M.: A 44 64-QAM reduced-complexity K-Best MIMO detector up to 1.5Gbps. In: *IEEE International Symposium on Circuits and Systems*, pp. 3953–3956 (2010)
8. Wu, B., Masera, G.: Efficient VLSI implementation of soft-input soft-output fixed-complexity sphere decoder. *IET Commun.* **6**(9), 1111–1118 (2012)
9. Wey, C.L., Shieh, M.D., Lin, S.Y.: Algorithms of finding the first two minimum values and their hardware implementation. *IEEE Trans. Circuits Syst. I* **55**(11), 3430–3437 (2008)
10. Amaru, L.G., Martina, M., Masera, G.: High speed architectures for finding the first two maximum/minimum values. *IEEE Trans. VLSI* **20**(12), 2342–2346 (2012)

Chapter 6

Design and Implementation of a Portable fNIRS Embedded System

Diego Agrò, Riccardo Canicatti, Maurizio Pinto, Giuseppe Morsellino, Alessandro Tomasino, Gabriele Adamo, Luciano Curcio, Antonino Parisi, Salvatore Stivala, Natale Galioto, Alessandro Busacca and Costantino Giaconia

Abstract We report on the design, development and operation of a portable, low cost, battery-operated, multi-channel, functional Near Infrared Spectroscopy embedded system, hosting up to 64 optical sources and 128 Silicon PhotoMultiplier optical detectors. The system is realized as a scalable architecture, whose elementary leaf consists of a probe board provided with 16 SiPMs, 4 couples of bi-color LED, and a temperature sensor, built on a flexible stand. The hardware structure is very versatile because it is possible to handle both the switching time of the LED and the acquisition of the photodetectors, via an ARM based microcontroller.

Keywords Silicon photomultiplier · Infrared spectroscopy · fNIRS embedded system

D. Agrò (✉) · R. Canicatti · M. Pinto · G. Morsellino · A. Tomasino · G. Adamo
L. Curcio · A. Parisi · S. Stivala · N. Galioto · A. Busacca · C. Giaconia
DEIM, Department of Energy, Information Engineering and Mathematical Models,
University of Palermo, Viale Delle Scienze, Bldg. 9, 90128 Palermo, Italy
e-mail: diego.agro@unipa.it

R. Canicatti
e-mail: canicatt.riccardo@yahoo.it

M. Pinto
e-mail: mauripin@hotmail.com

G. Morsellino
e-mail: g.morsellino84@gmail.com

A. Tomasino
e-mail: alessandro.tomasino@unipa.it

G. Adamo
e-mail: gabriele.adamo@unipa.it

L. Curcio
e-mail: luciano.curcio@unipa.it

6.1 Introduction

Functional Near InfraRed Spectroscopy (fNIRS) is an imaging technique for human brain monitoring. It is employed both in clinical medicine and in research laboratories in order to measure the oxygenation status of the brain. Whereas light in the fNIRS range (650–900 nm) is poorly absorbed by organic tissues, water and lipid, the main bio-molecular absorbers are blood chromophores, as the case of oxygenated and deoxygenated haemoglobin (HbO_2 and Hb, respectively). In particular, both concentration variations of the latter, provide important information on brain activity. If more than one wavelength is used, oxy-haemoglobin and deoxy-haemoglobin changes can be recovered by using the modified Beer-Lambert law [1].

Different types of fNIRS devices have been developed and reported in literature [2]: continuous wave (CW), time domain (TD) and frequency domain (FD) spectroscopy systems. Among such methods, CW-NIRS system is practically easier to develop than the other two techniques and, therefore, it allows to remarkably reduce the cost, especially when a high number of sensors should be employed. In CW systems, light is commonly emitted at a constant power level, but in some cases, it could be modulated at a few kilohertz. Anyway, such systems cannot give absolute concentration values—unlike the FD and TD systems—but could only provide relative changes [3].

Hardware novelty mainly consists in the adoption of Silicon PhotoMultiplier (SiPM) as optical sensors, which show significant properties in terms of gain (around 6×10^6), signal-noise ratio (SNR, around 45 dB at 11 pW and 28 °C), low operating voltage (<30 V), ruggedness, compactness, reduced sensitivity with temperature, voltage fluctuations, and magnetic fields versus conventional photo-multiplier tubes. With respect to Avalanche PhotoDiode, PIN or Single Photon Avalanche photodiodes, a SiPM matrix ensures better gain and SNR [4–8]. The use of Silicon Photomultipliers in a fNIRS system could potentially increase the spatial resolution.

A. Parisi
e-mail: antonino.parsi@unipa.it

S. Stivala
e-mail: salvatore.stivala@unipa.it

N. Galioto
e-mail: natale.galioto@unipa.it

A. Busacca
e-mail: alessandro.busacca@unipa.it

C. Giaconia
e-mail: costantino.giaconia@unipa.it

6.2 Hardware

The purpose of this work is to implement an embedded CW-fNIRS system capable of monitoring haemodynamic signals during brain activity [9–12]. The system has been designed in order to cover the entire skull surface, by employing a high number of optical components. Thus, it hosts up to 64 LED sources and 128 SiPM detectors. Moreover, a multiplexing technique has been exploited in order to realize a compact and portable system, with a total power consumption suitable for a battery-operated equipment.

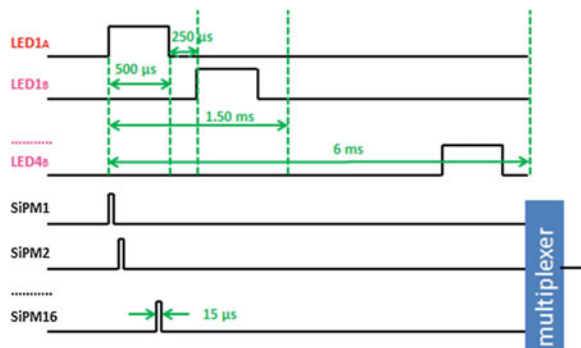
The designed system is based on a scalable architecture, in which every leaf consists of 8 modular probe boards (henceforth, simply named probes) built on a flexible stand, hosting 4 bi-color LED (wavelengths equals to 735 and 850 nm) as light sources, 16 SiPMs as photo-detectors and a temperature sensor. The used SiPM is relied on n/p technology developed at STMicroelectronics R&D (Catania, Italy), showing a breakdown voltage of about 28.0 V and a $3.0 \times 3.0 \text{ mm}^2$ active area (2500 microcells, 62 % fill factor, and $60 \text{ }\mu\text{m}$ cell pitch) enclosed in a $5.1 \times 5.1 \text{ mm}^2$ package.

Each probe is connected to the main board through a flexible flat cable connector, where a powerful ARM microcontroller (μC), able to handle both the LED switching time and the acquisition of SiPM, is hosted. Furthermore, on the main board all the needed supply voltages both for the analog and the digital circuitry are implemented, starting from a unique battery pack. Finally, both a UART/RS232 and UART/USB interfaces are realized in order to deliver the digital data to a personal computer.

In our CW-fNIRS system, the LED sources illuminate the tissue under test with a constant optical power. Light beam propagates inside the various layers constituting the head, up to the grey matter, where it mostly bends thus the SiPMs are able to capture the scattered and outgoing optical signal. The μC also performs a preliminary data filtering in view of a next post-processing and plotting.

The acquisition process is implemented by following the time sharing polling schedule, presented in Fig. 6.1. The light sources are enabled one by one and they emit a fixed optical power for $500 \text{ }\mu\text{s}$, that is the time which the measure process

Fig. 6.1 Timing diagram of the overall fNIRS polling schedule for a single probe sub-platform



takes. During this time interval, SiPM selection is sequentially carried out. Each selected SiPM—biased to the working voltage (see later)—gives rise to a photocurrent proportional to the detected optical power. The latter is then acquired through the Analog to Digital Converter (ADC) of the microcontroller, in a time window of $15\ \mu\text{s}$. Both SiPM selection and readout are performed by a multiplier-demultiplexer couple, which is handled by the same selection signals coming from the microcontroller. A dead time of $250\ \mu\text{s}$, in which all LEDs are off and no readout operations are performed, spaces the lighting of consecutive light sources.

However, it is worth highlighting that if more than one probe is connected to the main board, the system is capable to handle several LED-SiPM couples even if such components are positioned in different probes. Such a feature allows an user to optimally handle all LED-SiPM couples, covering the head area under investigation. A relevant property of the probe geometric structure is its repeatability, i.e., the possibility to realize a structure covering the entire head surface of the patient by simply joining all the identical probes. As a result, the chosen geometry is sketched in Fig. 6.2.

Each of the eight probes is driven by a dedicated circuit, as shown in Fig. 6.3a. The latter selects, by means of a de-multiplexer, the correct sequence of LEDs, according to the selection signals coming from the μC , as stated by the previous time diagram scheme. The current of the selected LED is controlled via the voltage signal (V_{DAC}) generated by the Digital to Analog Converter (DAC), integrated within the microcontroller, allowing a very fine tuning of the optical power emitted by each LED.

As shown in Fig. 6.3b, an analog voltage adder generates the SiPM working voltage, by adding a fixed voltage of $-31\ \text{V}$ with a programmable voltage value

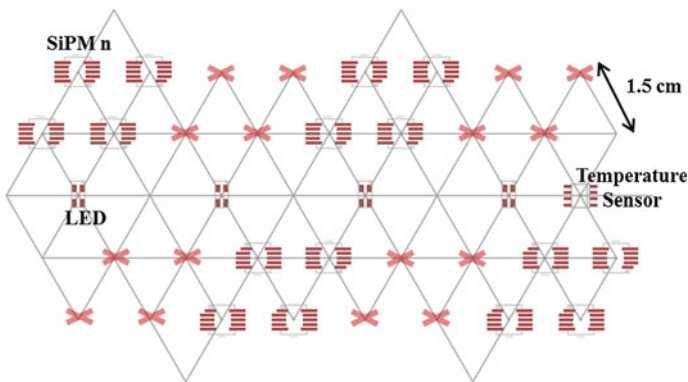


Fig. 6.2 Probe geometric structure. The red crosses, visible in the picture, represent the positions where sensors of an adjacent probe can be stitched together

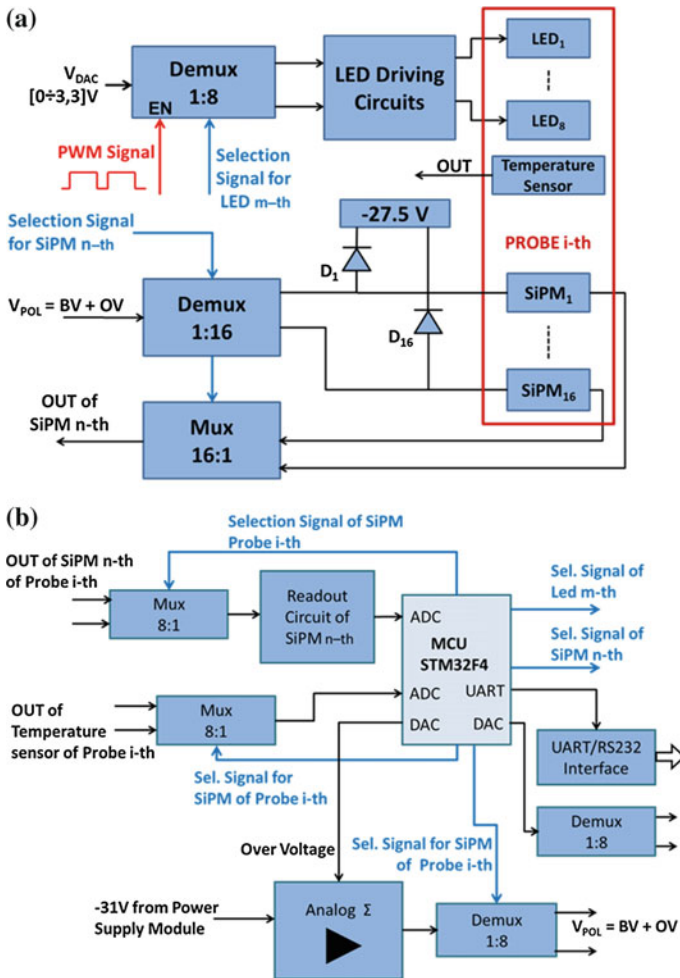


Fig. 6.3 Block diagram of **a** the circuit handling of i th probe; **b** the circuit handled by the microcontroller

coming from the DAC module. A dynamic voltage range of 3.3 V is thus obtained in order to finely control the SiPM Overvoltage ($-27.7 \div -31$ V). It is worth noticing that a -27.5 V voltage reference is needed in order to keep all the SiPMs at a voltage level just below their threshold, thus obtaining a faster response than the one achieved when SiPMs are firstly biased at 0 V.

Finally, the whole realized embedded system is shown in Fig. 6.4.

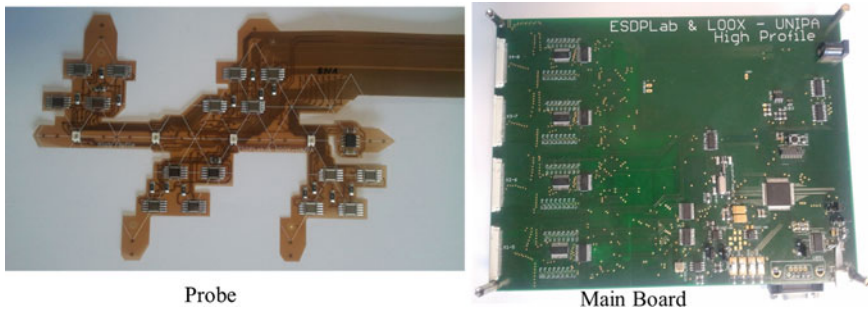


Fig. 6.4 Picture of the whole fNIRS prototype

6.3 Experimental Results

A first test has been carried out on a volunteer subject. A probe board was fixed to his forehead by using a black elastic band. Working in these conditions, a set of 200 s monitoring trials was recorded. During this time window, the subject under test was normally breathing with the exception of the time interval between 70 and 100 s, during which a breath holding phase was set up. Data shown in Fig. 6.5b, coming from the SiPM 11 detector (see Fig. 6.5a), which is positioned at 2.6 cm from the light source LED 3, have been elaborated by the modified Beer-Lambert law and then they have been filtered by a 300 mHz low-pass filter, in order to cut out the unwanted cardiac pulse (located at about 1.1 Hz) and emphasize the oxygenated (red curves, HBO_2), de-oxygenated (blue curves, HB) and total (yellow curves, blood volume, BV) haemoglobin concentration levels. A significant variation of the total haemoglobin is experienced during the breath holding phase. This fact proves that the grey matter of the brain has been reached. In particular, this experimental result is in agreement with other similar tests carried out by different researchers and with a completely different equipment [13].

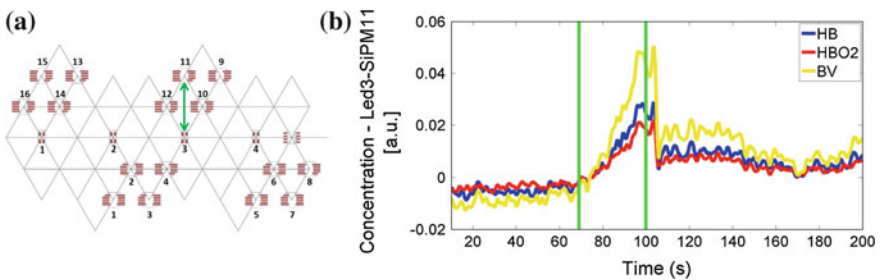


Fig. 6.5 **a** Probe board pointing out the LED3-SiPM11 couple selected during the experiment; **b** oxygenated (red curves), de-oxygenated (blue curves) and total (yellow curves) haemoglobin levels coming from LED3-SiPM11 couple

6.4 Conclusion

A CW-fNIRS embedded system prototype has been realized and presented. It hosts up to 64 LED sources and 128 SiPM sensors and it is based on a scalable solution, performing a high level modularity and keeping a low consumption level. Several preliminary functional tests were successfully carried out, thus achieving very encouraging results to be confirmed in the proper validation follow-up. The maximum voltage generated within the prototype is 40 V, so it can be classified as a low voltage device according to the proper regulation ($V_{\max} < 50$ Volts @ IEC EN 60601-1). The obtained results by our prototype confirms the high detection performance of the Silicon PhotoMultipliers.

Acknowledgments This work has been developed in the framework of the ARTEMIS “High Profile” European Funded Project (grant agreement 269356) and it was supported by the Telecom Italia for the Ph.D. program of D. Agrò. We would also like to thank STMicroelectronics of Catania for the SiPMs provision.

References

1. Bakker, A., Smith, B., Ainslie, P., Smith, K.: Near-Infrared Spectroscopy. In: Ainslie, P. (ed.) *Applied Aspects of Ultrasonography in Humans*, pp. 65–89 (2012)
2. Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Murray, T., Ovetsky, Y., Pidikiti, D., Thomas, R.: A novel method for fast imaging of brain function, non-invasively, with light. *Opt. Express* **2**, 411–423 (1998)
3. Rolfé, P.: In vivo near infra-red spectrophotometry. *Ann. Rev. Biomed. Eng.* **2**, 315–354 (2000)
4. Adamo, G., Agrò, D., Stivala, S., Parisi, A., Giaconia, C., Busacca, A.C., Fallica, G.: SNR measurements of silicon photomultipliers in the continuous wave regime. *Photonics West 2014, Silicon Photonics IX*, paper no. 8990-43 (2014)
5. Adamo, G., Agrò, D., Stivala, S., Parisi, A., Giaconia, C., Busacca, A.C., Mazzillo, M.C., Sanfilippo, D., Fallica, G.: Measurements of silicon photomultipliers responsivity in continuous wave regime. *IEEE Trans. Electron Devices* **60**(11), 3718–3725 (2013)
6. Adamo, G., Agrò, D., Stivala, S., Parisi, A., Curcio, L., Ando', A., Tomasino, A., Giaconia, C., Busacca, A.C., Mazzillo, M.C., Sanfilippo, D., Fallica, G.: Responsivity measurements of 4H-SiC Schottky photodiodes for UV light monitoring. *Photonics West 2014, Silicon Photonics IX*, paper no. 8990-41 (2014)
7. Pernice, R., Adamo, G., Stivala, S., Parisi, A., Busacca, A.C., Spigolon, D., Sabatino, M.A., D'Acquisto, L., Dispenza, C.: Opals infiltrated with a stimuli-responsive hydrogel for ethanol vapor sensing. *Opt. Mater. Express* **3**(11), 1820–1833 (2013)
8. Adamo, G., Agrò, D., Stivala, S., Parisi, A., Giaconia, C., Busacca, A.C., Mazzillo, M.C., Sanfilippo, D., Fallica, G.: Responsivity measurements of N-on-P and P-on-N silicon photomultipliers in the continuous wave regime. In: *Proceedings SPIE 8629, Photonics West 2013, Silicon Photonics VIII*, 86291, pp. 86291A-1–86291A-9. San Francisco, USA (2013)
9. Sanfilippo, D., Valvo, G., Mazzillo, M., Piana, A., Carbone, B., Renna, L., Fallica, P.G., Agrò, D., Morsellino, G., Pinto, M., Canicattì, R., Galioto, N., Adamo, G., Stivala, S., Parisi, A., Curcio, L., Giaconia, C., Busacca, A.C., Pagano, R., Libertino, S., Lombardo, S.: Design and development of a fNIRS system prototype based on SiPM detectors. *Photonics West 2014, Silicon Photonics IX*, paper no. 8990-40 (2014)

10. Xue, H., Bestonzo, M., Acharya, U.R., Molinari, F.: Design and implementation of a continuous wave near infrared spectroscopy system for bedside and home monitoring. *J. Med. Imaging Health Inform.* **1**, 317–324 (2011)
11. Bozkurt, A., Rosen, A., Rosen, H., Onaral, B.: A portable near infrared spectroscopy system for bedside monitoring of newborn brain. *BioMed. Eng. Online* **4**, 1–11 (2005)
12. Zimmermann, R., Braun, F., Achnich, T., Lambercy, O., Gassert, R., Wolf, M.: Silicon photomultipliers for improved detection of low light levels in miniature near-infrared spectroscopy instruments. *Biomed. Opt. Express* **4**, 659–666 (2013)
13. Vikrant S.: Near infrared spectroscopy: a study of cerebral hemodynamics during breathholding and development of a system for hotflash measurement. M.S. thesis, University of Arlington Texas (2005)

Chapter 7

Advancements on Silicon Ultrasound Probes (CMUT) for Medical Imaging Applications

Giosuè Caliano and Alessandro S. Savoia

Abstract Capacitive micromachined ultrasonic transducers (CMUTs) are micro-electromechanical devices (MEMS) fabricated using silicon micromachining techniques. The interest of this technology relies in its full compatibility with the microelectronic technology that makes possible to integrate on the same chip the transducer and the controlling/conditioning electronics, so as to achieve low-cost and high-performance devices. The design and fabrication of a 192-element linear array CMUT probe operating in the range 6–18 MHz is here presented. The CMUT array is micro-fabricated and packed using a novel fabrication concept specifically conceived for imaging transducer arrays. The performance optimization of the probe is performed by connecting the CMUT array with multichannel analog front-end electronic circuits housed into the probe body. Characterization and imaging results are used to assess the performance of CMUTs with respect to conventional piezoelectric transducers. This paper is a review on the activities of our group in this field.

Keywords CMUT · Capacitive transducer · Medical imaging · Micromachined

7.1 Introduction

Capacitive Micromachined Ultrasonic Transducers (CMUTs) were first researched at Stanford University in 1994 and then in other research laboratory [1–3]. Unlike of piezoelectric transducers, which use thickness-mode vibration of piezoceramic materials, CMUTs are based on flexural vibration of electrostatically-actuated micromachined plates. The enhanced compatibility between micro-electro-mechanical system and standard integrated circuit technologies, with respect to piezoelectric technology, led in a very short time to a plethora of applications and to the development of

G. Caliano (✉) · A.S. Savoia
Dipartimento di Ingegneria, Università Roma Tre, via della Vasca Navale 84,
00146 Rome, Italy
e-mail: giosue.caliano@uniroma3.it

a variety of manufacturing technologies in various research laboratories. In 2003, the fulfilment of the first real time in vivo ultrasound images in USA [4, 5], and one year later in Italy by the Acousto-Electronics Laboratory of Roma Tre University [6–8], where the work here presented was carried out, represented the turning point in studying of these devices. Several patents have been filed by many companies and universities to date: among those, a patent concerning the novel CMUT fabrication process [9] used in the work here presented, was filed by the authors of this paper [10].

The aim of this paper is to present a review of the state of art in this field by our research group, showing the development of a high frequency ultrasound probe for medical echography allowed by the availability of a well-established CMUT micro-fabrication and packaging technology. The 192-elements CMUT linear array design and fabrication as well as the probe engineering and the integration in a commercial ultrasound medical imaging system is shown [11]. A performance evaluation of the CMUT probe is operated by electrical and acoustic measurements [12].

7.2 Probe Fabrication and Packaging

The 192-elements array was fabricated using the Reverse Fabrication Process (RFP) developed and showed in [9]. The RFP process is a sacrificial release micro-fabrication technology for the CMUT fabrication on silicon wafers. CMUTs are fabricated using a down-top approach (unlike the classical top-down approach used in microelectronics), namely that the microfabrication is performed starting from the top layers of the structure, i.e. the transducer vibrating surface, and ending with the device sealing and the delimitation of the pads for the electrical interconnection. After the micro-fabrication, the device is electrically interconnected and packed and, ultimately, the silicon substrate (i.e. the carrier) is completely removed to release the CMUT vibrating surface. In Fig. 7.1a–i, the main RFP flow-steps are described.

The micro-fabrication starts on top of a low-stress and ultra-precise thickness controlled LPCVD silicon nitride film grown over a silicon wafer (Fig. 7.1a). The top electrode is deposited and patterned using an Al-Ti-W multilayer (Fig. 7.1b). The use of a thin-layer of Ti-W over the Al film avoids the degradation of the metallization during the subsequent PECVD silicon nitride deposition (Fig. 7.1c), which acts as a passivation layer of the top electrode. Following, a sacrificial Cr film is deposited and patterned (Fig. 7.1d). A PECVD silicon nitride layer (Fig. 7.1e) is deposited to passivate the bottom electrode that is defined by depositing and patterning a second Al-Ti-W multilayer film (Fig. 7.1f). After an additional PECVD silicon nitride passivation layer, the etching holes for the removal of the sacrificial Cr layer underneath the cells are created (Fig. 7.1g). The successive step is the sacrificial release using the Cr wet- etching (Fig. 7.1h). Ultimately, the device is sealed by depositing a 4 μm -thick PECVD silicon nitride film that is succeeding etched to define interconnection pads (Fig. 7.1i). A photo of a realized CMUT device, fabricated using the RFP, is in Fig. 7.1j where a detail of the cell layout of one array element is shown.

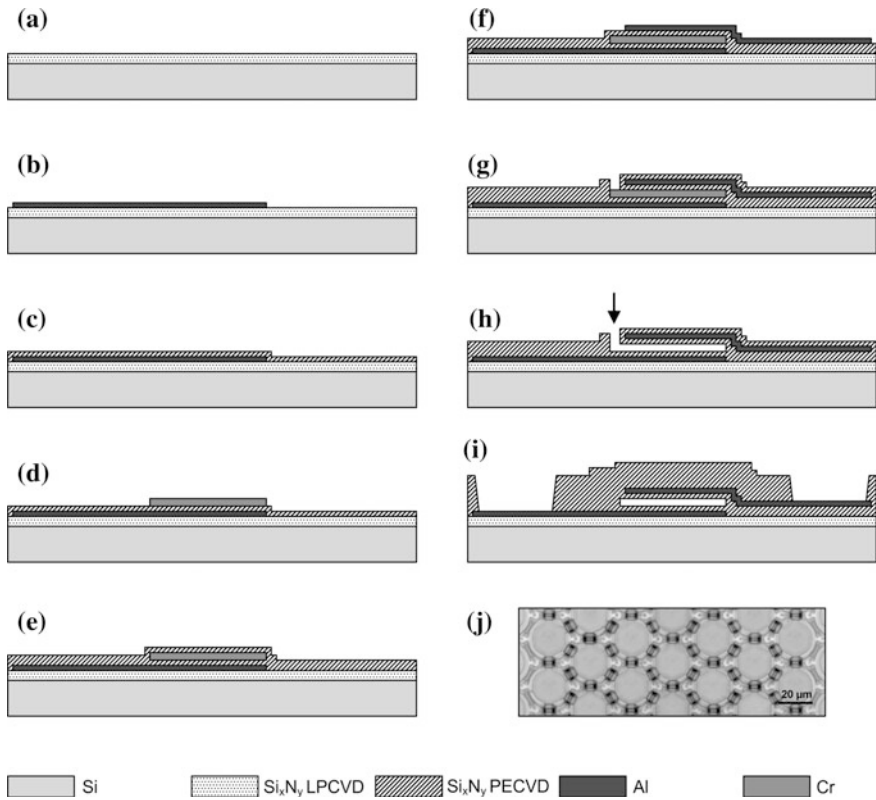


Fig. 7.1 a–i Reverse Fabrication Process (RFP) technology flow. **j** Photo of the cell layout of a microfabricated CMUT device [11, 13]

Meaningful benefits result from the RFP fabrication method: first of all, the CMUT vibrating structure is almost entirely fabricated of LPCVD silicon nitride: this results in a high accuracy of membrane thickness over the silicon wafer, and in a highly uniform mechanical behavior of the CMUT device. Another important feature is that the use of PECVD silicon nitride as passivation layer, which can be deposited at relatively low temperatures (350 °C) as compared to LPCVD silicon nitride (800–900 °C), and makes it possible to use Al as metal layer, whose electrical properties ensure low resistivity of the electrodes. Ultimately, thanks to the down-top nature of the fabrication method, the pads are positioned on the back of the final device. In other micro-fabrication approaches, this task is obtainable exclusively by using through-silicon vias (TSV) processes that are expensive and often difficult to achieve. In order to equip the CMUT device with a custom acoustic-backing, and simultaneously enable the electrical access to the elements, release the vibrating membrane structure, and apply an acoustic lens for protection and elevation focusing, a special packaging procedure was developed and implemented as in [11]. The CMUT device is placed and held on a vacuum fixture in a

way that the interconnection pads are facing up-wards. A rigid-flex (FR4-polyimide) PCB is placed close and aligned to the chip at a different vertical position and a wire bonding is performed, using Al wire. A pre-shaped custom acoustic-backing, obtained by a previously cured part of composite material made of Epotek 301 (Epoxy Technology, Bellerica, USA) epoxy, filled with W and Al_2O_3 powders, is then glued on the bonded assembly. The resulting acoustic-electric device is then dismantled from the vacuum fixture and placed in a PTFE holder, to perform the chemical etching of the silicon substrate (the carrier) and to release the CMUT active membranes. The silicon etching is performed using a mixture of acetic, nitric and hydrofluoric acid (HNA). Ultimately, an acoustic lens, made of RTV silicone rubber, filled with metal oxide nano-powders, is applied onto the CMUT device, using a stainless steel mold.

7.3 Probe Characterization and Ultrasound Imaging

The first characterization of the acoustic device is the measurement of the electrical impedance of the elements of the CMUT array, using an HP4194A (Agilent Technologies, Inc., Santa Clara, CA) impedance analyzer. Further, electrical impedance measurements at a fixed high bias voltage (200 V) were performed on all the array elements, using an external fixture, to estimate the capacitance and the resonance frequency variation across all the elements of the array. All the measurement data were compensated, taking into account the parasitic capacitance and inductance of the cables and of the fixture used for the electrical interconnection of the CMUT to the impedance analyzer. Acoustic characterization (pulse-echo) was carried out by placing the probe in front of a water immersed planar reflector to a depth of 15 mm, equal the elevation focus of the probe's lens. The probe characterization was carried out using a piezoelectric commercial probe, for comparison, with the same geometrical characteristics, and the same nominal frequency (LA435, Esaote spa, Firenze, Italy). The CMUT probe was biased at 200 VDC. The pulse-echo measurements were carried out with the scope to find the center frequency and bandwidth of the two-ways transfer function, as well as the pulse length, quantities that as is known, featuring the probe. All the probe elements were excited using a -100 V broadband pulse, generated by a Pulser/Receiver 5800 (Panametrics Inc., Waltham, MA). The echo signals received by the central element (96) of the CMUT and the piezoelectric arrays, normalized to their maximum absolute value, are depicted respectively in Fig. 7.2a, b. The FFT magnitudes and the envelopes of the two echo signals are reported in Fig. 7.2c, d. The CMUT probe shows larger bandwidth, and shorter pulse length. The envelope of the echo-signals (Fig. 7.2d) demonstrates that, unlike the piezoelectric probe element that shows residual oscillations in its time-domain response, the CMUT probe element exhibits a nearly ideal behavior, resulting in a better detail and contrast resolution during B-mode operation.

B-mode images were generated by connecting the CMUT and the piezoelectric probes LA435 to the Technos ultrasound imaging system (Esaote spa, Firenze, Italy). The frequency of the driving signal was set at 13 MHz and the burst count at 1. Due to

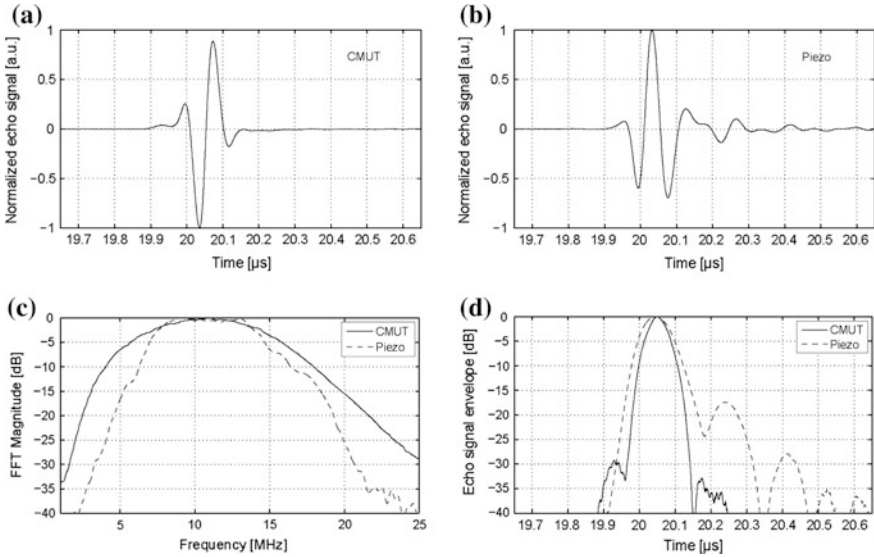


Fig. 7.2 Pulse-echo time-domain normalized response from the central (96) array element of the **a** CMUT probe biased at 200 V and **b** piezoelectric probe (LA435). Normalized **c** FFT magnitude and **d** time-domain envelope of the pulse-echo responses [11]

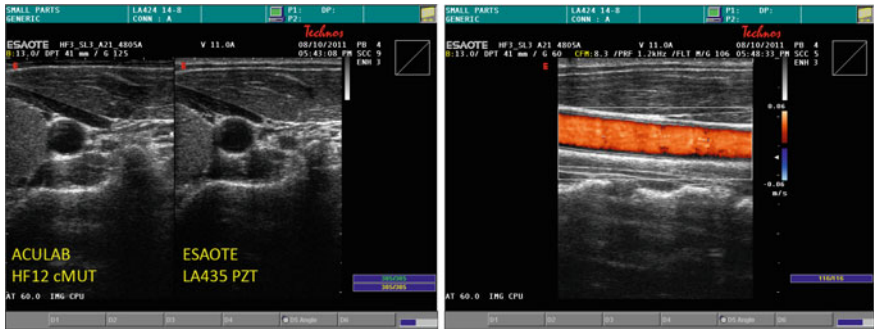


Fig. 7.3 In-vivo ultrasound images of a carotid artery obtained by an Esaote Technos imaging system. In the B-mode image, the carotid artery is scanned using the CMUT probe (left) and the Esaote LA435 piezoelectric probe (right). The right image reports a color-doppler analysis of the longitudinal section of the same carotid artery

the maximum voltage limitation of the LA435 probe, the driving signal amplitude was set in both cases at ± 60 V while the CMUT probe was biased at 200 VDC. Additionally, having the CMUT probe higher sensitivity, the B-mode gain control on the scanner system was adjusted in order to achieve equivalent brightness of the images generated with the two probes. The two probes were placed in the same position in order to scan the same portion of tissue, as you can see in Fig. 7.3a. As matter of fact, Fig. 7.3 reports the B-mode and the color-doppler image of a human

carotid artery. In the figure, the represented B-mode images were obtained with the CMUT probe (on the left) and by the LA435 piezoelectric probe (on the right), while the color-doppler image was obtained only with the CMUT probe. An improved axial resolution and penetration depth is observable in the image generated using the CMUT probe. Such improvement however, which is ascribable to the CMUT larger bandwidth respect to piezoelectric probe, was limited by the limited bandwidth of the reception filters implemented in the Technos imaging system.

7.4 Conclusions

This review paper has presented the state of art of the CMUT probe carried out in the University of Roma Tre. The design, fabrication and characterization of a CMUT probe for medical ultrasound imaging were shown. The heart of the probe, i.e. the CMUT array, was microfabricated using the Reverse Fabrication Process, and successfully assembled using an efficient and reliable packaging procedure. The CMUT array was provided with an acoustically and thermo-mechanically engineered material as backing and the acoustic lens designed to optimize the performance of the probe. The integration of the CMUT probe-head in a commercial ultrasound imaging system was supported by the introduction of electronics inside the probe that made it possible to use a standard cable and the system connector, resulting in a full compatibility with the system itself. Electric and acoustic characterizations were performed to assess the probe-head fabrication process reliability and the probe performance, respect to a piezoelectric commercial probe, as comparison. In vivo ultrasound images using various ultrasound modes were shown.

The results obtained showed the advantage of CMUT respect to piezoelectric technology, as regards both transmission and two-ways time and frequency response shape. The CMUT probe showed lower transmission sensitivity (-4.2 dB at the nominal center frequency) and higher two-ways sensitivity ($+5$ dB at the nominal center frequency) resulting from its higher reception sensitivity.

References

1. Haller, M.I., Khuri-Yakub, B.T.: A surface micromachined electrostatic ultrasonic air transducer. *Proc. IEEE Ultrason. Symp.* **2**, 1241–1244 (1994)
2. Caliano, G., Galanello, F., Caronti, A., Carotenuto, R., Pappalardo, M., Foglietti, V., Lamberti, N.: Micromachined ultrasonic transducers using silicon nitride membrane fabricated in PECVD technology. *Proc. IEEE Ultrason. Symp.* **1**, 963–968 (2000)
3. Caliano, G., Carotenuto, R., Caronti, A., Pappalardo, M.: cMUT echographic probes: design and fabrication process. *Proc. IEEE Ultrason. Symp.* **2**, 1067–1070 (2002)
4. Mills, D.M., Smith, L.S.: Real-time in-vivo imaging with capacitive micromachined ultrasound transducer (cMUT) linear arrays. *Proc. IEEE Ultrason. Symp.* **1**, 568–571 (2003)

5. Panda, S., Daft, C., Wagner, P., Ladabaum, I., Pellegretti, P., Bertora, F.: Microfabricated ultrasonic transducer (cMUT) probes: imaging advantages over PZT probes. In: Proceedings WFUMB (hosted by AIUM) Conference, Montreal, Canada (2003)
6. Caliano, G., Carotenuto, R., Cianci, E., Foglietti, V., Caronti, A., Pappalardo, M.: A cMUT linear array used as echographic probe: fabrication, characterization, and images. *Proc. IEEE Ultrason. Symp.* **1**, 395–398 (2004)
7. Savoia, A., Caliano, G., Carotenuto, R., Longo, C., Gatta, P., Caronti, A., Cianci, E., Foglietti, V., Pappalardo, M.: Enhanced echographic images obtained improving the membrane structural layer of the cMUT probe. *Proc. IEEE Ultrason. Symp.* **4**, 1960–1963 (2005)
8. Caronti, A., Caliano, G., Gatta, P., Longo, C., Savoia, A., Pappalardo, M.: A finite element tool for the analysis and the design of capacitive micromachined ultrasonic transducer (cMUT) arrays for medical imaging. *J. Acoust. Soc. Am.* **123**, 3375 (2008)
9. Caliano, G., Caronti, A., Savoia, A., Longo, C., Pappalardo, M., Cianci, E., Foglietti, V.: Capacitive micromachined ultrasonic transducer (cMUT) made by a novel “Reverse Fabrication Process”. *Proc. IEEE Ultrason. Symp.* **1**, 479–482 (2005)
10. Caliano, G., et al.: Surface micromechanical process for manufacturing micromachined capacitive ultra-acoustic transducers and relevant micromachined capacitive ultra-acoustic transducer. United States Patent 7790490, 7 Sept 2010
11. Savoia, A.S., Caliano, G., Pappalardo, M.: A CMUT probe for medical ultrasonography: from microfabrication to system integration. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **59**, 1127–1138 (2012)
12. Savoia, A., Caliano, G., Mauti, B., Pappalardo, M.: Performance optimization of a high frequency CMUT probe for medical imaging presented at the *IEEE Ultrason. Symp.*, 2011
13. Savoia, A., Caliano, G.: MEMS-based Transducers (CMUT) For medical ultrasound imaging. In: Chen, C.H. (ed.) *Frontiers of Medical Imaging*, pp. 445–464. World Scientific, 2014

Chapter 8

Open Platforms for the Advancement of Ultrasound Research

Enrico Boni, Luca Bassi, Alessandro Dallai, Gabriele Giannini, Francesco Guidi, Valentino Meacci, Alessandro Ramalli, Stefano Ricci and Piero Tortoli

Abstract The implementation and experimental test of new imaging methods has been hampered by the closed architecture of clinical ultrasound scanners for many years. The so-called open platforms, i.e. flexible scanners with unlimited access to raw echo-data, allow overcoming this limitation and are increasingly used in ultrasound research laboratories. In this paper, a family of open platforms developed in the MSD laboratory in Florence is described. The first system was designed by taking into consideration the need of accurately balancing computational power with cost, dimensions with programmability. A compact and flexible 64-channel system was thus implemented, and is presently adopted by more than 20 worldwide research centers. In the new version, which is in advanced development phase, emphasis is put on the capability of independently controlling a high number (256) of channels as well as on the computational power and memory size.

Keywords Research scanner · High frame rate · High performance computing

8.1 Introduction

Experimental research in ultrasound Imaging/Doppler frequently involves new transmission strategies, non-conventional beamforming techniques, and custom data processing. This flexibility is not available on commercial equipment designed for clinical use, while full control of the transmit/receive (TX/RX) process and full access to raw echo data are possible in few open research platforms. Among these, the ULtrasound Advanced Open Platform (ULA-OP), which was entirely developed in our laboratory, is characterized by an accurate balance among computational power, cost, dimensions, programmability and friendly use. As described in

E. Boni (✉) · L. Bassi · A. Dallai · G. Giannini · F. Guidi · V. Meacci · A. Ramalli · S. Ricci · P. Tortoli
Information Engineering Department, University of Florence, Florence, Italy
e-mail: enrico.boni@unifi.it

the next Section, by using five Field Programmable Gate Arrays (FPGAs) and one Digital Signal Processor (DSP), all electronics was integrated in two programmable boards, coupled to a host PC through USB 2.0. A powerful, compact, portable and relatively cheap architecture was thus obtained. ULA-OP features include the transmission of arbitrary waveforms to 64 elements selected out of a 192-element array probe, the storage of radiofrequency (RF) echo data, the possibility of implementing custom data processing algorithms on the DSP and of displaying the results in real-time.

ULA-OP was duplicated on request, and delivered to external users. An international network of laboratories that employ this ultrasound platform has thus been created. The feedback from these laboratories has pushed the implementation of additional hardware and software modules addressed to make possible the test of different new US methods, described in Sect. 8.4.

This successful experience, together with the need of further expanding the range of possible new applications, has recently pushed the development of a more complex ultrasound platform called ULA-OP 256. ULA-OP 256 has been designed to directly and independently control up to 256 TX-RX channels connected to linear or matrix probes, either Capacitive Micromachined Ultrasound Transducers (cMUT) or piezoelectric (Sect. 8.3). Real-time user-definable operation modes including innovative beamforming algorithms and strategies (e.g. multi-beam, parallel beam, plane-wave imaging etc.) are supported. Access to raw data at any point of the RX chain, including, in particular, data acquired by each probe element, is possible. High computational power and storage capability are achieved through extensive and optimized use of high-end devices. At the same time, particular efforts have been made to integrate the electronics in a limited number of boards, in order to facilitate system transportability and further expand the international network of laboratories and clinical centers which can benefit from ULA-OP features.

8.2 ULA-OP Description

ULA-OP is contained in a compact rack with dimensions $33 \times 23 \times 18$ cm and weight 5 kg. The rack is connected to a PC. The hardware of ULA-OP includes two specialized boards (an analog and a digital board) linked by a backplane (see Fig. 8.1). A third board provides the power supply. External connections are limited to a connector for 192-element array probes, the 18 V ac power input and a USB 2.0 channel toward the host PC.

The analog board includes the RF front-end while the digital board carries out the required numeric signal processing. In particular, the latter is based on 5 FPGAs from Stratix II family (Altera, San Jose, CA, USA) and a DSP from TMS320C64XX family (Texas Instruments, Austin, TX, USA). The TX section integrates a bank of arbitrary waveform generators (AWGs) that synthesize 64 independent excitation signals through the sigma-delta technique. The signals, ranging between 1 and 16 MHz, are passed through 64 linear amplifiers.

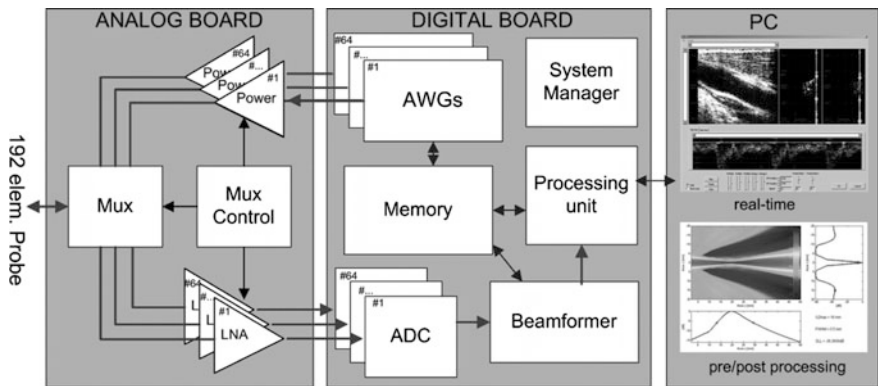


Fig. 8.1 ULA-OP architecture

A programmable switch-matrix maps the excitation signals to a selection of the 192 array elements. During the TX and RX phases the switch-matrix can select either the same or a different group of array elements. The received signals are processed through Low Noise Amplifiers (LNA) and Programmable Gain Amplifiers (PGA) with Time-Gain Compensation (TGC) control. The gain of each channel is programmable in the range 6–46 dB. The 64 channels feed the digital board where a bank of Analog-to-Digital (AD) converters works at 50 Msps. Four FPGAs are devoted to the beamforming, with programmable apodization and dynamic focusing. The 5th FPGA, together with the DSP, elaborates the beamformed signal. In particular, this FPGA contains 4 coherent demodulators, filters, decimators and other calculation-intensive modules which can be optionally inserted in the signal processing chain. The firmware architecture running on the DSP supports several concurrent processing modules each devoted to a standard elaboration, like B-mode or spectral Doppler processing, or to novel custom experimental methods. Each module arranges the results in a frame that is transferred to the host PC through the USB channel.

During system operation, at each Pulse Repetition Interval (PRI), the DSP programs all the ULA-OP aforementioned functions to obtain the TX/RX desired behavior. For example, for each PRI, the TX waveforms and the relative delays, the TX probe aperture, the RX aperture, the channel gains, the beamformer dynamic apodization and focus, the digital processing chain, etc. can be changed by a software control.

One of the most requested features in a research platform is the capability of accessing data in multiple points of the RX chain. The digital board reserves up to 1 GB of SDRAM memory for saving the raw data acquired from the 64 input channels. Here, the data flow rate is near 40 Gb/s. Hence, depending on the PRI and ROI extension, up to 1 s of raw data can be saved. The DSP can access a 256 MB memory, where beamformed and/or DSP-processed data can be saved. Since the throughput rate is here much lower, several seconds of data can be saved. Finally, the video data moved to the PC can be continuously saved, with the only limit of hard-disk capacity. At any moment the on-board memories can be downloaded to the PC into an open-format file, provided that the real-time acquisition is temporarily halted.

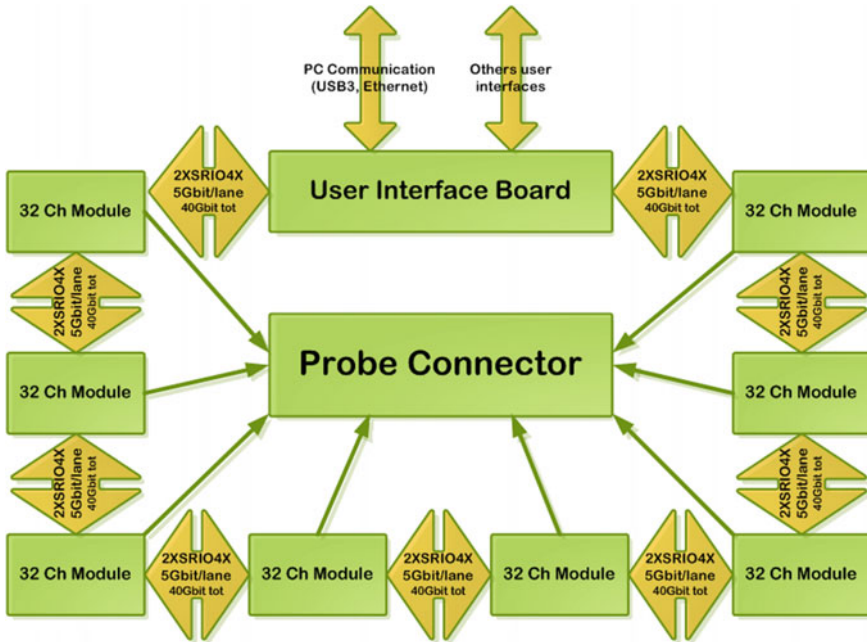


Fig. 8.2 ULA-OP 256 architecture

8.3 ULA-OP 256 Design Specifications

The new ULA-OP 256 platform is fully enclosed in a $35 \times 30 \times 27$ cm rack. The system architecture has been designed in order to gain in flexibility, computational power and bandwidth with respect to the previous instrument (Fig. 8.2).

256 ultrasound channels are managed by 8 FrontEnd (FE) boards. One additional board, the User Interface (UI) board, manages the interaction between a host PC and the system using an USB3.0 connection. The boards are interconnected by a SerialRapidIO (SRIO) link running at 5 Gbit/s on 4 lanes. Each board has 4 SRIO interfaces interconnected through a ring running on the system backplane. The total I/O bandwidth available for each board in the ring is thus 80 Gbit/s full-duplex.

Each FE board integrates all of the electronics to manage TX, RX, beamforming and elaboration of 32 probe channels. In particular, a single FPGA from ARRIA V GX Family (Altera, San Jose, CA, USA) generates the 32 independent TX signals, and then beamforms the received ultrasound echoes according to a programmable strategy. A multi-line beamformer has been implemented inside the FPGA in order to ease the development of high frame rate techniques. The board hosts two DSPs from TMS320C6678 family (Texas Instruments, Austin, TX, USA) featuring eight cores running at 1.2 GHz and 8 GBytes of DDR3 memory each. The DSPs are in charge of receiving the beamformed data and applying coherent demodulation, filtering and custom elaboration modules. The DSPs and the FPGA can

communicate each other and with the rest of the system through a SRIO switch (CPS-1432, Integrated Devices Technology Inc., San Jose, CA, USA) that interconnects all the devices on-board and to the backplane ring.

Full 256 channels beamforming is achieved by summing and elaborating the contribution from the 8 FE boards. The summing can be performed either on one of the DSPs from a FE board or on the UI board.

Data storage capability has been expanded and generalized in this new system. Up to 128 GB of memory are available to store both the raw pre-beamforming data and the results of elaboration. Now, depending on the PRI and ROI extension, up to 30 s of raw data can be saved. The new platform will allow extending the range of applications to real-time non-conventional methods like plane wave imaging and high frame-rate vector Doppler.

8.4 ULA-OP Applications

ULA-OP has been so far used in several non-standard ultrasound applications thanks to its high level of programmability. Specific real-time software modules have been developed for elastography, Doppler and coded imaging. In the former case a novel frequency domain based strain estimation algorithm for freehand elastography [1] has been implemented in real-time. Preliminary in vivo results demonstrate the potentiality of such method as valid support for ultrasound diagnosis of breast lesions [2]. In addition, by processing the multiple echoes received from a selected investigation line through high-speed FFTs performed by the DSP, the blood velocity profiles in human arteries are detected for application in the early diagnosis of possible vascular pathologies [3]. ULA-OP has also been used to implement original vector Doppler methods [4, 5], which allow detecting both blood velocity components in the scan plane, while classic methods only detect the axial component. In these cases, specific modules for the front-end FPGAs, the DSP and the PC software, have been implemented.

Coded imaging is an emerging technique imported from the radar field, in which amplitude and/or phase modulated signals are transmitted and the echoes are compressed in a matched filter. In this application, the sigma-delta AWGs of ULA-OP are properly programmed in TX, while the DSP performs matched filtering in the frequency domain to obtain significantly improved penetration depths [6].

The capability of using non standard TX strategies has also been exploited for high frame-rate (HFR) imaging applications [7]. Here, multiple beams or plane waves are transmitted and by properly processing the corresponding echoes, multiple scan lines are simultaneously beamformed (Parallel beamforming). In this way, a frame rate coincident with the pulse repetition frequency is obtained. In preliminary studies, plane waves have been transmitted and the echoes stored in the memory on board ULA-OP. Suitable off-line processing allowed reconstructing the images of human vessels at a rate of several hundred Hertz [8]. The combination of HFR with vector Doppler was

finally used in studies addressed to reconstruct, off-line, vector velocity maps over extended regions in the scan plane [9].

The new ULA-OP 256 will allow implementing the aforementioned applications in real-time, and further extending the range of possible applications. In particular, the connection to matrix (2D) CMUT probes has been scheduled in collaboration with other Units of a Project of relevant national interest (PRIN 2010–2011).

Acknowledgment This work was partially funded by the Italian Ministry of Education, University and Research (PRIN 2010–2011).

References

1. Ramalli, A., Basset, O., Cachard, C., Boni, E., Tortoli, P.: Frequency domain-based strain estimation and high frame-rate imaging for quasi-static elastography. *IEEE Trans. Ultrason. Ferroelectr. Freq. Contr.* **59**(4), 817–824 (2012)
2. Ramalli, A., Ricci, S., Giannotti, E., et al.: Fourier domain and high frame rate based elastography for breast nodules investigation. In: *IEEE Ultrasonics Symposium Proceedings*, pp. 2241–2244 (2011)
3. Tortoli, P., Palombo, C., Ghiadoni, L., Bini, G., Francalanci, L.: Simultaneous ultrasound assessment of brachial artery shear stimulus and flow-mediated dilation during reactive hyperemia. *Ultrasound Med. Biol.* **37**(10), 1561–1570 (2011)
4. Ricci, S., Bassi, L., Tortoli, P.: Real time vector velocity assessment through multigate Doppler and plane waves. *IEEE Trans. Ultrason. Ferroelectr. Freq. Contr.* **61**(2), 314–324 (2014)
5. Tortoli, P., Dallai, A., Boni, E., Francalanci, L., Ricci, S.: An automatic angle tracking procedure for feasible vector Doppler blood velocity measurements. *Ultrasound Med. Biol.* **36**(3), 488–496 (2010)
6. Ramalli, A., Guidi, F., Boni, E., Tortoli, P.: Real-time base-band pulse compression imaging. In: *IEEE Ultrasonics Symposium Proceedings*, pp. 2002–2005 (2013)
7. Tong, L., Ramalli, A., Jasaityte, R., Tortoli, P., D’hooge, J.: Multi-transmit beam forming for fast cardiac imaging—experimental demonstration and in-vivo application, accepted for publication in *IEEE Transaction on Medical Imaging* (2014)
8. Boni, E., Cellai, A., Ramalli, A., Tortoli, P.: A high performance board for acquisition of 64-channel ultrasound RF data. In: *IEEE Ultrasonics Symposium Proceedings*, pp. 2067–2070 (2012)
9. Lenge, M., Ramalli, A., Boni, E., Cellai, A., Liebgott, H., Cachard, C., Tortoli, P.: Frequency-domain high frame-rate 2D vector flow imaging. In: *IEEE Ultrasonics Symposium Proceedings*, pp. 643–646 (2013)

Chapter 9

A Robust Tracking Algorithm for Super-Resolution Reconstruction of Vehicle License Plates

Stefano Marsi, Sergio Carrato and Giovanni Ramponi

Abstract We propose a novel, very robust method for tracking a vehicle license plate in a sequence of low-resolution frames acquired by a video surveillance camera in order to reconstruct the license plate view in a super-resolution image. The tracking method is able to follow the license plate corners position with sub-pixel resolution and to compensate for small non translational spatial movements of the target during the motion by adopting a perspective transformation. In the reconstruction of the target each frame is perspectively transformed, aligned, cropped, de-convolved and interpolated to higher resolution. Eventually the data are combined into a super-resolution image.

Keywords Perspective tracking · Super-resolution · License plate

9.1 Introduction

License plate recognition (LPR) is widely recognized as a fundamental but difficult task in forensic applications. Many factors indeed affect the result e.g., a wide variety of environmental conditions, the relative orientation of license plate and camera, and the angular width of the target as seen by the camera. A recent survey paper indicates in particular that future research in this field should concentrate, among other topics, also on exploiting the temporal information [1].

Multiframe super-resolution (SR) techniques are particularly well suited for this application, since they permit to recover valid results from poorly acquired data,

S. Marsi (✉) · S. Carrato · G. Ramponi
Department of Engineering and Architecture, University of Trieste,
Via A. Valerio 10, Trieste, Italy
e-mail: marsi@units.it

S. Carrato
e-mail: carrato@units.it

G. Ramponi
e-mail: ramponi@units.it

gathering information not only from the target itself but also from the motion of the vehicle, and in particular from the constraints that this motion must comply to. According to [2], “enhancing license plate text in real traffic videos is a challenging problem for LPR which is not sufficiently addressed in the literature and still has plenty of room for research”: SR is most profitable for this task since the target object smoothly moves in the video.

For general-purpose SR, frequency-domain techniques can be used [3, 4], which are however limited to translation motion [5]. Spatial-domain techniques instead do not suffer from this limitation [6, 7], and are used in the technique we propose. To build the super-resolved image, we propose a tracking method that is able to follow the license plate corners position with a sub-pixel resolution and to compensate for small spatial movements of the target during the motion. To reconstruct the target, each frame undergoes a perspective transform, and is then aligned, cropped, de-convolved and interpolated to higher resolution.

9.2 Overview of the Proposed Method

The algorithm we describe is innovative in particular for what concerns the registration process. It is to be noted that real applications of SR require having a nearly perfect registration of the native images. In the case of license plates, the registration can be done knowing the correct position of the four corners of the license plate itself. However, tracking these points using conventional methods (e.g. block matching algorithms) is critical, since the result can be greatly influenced by many external factors (noise, target scale and orientation, lighting conditions etc.).

In order to make the tracking more robust, we exploit the fact that the license plate (and the vehicle to which it is attached) is of course a rigid body, whose motion can be represented via a perspective transformation applied to the positions it had in successive frames. Since this transformation is described by eight independent parameters, it can be completely defined when the displacement of four not-aligned coplanar points is known; however, to make the process more robust, we over-constrain it by tracking a larger number of points. Indeed, we estimate the displacement of a dozen of different “points of interest” (POIs) located in the front or back part of the vehicle. In this preliminary version of the algorithm, the POIs are defined by the final user, who should select points that can be easily tracked; good candidates for instance are points that belong to regions with a particular shape or with a high local contrast.

It has to be noted that we do not estimate an independent motion vector for each POI; rather, we constrain the vectors to obey to the same perspective transformation. We have thus to solve a minimization problem, i.e. we find the frame-by-frame perspective transformation which minimizes the block matching error for all the considered POIs.

Following this method we are able to track the license plate corners along the video sequence in a very effective way. The next step is the registration of all the

low-resolution frames; the perspective transformation can be fruitfully used for this task too, since it outperforms other registration methods based on linear transformations.

After completing the image registration process, we merge the various frames into a single high-resolution frame. For this purpose, we tested two different algorithms. In the former, all the low-resolution frames are cropped (around the license plate), perspectively transformed and interpolated via a bicubic operator; then, all the resulting images are combined using a temporal α -trimmed mean filter [8, 9]. In the latter, the images are perspectively transformed into a higher-resolution domain and then combined via an iterative process of gradient descent of a suitable cost function as described in [10].

Both the above techniques provide results of acceptable quality. It should be noted that an objective comparison method is however lacking: only subjective examination can be used, because no ideal reference image is available in real-world test conditions.

9.3 License Plate Tracking and Registration

The aim of this part of the algorithm is to best identify the location and the orientation of the license plate.

At the beginning, the user marks the corners of the license plate and the POIs in the first (reference) frame of the sequence. Then, the system automatically computes, frame by frame, the new positions for all the POIs. This identification is based on a maximum likelihood criterion, with an algorithm that could be described as a “constrained multiple block matching” (CMBM). More in particular, we consider a block around each POI in the reference frame and—like in classical block matching (BM)—we search, among all the blocks in the present frame, the displacement that minimizes the mean squared difference (MSD) with respect to the corresponding block in the first frame. However, while in classical block matching each block position is estimated independently of all the other ones, in our method each position is constrained by those assumed by all the other blocks: i.e., they all must be obtained by the same perspective transformation from the original position. Moreover, to make the BM process less sensitive to possible luminance variations, the mean value of the pixels that compose each block is preliminarily subtracted to all the block’s pixel values.

The idea which the proposed CMBM process is based on is depicted in Figs. 9.1 and 9.2. In each figure two frames of the sequence are considered. The red squares represent the blocks surrounding the POIs, and the yellow grid represents the plane that is supposed to contain the POIs. Moving from one frame to the following one, the plane grid will be stretched by a perspective transformation while the relative position of the POIs in the plane grid is preserved. An iterative algorithm based on genetic algorithms is used to solve this optimization problem. In Fig. 9.2, in particular, it can be seen that the system is able to track a perspective change due to a rotation of the object.



Fig. 9.1 Example of the CMBM algorithm applied to two frames of the sequence. The block matching is performed under the constraint that all blocks must maintain their relative positions

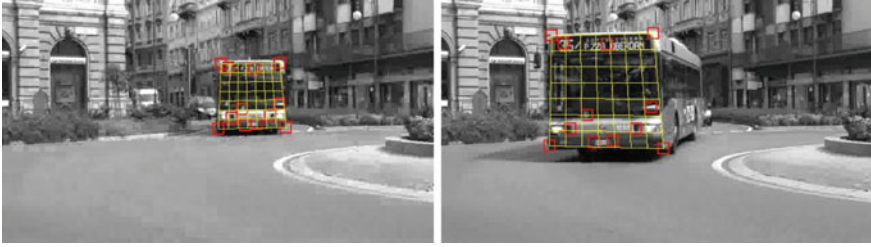


Fig. 9.2 A different application of the CMBM algorithm. It can be noted that the system is able to track the bus face even during a rotation

If $\bar{P}(x, y)$ represents the value of the pixel after subtraction of the block mean value, for each block surrounding a POI we can write

$$MSD^k(x_k, y_k) = \frac{1}{(2N + 1)} \sum_{i=-N}^N \sum_{j=-N}^N \left(\bar{P}_o^k(i - x_k, j - y_k) - \bar{P}_r^k(i - \hat{x}_k, j - \hat{y}_k) \right)^2 \quad (9.2)$$

and:

$$MSD = \sum_{k=1}^K MSD^k(x_k, y_k) \quad (9.3)$$

where k identifies one of the K POIs, P_o are the pixels values in the frame under evaluation and P_r are the pixels in the reference frame, (x_k, y_k) are the coordinates of the present position of the k -th POI and (\hat{x}_k, \hat{y}_k) are the corresponding coordinates in the reference frame; finally, MSD^k is the mean squared difference associated with the position (x_k, y_k) . The goal of the CMBM algorithm is to find a complete set of coordinates $\{x_k, y_k\}$ where $k \in K$ (i.e. for all the POIs) that both minimize the MSD in Eq. (9.3) and are related to $\{\hat{x}_k, \hat{y}_k\}$ by a suitable perspective transformation f_p :

$$\{x_k, y_k\} = f_p\{\hat{x}_k, \hat{y}_k\} \quad (9.4)$$

The perspective transformation is described by the following equations:

$$\begin{aligned} x_k &= \frac{t_{11}\hat{x}_k + t_{12}\hat{y}_k + t_{13}}{t_{31}\hat{x}_k + t_{32}\hat{y}_k + 1} \\ y_k &= \frac{t_{21}\hat{x}_k + t_{22}\hat{y}_k + t_{23}}{t_{31}\hat{x}_k + t_{32}\hat{y}_k + 1} \end{aligned} \quad (9.5)$$

9.4 Observations About the Implementation of the Genetic Algorithm

As already mentioned the proposed algorithm operates in two phases: first, the car face is tracked along the frames; then, all the frames are cropped around the license plate, are registered and recombined in a super-resolution image.

For the tracking phase, a genetic algorithm [11] has been adopted. This algorithm starts from a random population of elements called chromosomes; each chromosome represents a possible solution of the problem. It then evaluates a suitable fitness function to estimate the reliability of the solution, and iteratively creates new populations composed only by the best performing elements. The algorithm discards all the chromosomes that show low fitness function value and creates new elements through suitable operations of crossover and mutation applied to the best chromosomes. Genetic algorithms usually permit to reach effective solutions, avoiding getting stuck in local minima, with a limited number of iterations.

The algorithm described above can operate in real time on a suitable embedded system composed by a processor with custom hardware resources, exploiting a master-slave strategy [12, 13]. The processor can be used to control the chromosome generation, supervising the crossover and the mutation operations together with the selection process. In turn, a dedicated hardware adopting a suitable direct memory access (DMA) architecture can directly read the pixel values stored in the frame buffer and can compute the fitness function. The creation of new chromosomes is not very demanding in terms of computational load, and can easily be performed in a serial way; on the contrary, the evaluation of the fitness function requires a more careful design.

In our specific case, the fitness function computes the sum of the absolute differences (SAD) of a dozen blocks. To reach convergence, we estimated that the algorithm must pass through about a hundred of generations, each consisting of at least 25 chromosomes. This process requires approximately 2500 calls to the fitness function. In each of these calls the system must calculate the SAD of a dozen blocks that in our experiments have a 32×32 pixels size, for a total of about 30×10^6 accesses to the frame buffer. Considering a rate of 10 frames/s that is typical of a video surveillance system, and assuming that the images are encoded with 8 bit/pixel, real time operation would require 300 MT/s and a bit-rate of about

2.4 Gb/s to access the frame buffer. Considering a resolution of 1024×768 pixels, 63 Mb/s are further needed to store the frames into the frame buffer. This data transfer frequency is compatible only with DDR2 memory or superior [14].

Two strategies are often used [15, 16] to improve the speed of genetic algorithms: cache memory and multiprocessor systems. In our case, they are not feasible solutions unfortunately. Indeed, cache memories can limit the memory bandwidth in traditional block matching approaches that use a regular window scan: they store the pixels involved in the matching computations and reduce the access to the frame buffer to what is needed to update few pixels at each iteration. In our case, however, the positions of the blocks change over a wide spatio-temporal area in an unpredictable way at each iteration, and we could not devise any strategies to select pixels to be cached.

In multiple-processor systems, several fitness functions associated to different chromosomes are evaluated in a parallel way, thus reducing the time needed to manage each new generation. This approach is effective when the fitness function for different chromosomes can be evaluated independently and without sharing hardware resources. In our case each fitness function accesses the pixel values, and even if the operation itself could be split among a number of parallel processors, the access to the memory must be performed in a serial way. The memory bandwidth is thus the main factor that limits the processing speed.

9.5 Super-Resolution Image Generation

The second part of the algorithm, which consists in the recombination of all the images in a super-resolved one, can be carried out only when all the low-resolution images are available and effectively registered. This part typically operates off-line, but it can easily be changed into a delayed real-time process: it is sufficient to store all the required images in a suitable FIFO memory bank.

Once we have a valid registration of images with subpixel accuracy, we must exploit all the available information in order to build the SR frame.

This process goes through several steps that typically include:

- Interpolation to higher resolution
- Merging of the various images into a single higher resolution version
- De-blurring of the final result

These steps can be carried out separately or can be combined within an optimization algorithm that finds the optimal solution that minimizes a suitable cost function. In this paper, we consider both these approaches.

In the first approach, after the identification of the license plate corners coordinates, the process applies to each frame a suitable perspective transformation followed by a bicubic interpolation algorithm to map the license plate image into a higher resolution domain. This process produces a series of higher resolution images of the license plate, which are well aligned to each other. Then, each pixel of the final super-resolution image is estimated by applying a temporal α -trimmed

filter to all the pixels at the same spatial coordinates in the series of images. The main advantage of such a filter is its ability to discard the outliers and to calculate the average only of the pixels that have values close to each other. A blind deconvolution operator is finally applied to partially recover the blurring present in the acquisition phase due to the PSF of the camera and of the atmosphere.

The second approach is based on the assumption that the low-resolution images actually derive from a high-resolution image (i.e., our target) that in the different frames generation has undergone a sequence of alterations, namely spatial transformation, blurring, filtering, decimation and noise corruption. This approach tries to reverse this process and aims at estimating the image that can best generate the low-resolution image sequence. In order to do this, an iterative process is performed which minimizes a suitable cost function consisting of the weighted sum of two factors. The first one takes into account the difference between the low-resolution data and the data coming from the target that supposedly yielded the available sequence of low-resolution frames when the decimation process is applied. The second part of the cost function takes into account the difference between the target and its “ideal” version, which consists in a noise-free image with sharp edges; the latter is estimated via a suitable edge-preserving smoothing. An iterative process of gradient descent along the error function is used to find the desired result. It should be noted that this optimization process is rather critical, as it requires the a priori knowledge of the proper decimation process that generated the low-resolution images and the PSF of the acquisition system, and is quite sensitive to the tuning on the various parameters.

The first solution is quite trivial, but it has the important advantage that it does not require any fine parameter tuning. On the contrary, the second solution depends on a correct estimation of the point spread function of the camera; moreover, the result is guaranteed only in case of Gaussian noise, and the iterative process used to find the minimum in the cost function is critical with respect to the choice of the convergence parameters.

As an example, in Fig. 9.3 we show a detail of a few low-resolution frames (top), and the reconstructed license plate. It can be noticed that a readable license plate has been obtained, while very few characters were recognizable in the original images.



Fig. 9.3 Detail of a few low-resolution frames (*top*), and the reconstructed license plate using our algorithm (*bottom*)

9.6 Conclusions

We have presented in this contribution a robust method for multiframe super-resolution, and its application to license plate recognition. Some observations about the implementation of the genetic algorithm have been provided. The results that have been achieved are promising, but further effort needs to be devoted to a deeper analysis of the most effective choice of the POIs, and of the possible automation of their selection. Tackling the mentioned issues would permit to exploit the proposed method not only in (typically off-line) forensic studies but also in the real-time context of surveillance tasks.

References

1. Shan, D., Ibrahim, M., Shehata, M., Badawy, W.: Automatic license plate recognition (ALPR): a state-of-the-art review. *IEEE Trans Circuits Syst. Video Technol.* **23**(2), 311–325 (2013)
2. Anagnostopoulos, C.-N.E.: License Plate Recognition: A Brief Tutorial. *IEEE Intell. Transp. Syst. Mag.* **6**(1), 59–67 (2014)
3. Rhee, S.H., Kang, M.G.: Discrete cosine transform based regularized high-resolution image reconstruction algorithm. *Opt. Eng.* **38**(8), 1348–1356 (1999)
4. Kim, S.P., Su, W.Y.: Recursive high-resolution reconstruction of blurred multiframe images. *IEEE Trans. Image Process.* **2**(4), 534–539 (1993)
5. Zeng, W., Lu, X.: A generalized DAMRF image modeling for superresolution of license plates. *IEEE Trans. Intell. Transp. Syst.* **13**(2), 828–837 (2012)
6. Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the nonlocal means to superresolution reconstruction. *IEEE Trans. Image Process.* **18**(1), 36–51 (2009)
7. Cortijo, F.J., Villena, S., Molina, R., Katsaggelos, A.: Bayesian superresolution of text image sequences from low-resolution observations. In: *Proceedings IEEE International Symposium on Signal Process Applications*, pp. 421–424 (2003)
8. Bednar, J., Watt, T.: Alpha-trimmed means and their relationship to median filters. *IEEE Trans. Acoust. Speech Signal Process.* **32**(1), 145–153 (1984)
9. Peterson, S.R., Lee, Y.H., Kassam, S.A.: Some statistical properties of alpha-trimmed mean and standard type M filters Acoustics. *IEEE Trans. Speech Signal Process.* **36**(5), 707–713 (1988)
10. Farsiu, M., Robinson, M.D., Elad, M., Milanfar, P.: Fast and robust multiframe super resolution. *IEEE Trans. Image Process.* **13**(10), 1327–1344 (2004)
11. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley, Boston (1989)
12. Ke, Y., Li, Y., Li, D.: Image matching using genetic algorithm on GPU control. In: *2011 International Conference on Automation and Systems Engineering (CASE)*, pp. 1–4 (2011)
13. Paulinas, M., Ušinskas, A.: A survey of genetic algorithms applications for image enhancement and segmentation. *Inf. Technol. Control* **36**(3), 278–284 ISSN 1392-124X (2007)
14. JEDEC Solid State Technology Association—JEDEC Standard—DDR2 SDRAM Specification—JESD79-2E—April 2008 (www.jedec.org)

15. Johar, F.M., Azmin, F.A., Suaidi, M.K., Shibghatullah, A.S., Ahmad, B.H., Salleh, S.N., Aziz, M.Z.A.A., Md Shukor, M.: A review of genetic algorithms and parallel genetic algorithms on graphics processing unit (GPU) Control System. In: 2013 IEEE International Conference on Computing and Engineering (ICCSCE), pp. 264–269 (2013)
16. Caifeng, T., Anguo, M., Zuocheng, X.: Research on the parallel implementation of genetic algorithm on CUDA platform. *Comput. Eng. Sci.* **31**, 68–72 (2009)

Chapter 10

c-Walker: A Cyber-Physical System for Ambient Assisted Living

**Luca Rizzon, Federico Moro, Roberto Passerone, David Macii,
Daniele Fontanelli, Payam Nazemzadeh, Michele Corrà,
Luigi Palopoli and Domenico Prattichizzo**

Abstract The c-Walker is a smart rollator that provides physical sustain to people with mobility difficulties together with a cognitive support to overcome disabilities related to the decrement of sensorial abilities. The proposed system is made of a conventional walker equipped with a variety of sensors, actuators, user interfaces, and computing units. Various algorithms monitor environmental data. The system processes them to define the safest path for the user, and transmits useful information for navigation via multiple interfaces to the assisted person. Moreover, the system can take control of the direction to avoid hazards. To design and develop the c-Walker, we adopt state-of-the-art design methodologies that assist the designers in

L. Rizzon (✉) · F. Moro · R. Passerone · D. Macii · D. Fontanelli ·
P. Nazemzadeh · L. Palopoli
Università di Trento, via Sommarive, 9, 38123 Povo di Trento, Italy
e-mail: luca.rizzon@unitn.it

F. Moro
e-mail: federico.moro@unitn.it

R. Passerone
e-mail: roberto.passerone@unitn.it

D. Macii
e-mail: david.macii@unitn.it

D. Fontanelli
e-mail: daniele.fontanelli@unitn.it

P. Nazemzadeh
e-mail: Payam.nazemzadeh@unitn.it

L. Palopoli
e-mail: luigi.palopoli@unitn.it

M. Corrà
TRETEC S.r.l c/o Polo Tecnologico, via Solteri 38, 38121 Trento, Italy
e-mail: michele.corra@3tec.it

D. Prattichizzo
Università degli Studi di Siena, via Roma, 56 Siena, Italy
e-mail: prattichizzo@dii.unisi.it

the integration phase. In this work we describe the technology of hardware and software components included in the prototype device.

Keywords Ambient assisted living (AAL) • Cyber-physical systems • Indoor navigation • Robotics

10.1 Introduction

Adults with reduced physical or mental ability, particularly elderly people, may find it difficult to perform simple daily tasks, such as shopping, or moving in crowded areas. As a result, they tend to avoid going out, and suffer from a consequent loss of physical and mental wellbeing, arising from reduced exercise and reduced socializing [1]. We address this problem by developing a portable assistive device for guiding users with reduced cognitive abilities to navigate autonomously. The device we developed can guide the user towards the safest path in the local environment, by providing him/her with additional cues for orientation. People with cognitive disabilities cannot take advantage of a conventional walker. On the contrary, our device combines a physical support for gait with a guidance device that augments spatial information and navigation ability of the assisted person, thus increasing her/his possibilities of conducting independent life.

This work is developed in the context of the European project DALi (Devices for Assisted Living). Our system, called c-Walker, consists of a “smart-rollator”. The c-Walker localizes itself inside the environment, and acquires dynamic information from the surrounding space. This information is used to compute the safest path the user has to follow in order to reach the destination point he/she has previously selected on a touchpad. The user receives information to follow this path via multiple interfaces, and the system is able to correct his/her behavior to avoid dangerous situations. The construction of a system of such a complexity represents a major scientific and technological effort bringing together expertise from different disciplines. The proposed device represents a case study for Cyber-Physical Systems, as it is a robotic platform that closely interacts with the user (human in the loop), and requires a seamless integration of different subsystems [2]. The physical entities of the system are the mechanical components of the c-Walker, and the actuators used to provide stimuli to the users.

In this paper we describe the technology we chose to build such a system, and describe the design process, and the development of the functional prototype. Given the peculiarity of the system some of the requirements are not only driven by technological aspects, but can arise from the mechanism of perception and physiological features of people with special needs.

Fig. 10.1 Picture of the c-Walker prototype



10.2 System Description

A conventional four-wheeled walker has been equipped with sensors, actuators, and computing units in order to create a robotic platform with the desired capabilities. A picture of the prototype is shown in Fig. 10.1. A pair of incremental encoders is mounted on the back wheels, together with a pair of electrical brakes. Each of the front wheels is equipped with a stepper motor that allows for steering actions, and with an absolute encoder that monitors the steering angle of the wheels. Every wheel is driven by a microcontroller that implements all the wheel functionalities, such as readings of the encoders (incremental or absolute), actuation on the brakes, and actuation/readings from the stepper motors. The walker has also an on-board inertial platform equipped with gyroscopes, and accelerometers. The four microcontrollers, and the inertial platform are five nodes of a CAN bus network, with a Beaglebone being the master of the network.

The Cyber part of the c-Walker has the role of localizing the system inside the environment. Also, it computes the route to the destination that minimizes the risk of accidents, and show to the user the direction to take to reach a given point of interest. The user receives tactile, visual, and audio stimuli that suggest the direction he/she has to follow. Localization, control algorithms, and human-machine interface (HMI) are described in the following subsections.

10.2.1 *Localization and Tracking*

One of the basic functions implemented in the smart walker is its ability to estimate in real-time both user position and direction of motion with respect to a known reference frame. Walker localization is indeed essential not only to support user navigation, but also to enable higher-level functions, such as long-term and short-term trajectory planning. The position-tracking algorithm implemented in the walker relies on multi-sensor data fusion based on an Extended Kalman Filter (EKF), built upon a classic unicycle kinematic model [3, 4]. As known, the EKF algorithms are iterative and consist of two steps, i.e. prediction and update. In the prediction step, the state of the walker (namely its x-y Cartesian coordinates on the map of the chosen environment as well as its orientation with respect to x-axis) is obtained by integrating the angular displacements measured by the two incremental encoders installed on the rear wheels. However, the initial values of orientation and position (i.e. at time 0) cannot be observed when just the odometers are used. Moreover, the uncertainty (both in position and orientation) resulting from the integration of encoders' data accumulates as the traveled distance grows, thus leading to unacceptable errors in the long run.

In order to solve the initial observability problem, and to keep positioning uncertainty below 1 m with 95 % probability, in the update step, the EKF adjusts the estimated state by detecting a grid of low-cost passive RFID tags (for position measurements), and visual QR markers (for orientation measurements) stuck on the floor. The orientation update is also assured by a constantly active gyroscope, which improves performance and reduces the algorithm convergence time. RFID tag and marker detection are generally sporadic because they depend on the user trajectory. Also, the number of tags and markers in the environment has been minimized to reduce deployment complexity and the related costs.

Once the c-Walker position is estimated, the user can tap on the screen of a touchpad the point of interest that he/she wants to reach by choosing it among a list of possible destinations. At that point, the c-Walker computes the path the user has to follow to reach the desired location, and continuously perform self-localization during the walk. The system can also keep track of human agents populating the environment, and the motion-planning algorithm can eventually modify the suggested path according to scene's dynamics. The motion planner has been modeled according to psychological, and sociological mechanisms that rule human motion in crowded environments. The mathematical model is used to predict the long-term motion of human agents as well as the assisted person's intent. The system operates according to this model that represents an extension of the social force model relying on statistical model checking techniques [5]. The algorithm goal is to minimize the probability of collisions or hazards.

10.2.2 *Guidance Systems*

The HMI is responsible for showing to the user navigational information through intuitive cues that stimulate the assisted person at a high level of cognition. The

interface consists of an audio synthesis algorithm for headphones, vibro-tactile bracelets, and a tablet mounted over the c-Walker handlebar.

The audio interface gives to the user spatial information exploiting human sound localization abilities. Humans are able to estimate the position of sound sources because brain is trained to interpret binaural difference and spectral modification that depend on the direction of arrival of sound waves [6]. By synthesizing those phenomena it is possible to give the illusion of spaciousness and direction through headphones [7]. However, sound spatialization effects depend on anthropometric quantities. Therefore, a parametric model of the listener head and pinna is used to generate synthesized stimuli [8]. The c-Walker acoustic interface generates audio signals by associating a virtual sound source to the spatial coordinates corresponding to the relative position of the suggested path, thus giving the sensation that a speaker is placed on the track [9]. The listener is asked to move toward the direction from which he/she interprets that the sound is coming from.

The audio interface is based on a software able to generate positional sounds which depend on the direction computed by the motion planner. Sound stimuli are played back by traditional full-size headphone equipped with an inertial platform mounted on top of the headphone's arches. The inertial platform monitors user's head orientation to ensure a correct displacement of the virtual sound when the user moves his/her head. This allows the user to benefit from head movements in order to decrease the localization blur in case the sound stimuli are difficult to interpret.

The c-Walker also conveys information to the user through haptic interfaces. Tactile stimuli assist the user to follow a path or to move within a safe area, avoiding actual or potential obstacles. The haptic actuators consist of two wearable bracelets. In order to invite the user to steer left (or right), the left (right) bracelet vibrates at a constant rate [10]. The control software changes the vibration as a function of the direction recommended by the motion planner. On the screen of the c-Walker's touchpad, a dedicated application shows the current position of the user on the map, and a list of possible destinations. When the user selects a destination, a red arrow shows the direction to follow in correspondence of decision points such as doors and intersections.

In case the user does not correctly interpret the navigation information, the system is able to take control of brakes, and steering wheels to actively drive the user toward the right path [11]. In fact, if the actual user's trajectory departs significantly from the suggested route, the c-Walker gently guides the user toward the safest path without overriding his/her responsibility of autonomous navigation. However, the brakes can be also used to stop the walker to prevent the user to fall.

10.3 Hardware/Software Architecture

While the c-Walker control algorithm runs on a *Beaglebone* mounted on the walker, localization task and HMI are temporarily executed on a laptop for refinement purposes. The *Beaglebone* is provided with a Linux system with *Preempt_RT*

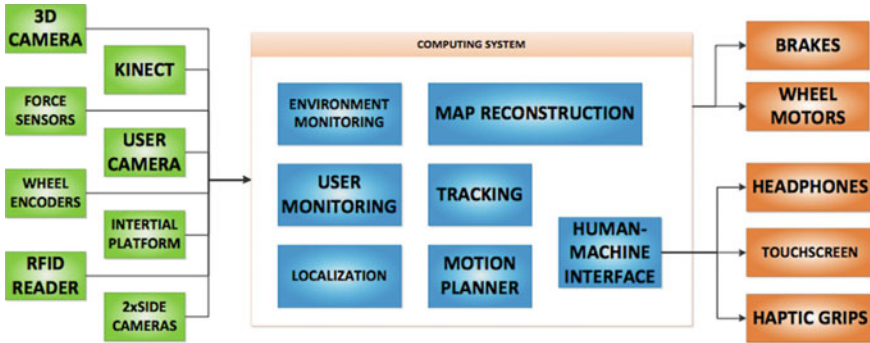


Fig. 10.2 Block diagram of the c-Walker system architecture

patches for real-time support. The same platform collects data from the CAN bus and is connected via USB to the RFID reader. An application interface over Ethernet provides a connection between the hardware at the low level and a potential computing network implementing high-level algorithms that exploit the sensors data. Figure 10.2 represents a simplified block diagram of the systems' components. To accommodate all these algorithms together, we have designed a middleware architecture that offers a seamless integration of the software components deployed across different embedded platform and the PC [12].

Much of the hardware composing the c-Walker is based on off-the-shelf components. This choice is driven by both cost reasons and the will to focus the research resources on cognitive navigation. Designing Cyber-Physical Systems such as the c-Walker cannot be done with a monolithic approach. Designers have to adopt compositional methods that allow researchers to realize a large complex system by assembling simpler components in a structured manner [13].

10.4 Preliminary Results

We performed a series of experiments with elderly users in order to validate the proposed approach as well as to test the developed prototype. During the experimental sessions several people used the c-Walker to reach some predefined points of interest within an ad hoc environment. An operator selected the target location without notifying it to the actual user. The target locations were not associated with a real object in the environment in order to emulate a complete unknown scenario for the user. He/she was required to perform multiple trials, in order to reach the points of interest by means of the assistance guidance. During such trials, the c-Walker was continuously performing self-localization with no other people in the surroundings.

Users were asked to test different guidance techniques, the first one being a set of visual indications shown on the touchpad screen. Other three guidance techniques were tested and they relied on a mix of visual indications and one between haptic

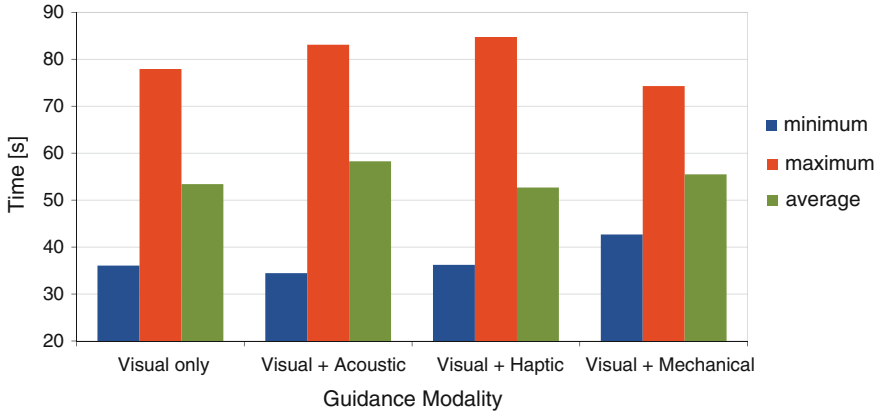


Fig. 10.3 Time (in seconds) required by users navigating in a predefined path. *Blue* and *orange bars* represent respectively the fastest and slower time required by the participants. The *green bars* represent the average time. This data has been collected with 18 participants

bracelets, headphone and front wheels steering (mechanical guidance). Usually, all these guidance techniques are not continuously active. On the contrary, they can give indications about the direction to follow only when the user reaches a decision point, namely whenever the direction to follow is unclear (i.e. at intersections). This series of experiments allowed us to test data acquisition from sensors and communication features as well as to validate the localization algorithm and the different subsystems for guidance. Moreover, we received important feedbacks about usability and correct system integration.

In Fig. 10.3 it is possible to see that there are no remarkable differences between the effectiveness of various guidance techniques. In fact, the users took approximately the same minimum, average and maximum time to reach the target destination. Apparently, the joint visual-haptic approach offers the best average performance. The visual-mechanical, even though looks the best one in the worst case, it is also the worst one in the best case. Therefore, its usability is characterized by a lower variance. This maybe can be due to the fact that the mechanical guidance reduces the probability that the user interprets wrongly the suggestions provided by the system.

10.5 Conclusions

In this paper we have described the hardware and software technologies underlying the development of a functional prototype of an assistive wheeled robotic walker for people with cognitive impairments. We have performed an early evaluation of the prototype by asking potential users to test the guidance systems moving inside a structured environment. Further extensive tests with users with different levels of

disabilities, and technological background will be performed in the near future. The experimental activities will enable us to evaluate the benefits of the c-Walker for the user, when different configurations and combinations of interfaces are adopted. Moreover, we will consider different implementation options to optimize both the hardware and the software system architecture.

Acknowledgments This work is supported by the European FP7 Project Devices for Assisted Living (DALi), grant number ICT-2011-288917, <http://www.ict-dali.eu/dali>.

References

1. Drewnowski, A., Evans, W.J.: Nutrition, physical activity, and quality of life in older adults summary. *J. Gerontol. Ser. A: Biol. Sci. Med. Sci.* 56(suppl 2): 89–94 (2001)
2. Sangiovanni-Vincentelli A., Damm, W., Passerone, R.: Taming Dr. Frankenstein: contract-based design for cyber-physical systems. *Eur. J. Control* 18(3), 217–238 (2012)
3. Nazemzadeh, P., Fontanelli, D., Macii, D.: An indoor position tracking technique based on data fusion for ambient assisted living. In 2013 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pp. 7–12. IEEE (2013)
4. Nazemzadeh, P., Fontanelli, D., Macii, D., Palopoli, L.: Indoor positioning of wheeled devices for ambient assisted living: a case study. In: Instrumentation and Measurement Technology Conference (I2MTC) Proceedings. IEEE (2014)
5. Colombo, A., Fontanelli, D., Legay, A., Palopoli, L., Sedwards, S.: Motion planning in crowds using statistical model checking to enhance the social force model. In 2013 IEEE 52nd Annual Conference on Decision and Control (CDC). IEEE (2013)
6. Blauert, J.: *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT press, Cambridge (1997)
7. Rizzon, L., Passerone, R.: Spatial Sound Rendering for Assisted Living on an Embedded Platform. *Applications in Electronics Pervading Industry, Environment and Society*, pp. 61–73. Springer, Berlin (2014)
8. Spagnol, S., Geronazzo, M., Avanzini, F.: Structural modeling of pinna-related transfer functions. In: Proceedings of the International Conference on Sound and Music Computing (SMC), pp. 422–428 (2010)
9. Rizzon, L., Passerone, R.: Embedded soundscape rendering for the visually impaired. In: 2013 8th IEEE International Symposium on Industrial Embedded Systems (SIES), pp. 101–104. IEEE (2013)
10. Scheggi, S., Chinello, F., Prattichizzo, D.: Vibrotactile haptic feedback for human-robot interaction in leader-follower tasks. In Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments. ACM (2012)
11. Fontanelli, D., Giannitrapani, A., Palopoli, L., Prattichizzo, D.: Unicycle steering by brakes: a passive guidance support for an assistive cart. In: 2013 IEEE 52nd Annual Conference on Decision and Control (CDC). IEEE, (2013)
12. Rizano, T., Abeni, L., Palopoli, L.: Middleware for robotics in assisted living: a case study. In Proceedings of the 15th Real-Time Linux Workshop (2013)
13. Davare, A., Densmore, D., Guo, L., Passerone, R., Sangiovanni-Vincentelli, A.L., Simalatsar, A., Zhu, Q.: MetroII: a design environment for cyber-physical systems. *ACM Trans. Embed. Comput. Syst.* 12(1s), 49:1–49:31 (2013)

Chapter 11

2D and 3D Palmprint Extraction by an Automated Ultrasound System

Antonio Iula, Gabriel Hine, Alessandro Ramalli, Francesco Guidi
and Enrico Boni

Abstract In this work, some possible procedures to extract both 2D and 3D palmprints from the same experimental 3D ultrasound image of the human palm are presented. The ultrasound system used to achieve the 3D images is composed of a CNC commercial pantograph, which moves a high frequency (12 MHz) ultrasound probe along its elevation direction to cover the desired area of the human palm. The ULtrasound Advanced Open Platform (ULA-OP) is employed as ultrasound imaging system.

Keywords Ultrasound system · Biometrics · Palmprint · Acoustic imaging

11.1 Introduction

Biometrics refers to methods for uniquely recognizing humans based upon one or more physical or behavioral traits.

Although biometrics emerged from its extensive use in law enforcement to identify criminals it is being increasingly used to establish person recognition in a large number of civilian applications [1]. Most employed biometric characteristics include DNA, face, hand vein, fingerprint, hand geometry, iris, palmprint, retina, signature, and voice [1].

Multimodal biometric systems, which use more than one independent source of information to recognize individuals in order to improve recognition accuracy, have been successfully developed as well and are more and more exploited in applications [1]. Ultrasounds have several advantages over other technologies (optical, capacitive, NIR) employed for biometric recognition purposes, including the

A. Iula (✉) · G. Hine
School of Engineering, University of Basilicata, Viale dell'Ateneo Lucano 10,
85100 Potenza, Italy
e-mail: Antonio.iula@unibas.it

A. Ramalli · F. Guidi · E. Boni
University of Firenze, Florence, Italy

capability of proving a 3D representation of the biometric characteristics. Furthermore, ultrasounds are able to detect life (Doppler mode) and therefore are difficult to counterfeit, and finally, are not sensitive to the surface contaminations. The main explored ultrasound technique in the biometric field consists in a XY mechanical scan [2–4] for livescan fingerprint: a high frequency (high resolution) A-Scan ultrasonic transducer scans each dot of the XY area and stores in the memory all the reflected acoustic signals from each dot. The result is a 3D image of the under skin volume. The main drawback in the this technique relies in the very high scanning time, which could be prohibitive for practical applications.

In the last years, some of the authors have experimented the possibility to acquire a volume of the human hand or finger with a linear array; in this way only one mechanical scanning, in the elevation direction, has to be performed.

A commercial ultrasound imaging systems and a motorized stepper sledge were employed. During the time interval between two consequent steps, a B-scan is acquired and stored.

The technique has been exploited for extracting and evaluating different biometrics characteristics based on ultrasound by exploiting both piezoelectric and cMUT linear arrays: the internal hand geometry [5]; hand vein pattern [6], fingerprint [7, 8] and palmprint [8, 9].

In a recent work [10], a new system for the acquisition of the ultrasound images is proposed; it is based on the open ultrasound platform ULA OP [11], which allows nonstandard and dedicated transmit/receive strategies and is able of providing a very quick acquisition. An improved automated system, which exploit a numeric controlled pantograph, is employed to guarantee stable and repeatable measurements.

A similar experimental set up has been exploited in this work for the extraction of 2D and 3D palmprints.

11.1.1 2D and 3D Palmprint Extraction

The palms of the human hands contain pattern of ridges and valleys much like the fingerprints [12]. The area of the palm is much larger than the area of a finger and, as a result, palmprints are expected to be even more distinctive than the fingerprints. Since palmprint scanners need to capture a large area, they are bulkier and more expensive than the fingerprint sensors. High-resolution palmprint are suitable for forensic applications such as criminal detection. Human palms also contain additional distinctive features such as principal lines and wrinkles that can be captured even with a lower resolution scanner (150 dpi or less). Most research attention has recently focused on low resolution images for civil and commercial applications. Classical palmprint recognition is based on the acquisition of a 2D image of the palm. Then, a region of interest (ROI) is selected and post elaborated for feature extraction and matching.

Palmprint recognition using 3-D information of the palm surface is the most recently explored technique [13] in order to overcome 2-D palmprint difficulties

like scrabbling problems due to illumination variations and the possibility to be counterfeited. This technique is based on the “light structural imaging” and allows to extract the features by taking account for the palm curvature.

11.2 Experimental Setup

Figure 11.1 shows the experimental setup used for get 3D ultrasonic images of the human hand. The palm of the user is placed on a plastic jig, which has guidance marks on it. The hand is properly aligned by the marks and is completely immersed in water with the palm facing upwards. The hand is not clamped to the jig; in this way, users with hand of different dimensions can assume a position as much natural as possible.

The Region Of Interest (ROI) of the human palm is illuminated by an ultrasound ultrasonic probe, which is connected to an ultrasound imaging system.

The ultrasonic probe is a commercial 192 elements linear array by Esaote S.p.A., Genova—ITALY, model LA435, which is based on piezocomposite technology and is used for several clinical applications investigating small parts. It has a central frequency of 12 MHz with a bandwidth of about 75 %, a pitch of 200 μm , a 3.5 mm elevation aperture, a total aperture of 38.4 mm. It is partially immersed in order to be faced and acoustically coupled to the hand palm.

The probe is mechanically shifted along the elevation direction by means of a numeric controlled pantograph by Delta Machine CNC, Vazia (RI)—ITALY, which is able to guarantee a precision better than 20 μm .

As ultrasound imaging system an advanced open platform for ultrasound research (ULA OP) was employed. The system is extremely compact as all electronics are integrated in two dedicated boards exploiting the latest digital electronic technology., which are contained in a box and connected to a PC through USB 2.0. The system is characterized by the full programmability of each critical section and allows to implement TX-RX strategy to control simultaneously a maximum number of 64 elements.

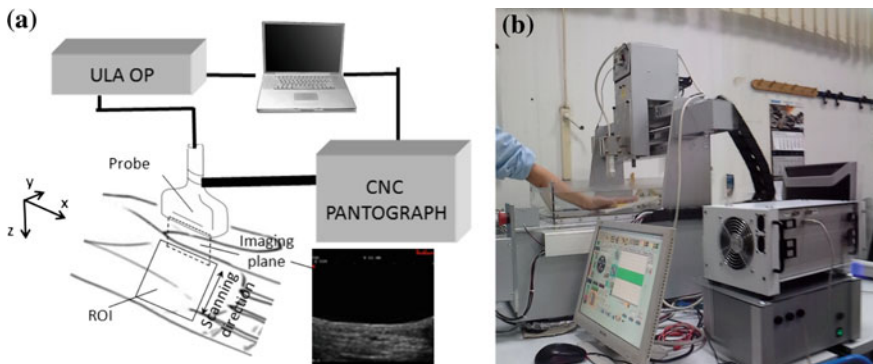


Fig. 11.1 Experimental set up use for the acquisition of the 3D ultrasonic image: **a** schematic; **b** photo

Both the pantograph and ULA OP are controlled in MATLAB environment (The MathWorks, Inc., MA, USA).

In order to make the acquisition process approximately as fast as that achievable with conventional techniques, 250 BScan images (see Fig. 11.1a) have been acquired at regular time intervals during the continuous motion of the probe. Following this approach, the 3D ultrasound image corresponding to a volume of $40 \times 50 \times 15$ mm was acquired in about 5 seconds. An ad hoc software written in MATLAB code was successively used to process the data provided by ULA OP. The software firstly reconstruct the 3D matrix of pixels representing the acquired volume. Then, it is able to provide several renderings that can be exploited to extract both 2D and 3D palmprints.

11.3 Results

Figure 11.2 shows some possible renderings that allow to appreciate two different features that characterize the 3D palmprint: (a) the curvature of the palm together with the principal traits and (b) the projection of the palmprint on a plane (2D palmprint) superimposed to the corresponding optical one. The last one could be immediately exploited in classical enrolment procedures by adapting one of the well established algorithms for 2D template extraction.

Figure 11.3 shows (a) and (b) two 2D ultrasound images acquired at different under skin depths, respectively, (c) the template extracted by using a conventional algorithms, and (d) the template superimposed on the ultrasound image.

It should be highlighted that the 2D ultrasonic images acquired at various depths can be combined to provide a more accurate template, which can improve the distinctiveness of this biometric characteristic. Also, they allow to achieve 3D information (under skin) of the main traits of the human palmprint; this last feature could be used to define a 3D template for palmprint.

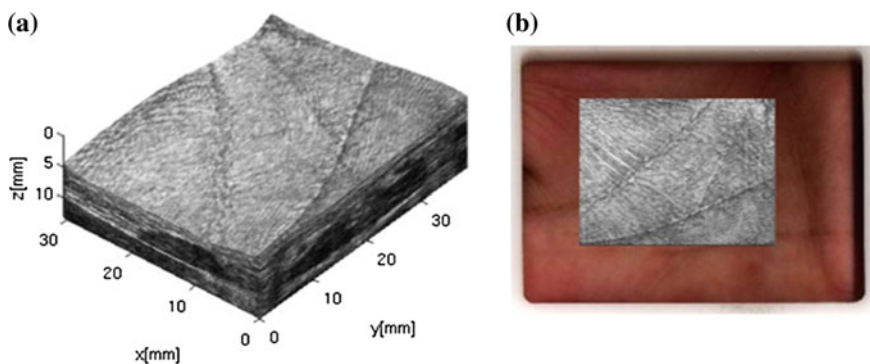


Fig. 11.2 Two renderings of the Palmprint: **a** a 3D view which allows to appreciate the curvature of the palm and the principal traits **b** the projection of the skin surface on a plane superimposed

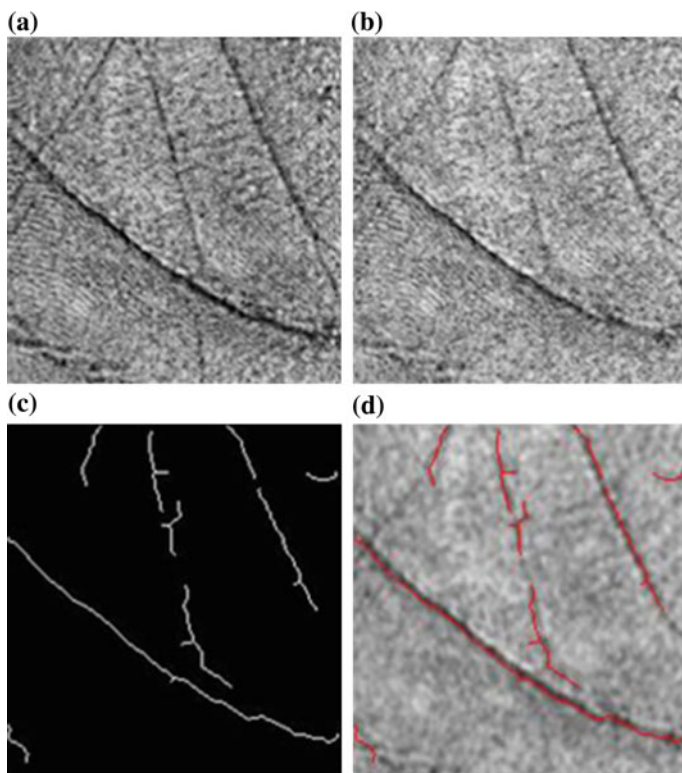


Fig. 11.3 An example of 2D Palmprint: **a** image acquired at 0.03 mm under the skin; **b** image acquired at 0.09 mm under the skin; **c** extracted template; **d** template superimposed on the ultrasound image

Figure 11.4 shows the main steps for template extraction. Starting from (a) a 3D rendering of a ultrasound palmprint image, the numerical profile (b) of the surface of the palm is achieved and, by exploited a known method [12], the mean curvature image, which highlights the main traits, is (c) obtained and (d) binarized.

As the two biometrics characteristics are obtained by the same acquired volume, they can be easily fused. The proposed technique represents, as a matter of fact, a multimodal biometric system.

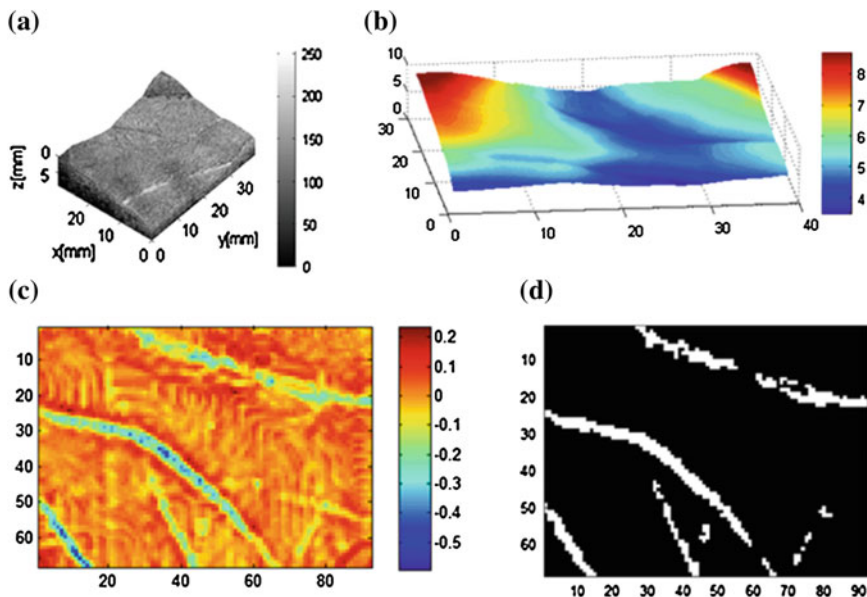


Fig. 11.4 An example of 2D Palmprint: **a** 3D ultrasound palmprint; **b** profile of the surface of the palm; **c** mean curvature image; **d** binarized mean curvature image

11.4 Conclusions

In this work, preliminary results of 2D and 3D palmprints, obtained by an automated ultrasound system are presented. 3D ultrasonic images of the human hand were obtained by moving a high frequency probe along its elevation direction to cover a desired ROI by a CNC pantograph. ULA OP was used as imaging system.

Possible procedures to extract templates from both 2D and 3D data have been shown.

Acknowledgments This work has been partially funded by the Italian Ministry of Education, University and Research (PRIN 10/11).

References

1. Jain, A.K., Ross, A., Prabhakar, S.: An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Technol.* **14**, 4–20 (2004)
2. Schneider, J.K., Gojevic, S.M.: Ultrasonic imaging systems for personal identification. *Proc. IEEE Ultrason. Symp.* **1**, 595–601 (2001)
3. Schmitt, R.M., Zeichman, J., Casanova, A.C., Delong, D.: Model based development of a commercial, acoustic fingerprint sensor. In: *Proceedings of the IEEE Ultrasonics Symposium*, pp. 1485–1488 (2012)

4. Maev, R.G., Severin, F.: High-speed biometrics ultrasonic system for 3D fingerprint imaging. In: Proceedings of SPIE—The International Society for Optical Engineering, vol. 8546, art. no. 85460B (2012)
5. Iula, A., De Santis, M.: Experimental evaluation of an ultrasound technique for the biometric recognition of human hand anatomic elements. *Ultrasonics* **51**(6), 683–688 (2011)
6. Iula, A., Savoia, A., Caliano, G.: 3D Ultrasound palm vein pattern for biometric recognition. In: IEEE International Ultrasonics Symposium, pp. 2442–2445 (2012)
7. Lamberti, N., Caliano, G., Iula, A., Savoia, A.S.: A high frequency cMUT probe for ultrasound imaging of fingerprints. *Sens. Actuators A Phys.* **172**, 561–569 (2011)
8. Iula, A., Savoia, A., Caliano, G.: Capacitive microfabricated ultrasonic transducers for biometric applications. *Microelectron. Eng. J.* **88**, 2278–2280 (2011)
9. Iula, A., Savoia, A.S., Caliano, G.: An ultrasound technique for 3D palmprint extraction. *Sens. Actuators A Phys.* **212**, 18–24 (2014)
10. Iula, A., Hine, G., Ramalli, A., Guidi, F., Boni, E., Savoia, A.S., Caliano, G.: An enhanced Ultrasound technique for 3D palmprint recognition. In: IEEE International Ultrasonics Symposium, pp. 978–981 (2013)
11. Tortoli, P., Bassi, L., Boni, E., Dallai, A., Ricci, S.: ULA-OP: an advanced open platform for ultrasound research. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **56**(10), 2207–2216 art. no. 5306767 (2009)
12. Kong, A., Zhang, D., Kamel, M.: A survey of palmprint recognition. *Pattern Recogn.* **42**(7), 1408–1418 (2009)
13. Zhang, D., Lu, G., Li, W., Zhang, L., Luo, N.: Palmprint Recognition Using 3-D Information. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **39**(5), 505–519 (2009)

Chapter 12

AA-Battery Sized Energy Harvesting Power Management Module for Indoor Light Wireless Sensor Applications

Alessandro Vinco, Rashid Siddique, Davide Brunelli and Wensi Wang

Abstract Wireless sensor nodes and pocket devices are mostly supplied by batteries. These storage units are efficient but often not rechargeable or recycled. This leads to environmental and sustainability problems, which are avoided by implementing new types of environmentally-powered systems. In this work we present a new high-efficiency power management module for indoor light applications, which is fitted into an AA battery size adapter. This prototype can be used instead of AA-batteries to supply many low power devices by storing the energy collected from a solar cell into a built-in supercapacitor.

Keywords Indoor light · Energy harvesting · Supercapacitor · Boost converter

12.1 Introduction

Energy harvesting consists in using micro-power management technologies to collect energy from ambient sources in order to supply electronic devices. Recent developments in the area of low power electronics, particularly wireless sensor networks (WSN) and low power Bluetooth (LPB) enabled devices, demonstrated the needs of designing power management circuits of the 1mW or sub-1mW power scales. Battery power supply is able to fulfil the power management requirements for certain types of applications. However, battery replacements also generate substantial environmental and maintenance cost issues for WSN [1]. Energy harvesting then becomes a valid solution to solve this issues by substituting batteries in many low-power electronic devices, or extending the battery life of these systems.

A. Vinco (✉) · D. Brunelli
University of Trento, Department of Engineering and Computer Science,
Via Sommarive, Trento, Italy
e-mail: alex.vinco@gmail.com

R. Siddique · W. Wang
Tyndall National Institute, Dyke Parade, Cork, Ireland

In contrast to battery power management, energy harvesting sources are sporadic and their energy intensities depend on ambient environment. Thus, micro-power management circuits for energy harvesting must be designed to operate with a very low input power. Due to the inherent small input power, a high conversion efficiency is required in order to harvest sufficient power for WSN. Most wireless sensor networks utilize small form factor sensor nodes. This also derives certain requirements to develop miniaturized energy harvesting and power management solutions. High energy density energy storage units with small form factors such as supercapacitors have been proposed to be used as the main storage element instead of batteries [2].

This paper presents the simulation, development and characterizations concerning the behaviour of an energy harvesting micro-power management circuit design for indoor-light applications. A high efficiency boost converter based energy harvesting power management circuit has been designed. It consists of a supercapacitor energy storage unit, a cold start circuit and a DC-DC converter. The complete micro-power management device was designed with a clear target of miniaturization and can be fitted into an AA battery-sized adapter together with the supercapacitor energy storage unit. This device can either work as a standalone harvesting system or as the output voltage regulation stage for multi-source energy harvesting modules.

12.1.1 Energy Harvesting in Embedded Systems

Power management is a fundamental aspect to consider when designing embedded electronic systems. Typical examples of these systems are wireless motes in Wireless Sensor Networks or aforementioned Bluetooth Low Energy (BLE) enabled devices. Typically an energy harvester cannot be directly connected to the electronic device [3]. Instead, a power management module is essential for the continuous operation of the electronic device (Fig. 12.1).

In this work, an amorphous silicon solar cell Sanyo AM-1815 has been used due to its adjusted spectrum response towards indoor light (fluorescent light). This cell generates between 30 and 40 μA at 3 V when exposed to a luminance level of 200 lux.

The power consumption of the Tyndall wireless sensor node is shown in Fig. 12.2. This power profile indicates two characteristics: (1) High “active” mode current during microcontroller data processing and wireless communications. The current

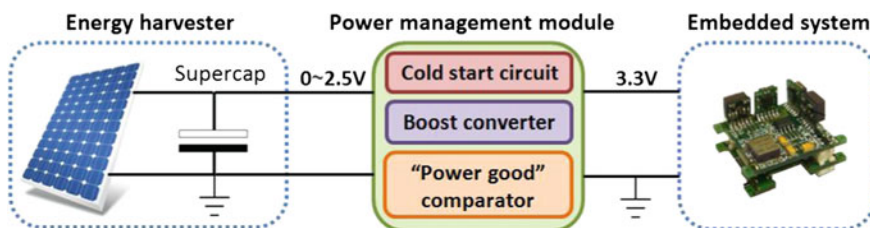


Fig. 12.1 Architecture of a conventional energy harvesting system

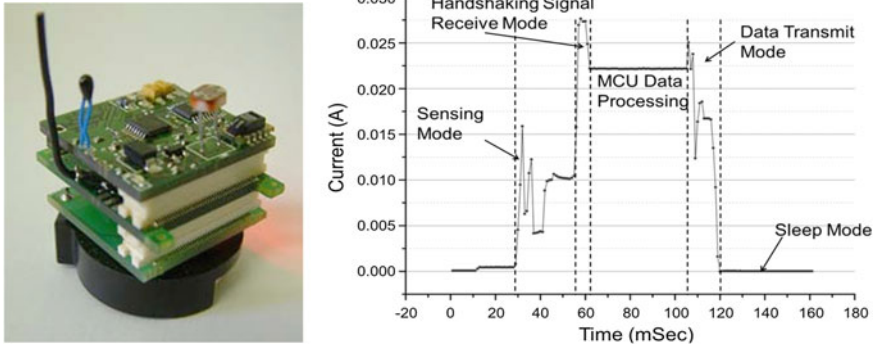


Fig. 12.2 Power consumption profile of Tyndall WSN mote during temperature monitoring tasks

consumption is 20–30 mA during the “active” mode; (2) Low power “sleep” mode current. The “sleep” mode current is reduced to less than 10 μA when power-down mode is selected. Due to the power consumption differences (3000 times, i.e. 30 mA active current versus 10 μA sleep current), it is obvious that in order to conserve energy, the mote has to operate in sleep mode for most of the time [4]. However, for power management circuit design, the maximum load current of the module must meet both the requirements, i.e. the maximum load current must be higher than the 30 mA limit and has also to efficiently operate with the 10 μA sleep current.

The power consumption profile of the WSN mote also indicates that the current consumption during the “active” mode (20–30 mA) is one order of magnitude higher than the solar cell generated power in indoor condition (30–40 μA). An energy storage element must be included to supply the “active” mode current consumption. In this design, a supercapacitor energy storage unit is used due to its long lifetime (>10 years) and high number of recharge cycles (>1 million cycles) [5]. These factors make it a suitable component for micro-power applications.

However, due to its capacitive characteristics, its voltage range is between 0 V and voltage rating (2.5 or 5 V). This wide output voltage range needs to be regulated, considering that the Tyndall WSN mote has a voltage range of 2.5 ~ 3.5 V. The supercapacitor voltage range can be both higher or lower than the WSN mote voltage range. This will require a wide voltage range boost converter or a buck-boost converter to perform the voltage regulation. In this work, the Texas Instruments TPS6122x series has been selected for its high conversion efficiency in 1 mW or sub-1 mW applications. The details of the circuit design will be given in the following sections.

12.2 Power Management Module Design and Simulations

In order to develop a practical power management module with high efficiency that could provide a constant output voltage regardless of the voltage available on the energy storage unit, several design goals have been identified: (1) self-start

capabilities; (2) output voltage has to be regulated as long as the voltage of the supercapacitor is within a certain range; (3) fully charged supercapacitor can supply the WSN mote for 24 h with only the stored energy [6–8].

The power management circuit is shown in Fig. 12.3: it consists of the main solar cell, which is directly connected to the supercapacitor. The supercapacitor is connected to the “power good” comparator, which is supplied by the reference solar cell. The load switch instead is controlled by the comparator and separates electrically the DC-DC converter from the input-end circuit. The selection of the DC-DC converter is due to its high performances in terms of minimal operational voltage (down to 0.7 V) and the very high efficiency (>90+ %).

To fully utilize the supercapacitor stored energy, a comparator has been used to create hysteresis to enable the DC-DC converter when the supercap voltage is higher than 1.5 V and disable it when the supercap voltage is <0.7 V. The comparator is the ultra-low power Maxim Integrated MAX9064 with internal voltage reference.

The comparator is supplied by a very small solar cell of 29.6 × 11.8 mm. It continuously compares the supercap voltage to the fixed reference and then keeps the DC-DC converter active as long as the supercap voltage is in the specified range.

The output of the voltage regulator is also connected to the supply pin (Vcc) of the comparator. Thus, the comparator can be powered when the reference cell generates zero power once the DC-DC converter is operating. In addition, the comparator Vcc pin has also been connected to the supercapacitor in order to supply the comparator when the supercapacitor voltage is below 1.5 V.

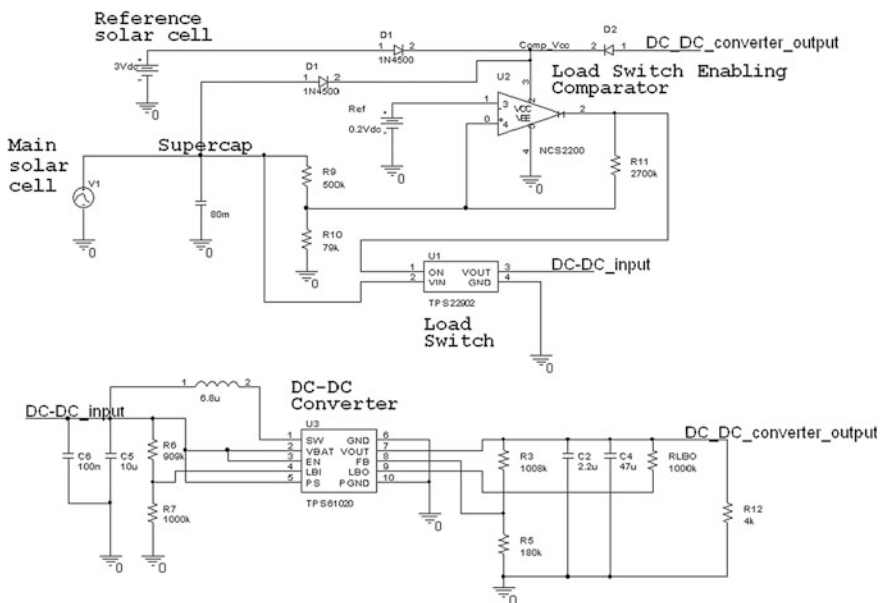


Fig. 12.3 SPICE schematic of the EH system

An issue exists in most low power boost converters. The boost converter has two operating voltages: (1) minimal operational voltage V_{OP} and (2) minimal start up voltage V_{SU} . It is often that $V_{OP} < V_{SU}$. Once V_{OP} is reached, the converter is enabled and the inductor current starts to increase. However, since minimal start up voltage V_{SU} has not been reached, the input voltage and current are insufficient to cold start the DC-DC converter. The boost converter then enters a phase that repeatedly attempts a cold start below its minimal start up voltage. This leads to a failing oscillation. To solve this problem, an integrated load switch has been placed between the supercapacitor and the output regulator to isolate the TPS61221 from the input-end circuit. Simulations have been created to test the behavior of the circuit before building and prototyping any board. In these simulations the main photovoltaic cell Sanyo AM-1815 is simulated as a voltage-controlled current source. The supercapacitor is here used as the energy storage unit of the system and provides a time-varying voltage to the whole circuit. The comparator is supplied by the reference cell. The cell is simulated using a 3 V constant voltage source which is consistent with the small solar cell behavior. The load switch is here used to separate the comparator from the DC-DC converter, avoiding any oscillations at the input of the converter during the start phase due to the inrush current flowing through the regulator.

Three Schottky diodes are here used to allow the comparator to be supplied from all the different sources of the circuit without interfering with each other.

In the simulation (Fig. 12.4), the comparator switches to the digital-high when the supercapacitor voltage rises above 1.5 V. At the same time (with 2.5 ms propagation delay) the DC-DC converter starts to regulate the output to 3.3 V. The output voltage of the comparator rises from 2.5 to 2.8 V when the DC-DC converter is active. Once the supercapacitor voltage falls below the programmed threshold of 1 V, it is possible to see the discharge of the supercapacitor and the shutdown phase of the DC-DC converter, together with the comparator switching to the low level and then the output decreasing to 0 V.

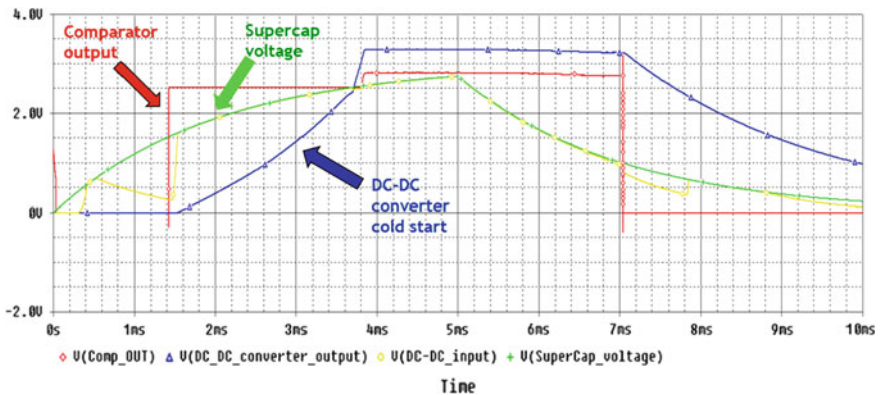
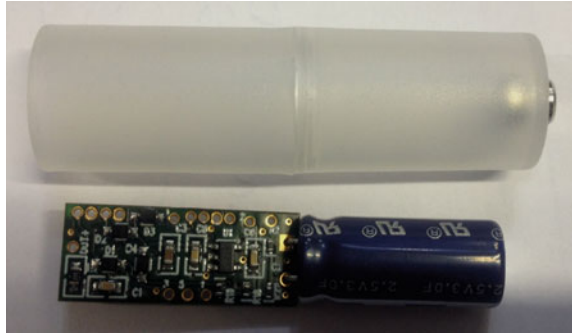


Fig. 12.4 Simulation results of the EH circuit in the time domain. The voltage levels of both the self-start and shutdown phases of the different blocks of the circuit are shown

Fig. 12.5 AAA-to-AA battery adapter here used as the chassis of our EH system, with the energy storage unit and the output regulator board



12.3 System Implementation and Test Results

In order to fit the power management module into a normal AA battery pack, a design goal has been set to fit the aforementioned system within an area of 44.5×10.5 mm. Miniaturization has been considered in all components selections. This also leads to a very precise choice for the energy storage unit: it has to be a reliable power source with a minimum required lifetime of 24 h in darkness. The supercapacitor used for this application is a PowerStor 3F 2.5 V capacitor with a $<10 \mu\text{A}$ average leakage current. With its small dimensions, low ESR (100 m Ω) and low leakage, it can be fitted into the AAA battery adapter, which makes this device perfectly suitable for our purposes (Fig. 12.5).

In the performed tests, the initial start-up time is approx. 2–3 h at a luminance level of 450 lux. The start-up occurs automatically at 1.55 V (supercap voltage), which is very similar to the behavior shown in the simulations of Fig. 12.3. After the cold start, the output voltage becomes stable at 3.3 V. The peak output current is instead 200 mA featuring a 90+ % conversion efficiency with a constant 20 mA load current.

12.4 Conclusions and Future Work

In this work a complete energy harvesting system with a new level of miniaturization and a very high conversion efficiency has been developed, which can be mounted inside an AA battery adapter. Its straightforward, fully analogue design based on ultra-low-power components makes it a very efficient and cost-effective solution to enable the autonomous operation of low power applications such as wireless sensor nodes and many other low-power devices where a battery-less system is preferred, without increasing the size of the overall system.

The amount of energy that is scavenged by the system can be further investigated by testing its performance in different daytimes and places to extend its range of possible applications. This could also create useful statistics that would be a good reference for the development of new advanced indoor solar energy harvesting systems.

References

1. David, J.: Nickel-cadmium battery recycling evolution in Europe. *J. Power Sources* **57**(1–2), 71–73 (1995)
2. Yang, H., Zhang, Y.: Self-discharge analysis and characterization of supercapacitors for environmentally powered wireless sensor network applications. *J. Power Sources* **196**(20), 8866–8873 (2011)
3. Roundy, S., Steingart, D., Frechette, L., Wright, P., Rabaey, J.: Power sources for wireless sensor networks. In: Karl, H., Wolisz, A., Willig, A. (eds.) *Wireless Sensor Networks*, pp. 1–17. Springer, Berlin (2004)
4. Wang, W., Wang, N., Jafer, E., Hayes, M., O’Flynn, B., O’Mathuna, C.: Autonomous wireless sensor network based building energy and environment monitoring system design. In: 2010 International Conference on Environmental Science and Information Application Technology (ESIAT), vol. 3, pp. 367–372 (2010)
5. Maxwell Corporate: Supercapacitors overview. http://www.maxwell.com/products/ultracapacitors/docs/Maxwell_CorporateBrochure.pdf
6. Brunelli, D., Dondi, D., Bertacchini, A., Larcher, L., Pavan, P., Benini, L.: Photovoltaic scavenging systems: modeling and optimization. *Microelectron. J.* **40**(9), 1337–1344 (2009)
7. Carli, D., Brunelli, D., Benini, L., Ruggeri, M.: An effective multi-source energy harvester for low power applications. *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp.1, 6, 14–18 March 2011
8. Rossi, M., Rizzon, L., Fait, M., Passerone, R., Brunelli, D.: Energy neutral wireless sensing for server farms monitoring. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **4**(3), 324–334 (2014)

Chapter 13

A Framework for Network-On-Chip Comparison Based on OpenSPARC T2 Processor

G. Causaprano, A. Audero, S. Tota and M. Ruo Roch

Abstract Network-on-Chip is gaining interest in these years thanks to its regular and scalable design. Several topologies have been proposed, and there is the need of a general framework for their test, validation and comparison. In this article a framework based on the OpenSPARC T2 processor is presented, where the NoC is used to replace the Cache Crossbar. With the introduction of protocol translators, it is possible to accommodate any NoC inside the T2. Processor regression tests can be used to validate the design and evaluate timing performance.

Keywords Network-on-Chip · Parallel architectures · Protocol translator · Simulation framework

13.1 Introduction

System-on-chip (SoC) architectures are getting communication-bound both from physical wiring and distributed computation point of view. In recent years always more attention has been dedicated to the development of new solutions for intra-chip communications [1]. In particular, Network-on-Chip (NoC) paradigm in multiprocessor chip is gaining interest since it is easily scalable, flexible, regular, and avoids long interconnections [2]. For this reason this approach has been proposed in a wide range of applications [3–5].

Several NoC structures have been designed, which differ in the topology and the routing algorithm. To compare these solutions, a framework that emulates a real application environment should be used. There is the need of a standardized approach

G. Causaprano (✉) · A. Audero · S. Tota · M. Ruo Roch
Politecnico di Torino, Dipartimento di Elettronica e Delle Comunicazioni,
Corso Duca Degli Abruzzi, 24, 10129 Torino, Italy
e-mail: giovanni.causaprano@polito.it

© Springer International Publishing Switzerland 2016
A. De Gloria (ed.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 351,
DOI 10.1007/978-3-319-20227-3_13

for testing and validation of NoCs, and for their performance evaluation [6]. For example in [7] an emulation framework is proposed, based on stochastic or real-life traffic generators, but the environment is limited to the NoC structure, hence it is not included in a more complicated and realistic SoC architecture.

Without presuming to provide the definitive answer to the need of a standardized approach, we present a framework based on the OpenSPARC T2 processor [8]. Several research projects have been carried on this processor [9] since its source code is accessible along with a development kit. In this study the original Cache Crossbar of the T2 is substituted with a NoC, thanks to the introduction of protocol translators. This architecture can accommodate any NoC structure, and performance can be evaluated running a given set of algorithms on the processor.

13.2 OpenSPARC T2 Crossbar

In this paragraph we provide a short description of the Cache Crossbar (CCX) of the OpenSPARC T2 processor. The aim is to detail fundamental characteristics of this structure in order to compare it with our proposed solution.

The CCX connects the eight SPARC cores to the eight banks of L2 cache and the IO subsystem. The basic element managed by the CCX is the transaction. A transaction can be thought as a complete data transfer cycle from a source to a destination. Each transaction is managed independently from the others, on source-destination pair basis. Data is exchanged in form of packets, that are 130-bit wide if packet is transmitted from cores to cache, 146-bit wide in reverse direction. As shown in Fig. 13.1a, CCX is made of two different blocks to manage separately Processor to Cache (CPX) and Cache to Processor (PCX) transactions. Since the same target can be requested by different sources at the same time, a mechanism to prevent colli-

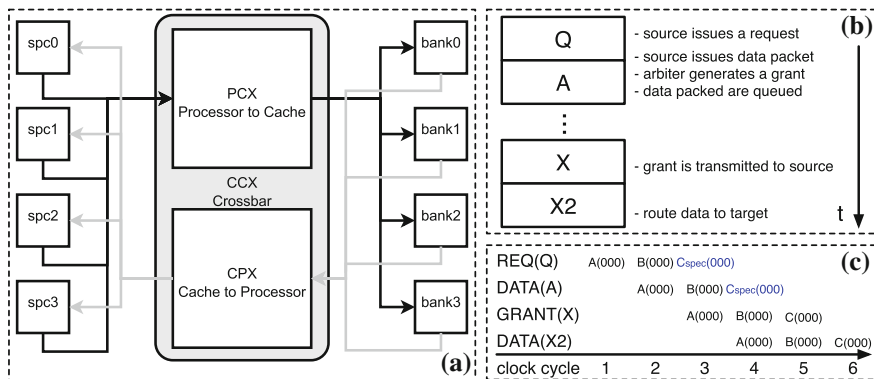


Fig. 13.1 a General architecture of the T2 Cache Crossbar (CCX), with a reduced number of processor (spc) and banks of cache for simplicity. b Packet transmission stages. c Example of accepted speculative transmission: grant for C(000) is generated at cycle 5

sion is provided: for each source there is an arbiter that can process requests queue without biasing. It generates a “grant” when the request has been issued.

Further, the CCX can virtually manage one transaction per clock cycle, using a pipelined structure. The limit of one transaction per clock cycle can be reached if there are no simultaneous requests for the same destination. In that case a source must be stalled until the destination becomes available.

The whole transmission is split in four stages, as shown in Fig. 13.1b. There must be no delay between stage Q and A, and between stage X and X2. The delay between stage A and X is instead given by the collision preventing mechanism operated by arbiters. Each source can send up to two requests, then it must wait for a grant before sending a new request. A speculative approach can be used, shown in Fig. 13.1c: the source can send a new request, if it is expecting a grant in the same clock cycle. Notice that if the grant is not received, it is up to the source to send again the request, otherwise it would be lost.

Transactions can occur in two different ways: normal or atomic. In normal transaction one packet is sent to the CCX and managed according to the rules previously described. When an atomic request is issued, instead, two successive clock cycles are used by the source to put two data packets on the bus. Arbiter must generate two grants, one for each packet, that must come without delay between them. Moreover, data packets must be issued to the destination sequentially, without any empty cycle between them.

13.3 Network-on-Chip Replacement

To introduce the NoC in the T2 architecture it is necessary to insert a structure that translates the original CCX protocol into the NoC one. This is done with a pair of wrappers shown in Fig. 13.2, called L_rail and R_rail. The structure composed of the two wrappers and the torus NoC is called NoX. The wrappers take requests and data from the sources (cores or banks of cache) and provide the interface to the NoC along with addressing and control signals for the switches. Further, the NoC routes the packets to their destination switches, where the other protocol translator provides the signals for the real destination, accordingly to the CCX protocol specifications. The NoC used for this implementation, is in a first moment a grid of 16×16 switches linked in a torus configuration (Fig. 13.2).

All the properties of the CCX must be preserved during the replacement with the NoX. In the following the most important are analyzed.

Access Arbitration. CCX provides arbitration for each destination. In the NoX, arbitration is provided by the NoC itself: boundary switches that interface with wrappers receive and transmit one packet at a time, therefore avoiding collisions. However, while the maximum delay in the CCX due to arbitration can be evaluated in advance, with the NoC it depends on the status of the network when the packet is sent.

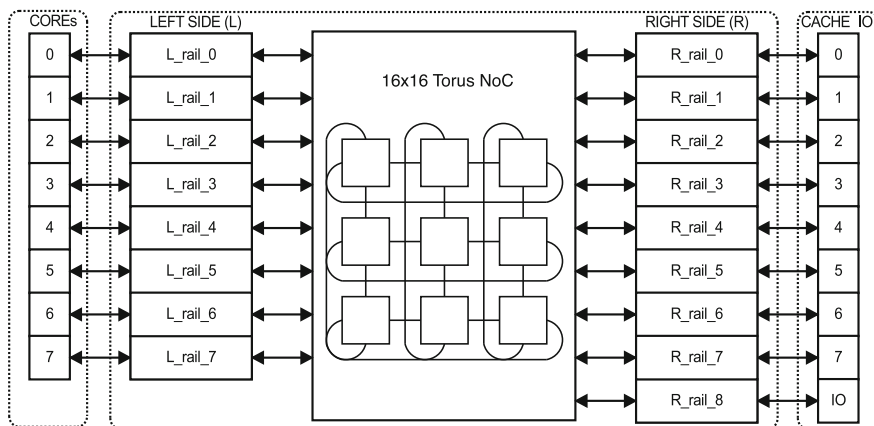


Fig. 13.2 NoC replacement for CCX: two wrappers (L_rail and R_rail) are added to interface the NoC with cores, banks of cache and IO

Packets delivery order. CCX guarantees in-order delivering of packets belonging to the same pair of source and destination. To achieve this feature in the NoX, a sequence number (SN) is added to the header of the data packet (Fig. 13.3c). The SN is added to the packet in the source wrapper and checked at destination wrapper. Both wrappers maintain synchronized counters for this purpose. When a packet for a given destination is sent, the current value of a counter is included as the SN into the header of the packet and the counter itself is increased by one. When the packet arrives at destination, that value is checked against the value of the destination counter. If the values match, then the destination counter is incremented by one and the packet delivered, otherwise the packet is buffered into a reordering buffer. The maximum number of packets (grant or data) that can be issued by a source to a destination and that can be routed by the NoC at the same time is 4. Therefore, the reordering buffer allocated for each destination must be able to store up to 4 packets.

Atomic transactions. Atomic transactions are managed in the NoX using an holding register at each receiving wrapper. Being sent in two successive clock cycles, the two packets will have successive SN as well. A bit in the packet (Fig. 13.3c) indicates when it is part of an atomic transaction. The first received packet is therefore hold in a register until also the second packet arrives, then they can be delivered to the destination in order.

Packets buffering. Besides packet delivery order and atomic transactions, there are other two conditions in which packet buffering may be required: (1) transmission switch is full—therefore the transmitting wrapper cannot put the packet on the bus, so a queue is implemented at the end of the transmitting process for buffering packets that cannot be delivered; (2) cache asserts the stall signal—therefore data packets, instead of being delivered to the destination, must be hijacked to a parking queue.

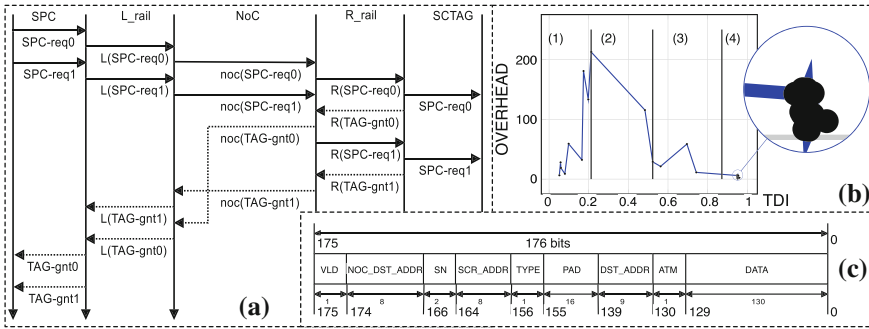


Fig. 13.3 **a** Timing diagram of a transaction with grant reordering. **b** Overhead due to the replacement of CCX with NoX, in function of Threads Density Index (TDI). **c** Data packet organization for core to cache transaction

In Fig. 13.3a the timing diagram relative to an example of transaction with grant reordering is shown. The SPARC Core (SPC) sends two requests to the right side (SCTAG), one per clock cycle. Since on the third clock cycle it does not receive any grant, the core cannot send any more requests. The requests are then sent through the NoC and received in order by the right side wrapper (R_rail). The right side wrapper delivers the packets to the destination and in the meanwhile generates the grants for the left side. These grants reach the left side out of order. The L_rail is in charge of reordering the grants using the reordering buffer and it delivers them to the SPC in the right order.

13.4 Results

In this section we summarize results of synthesis with a 45 nm library and performance evaluation derived from regression tests for CCX and NoX. Although the final objective is not to compare these two solutions, but to design a framework in which the designer can place any kind of NoC, with wrappers that take care of packet translation, this discussion is useful to understand the overhead that the points discussed in Sect. 13.3 lead to. In this context a 16×16 torus NoC represents a too big structure with respect to CCX. The NoC must interface with 8 cores, 8 banks of cache and 1 IO structure. Therefore a structure with at least 17 boundaries switches is necessary. For this reason the synthesis has been deployed using a reduced NoC of 4×8 switches. Notice that this impacts on the number of FIFO queues allocated, but also on the size of each register, since NOC_DEST_ADDR (Fig. 13.3c) can be reduced from 8 to 5.

Synthesis results reported an area of $0.378 \mu\text{m}^2$ for the CCX, $15.986 \mu\text{m}^2$ for the NoX. The reason for this great increase in area can be justified by the fact that the original CCX can rely on following aspects: packets exchanged between cores

and banks of cache are always delivered and received in order by design; there are no holes between atomic transaction packets; grants packet travel through reserved paths; data and grants are decoupled by design for stall management. To cope with the challenges posed by the items above, a great amount of circuitry has been added in the NoX.

We used regression tests as a benchmark to compare performance, in terms of number of clock cycles required to complete each test, of NoX with respect to the original CCX implementation. In no case the NoX performs better than the CCX. This is mainly due to a longer latency in delivering a packet, caused by the presence of the wrappers and the NoC itself. When the test requires a high number of threads running together, requiring a great amount of packet exchange, the overhead of the NoX is not so relevant. Conversely, when the number of concurrent threads is small, the overhead can reach values even greater than 200.

These results are summarized in Fig. 13.3b: on the x-axis is represented the *Threads Density Index* (TDI), evaluated for each test multiplying the number of cycles in which a number of threads have been run concurrently, by the number of threads in that time slice; on the y-axis is represented the overhead of the NoX against the CCX. The plot is divided in four different zones: zone (2) clearly shows the inverse relation between the TDI and the overhead; similarly, in zone (4), where the TDI is close to 1, the overhead is very small. Zone (1) and (3) group tests that run for a little amount of time, therefore the time required by the NoC has a greater degree of uncertainty. In conclusion, we can state that when a high number of threads run concurrently, the NoX produces delays comparable to those of the CCX.

13.5 Conclusions

We presented a strategy and a framework for the replacement of the Cache Crossbar of the OpenSPARC T2 processor with a Network-on-Chip. This has been achieved inserting two wrappers to connect cores, banks of cache and IO to the NoC (Fig. 13.2). This study can be considered a starting point for a comparative analysis of different NoC solutions, employed in a real scenario with meaningful results. As a matter of fact, with this framework it is possible to place any kind of NoC, of any dimension, in the T2 processor to evaluate its performance.

References

1. Tota, S., Casu, M., Ruo Roch, M., Macchiarulo, L., Zamboni, M.: A case study for NoC-based homogeneous MPSoC architectures. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **17**, 384–388 (2009)
2. Benini, L., De Micheli, G.: Networks on chips: a new soc paradigm. *Computer* **35**, 70–78 (2002)

3. Quaglio, F., Vacca, F., Castellano, C., Tarable, A., Masera, G.: Interconnection framework for high-throughput, flexible ldpc decoders. In: Proceedings Design, Automation and Test in Europe (DATE), vol 2. (2006)
4. Condo, C., Martina, M., Masera, G.: Vlsi implementation of a multi-mode turbo/ldpc decoder architecture. *IEEE Trans. Circuits Syst.* **60**(1), 1441–1454 (2013)
5. Tota, S., Casu, M., Ruo Roch, M., Rostagno, L., Zamboni, M.: Medea: A hybrid shared-memory/message-passing multiprocessor noc-based architecture. In: Design, Automation Test in Europe Conference Exhibition (DATE) pp. 45–50 (2010)
6. Pande, P., Grecu, C., Jones, M., Ivanov, A., Saleh, R.: Performance evaluation and design trade-offs for network-on-chip interconnect architectures. *IEEE Trans. Comput.* **54**, 1025–1040 (2005)
7. Genko, N., Atienza, D., De Micheli, G., Mendias, J., Hermida, R., Catthoor, F.: A complete network-on-chip emulation framework. In: Proceedings Design, Automation and Test in Europe, vol. 1, pp. 246–251 (2005)
8. Parulkar, I., Wood, A., Microsystems, S., Hoe, J.C., Falsafi, B., Adve, S.V., Torrellas, J.: Opensparc: an open platform for hardware reliability experimentation (2008)
9. Pulimeno, A., Graziano, M., Piccinini, G.: Udsm trends comparison: from technology roadmap to ultrasparc niagara2. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **20**, 1341–1346 (2012)

Chapter 14

A GPU 3D Segmentation Framework for Medical Imaging

Francesca Galluzzo, Luca De Marchi, Nicola Testoni and Guido Masetti

Abstract In this work we propose a fast and flexible GPU 3D level-set segmentation framework able to handle different segmentation tasks. Experiments on simulated and real images demonstrate the method ability at achieving high computational efficiency with no reduction in segmentation accuracy compared to its sequential counterpart. The method clinical applicability is demonstrated by addressing the task of Left-Ventricle myocardium segmentation in Real-Time 3D Echocardiography.

Keywords Segmentation · GPU computing · Level-set · Sparse field

14.1 Introduction

Segmentation is an important task in computer vision and medical imaging. In medical imaging it is usually performed manually and it allows to detect the boundaries of anatomical structures to measure organ dimensions and compute clinically relevant metrics. However, manual segmentation is a complicated and time consuming task and its accuracy is extremely operator-dependent. Fast and automated segmentation procedures are therefore highly desirable to speed up the segmentation process and reduce its subjectivity, thus enabling and improving different clinical evaluations. This is particularly true for the ultrasound (US) imaging, where the data quality influences the segmentation accuracy and fast image processing is essential to preserve the real-time nature of the examination.

To handle the segmentation of 3D images, many approaches have been proposed among which Level-Set (LS) based methods have received increasing attention [1].

F. Galluzzo (✉) · L. De Marchi · N. Testoni · G. Masetti
DEI—Department of Electrical, Electronic and Information Engineering
“Guglielmo Marconi”, University of Bologna, Bologna, Italy
e-mail: francesca.galluzzo@unibo.it

© Springer International Publishing Switzerland 2016
A. De Gloria (ed.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 351,
DOI 10.1007/978-3-319-20227-3_14

Unfortunately the flexibility of the LS method results in long computational times and therefore limits its attractiveness in medical applications.

In this work, we propose a GPU LS 3D segmentation framework based on a rigorous GPU implementation of the LS Sparse Field (SF) algorithm [2] that presents an interesting compromise between very high computational efficiency and application flexibility obtained with optimal segmentation accuracy. Thanks to its flexibility and to an efficient implementation of different segmentation approaches, our framework is able to handle different segmentation tasks and can be specialized to deal with challenging medical applications. To demonstrate the method applicability in clinical environment, we focus on the segmentation of the left ventricle (LV) of the heart in real-time 3D echocardiography (RT3DE) images.

14.2 Background

Level-set techniques correspond to a class of deformable models where the shape of interest is captured by propagating an interface implicitly represented as the zero LS of a smooth function. The interface evolution is generally derived through a variational formulation: the segmentation problem is expressed as the minimization of an energy functional that reflects the properties of the target object. Formally, this minimization allows deriving a speed function (spf) which makes the LS evolve along a path normal to the interface. Let ϕ be the function embedding the evolving interface (LS function), the segmentation problem is handled by the evolution of one LS driven by the LS equation: $\frac{\partial \phi}{\partial t} = F \cdot \|\nabla \phi\|$, where $F(x, t)$ is the spf. A wide range of segmentation solutions can be implemented by defining appropriate speed functions (spfs). Hereto, to illustrate the flexibility of our framework we propose the use of four spfs corresponding to different segmentation approaches. The first one, proposed by Lefohn [3], implements an intensity thresholding method. The second one corresponds to the Chan-Vese region-based model [4] that uses global statistics of the foreground and background regions to drive the LS evolution. Then, we use two spfs based on the Lankton localized region-based model [5] that drives the LS evolution by using local statistics computed on a neighborhood of each interface point. In particular, we use the localized version of respectively the Chan-Vese energy and the modified Yezzi energy [6]. Note that, for these spfs our neighborhood consists in a set of points belonging to the normal direction to the interface [6].

To overcome the inherent LS long computational times the narrow band (NB) methods make the LS evolve only into a few voxel wide region around the zero LS (see Fig. 14.1). The SF method takes this strategy to the extreme by restricting the computation on a one voxel wide domain. Despite SF gives the most efficient results as CPU (sequential) LS solver, it remains too time consuming for effective use in 3D applications. The implementation on highly parallel architectures allows to achieve this goal. Due to the high application parallelism and the large computational and real-time performance requirements, GPUs are the best architectural choice in terms

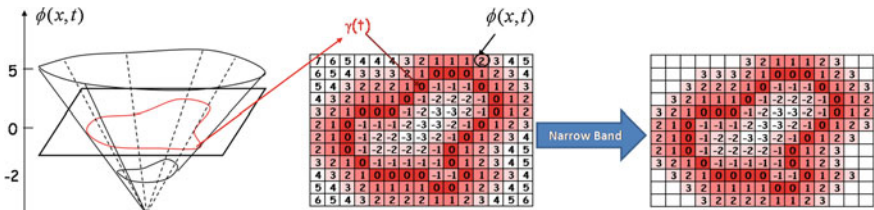


Fig. 14.1 Example of LS function ϕ embedding a 2D contour, its discretization on a grid of points and the narrow band domain restriction

of memory bandwidth, cost, flexibility and developing time. This motivates our SF GPU implementation. Among the existing NB GPU LS solvers [3, 7, 8] only one of them borrows from both the NB and the SF [8], which however is a temporally coherent SF version.

14.3 Parallel Implementation

In the SF algorithm, the narrow band points (active points) are initialized with a signed distance function computed from an initial interface (the zero LS). Then, both the LS values and the computational domain are updated at each algorithm iteration according to (1). To do this, a neighborhood (5 voxel wide) around the active points must be maintained. Both the active set (L_0) and the neighborhood are maintained in layers L_i for $i \in \{\pm 1, \pm 2\}$ (see Fig. 14.2). The LS evolution is then realized by updating ϕ thanks to (1) and consequently moving the corresponding points in the appropriate L_i . Five additional status lists S_i are also used to track points that move from one layer to another one [2].

14.3.1 Parallel Implementation of the Sparse Field Algorithm: P-SF

Efficient sequential implementations of the SF algorithm use doubly-linked-lists to implement both L_i and S_i . Since these structures are poorly suited for parallel programming, we propose to treat each list as a 1D buffer, whose size is equal to the support of ϕ , and using the voxel coordinates as array indices. Each list contains only values '1' or '0' to indicate the presence (or not) of a voxel in the list. Two further arrays are used to maintain the LS values (ϕ) and to record the status of each point providing a way to track which list a point belong to (*label*). The implementation parallelism has the granularity of the individual voxel. It means that each thread manages an element of each buffer (see Fig. 14.2).

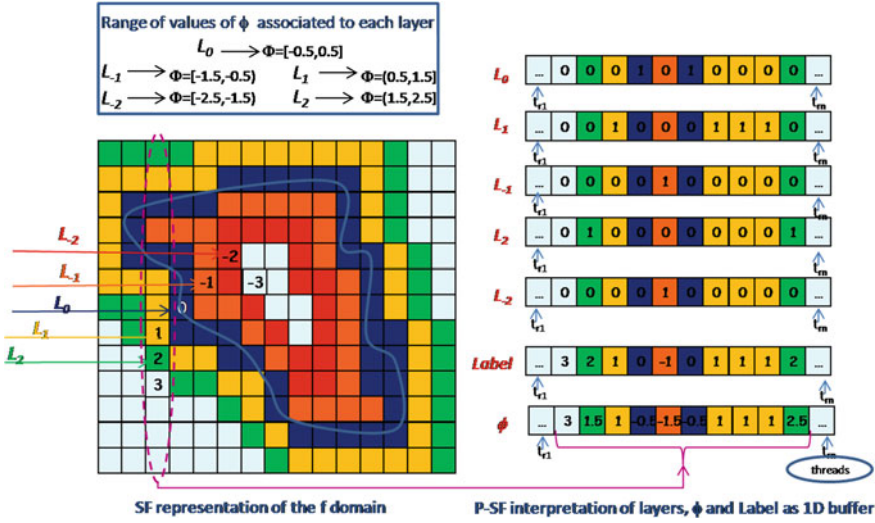


Fig. 14.2 SF representation of the ϕ domain and the P-SF interpretation of layers, ϕ and label, as 1D array. An example of the domain mapping on them is highlighted

The P-SF workflow can be summarized as follows:

1. Initialization procedure

- From an input binary mask in parallel initialize ϕ , label and L_0 and populate $L_{\pm 1}, L_{\pm 2}$ based on label values of the neighbors in the next layer closest to L_0 (sequentially analyzed).

2. Updating procedure

- In parallel for each point in L_0 compute $F(\cdot)$ and update ϕ according to equation (1) assuming that $\|\nabla\phi\| = 1$; Update L_0 and $S_{\pm 1}$ in parallel;
- Traverse each list L_i in parallel to identify points whose status changed and update ϕ so that the updated values are one unit from their nearest neighbors in the next layer closest to L_0 ; Update L_i and S_i in parallel;
- Deal with S_i exploiting the parallelism as done for L_i .

Note that P-SF has a modular structure: each step of the workflow is split into different CUDA kernels to manage data dependencies between different L_i (and S_i) thus ensuring synchronization. Each kernel is launched with 2D thread blocks with fixed size for efficiency purposes. These blocks are organized in 2D grids whose size is computed to traverse the entire volume domain (each 3D coordinate is converted into a 1D memory address). Atomic operations are used to safely update data structures read and written by neighbors threads. Moreover, specific CUDA kernels are devoted to the ϕ computation, thus allowing to exploit many segmentation approaches without affecting the rest of the implementation.

14.3.2 Parallel Implementation of Speed Functions

While the parallel implementation of the Lefohn spf is straightforward, region-based approaches require efficient computations of the intrinsic statistical parameters. In particular, to efficiently implement the Chan-Vese spf we use four auxiliary buffers to track the points that cross the zero LS during the evolution maintaining both their status and intensity value. These buffers are updated basing on ϕ values and parallel reduction techniques are used for statistics computation. Combining the localized region-based approach with the SF method implies that, at each algorithm iteration, a local neighborhood must be identified on the normal direction to the interface for each point in L_0 and local statistics must be computed on it. Since both the normal direction and the neighborhood change during the evolution they must be computed at each iteration, without requiring additional buffers. For an efficient parallel implementation we associate one thread to one local neighborhood. Thus each thread manages a voxel of L_0 and the local statistics are computed in parallel at each iteration, while computations in each neighborhood are sequentially performed by the associated thread. Note that P-SF also allows an efficient parallel computation of the time step involved in the discretization of (1) according to the CFL condition [1].

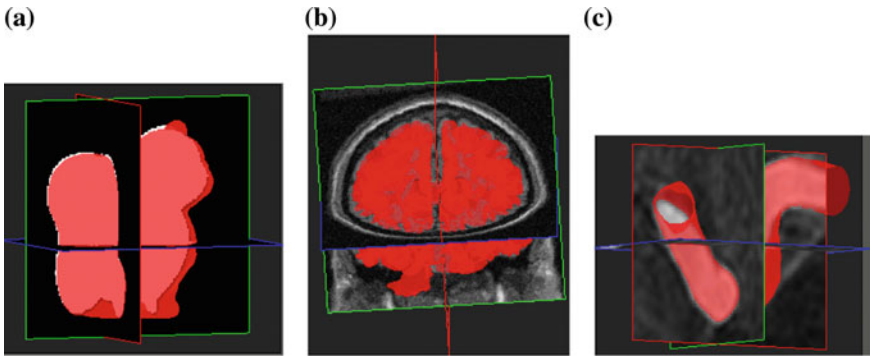
14.4 Experimental Results

To assess the P-SF segmentation accuracy and analyze its performance, we compared it with its sequential counterpart (S-SF). All the experiments were performed on a quad-core Intel 2.4 GHz Xeon processor with 8 GB of memory and an NVIDIA GTX580 graphic card with 512 streaming processors (CUDA cores) and 1.5 GB of DRAM. P-SF was implemented by using CUDA in Matlab environment. S-SF, developed basing on [9], runs on a single CPU core. We performed five segmentation tasks on volumetric images: a synthetic binary image representing a squirrel, an inhomogeneous and a noisy version of it, a simulated MRI brain image and a real rotational angiography vessel image with aneurysm. The comparison metric was the standard Dice coefficient computed *w.r.t.* a reference. The reference was known for all images except for the vessel where the S-SF segmentation result was used. P-SF and S-SF were initialized using the same binary mask: a sphere manually drawn by a user. Table 14.1 reports the used segmentation approaches along with image sizes and segmentation results. From these results, we observe that for each segmentation task P-SF dramatically reduces the time consumption *w.r.t.* S-SF while maintaining the same segmentation accuracy. As expected, the speed up (spu) increases with the volume dimension. It also increases with the spf complexity, as demonstrated by the impressive spu obtained when a localized spf is used (many computations for each L_0 point at each iteration). Figure 14.3 shows some P-SF segmentation results.

To verify the method clinical applicability we addressed the task of 3D LV segmentation in RT3DE images. To this end, by using the localized Yezzi energy [6],

Table 14.1 Segmentation results on simulated and real images

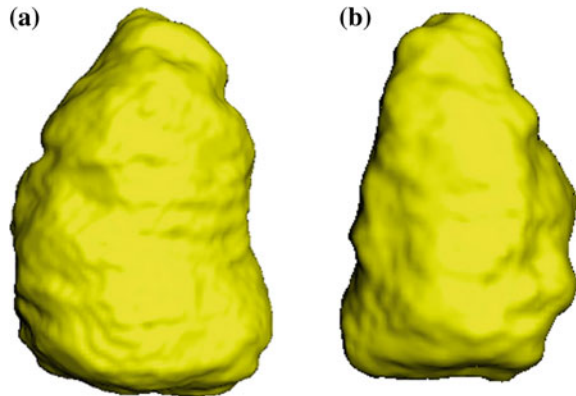
	Size	DICE		Execution Time (s)		Speed Up (times)	# iter.	Seg. Approach
		S-SF	P-SF	S-SF	P-SF			
Vessel	$66 \times 97 \times 80$	1	1	33.03	1.77	19	1800	Chan-Vese
Binary squirrel	$148 \times 108 \times 147$	1	1	190.74	0.55	347	450	Lefohn
Noisy squirrel	$148 \times 108 \times 147$	0.99	0.99	459.39	2.72	169	1000	Chan-Vese
Inhomo. squirrel	$148 \times 108 \times 147$	0.96	0.99	17739.00	15.32	1158	3000	Localized Chan-Vese
MRI brain	$181 \times 217 \times 181$	0.92	0.92	5093.27	7.76	657	2000	Lefohn

**Fig. 14.3** P-SF segmentation results along with the corresponding execution time. **a** 0.55 s. **b** 7.76 s. **c** 1.77 s

we let our framework segment two volumetric frames, corresponding to the end-diastolic (ED) and end-systolic (ES) instants, of three real RT3DE sequences (three patients). Then we compared our segmentation results with the average results of manual segmentation performed by three expert physicians. The measured quantities are the end-diastolic, end-systolic volumes (EDV, ESV) and the ejection-fraction (EF). In this task P-SF was initialized with an automated initialization procedure [6]. From the results in Table 14.2 we observe a good agreement between manual and automated segmentation methods. P-SF took about 0.82 s on average to segment each volume (250 iterations), outperforming S-SF with significant spu. An example of 3D LV segmentation performed by P-SF is shown in Fig. 14.4.

Table 14.2 3D LV segmentation results

	Volumetric Index		Execution Time (s)		Speed Up (times)
	Manual	P-SF	P-SF	S-SF	
EDV (ml)	82.27	83.01	0.83	10.16	12.24
ESV (ml)	34.70	36.22	0.81	3.64	4.49
EF (%)	47.57	46.56	–	–	–

Fig. 14.4 P-SF RT3DE exam segmentation. **a** EDV 0.99 s. **b** ESV 0.94 s

14.5 Conclusions

In this work a fast and flexible GPU LS 3D segmentation framework is proposed. It is based on a rigorous GPU implementation of the LS SF algorithm and on an efficient parallel implementation of different segmentation approaches. A comparison with its sequential counterpart demonstrates the solver effectiveness at achieving high computational performance without losing segmentation accuracy and its ability in handling different segmentation tasks. A comparison with manual contouring from expert physicians demonstrates the method effectiveness at achieving accurate LV segmentation in RT3DE near real-time, confirming its usefulness as tool for quantitative cardiac morphology and function analysis.

Acknowledgments This research was partially funded by the Italian Ministry of Education, University and Research (PRIN 2010–2011).

The authors would like to thank Jan D’hooge and Daniel Barbosa (KU Leuven) for providing RT3DE data, and Olivier Bernard (Creatis-INSA, Lyon) for his helpful comments and suggestions.

References

1. Osher, S., et al.: *Level Set Methods and Dynamic Implicit Surfaces*. Springer (2003)
2. Whitaker, R.: A level-set approach to 3d reconstruction from range data. *Int. J. Comput. Vision* **29**(3), 203–231 (1998)
3. Lefohn, A., et al.: A streaming narrow-band algorithm: interactive computation and visualization of level sets. *IEEE Trans. Visual Comput. Graphics* **10**(4), 422–433 (2004)
4. Chan, T., et al.: Active contours without edges. *IEEE Trans. Image Process.* **10**, 266–277 (2001)
5. Lankton, S., et al.: Localizing region-based active contours. *IEEE Trans. Image Process.* **17**(11), 2029–2039 (2008)
6. Barbosa, D., et al.: Fast and fully automatic 3D echocardiographic segmentation using B-spline explicit active surfaces: feasibility study and validation in a clinical setting. *UMB* **39**(1), 89–101 (2013)
7. Jeong, W., et al.: Scalable and interactive segmentation and visualization of neural processes in em datasets. *IEEE Trans. Visual Comput. Graphics* **15**(6), 1505–1514 (2009)
8. Roberts, M.: A Work-Efficient GPU Algorithm for Level Set Segmentation, HPG, pp. 122–132 (2010)
9. Lankton, S.: Sparse field methods-technical report. <http://www.shawnlankton.com>

Chapter 15

Augmented Reality Tools for Structural Health Monitoring Applications

L. De Marchi, A. Ceruti, N. Testoni, A. Marzani and A. Liverani

Abstract A novel Augmented Reality (AR) tool for structural health monitoring is illustrated in this work. It provides maintenance operators with the results of an impact detection methodology. It interacts with an eyepiece allowing the inspector to see the estimated impact position on the structure. Electric signals are collected by a network of piezosensors bonded on the structure to be monitored. Dispersive propagation compensation is performed to improve estimation robustness. Hyperbolic beamforming is exploited to locate the impact. Real-time impact data are finally fed to the AR eyepiece. The proposed approach is tested on a Cessna 150 engine cowling. Experimental results confirm the feasibility of the method and its exploitability in maintenance practice.

Keywords Augmented Reality · Guided waves · Structural health monitoring

15.1 Introduction

Air transportation infrastructure reliability and level of quality are founded on continuous innovation and aircraft maintenance, which is generally performed on a fixed time basis. The main limitation of this paradigm is related to the costs due to unnecessary inspections and long aircraft unavailability for commercial operation. To tackle these problems, Structural Health Monitoring (SHM) systems can be adopted. If properly designed and operated, these devices will autonomously provide an alert when the aircraft has been damaged, thus allowing for a shift to an on-demand based maintenance. This is crucial for aircraft structures based on composites such as carbon and glass fibers. Using SHM, invisible or barely visible damages like delaminations, potentially dangerous for the structural resistance, can be located in much shorter times and with much less effort. This need is driving recent research efforts in

L. De Marchi (✉) · A. Ceruti · N. Testoni · A. Marzani · A. Liverani
DEI-DICAM-DIN University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy
e-mail: l.demarchi@unibo.it

© Springer International Publishing Switzerland 2016
A. De Gloria (ed.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 351,
DOI 10.1007/978-3-319-20227-3_15

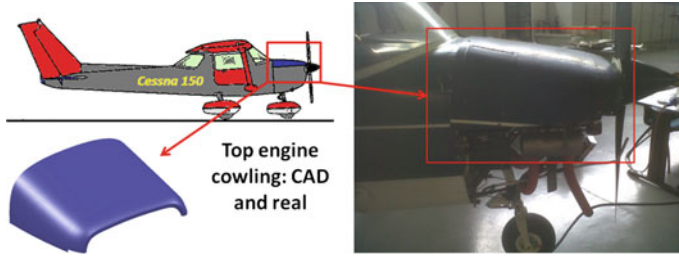


Fig. 15.1 Side view of a Cessna 150 with a real and CAD model of the top engine cowling

the development of SHM systems for aircraft components, such as wing and fuselage skins [1, 2].

Ultrasonic guided waves are a promising base on which such systems can be built on. When an impact occurs, a network of piezoelectric sensors will capture the wave. Gathered information is then locally processed to reduce redundancy and sent to a central unit where impact position is estimated. An alert is then generated so that operators can decide on the most appropriate maintenance strategy. In case of intervention, a narrow area will be inspected via a visual inspection and eventually by means of a detailed ultrasonic C-scan. Augmented Reality (AR) is crucial during this phase to support inspection crews with the estimated impact location superimposed to the structure under testing. The reduction in the area to be inspected is best appreciated on large structures, such as wings and fuselage (e.g. an Airbus A380 features an 80 meters wingspan and a 72 m fuselage length). Finally, SHM system can locate defects on parts with a complex geometry, in positions difficult to inspect or not reachable from external inspections.

This work proposes the validation of an SHM methodology on the top engine cowling of a Cessna 150. This part, shown in Fig. 15.1, is considered representative of real world larger aircraft structures as it features many of the complexities such as multiple curvatures, stiffeners, rivets and hinges. Other than aircraft components, a large range of industrial and civil structures based on thin metal or composite could also take benefit from this technique such as train coaches, storage silos, ship structures, competition cars.

15.2 Impact Localization

Dispersive propagation of guided waves on thin metal or composite plates is well documented [3]. To compensate recorded signals from dispersion due to the traveled distance, the Warped Frequency Transform (WFT) operator can be used. The WFT is a unitary transformation exploiting a flexible sampling of the time-frequency domain, chosen to match the spectro-temporal structure of a particular guided mode. The WFT operator \mathbf{W}_w transforms the original time-waveform $p(t)$ into a warped

version $p_w(t)$ by reshaping the periodic frequency axis using a proper function $w(f)$, called *warping map*, such as:

$$p_w(t) \doteq \mathbf{W}_w\{p(t)\}$$

$$\mathbf{F}\mathbf{W}_w\{p(t)\} = \sqrt{\dot{w}(f)} P(w(f))$$

where $\dot{w}(f)$ represents the first derivative of $w(f)$, and \mathbf{F} is the Fourier Transform operator. In our approach, $w(f)$ is defined through its functional inverse:

$$K \frac{dw^{-1}(f)}{df} = \frac{1}{c_g^m(f)} \quad (15.1)$$

where c_g^m is the group velocity of the m th guide mode and K is a normalization parameter selected so that $w^{-1}(0.5) = w(0.5) = 0.5$.

If a signal $p_{\chi,\tau}(t)$ is detected at distance $x = \chi$ from the actuator which generated it at time $t = \tau$, in force of Eq. (15.1), its frequency transform can be written as:

$$P_{D_x}(f) = P_0(f) \cdot e^{-j2\pi f\tau} \cdot e^{-j2\pi w^{-1}(f)K\chi}$$

Applying the WFT to $p_{D_x}(t)$ and exploiting the fundamental property of inverse functions ($w^{-1}[w(f)] = f$) yields a signal $p_w(t)$ whose frequency transform is:

$$P_w(f) = \mathbf{F}\mathbf{W}_w\{p_{D_x}(t)\} = \sqrt{\dot{w}(f)} P_0(w(f)) \cdot e^{-j2\pi w(f)\tau} \cdot e^{-j2\pi fK\chi}$$

where the phase term presents a linear dependence from the traveled distance.

Using signal correlation techniques this property can be fruitfully exploited. By computing the modulus of the Hilbert Transform of the cross-correlation of warped signals, a direct relationship between the difference in propagation distance and the shape of the function can be drawn. In particular, a peak which capture very well the difference in traveled distance between the two correlated signals is present (see Fig. 15.2).

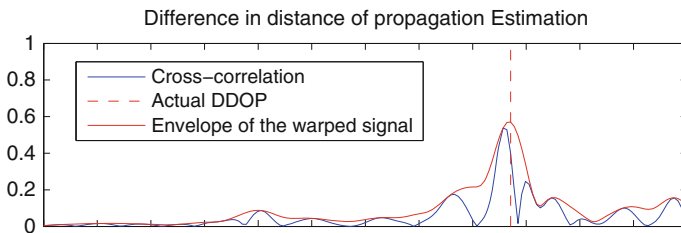


Fig. 15.2 Cross-correlation of dispersion compensated signals: the envelope peaks at a position which is directly related to the difference in distance of propagation

Knowing the recording locations and the difference in traveled distance of propagation (DDOP) estimated from the cross-correlation envelope, hyperbolic positioning methods (also called *multilateration*) can be applied to locate the wave source. At least three signals acquired from different transducers are required to locate a wave source (i.e. an impact) in a bi-dimensional space such as a plate. To increase localization robustness, four transducers are used in this work, at given coordinates (x_i, y_i) with $i = 1, \dots, 4$.

In conclusion, when an impact occurs, signals are recorded simultaneously from all the transducers. Next, WFT is applied to each signal and DDOP $\Delta_{i,j}$ are computed by locating the envelope peaks in the cross-correlations of signal couples (i, j) . Finally a Levenberg-Maquardt algorithm is used to solve the following system of equations to obtain the impact location estimation (x_p, y_p) .

$$\Delta_{i,j} = \sqrt{(x_i - x_p)^2 + (y_i - y_p)^2} - \sqrt{(x_j - x_p)^2 + (y_j - y_p)^2}$$

$$i = 1, \dots, 4 \quad j = i + 1, \dots, 4$$

15.3 Augmented Reality

Real-time techniques which superimposes images, sounds, graphics, or video data generated by the computer to real world images goes under the collective name of Augmented Reality (AR) [4]. Many application can take advantage for the use of AR techniques as reported for instance in the survey Wang et al. [3]. Recent studies [5, 6] showed how AR is a keen topic in very diverse research fields.

AR applications' work flow is usually divided into five steps:

1. *Image acquisition*, normally obtained by means of a camera embedded in a carried-on device placed on a subject's head. Mobile phones, tablets and digital cameras are also up to the task for some applications.
2. *Calibration*, performed to evaluate camera parameters and correct image distortion by means of a mathematical model of the camera. Each time the camera model or settings are changed, this step shall be repeated.
3. *Tracking*, accomplished exploiting sensor-based or vision-based techniques for the detection of the pose and position of the camera with respect to an external reference system.
4. *Registration*, necessary to synchronize the artificial image or scenario with the real world external view, dynamically following in real-time the movements of the experimenter head.
5. *Display*, where the external world view together with the symbols added by AR is presented to the experimenter. Usually Head-Mounted Display (HMD) systems or see-through devices are employed.

In this work, a vision based approach to tracking was adopted. Such method requires information about the shape and size of a marker, which is stitched to exter-

nal surface of the engine cowling. Each experimenter movement causes the marker representation to distort in the camera image. Rotation and translation are recovered from the distorted marker projection and its position in the real world with respect to the camera is computed in real-time.

Once the position of the marker has been referenced to the camera, it is possible to draw on the screen any object whose position and orientation with respect to the marker are known. In this work, a sphere is used to show the damage location.

The choice to use a see through device is motivated by the necessity to maintain a visual contact between the real the virtual world. In fact, as the synthetic image of the target location is projected onto transparent lenses, the user never lose contact with the structure under inspection. This is beneficial to non trained technicians who won't need to significantly change their day-to-day inspection routine since AR information will be shown only when the impact zone is focused on.

15.4 Experimental Setup and Results

To validate the use of AR in aircraft maintenance operations, a top engine cowling of a Cessna 150 has been considered. The average thickness of the structure is 0.8 mm. The nominal properties considered for the aluminum are: Young's modulus $E = 69$ GPa, Poisson's coefficient $\nu = 0.33$ and material density $\rho = 2700$ kg/m³. Four piezoelectric sensors (PZT discs PIC181, diameter 10 mm, thickness 1 mm) were bonded to the structure.

Tracking is performed by means of AR toolkit [7], an available library for Augmented Reality which is used to detect the marker, to measure image distortion as well as to relate in time the marker reference system to the image.

The experimenter wears a pair of Vuzix STAR 1200 glasses a see-through device specifically designed for AR applications [8] which features a camera, two miniaturized projectors for image display on the lenses and three gyroscopes that can provide pose angles (pitch, roll, yaw). The AR toolkit macros automatically detect the marker in the camera image and perform the calibration step whenever an image is taken.

Pencil breaks are used to simulate impacts and generate guided waves. Signals gathered by the sensors are recorded at a sampling frequency of $f_s = 300$ kHz.

To process the acquired data, the warping map $w(f)$ has to be defined according to Eq. (15.1). It has been verified that the low curvature of the panel does not affect the dispersion of the propagating guided waves. In force of this the warping map $w(f)$ was designed through the dispersion curves of a flat plate 0.8 mm thick. The fundamental A_0 guided wave mode group velocity was considered for c_g^m . Its profile was computed according to the procedure described in [9], taking into account the material and geometrical properties of the considered specimen.

The proposed approach was tested to locate 20 impact events on the plate surface. The mean localization error on the chosen sites is 3 cm. Processed data are finally passed to the AR algorithm to build the contents that will be projected over the real world image in real-time as shown in Fig. 15.3.

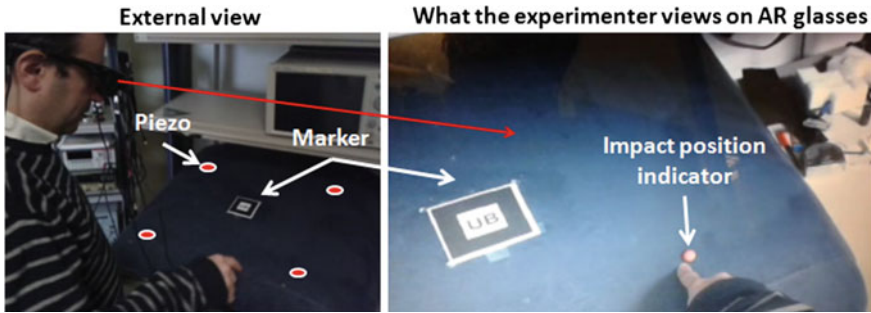


Fig. 15.3 Image on the Vuzix glasses: the impact indicator is superimposed to the external real world image.

15.5 Conclusions

An Augmented Reality (AR) tool for structural health monitoring was designed in this work. It is meant for providing maintenance operators with the results of an impact detection methodology by mean of a smart eyepiece. Electric signals are collected by a network of piezosensors bonded on a Cessna 150 engine cowling. Dispersive propagation compensation is performed to improve estimation robustness. Hyperbolic beamforming is exploited to locate the impact. Experimental results confirm the feasibility of the method and its exploitability in maintenance practice. It is worth noticing that the robustness of the wave traveled distance estimation allows to achieve such performances with sparse arrays of conventional transducers. Optimized and adaptive selection of the array shape and composition is under investigation to further improve the accuracy of the proposed approach.

References

1. Kundu, T., Das, S., Jata, K.V.: Detection of the point of impact on a stiffened plate by the acoustic emission technique. *Smart Mater. Struct.* **18**(3), 035006 (2009)
2. Perelli, A., De Marchi, L., Marzani, A., Speciale, N.: Acoustic emission localization in plates with dispersion and reverberations using sparse pzt sensors in passive mode. *Smart Mater. Struct.* **21**(2), 025010 (2012)
3. Wang, X., Kim, M. J., Love, P.E., Kang, S.C.: Augmented reality in built environment: classification and implications for future research. *Autom. Constr.* **32**, 1–13 (2013)
4. Liverani, A., Ceruti, A., Caligiana, G.: Tablet-based 3d sketching and curve reverse modelling. *Int. J. Comput. Aided Eng. Technol.* **5**(2–3), 188–215 (2013)
5. Chi, H.L., Kang, S.C., Wang, X.: Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. *Autom. Constr.* **33**, 116–122 (2013)
6. Debernardis, S., Fiorentino, M., Gattullo, M., Monno, G., Uva, A.E.: Text readability in head-worn displays: Color and style optimization in video vs. optical see-through devices. *IEEE Trans. Vis. Comput. Graphics* **20**(1), 125–139 (2014)
7. <http://www.hitl.washington.edu/artoolkit/>

8. http://www.vuzix.com/augmented-reality/products_star1200.html
9. Bocchini, P., Marzani, A., Viola, E.: Graphical user interface for guided acoustic waves. *J. Comput. Civil Eng.* **25**(3), 202–210 (2011)

Chapter 16

Squeeze the Lemon: Balancing as a Way to Use Every Drop of Energy in a Lithium-Ion Battery

Federico Baronti, Roberto Roncella and Roberto Saletti

Abstract This work discusses recent research results obtained in tackling one of the most limiting factor for an effective use of a Lithium-ion battery: the charge unbalance between the cells constituting the battery. First, it is recalled how unbalancing affects the performance of a battery consisting of series-connected cells, then some possible techniques to balance the battery are described and compared to each other. The comparison is made by modeling the balancing circuit topologies and by performing statistical simulations. Finally, we describe two balancing circuits that efficiently address the problem and we report on the experimental results that validate the circuits.

Keywords Lithium-ion battery · Cell balancing

16.1 Introduction

Lithium-ion battery technology has determined the explosion of the portable electronic device market (laptop, personal digital assistant, smart-phones), as it provides very high energy and power densities if compared to previously adopted technologies [1]. The introduction of Lithium-ion batteries has made it viable applications that seemed inconceivable before. The Electric Vehicle (EV) automotive market also seems to have migrated from Lead-acid or NiMH batteries to the more performing Lithium-ion batteries as new EVs are now equipped with the latter. Therefore, Lithium-ion is becoming the most widely adopted technology for rechargeable energy storage systems, even in medium- high-power applications. Unfortunately, Lithium-ion batteries are very fragile as they are sensitive to the operating conditions. Permanent damages and even flames may occur if the battery is overcharged, undercharged or operated outside the safe temperature range. Thus, Lithium-ion batteries

F. Baronti · R. Roncella · R. Saletti (✉)
Dipartimento Ingegneria Dell'Informazione, University of Pisa,
Via G.Caruso 16, 56126 Pisa, Italy
e-mail: r.saletti@iet.unipi.it

© Springer International Publishing Switzerland 2016
A. De Gloria (ed.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 351,
DOI 10.1007/978-3-319-20227-3_16

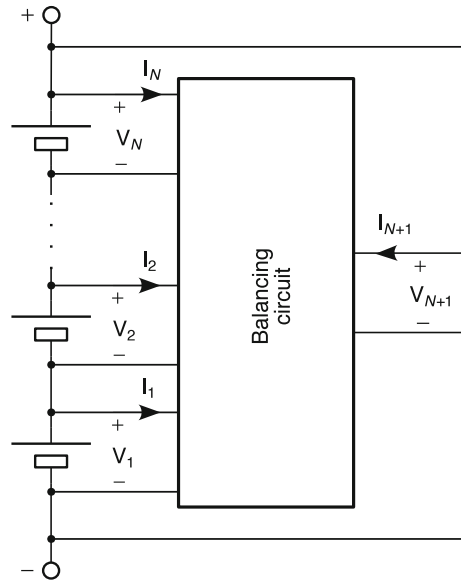
are always provided with an electronic control system that is in charge of monitoring, managing and preserving the safety of the battery. This circuit is generally called Battery Management System (BMS). A BMS is usually interfaced to upper level control systems in the considered application to provide valuable information such as the State of Charge (*SoC*) and State of Health (*SoH*) of the battery cells [2]. Besides some accidents reported on vehicles or planes and the issues that research efforts are trying to solve, the growth of the Lithium-ion battery market continues. One of the problems not efficiently solved yet for an effective use of a Lithium-ion battery is the charge unbalance between the series-connected cells building up the battery. As the Lithium-ion battery technology provides elementary cells with voltages of the order of $3 \div 4$ V, many cells are arranged in series to reach the voltage level required by the application, even hundreds of Volt in EVs. Possible mismatches in the physical properties or in the operating conditions of the cells determine a charge imbalance that makes it impossible to utilize all the energy in the battery without exceeding the safety limits of the cells [3]. Thus, the BMS must provide the battery pack with a balancing function also. This paper reviews and compares by means of statistical simulations some of the most popular balancing techniques. The implementations of two different balancing topologies are finally described, together with the experimental results measured on the related balancing circuits.

16.2 Battery Balancing Circuit Topologies

Cell unbalance is a major issue in a Li-ion battery consisting of series-connected cells, because battery recharge must be stopped when one cell reaches the charge cut-off voltage, even if the other cells are not completely charged. On the other hand, the discharge is interrupted when the least charged cell reaches the discharge cut-off voltage before the others, even if there is still some residual energy stored in the battery. The basic idea behind balancing is to bring all the unbalanced cells to the same charge level at the end of the balancing process. In this way, all the cells (assuming that they have the same capacity) will reach the charge/discharge cut-off voltage limits simultaneously, thus allowing the use of every drop of energy than can be stored in the battery.

Thus, balancing requires that every cell must be individually accessed to modify its charge level. The easiest way to balance a battery is to dissipate the extra energy in a resistor placed in parallel to the cell until all the cells reach the same charge level. This is called passive balancing. Instead of dissipating energy, active balancing aims at saving energy by transferring charges from the most charged cells to the least charged ones. Several topologies are possible, according to the way by which charges are transferred between the cells [4, 5]. In this work, we define the *Cell to Cell*, *Cell to Pack* and *Pack to Cell* topologies if the charges are exchanged between two individual cells, from one particular cell to the entire battery pack, or from the entire battery pack to one particular cell, respectively. These active balancing methodologies are based on a matrix of switches by which each cell of the

Fig. 16.1 General model of a balancing network based on a DC/DC converter and $N + 1$ ports [6]



battery is individually accessed to modify its charge. Each topology can be modeled with a very simple model of charge transfer, in which the charges are moved with energy efficiency η by means of a bidirectional DC/DC converter with $N + 1$ ports [6], connected to the N cells and to the battery pack terminals. The general model of the balancing network is shown in Fig. 16.1.

As detailed in [6], the optimum sequence of charge transfers can be found for each topology that makes the battery balanced. Unfortunately, the charge transfers occur with a certain energy efficiency as every transfer dissipates energy. The optimum sequence comes from the minimization of the charges to be transferred during balancing. This means that every unbalanced cell must reach the final charge level, which is identical for all the cells, in a monotonic way, i.e., always donating or accepting charges only.

A comparison among the different balancing topologies in terms of balancing time and energy loss during balancing is carried out in [6], by applying the above described model and strategy to different distributions, statistically chosen, of unbalance states of the battery cells. The following assumptions are made. Q_{\max} is the maximum charge stored in any cell. The maximum SoC mismatch considered in the example is 10%, as the most charged cell is at Q_{\max} , while the least charged cells is at $0.9 \cdot Q_{\max}$. The other cells are charged at Q_h , with $0.9 \cdot Q_{\max} < Q_h < Q_{\max}$, $h \in 1 \dots N - 2$ and Q_h randomly distributed with a uniform distribution between the two limit values. This means that we find an unbalanced battery with 10% maximum SoC mismatch, with all the cells between 90 and 100% of the maximum charge, in every statistical experiment. The optimum balancing algorithm is then applied and the values of the balancing time and the energy losses are calculated. Two compari-

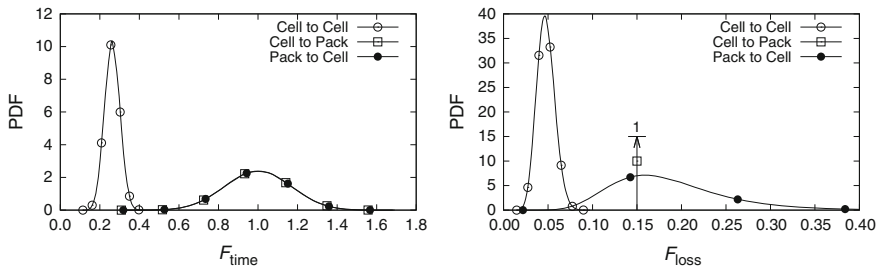


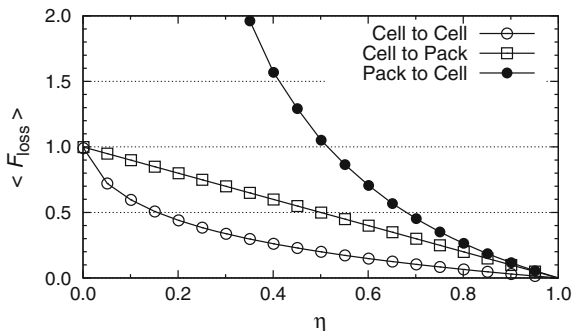
Fig. 16.2 Probability Density Function (PDF) of the balancing time figure F_{time} (left) and the balancing energy loss figure F_{loss} (right) [6]

son parameters, the balancing time figure F_{time} and the energy loss figure F_{loss} , are defined as the ratio between the balancing time and the energy loss, respectively, divided by the relevant values calculated for the passive balancing. Values of F_{time} and F_{loss} less than 1 mean that the active balancing techniques have performed better than the passive one. The experiment is repeated with 100,000 different unbalance configurations with the following additional parameters: the current I_{sh} in the shunt resistor of passive balancing is 200 mA, the DC/DC balancing current I_{bal} is 1 A and the converter energy efficiency η is 0.85. The probability density function (PDF) of the random variables F_{time} and F_{loss} are finally calculated. Figure 16.2 shows the results.

The diagrams show that active balancing performs better or equal to passive balancing in time, and by far better if energy losses are considered. However, the simulation results show that there are significant differences between the various active topologies. In particular, the *Cell to Cell* configuration clearly outperforms the other methods in terms of both balancing time and energy loss.

The energy loss comparison is also performed by varying the DC/DC converter efficiency η , as it is shown in Fig. 16.3, where the mean value of the energy loss figure F_{loss} is reported as a function of η . It is amazing to see that active balancing may even be worse than passive balancing (a value greater than 1) if the DC/DC

Fig. 16.3 Mean value $\langle F_{loss} \rangle$ of F_{loss} as a function of the converter efficiency η [6]



converter is of low quality and its efficiency drops below 50 %, for the *Pack to Cell* configuration, one of the most popular topologies.

In conclusion, the study shows that, besides the converter efficiency, the most important factor to efficiently achieve battery balancing and thus squeeze all the energy stored in a battery is the topology used to transfer charges between the battery cells.

16.3 Balancing Circuit Implementations

Two balancing circuits that implement the *Cell to Cell* and *Pack to Cell* balancing techniques are described hereafter.

16.3.1 Cell-to-Cell Balancing Circuit

The first circuit is shown on the left side of Fig. 16.4. The cells are connected by means of a switch matrix consisting of MOS transistors. The DC/DC converter is bidirectional and connects the cell selected by the switch matrix with a supercapacitor (SCAP) that acts as a temporary energy storage tank. The idea applied in this circuit is to extract energy from one cell, store it in the SCAP and then return it back to another properly selected cell. At the end of the two-way transfer we have succeeded in moving charges from one cell to another, thus applying the cell-to-cell balancing topology. The circuit was applied to a battery consisting of 11 40 Ah cells with Nickel Manganese Cobalt (NMC) cathode. Details on the circuit implementation and the experimental validation can be found in [7].

The right side of Fig. 16.4 shows the voltages of the battery cells as a function of time during a balancing experiment. The battery is strongly unbalanced, as one

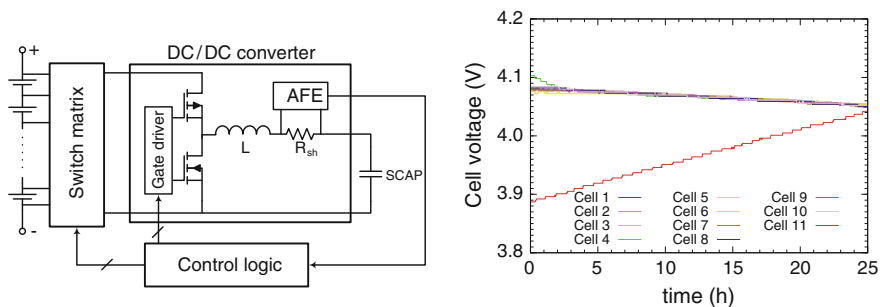


Fig. 16.4 Schematic circuit of the balancing circuit implementing the Cell-to-Cell topology (*left*). Cell voltages as a function of time, showing that balancing is finally achieved after starting with a large unbalanced configuration (*right*) [7]

cell (Cell 11) shows a *SoC* almost 20 % less than the other cells. It is worth noting how balancing is recovered by moving from the other cells the charges needed to equalize the battery. The balancing procedure lasts around 25 h and is very efficient as it only costs 1 % of the battery energy. In fact, the measured efficiency of the energy transfer is well over 75 %. Finally, an energy saving of a factor larger than 6 is obtained if compared to the passive balancing technique applied to the same unbalanced condition.

16.3.2 Pack-to-Cell Balancing Circuit

The circuit implementing the Pack-to-cell balancing topology is shown in Fig. 16.5 and is described in details in [8, 9]. The circuit consists of a DC/DC converter fed by the entire battery voltage that provides charges to one particular cell selected by the switch matrix. A peculiar feature of this circuit is the presence of an additional switch that enables the balancing between different battery modules connected with a circular balancing bus. The circuit performs the balancing of 4 60 Ah Lithium-ion with Lithium Iron Phosphate (LFP) cathode cells with balancing currents of 1.5 A. The switch matrix is realized with MOS switches characterized by an on-resistance of about 7.5 m Ω , thus leading to a rather low dissipation on them. The measured balancing efficiency is around 70 %.

Figure 16.5 also shows the *SoC* of the battery cells before and after a balancing process that starts with cell 3 unbalanced of 10 %. It is worth noting how a balanced configuration of the battery is restored, allowing the exploitation of the full capacity of the battery. The cell voltages are also reported to show the voltage relaxation phenomena occurring after the balancing end.

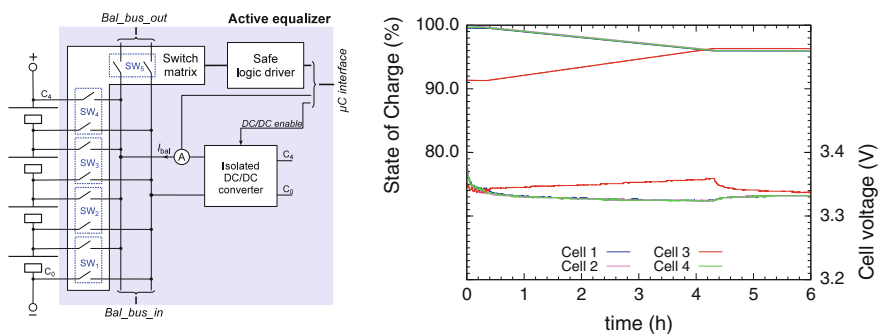


Fig. 16.5 Schematic circuit of the balancing circuit implementing the Pack-to-Cell topology (left). State of Charge and voltage of the cells as a function of time, showing that balancing is restored (right) [9]

16.4 Conclusions

The paper presents a survey of research efforts aimed to achieve the efficient balancing of a Lithium-ion battery. Balancing a battery allows the exploitation of every drop of energy stored in the battery, otherwise not fully usable. It is shown by statistical simulations that active balancing performed by transferring charges from the most charged cells to least charged ones is the most efficient technique for both balancing time and energy loss figures of merit. Two circuits that implement different balancing techniques are described and the results obtained during their experimental validation are finally presented.

References

1. Whittingham, M.S.: History, evolution, and future status of energy storage. *Proc. IEEE* **100**(Special Centennial Issue), 1518–1534 (2012)
2. Lu, L., Han, X., Li, J., Hua, J., Ouyang, M.: A review on the key issues for lithium-ion battery management in electric vehicles. *J. Power Sources* **226**, 272–288 (2013)
3. Zhong, L., Zhang, C., He, Y., Chen, Z.: A method for the estimation of the battery pack state of charge based on in-pack cells uniformity analysis. *Appl. Energy* **113**, 558–564 (2014)
4. Gallardo-Lozano, J., Romero-Cadaval, E., Milanés-Montero, M.I., Guerrero-Martinez, M.A.: Battery equalization active methods. *J. Power Sources* **246**, 934–949 (2014)
5. Daowd, M., Omar, N., Van Den Bossche, P., Van Mierlo, J.: Passive and active battery balancing comparison based on MATLAB simulation. In: 2011 IEEE Vehicle Power and Propulsion Conference, pp. 1–7. IEEE (2011)
6. Baronti, F., Roncella, R., Saletti, R.: Performance comparison of active balancing techniques for lithium-ion batteries. *J. Power Sources* **267**, 603–609 (2014)
7. Baronti, F., Roncella, R., Saletti, R., Zamboni, W.: Experimental validation of an efficient charge equalization system for lithium-ion batteries. In: 23rd IEEE International Symposium on Industrial Electronics (ISIE), Istanbul, vol. 2014, pp. 1–6 (2014)
8. Baronti, F., Fantechi, G., Roncella, R., Saletti, R.: High-efficiency digitally controlled charge equalizer for series-connected cells based on switching converter and super-capacitor. *IEEE Trans. Ind. Inf.* **9**(2), 1139–1147 (2013)
9. Baronti, F., Fantechi, G., Roncella, R., Saletti, R., Pede, G., Vellucci, F.: Design of the battery management system of LiFePO₄ batteries for electric off-road vehicles. In: 2013 IEEE International Symposium on Industrial Electronics ISIE, Taipei, pp. 1–6. IEEE (2013)

Chapter 17

Fully Integrated 60 GHz Transceiver for Wireless HD/WiGig Short-Range Multi-Gbit Connections

Sergio Saponara and Bruno Neri

Abstract The paper presents the design of a 60 GHz transceiver, with all active and passive devices integrated on-chip including the antenna, for multi Gbit short-range wireless communications. To minimize circuit complexity and cost an on-off-keying (OOK) modulation scheme is selected as well as a 65 nm bulk CMOS technology instead of more costly CMOS SOI or SiGe or III-V technologies. At transmitter side a differential 2 stage common source power amplifier allows for an output power of about 11 dBm. The receiver includes a cascode LNA with a gain of 11 dB and a noise figure of 4.6 dB, followed by a simple envelop detector. For the on-chip antenna, half-wavelength dipole and inverted-F topologies have been designed. For the transceiver prototype the half-wavelength dipole is selected since it has better gain performance (radiation efficiency of 38 %, a peak directivity of 1.76 and a gain of roughly -2 dBi) although for an higher area occupation. The fully integrated transceiver allows for a data rate of roughly 4 Gbit/s at distances of few meters, being compliant with physical-layer specifications of WirelessHD and WiGig alliances. *Work supported by NEWCOM # EU grant.*

Keywords 60 GHz · mm Wave electronics · Wigig · WirelessHD

17.1 Reasons for Integrated Transceivers at 60 GHz in Bulk-CMOS

To enable new wideband and interoperable communication services among consumer devices such as smartphones, tablet, PDA, multimedia players, the design of cost-effective single-chip transceivers working at multi Gbit/s is required.

S. Saponara (✉) · B. Neri

Dipartimento Ingegneria della Informazione, Università di Pisa, Via G Caruso 16, Pisa, Italy
e-mail: sergio.saponara@iet.unipi.it

B. Neri

e-mail: bruo.neri@iet.unipi.it

The connection distances amount to few meters for target applications such as wireless fast internet access at several Gbit/s to an host point, uncompressed video playing, wireless high-speed USB connections, wireless multimedia (video, audio) area networks. For high-speed board to board communications the connection distance can be in the order of tens of cm. The lower the link distance, the higher the maximum achievable data-rate.

To this aim the unlicensed band allocated world-wide at millimeter waves, around 60 GHz [1–4] (Extremely High Frequency part of the electromagnetic spectrum), can be exploited. Several countries in Europe, US, Asia, have allocated several GHz of free spectrum around 60 GHz for short range communications. As an example the FCC in US allocated 7-GHz band in the radio spectrum between 57 and 64 GHz [5]. The atmospheric absorption of 60 GHz energy by oxygen molecules limits undesired propagation over long distances and helps interference control with other systems and long distance reception, which is a concern to multimedia copyright owners.

To exploit the opportunities offered by the 60 GHz wireless technology, several consortia have been already created by industries such as the WirelessHD consortium, based on the IEEE 802.15.3c standard for multi-gigabit Wireless Personal Area Network, or the WiGig alliance, starting from the IEEE 802.11ad standardization effort. Such applications are complementary to well established blue-tooth or WLAN connections where operating frequencies (between 2.4 and 5 GHz) and data rates (from few to hundreds of Mb/s) are 1 or 2 orders of magnitude lower than the capabilities of 60 GHz wireless links.

As example the WiGig specifications foresee tri-band devices (operating at 2.4, 5 and 60 GHz) to deliver data transfer rates up to 7 Gbit/s, one order of magnitude faster than 802.11n connections, while maintaining compatibility with existing Wi-Fi devices. The 60 GHz signal cannot penetrate walls but can propagate off reflections from walls, ceilings, floors and objects. Hence the 60 GHz connection is foreseen for multi Gbit/s connections in short range home or office environments. Since the spectrum is freely available worldwide at 60 GHz there are no license costs to pay for the set up of new services at such frequency. When roaming away from the main room the protocol can switch from 60 GHz to lower bands at a lower rate, thus propagating through walls and reaching the connection distances of a Wi-Fi service.

Similar to WiGig alliance, the WirelessHD specification is based on a 7 GHz channel at 60 GHz allowing either lightly compressed or uncompressed digital transmission of high-definition video and audio and data signals, essentially making it equivalent of a wireless HDMI. First-generation implementations of WirelessHD reach data rates of about 4 Gbit/s, although theoretical data rates foreseen in the WirelessHD specification can be as high as 25 Gbit/s. The target application for Wireless HD products are in-room point-to-point connections up to few meters. Just as a comparison it is worth noting that the WHDI (Wireless Home Digital Interface) standard provides a wireless link with data rates up to 1.5 Gbit/s in a 20 MHz channel of the 5 GHz unlicensed band. Hence moving from 5 GHz of WHDI to 60 GHz of WirelessHD the achievable data rate grows by one order of magnitude.

To address the above issues the design of a 60 GHz OOK transceiver, with on-chip integrated antenna, in a 65 nm digital CMOS technology is discussed in the paper.

After this introduction Sects. 17.2 and 17.3 present the overall transceiver architecture and the main characteristics of the select technology. Section 17.4 reports implementation results and a comparison with the state of the art.

17.2 Target CMOS-Bulk Technology

Integrated 60 GHz transceivers in GaAs MMIC (monolithic microwave integrated circuits) or CMOS Silicon on Insulator (SOI) technologies with performances suitable for multi Gbit/s wireless communications have been already proposed in literature [1, 6]. However HEMT GaAs or CMOS SOI devices have a cost higher than conventional bulk CMOS nodes, used to integrate baseband analog or digital circuitry in complex system-on-chip solutions. Since the target consumer electronic market is a large volume one, the real challenge still to overcome [7–12] is the cost-effective integration of a 60 GHz transceiver in bulk CMOS while high-speed A/D converters and baseband digital processing systems are available in the state of art [13]. To further reduce system size and weight, also the antenna should be integrated on-chip. At millimeter waves the wavelength is few mm and hence an antenna can be integrated in 1 mm². Beside costs also power consumption is a key issue. The power consumption should be in the much less than 1 W so a data-rate of some Gbit/s will lead to an energy cost lower than 100 pJ/bit. To be compliant with System-on-Chip design flows for large volume consumer applications, the target technology selected in this work is a standard CMOS technology, typically used for digital and mixed-signal baseband designs: a UMC 65 nm bulk CMOS with 1.2 V voltage supply for the core, a low resistivity substrate (20 Ω cm) and 10 metal layers. In its GP (general purpose) version the F_{\max} is above 150 GHz and hence, although conceived mainly for digital and mixed-signal designs, is also suited to work at 60 GHz. The technology is also available through Europractice for multi-project wafer prototyping.

17.3 Low Complexity 60 GHz Transceiver Architecture

To keep low the circuit complexity a simple architecture has been selected for the transceiver, see Fig. 17.1, exploiting an On-Off Keying (OOK) modulation-demodulation scheme. At transmitter side, see block diagram in Fig. 17.2, a voltage controlled Local Oscillator (LO) generates a 60 GHz carrier which is an input of the OOK modulator.

The modulator, realized as a simple cascode switch, receives also as input the baseband digital code to be transmitted and as output the modulated OOK signal is obtained. The proposed 60 GHz oscillator has a cross-coupled topology, exploiting the negative resistance method, with on-chip integrated passive devices: inductors

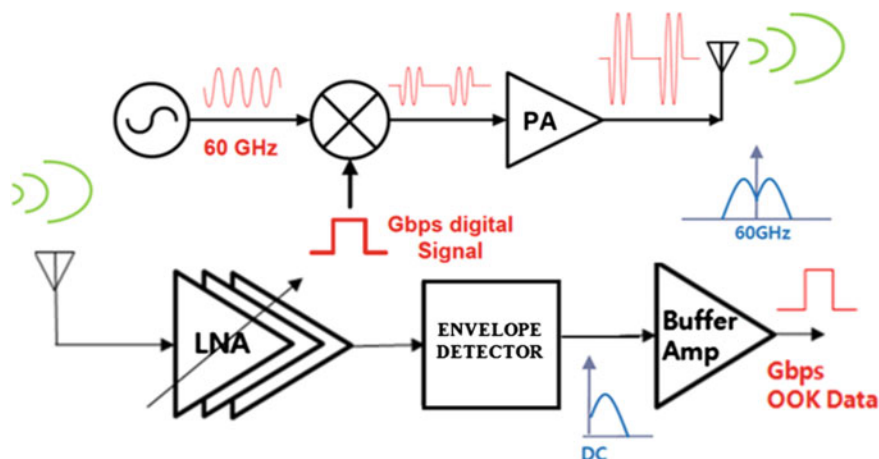


Fig. 17.1 Block diagram of the proposed integrated transceiver

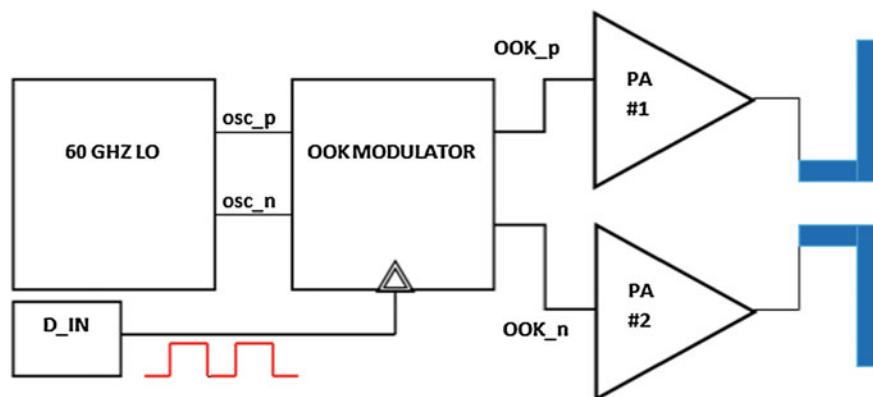


Fig. 17.2 Transmitter block diagram with differential PA and On-chip dipole antenna

(22 pH) and capacitors (26 fF) with quality factor (Q) of around 20. The designed oscillator in 65 nm bulk technology has low phase noise, -87 dBc/Hz at 1 MHz relative frequency, and an output power of 1.5 mW.

The OOK modulator is a cascode switch with a 20 dB input-output isolation between ON and OFF states. The transistors were biased with a current density J_{opt} of 0.3 mA/ μ m to maximize the OP1 dB compression point [8]. The transistor total width was set to achieve a low DC current (10 mA for each branch, 20 mA for the full differential OOK). For the HIGH and LOW states of the digital signal analog values of 1.2 V and 0 V were used. The input referred 1 dB CP is approximately 1.2 mW; since the gain is unitary it is equal to the output referred 1 dB compression point. The input oscillation power for each single ended output of the oscillator is 1.5 mW. A slightly

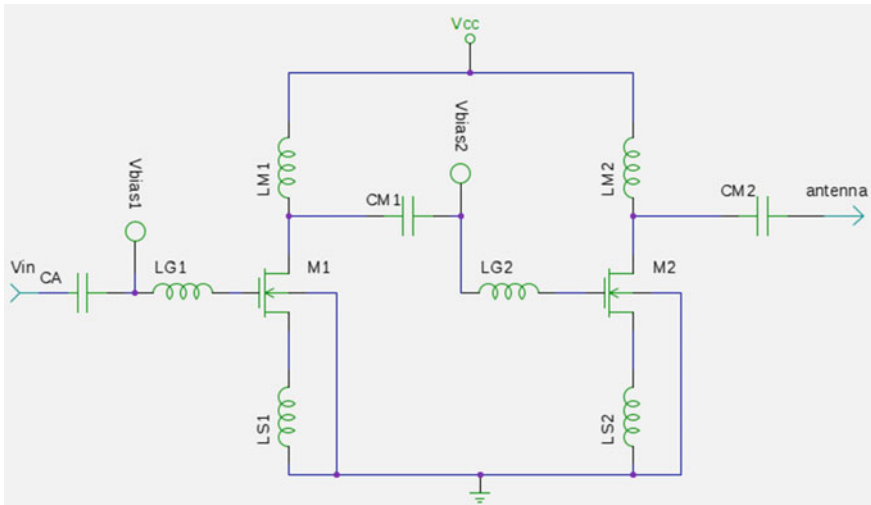


Fig. 17.3 Power amplifier circuit (class-A 2 stage CS)

smaller value is thus fed to the PA whose input referred 1dB CP is 1.5 mW. In this way the whole system is designed to exploit the maximum output power.

A differential power amplifier with on-chip integrated dipole antenna converts the OOK modulated power electrical signal in an electromagnetic wave with the required power level to ensure a link coverage of some meters. The power amplifier has a pseudo differential topology, see Fig. 17.2, where each of the PA blocks has a class-A 2 stage Common Source (CS) topology with inter-stage LC matching networks, see Fig. 17.3.

The power amplifier devices have been sized targeting a current density of 0.3 mA/ μm . While for other technologies, like SiGe HBT, the value of J_{opt} maximizing the OP1 dB increases almost with the square of the scaling factor, instead the CMOS peak F_t current density remains constant for different finger widths and technology nodes, also in a cascode topology [9]. The value of J_{opt} was experimentally found in [8] and it's approximately 0.3 mA/ μm as used in this work. With a 1.2 V power supply each PA circuit has a G_T gain of 6.8 dB, a 1 dB output compression point OP1 dB of 8.2 dBm. The peak PAE (power-added-efficiency) is 11.5 %. The power of the 2 PA blocks is combined in a differential topology and hence the whole transmitter provides to the antenna an output power of roughly 11 dBm (12 mW) and has a power consumption of 170 mW.

The integrated on-chip antenna has a half-wavelength dipole topology where the metallization has been realized using the upper metal layers (thick layer minimizing conductive losses). We proposed this integrated antenna in [10] in a 65 nm CMOS SOI technology. In this work it has been redesigned targeting a less expensive bulk 65 nm CMOS technology and achieves a maximum gain of roughly -1.76 dBi, a radiation efficiency of 38 %, a minimum S11 value of -14 dB and a -10 dB S11 Bandwidth of 8 GHz. To avoid interference with other on-chip devices,

from HFSS simulations we derived that other metallization should be kept at a distance of at least 0.16λ from the antenna. To this aim in the chip layout an area of roughly 2 mm^2 should be reserved to the antenna. The achieved gain in the 65 nm CMOS bulk technology is only few dB lower than the same design in the more expensive 65 nm CMOS SOI technology in [10]. The performance of the integrated antenna is enough for the link distances of our target applications. We also designed other antenna topologies such as the inverted-F integrated in the same 65 nm bulk CMOS technology. With respect to the dipole antenna the inverted-F design has halved area occupation but also a lower gain, -8 dBi , and radiation efficiency, 11 %.

At receiver side the signal is acquired through an active antenna realized by integrating on-chip a dipole antenna and a low-noise amplifier (LNA). The baseband digital information code is extracted from the RF received signal with a simple envelope detector thanks to the adopted OOK modulation scheme. With respect to more complex modulation schemes the OOK approach avoids complex circuitry for carrier frequency recovery; instead of a mixer and local oscillator at RF and some circuitry (automatic gain controlled amplifier and filter) at intermediate frequency (IF) before the demodulator, a simple envelope detector can be used to recover the baseband signal directly from the RF one. The limit of the OOK approach versus more complex modulation schemes is its lower spectral efficiency. However for wireless short range communications at 60 GHz, with several GHz of bandwidth available, the constrain on narrow band utilization is relaxed. For the LNA design in this work we adapted a single-stage cascode amplifier circuit we proposed in [10] for the 65 nm SOI technology CMOS. For a less expensive bulk 65 CMOS node with a single stage cascode we still achieve good results: a G_T gain of 11.5 dB, a NF of 4.7 dB, a -3 dB bandwidth of 8 GHz and a power consumption of 2.5 mW. Since both antenna and LNA are integrated on-chip, the impedance matching is not constrained to 50Ω , but its value has been found as an optimal trade-off between NF and gain of the receiver and resulted 75Ω . By adding another stage in the LNA its gain can growth beyond 20 dB although the power consumption growths up to roughly 10 mW.

The transceiver architecture is completed with a baseband digital signal processing (DSP) chain which includes a Reed-Solomon channel coding unit with 8 dB coding gain and a redundancy of 40 %. The RS approach has been preferred to more performing channel coding techniques such as the LDPC (Low Density Parity Check) codes to keep low the circuit complexity.

17.4 Implementation Results and Conclusions

Table 17.1 compares the performance of the designed OOK transmitter versus state-of-art solutions for 60 GHz wireless short range applications. The achieved results show that our design has good performance in terms of efficiency and provided output power level, roughly 11 dBm. Our transceiver is the only design providing the on-chip integration of the antenna while the others exploit a classic approach where the antenna is off-chip. Our fully integrated transceiver, without

Table 17.1 Performance of the whole OOK transmitter versus state of art

	[11]	[12]	Our
Output Power, dBm	1.76	7.6	11
Dc power, mW	156@1.2 V	183@1.2 V	172@1.2 V
TX efficiency, %	1	3.1	7
CMOS technology	90 nm	90 nm	65 nm
On-chip antenna	NO	NO	YES

coding redundancy of the baseband DSP technique, is capable of a data rate of roughly 4 Gbit/s at a distance of 2.5 m thus meeting the specifications of WirelessHD and WiGig consortia without requiring an off-chip antenna. In case of harsh environment conditions to improve the bit error rate (BER) in case of low SNR conditions the RS channel coding can be activated. In such case the transceiver can still provide 2 Gbit/s at 2.5 m distance with a BER of 10^{-6} at a low SNR of 12 dB.

The achieved results prove that 60 GHz high speed wireless communications can be achieved with fully integrated solutions in low cost bulk CMOS technology, suited for large volume markets as consumer electronics or mobile devices. Still work is needed to further increase the data rate integrating more complex LDPC channel coding schemes or integrating antennas with beam forming capabilities.

References

- Gunnarson, S., et al.: 60 GHz single-chip front-end MMICs and systems for multi-Gb/s wireless communication. *IEEE JSSC* **42**(5), 1143–1157 (2007)
- Vaughan, J.S.N.: Gigabit Wi-Fi is on its way. *IEEE Comput.* **43**(11), 11–14 (2010)
- Zhang, X., et al.: Physical layer design and performance analysis on multi-Gbps millimeter-wave WLAN system. *IEEE ICCS10* 92–96
- Singh, H., et al.: A 60 GHz wireless network for enabling uncompressed video communication. *IEEE Comm. Mag.* **46**(12), 71–78 (2008)
- FCC: Code of federal regulation, title 47 telecommunication, Chapter 1, part 15.524 (2004)
- Siligaris, A., et al.: CMOS SOI technology for WPAN. Application to 60 GHz LNA. *Emerg. Technol. Circuits LNEE* 2010 **2021**(5), 123–130
- Niknejad, A.: 0–60 GHz in four years: 60 GHz RF in digital CMOS. *IEEE Solid-State Circ. News.* **12**(2), 5–9 (2007)
- Dickson, T.O., et al.: The invariance of characteristic current densities in nanoscale MOSFETs and its impact on algorithmic design methodologies and design porting of Si(Ge) (Bi)CMOS high-speed building blocks. *IEEE JSSC* **41**(8), 1830–1845 (2006)
- Voinigescu, S., et al.: Algorithmic design of CMOS LNAs and PAs for 60 GHz radio. *IEEE JSSC* **42**(5), 1044–1057 (2007)
- Fonte, A., Saponara, S., et al.: Design of a low noise amplifier with integrated antenna for 60 GHz wireless communications. *IEEE IMWS* 160–163 (2011)
- Juntunen, E., et al.: A 60-GHz 38-pJ/bit 3.5-Gb/s 90-nm CMOS OOK digital radio. *IEEE Trans. MTT* **58**(2), 348–355 (2010)
- Chen, Y., et al.: A low-power low-cost fully-integrated 60-GHz transceiver system with OOK modulation and on-board antenna assembly. *IEEE JSSC* **45**(2) (2010)
- Saponara, S., et al.: Architectural exploration and design of time-interleaved SAR arrays for low-power and High Speed A/D Converters. *IEICE Trans. Electr.* **E92-C**, 843–851 (2009)

Chapter 18

Low Cost FMCW Radar Design and Implementation for Harbour Surveillance Applications

Sergio Saponara, Stefano Lischi, Riccardo Massini, Luca Musetti, Daniele Staglianò, Fabrizio Berizzi and Bruno Neri

Abstract The prototype of a low power radar with a coverage range from some hundred meters to a few kilometres is presented in this paper. A Frequency Modulate Continuous Wave (FMCW) X-band solution has been chosen from the architectural point of view, and an hybrid realization has been adopted. In fact, one of the main target of the design was to demonstrate the feasibility of high sensitivity radar sensor implemented in a low cost technology. Some original solutions have been adopted to reach this target. The radar prototype is composed by a radio frequency front end, entirely realized by using commercial components, and by a DSP platform implemented by using open source software. The prototype has been realized and tested. The main interesting results are: (i) the low cost of the radio frequency front end; (ii) the high sensitivity of the sensor which resulted capable of monitoring the movements of object and peoples in a range of some hundred meters with a transmitted power as low as a few mW; (iii) a very low level of EM pollution; (iv) a coverage range of 1 mile with an output power of 2 W; (v) the detection of range and velocity for targets with a radar cross section larger than 1 m inside the coverage range. The main application will be as a node of a radar network for harbour surveillance.

Keywords Radar · Surveillance system · Wireless electronics · X-band transceiver

18.1 Applications for MIC-Based Radar Transceivers

A radar sensing system allows for several advantages with respect to other competing technologies for contactless sensing based on ultrasounds, video or thermal camera and LIDAR: a radar is an “all wheatear” sensor that can work in severe

S. Saponara (✉) · S. Lischi · R. Massini · L. Musetti · D. Staglianò · F. Berizzi · B. Neri
Dipartimento Ingegneria della Informazione, Università di Pisa,
Via G. Caruso 16, Pisa, Italy
e-mail: Sergio.saponara@iet.unipi.it

environmental conditions; it can work during night and day; a radar gives multiple information about the target (if there is the target, at which distance, at which relative speed and direction). For such reasons there is a growing interest in designing radar sensing systems for civil applications such as obstacle detection and adaptive cruise control in automotive, trains, small yachts, and surveillance systems in banks, railway crossing, harbour [1–6]. In the above cited civil applications power consumption and implementation costs of the radar sensing systems are key issue. The radar architecture is mainly composed by a radio frequency (RF) front-end, an Analog-Digital converting board and a baseband digital signal processing (DSP) platform. The most challenging part of the radar design from an electronic point of view is the RF transceiver to meet the cost and power consumption budgets. For the baseband DSP the main focus is at algorithmic level. Two types of solutions are available for the realization of low power radar: Microwave Integrated (or Hybrid) Circuit and Monolithic Microwave Integrated Circuits (MMIC).

The term Microwave Integrated Circuits (MIC) refers to RF and Microwave circuits realized by integrating both passive and active devices in the same planar substrate. The other solution is the utilization of a Microwave Monolithic Integrated Circuits [MMIC] which consists of the integration of the entire front end in a single chip thus realizing a Radar-On-A-Chip (ROAC). With respect to MIC, the MMIC approach allows for better performances in terms of maximum operating frequency, dimensions, reliability and power efficiency, but due to high design time and development cost the MMIC solution is suitable just for large scale production. On the contrary, MIC can be realized in an electronic laboratory equipped with a photo-lithographic etching process and a work station for bonding.

To address the above issues the prototype of a low power radar with a coverage range from some hundred meters to a few kilometres is presented in this paper. A Frequency Modulate Continuous Wave (FMCW) X-band solution has been chosen from the architectural point of view. The main application will be as a node of a radar network for harbour surveillance. In the rest of the paper Sect. 18.2 presents the system requirements and the selected radar architecture. Section 18.3 presents implementation details and experimental measured results. Conclusions are drawn in Sect. 18.4.

18.2 System Specifications for Harbour Surveillance Radar

The main objective of this work was the design, simulation, realization and testing of a low power (a few Watts), low cost (less than 1.000 \$ for the radio front end) RADAR, capable of monitoring the presence, and tracking the movements, of any kind of ship inside and in the neighbouring of a harbour. The coverage range should be from few hundred meters to some kilometres. The overall design of the harbour surveillance system is based on a sensor network whose node are single Radar sensors distributed in the harbour area, allowing the data fusion at network level of the imaging coming from all of the nodes. The systems specifications are summarized hereafter:

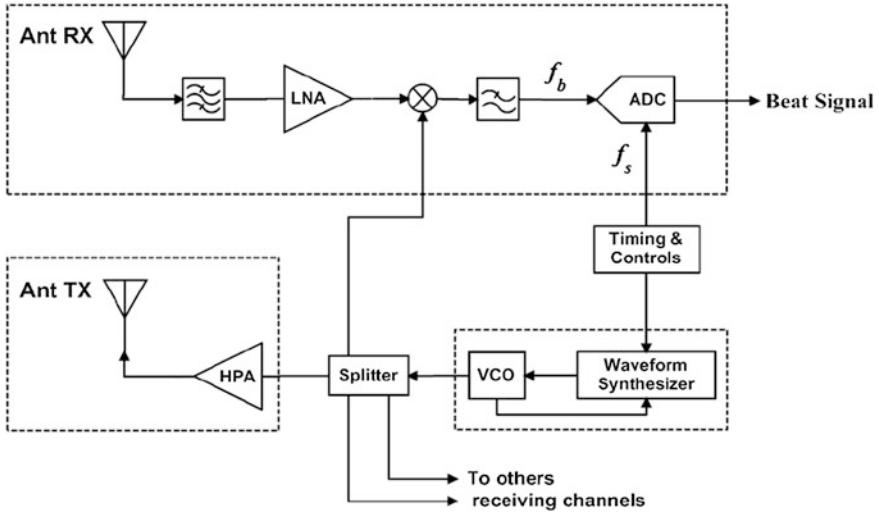


Fig. 18.1 Block diagram of a direct deramping FMCW RADAR front-end

Frequency:	X Band (8–12.5 GHz)
Waveform:	FMCW
Antenna:	60° azimuth
Range Coverage:	1.5 km
Range resolution:	0.5 m
Receiver dynamics:	>96 dB
Receiving channels:	3
SNR at the receiver output:	SNR ≥ 20 dB
Maximum Output power:	2 W

To meet the above specification, a direct deramping FMCW solution is the most suited one since it is capable of giving information both on the distance R and on the radial velocity V_R of the target, by using quite low values of the output power (a few watts) for some kilometres of coverage range and low cost components. The typical block diagram of a FMCW RADAR front-end is shown in Fig. 18.1 while the FMCW waveform adopted in the proposed transceiver is reported in Fig. 18.2.

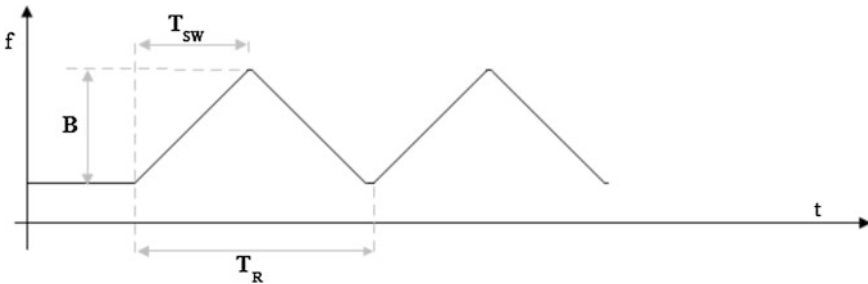


Fig. 18.2 Instantaneous frequency versus time

A system level simulation model of the whole Radar has been defined to derive from system-level specifications the requirements for all the Radar sub-blocks. As example of the results achievable using this model hereafter we report the requirements for the transceiver design.

Receiver NF (included ADC)	≤ 5 dB
LNA Gain	29 dB
Mixer conversion loss	8 dB
Minimum beat frequency (d = 25 m)	100 kHz
Maximum beat frequency(d = 1250 m)	5 MHz
IF band	100 kHz–5 MHz
IF Gain	43 dB

18.3 Low Complexity 10 GHz Transceiver Architecture

The design at system level has been carried out at CNIT who has also implemented the data acquisition and elaboration section, whereas the prototype of the radio-frequency front end and beat signal conditioning circuit has been realized at University of Pisa as well as the antennas. The details on system level dimensioning and on the DSP part have been published in [7]. Fast ADCs are available as COTS (commercial off the shelf) components, e.g. [8].

For the antenna design X-band Fabry-Perot resonator technology was adopted in order to realize compact and efficient antennas characterized by the total absence of side lobes on the azimuth plane and a Gain of 13 dBi.

18.3.1 Radio Front End

As far as the transmitting section, a hybrid continuous wave-form generator based on a Phase Locked Loop (PLL) has been utilized) in the range of 10–12 GHz. The PLL is capable of generating a FMCW output with a linearly varying instantaneous frequency $f(t) = At$, where A is the chirp rate. The maximum value of A is in the required range for our application of 10^{12}s^{-2} . A value of $6 \times 10^{11} \text{s}^{-2}$ has been really chosen which, together with $T_{SW} = 500$ us allowed to obtain the required value of $B = 300$ MHz corresponding to a range resolution of 0.5 m.

A maximum output power of 2 W is obtained thank to the stage marked as HPA (High Power Amplifier) in Fig. 18.3.

The frequency range of the beat signal depends on the maximum distance which has to be covered. For a coverage range of 1.250 km a $f_{IF} = 5$ MHz should be used.

A 6th order low pass Chebichev filter, with 5 MHz bandwidth, has been used as anti-aliasing filter before sampling the beat signal at a sampling rate of 12.5 MSa/s. Before the realization several simulations have been carried out both at the circuit and system level by using ADS (Agilent Technologies) CAD environment.

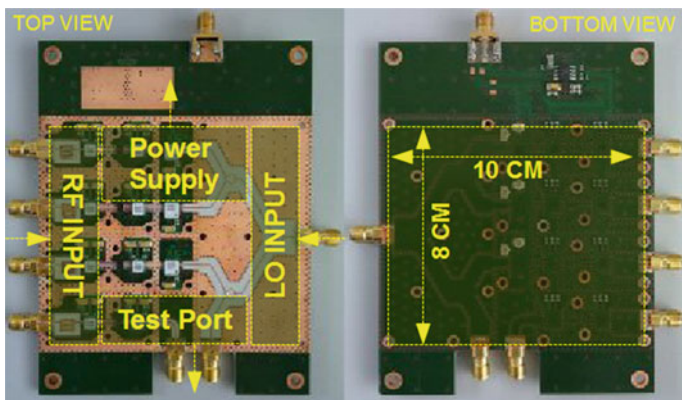
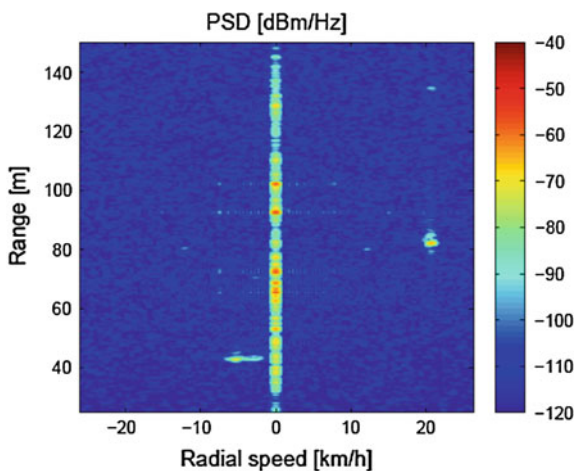


Fig. 18.3 The prototype of the receiver of the HS RADAR

In Fig. 18.3 the receiver section board is shown. Receiver and transmitter sections have been realized in separated boards to make easier the prototype test: they will be contained in the same compact shielded case in the final realization. The case will be connected to the antennas (receiving and transmitting) and to the power supply, whereas the low frequency output of the mixer, after low pass filtering and amplification (Test Port in Fig. 18.3), have been made available for sampling and DSP.

For a single sample of the Radar front end shown in Fig. 18.3, the cost of the devices, both active, passive, discrete and integrated and that for the realization of the board and of the case can be estimated around \$ 500 (€ 370). It is evident that in the case of series production of several hundreds or thousands of samples the unitary cost would drop in a significant way (at least 50 %).

Fig. 18.4 Range-Radial speed map of the experiment using a cooperative target



18.3.2 Experimental Measurements

Several tests have been carried out to assess the performance of the implemented Radar sensor. First experiments with cooperative targets were conducted observing a moving car in a short range scenario (100 m). Due to the reduced maximum range to cover (around 100 m), the power amplifier was not used in this scenario and the effective transmitted power was equal to -6 dBm. The obtained range-Doppler map is shown in Fig. 18.4. Two different targets have been detected: the one at about 80 m of distance and 20 km/h of speed was the car used for the test, while the one at 40 m and -5 km/h was a person walking around. The ground clutter line can also be noticed.

18.4 Conclusions

The prototype of a low power Radar, based on FMCW architecture, with a coverage range from some hundred meters to a few kilometres is presented in this paper. The radar prototype is composed by a RF front end, entirely realized by using commercial components, and by a DSP platform implemented by using open source software. The main interesting results achieved during testing of the whole radar system are: (i) the low cost of the RF front end; (ii) the extremely high sensitivity of the sensor which resulted capable of monitoring the movements of object and peoples in a range of some hundred meters with a transmitted power of -6 dBm; (iii) a very low level of EM pollution; (iv) a coverage range of 1 mile with an output power of 2 W; (v) the possibility to measure both range and velocity of targets with a radar cross section larger than 1 m inside the coverage range.

References

1. Hasch, J., et al.: Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band. *IEEE Tran. Microw. Theory Tech.* **60**(3), 845–860 (2012)
2. Li, C., et al.: High-sensitivity software-configurable 5.8-GHz radar sensor receiver chip in 0.13 μ m CMOS for noncontact vital sign detection. *IEEE Tran. Microw. Theory Tech* **58**(5), 1410–1419 (2010)
3. Li, Y.-A., et al.: A fully integrated 77 GHz FMCW radar transceiver in 65 nm CMOS, technology. *IEEE J. Solid-State Circuits* **45**(12), 2746–2756 (2010)
4. Menxel, W., et al.: Antenna concepts for millimeter-wave automotive radar sensors. *Proc. IEEE* **100**(7), 2372–2379 (2012)
5. Mitomo, T., et al.: A 77 GHz 90 nm CMOS transceiver for FMCW radar applications. *IEEE J. Solid-State Circuits* **45**(4), 928–937 (2010)
6. Neri, B., Saponara, S.: Advances in technologies, architectures and applications of highly-integrated low-power radars. *IEEE Aerosp. Electron. Syst. Mag.* **27**(1), 25–36 (2012)
7. Lischi, S., et al.: X-Band compact low cost multi-channel radar prototype for short range high resolution 3D-InSAR, EURAD Conference 2014, Rome, Italy
8. Saponara, S., et al.: Architectural exploration and design of time-interleaved SAR arrays for low-power and high speed A/D converters. *IEICE Trans. Electron.* **E92-C**, 843–851 (2009)

Chapter 19

Healthcare System for Non-invasive Fall Detection in Indoor Environment

Marco Mercuri, Carmine Garripoli, Peter Karsmakers, Ping Jack Soh, Guy A.E. Vandenbosch, Calogero Pace, Paul Leroux and Dominique Schreurs

Abstract Fall incidents and the sustained injuries represent the main causes of accidents for elderly people, and also the third cause of chronic disability. The rapid detection of a fall event can reduce the mortality risk, avoiding also the aggravation of injuries. In this paper an embedded healthcare system based on a microwave radar is presented. A Continuous Wave (CW) Doppler radar is used to detect the changes in speed of different persons experienced during daily activities, namely falling and normal/random movements. The resulted speed signals are then processed in real-time by a digital signal processor (DSP) in order to detect fall incidents. Experimental results, conducted on real human volunteers in a real room

M. Mercuri · C. Garripoli · P. Karsmakers · P.J. Soh · G.A.E. Vandenbosch · P. Leroux · D. Schreurs

Department of Electrical Engineering, KU Leuven, Kasteelpark Arenberg, 10 bus 2444, 3001 Heverlee (Leuven), Belgium
e-mail: Marco.Mercuri@esat.kuleuven.be

C. Garripoli
e-mail: Garripoli.Carmine@gmail.com

P. Karsmakers
e-mail: Peter.Karsmakers@esat.kuleuven.be

P.J. Soh
e-mail: PingJack.Soh@esat.kuleuven.be

G.A.E. Vandenbosch
e-mail: Guy.Vandenbosch@esat.kuleuven.be

P. Leroux
e-mail: Paul.Leroux@esat.kuleuven.be

D. Schreurs
e-mail: Dominique.Schreurs@esat.kuleuven.be

C. Garripoli · C. Pace (✉)
Dipartimento di Informatica, Modellistica Elettronica e Sistemistica, Università della Calabria, Via Pietro Bucci 42/C, 87036 Rende (CS), Italy
e-mail: Calogero.Pace@unical.it

setting, have shown a success rate of 100 % in detecting fall events. Moreover, no false positives have been reported.

Keywords Fall detection • Health monitoring • LS-SVM • Movement classification • Radar remote sensing • Telehealth systems

19.1 Introduction

The elderly population has been steadily increasing worldwide. This situation, together with the shortage of nursing homes and the natural desire to stay at home, has resulted in a growing need for healthcare approaches that emphasize routine long-term monitoring in the home environment. Elderly people who live alone are usually exposed to health risks that in some cases may cause fatality. In fact, fall incident among the elderly is considered one of the major problems worldwide, and often result in serious physical and psychological consequences [1]. Research pointed out that 30–45 % of the persons older than 60 years fall at least once a year. People who experience a fall event at home, and remain on the ground for an hour or more, may suffer from many medical complications, such as dehydration, internal bleeding, and cooling, and half of them die within 6 months [2]. The delay in hospitalization increases mortality risk. Studies have shown that the longer the person lies on the floor, the poorer is the outcome of medical intervention [3, 4]. For that reason, it is imperative to detect falls as soon as they occur such that immediate assistance may be provided.

Current health monitoring systems are based on necklace or wristwatch with a button that is activated by the patient in case of an accident. However, in emergency situations, this imposes an important risk factor. In fact, the person may forget to wear the device, or likely may no longer be able to press the button. The ideal solution is, therefore, a contactless approach that avoids the need for actions by the elderly person. Systems under investigation in the latter category are based on video cameras, floor vibration, and acoustic sensors. In the case of the video camera method, researchers are currently trying to address challenges related to low light, field of view, and image processing, but also privacy is a concern [5]. Floor vibration and acoustic sensors have limited success due to the environmental interference and background noise [6].

Due to the disadvantages of existing fall detection technologies, there is a need for further solutions. An alternative approach based on radar techniques has been demonstrated by the authors [7, 8]. The system uses a machine learning technique to distinguish fall events from normal movements as described in [9].

In this paper, an embedded healthcare system based on microwave radar measurements is described. As opposed to [8], a digital signal processor (DSP) platform has been used to process the motioning signals in order to detect fall emergencies in real-time.

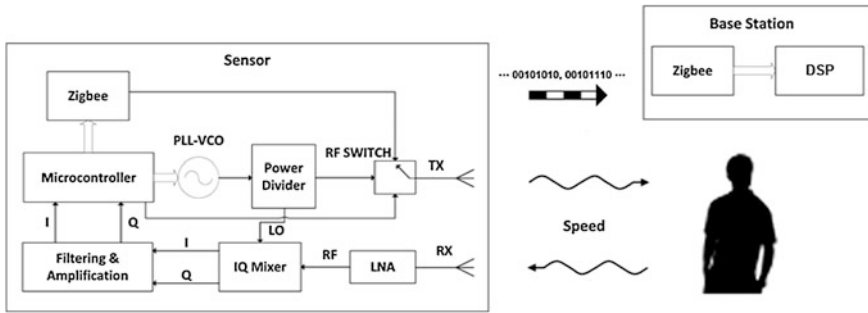


Fig. 19.1 Simplified block diagram of the embedded healthcare system

The embedded healthcare system and the data processing technique used to process the monitoring signals are introduced in Sect. 19.2. The implementation of the data processing technique by means of a DSP is detailed in Sect. 19.3. Experimental results are shown in Sect. 19.4.

19.2 Embedded Healthcare System

The embedded healthcare system used to design the real-time fall detector has been described by the authors in [8]. It consists of a sensor, combining both radar and wireless communications features, and a base station for data processing (Fig. 19.1). The sensor integrates a radar module, a Zigbee module, and a microcontroller, while the base station consists of a Zigbee module and the TMS320C6678 DSP platform.

A Continuous Wave (CW) waveform at 5.8 GHz is generated and used to detect the speed signals produced by the test persons during daily activities, namely falling and normal/random movement. The resulting baseband signals are digitized and transmitted to a base station to be processed.

A movement classification based on a Least Squares Support Vector Machines (LS-SVM) approach combined with Global Alignment (GA) kernel [9] is applied to analyze the digitized baseband speed signals in order to distinguish falls from other movements. The technique aims at assessing the changes in speed experienced during a fall or a normal movement. During a fall, in fact, the speed continuously increases until the sudden moment when the movement stops abruptly. During a normal movement, the Doppler signal experiences a controlled movement. More precisely, while a person is sitting down, the speed first gradually increases, and then decreases to a smooth stop, whereas, during a walk, the speed is quite constant over time.

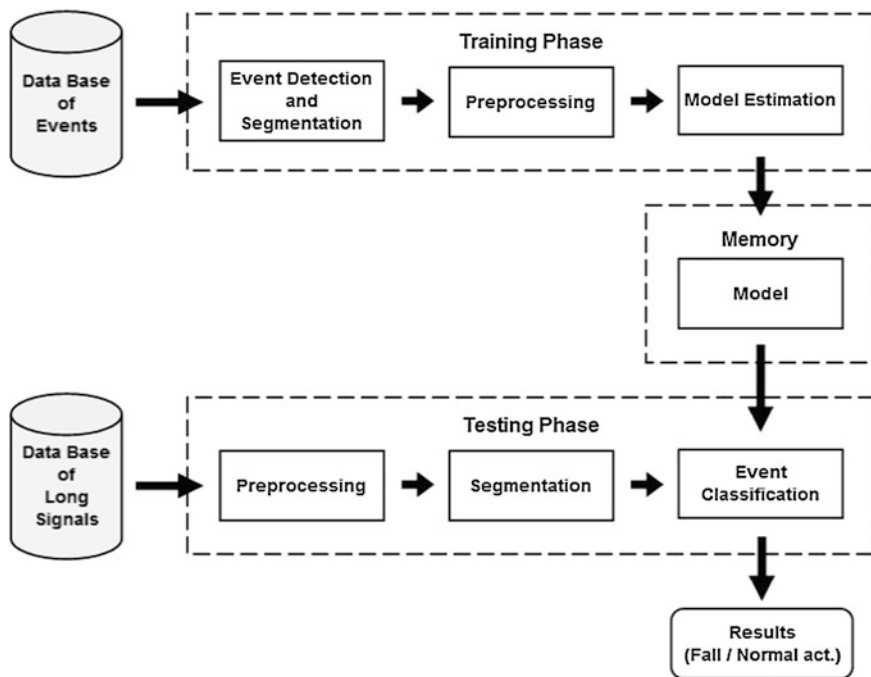


Fig. 19.2 Block diagram of the implemented classification technique

The developed algorithm consists of two stages of data analysis, namely the training phase and the testing phase (Fig. 19.2). Both phases use the digitized speed signal as input.

19.2.1 Training Phase

The training phase consists of activity detection and segmentation, feature extraction, feature selection, and model estimation. The training activities are divided in two main groups, namely fall events and normal movements. These acquired activities are used to build a data set. However, before learning a model, each activity is grouped in a segment of 2 s, considered sufficient to cover the details of the activities and mainly the fall event. This operation of segmentation consists in the detection of the activity energy's peak and in cutting the signal around this peak. Given such segments, the data is preprocessed, namely it is first standardized, such that each dimension has zero mean and unit standard deviation, and then transformed using the Short Time Fast Fourier Transform (STFT) from which only the magnitude spectrum is retained. Prior to the learning phase, the data is again

standardized. Once the learning process is finalized, the model is created and stored in a memory to be used in the validation stage.

19.2.2 Testing Phase

The algorithm performed in this phase presents a structure similar to the data processing of the training phase (Fig. 19.2). However, the main difference lies in the segmentation stage where the sliding window principle is applied due to the fact that the starting and ending points of the activities are unknown. The size of the sliding window has been fixed to 2 s in order to be consistent with the length of the activities' segments in the training phase, while the overlap among windows is fixed to 95 %.

19.3 Classification Implementation

In order to process a continuous stream of radar signals consisting of multiple activities invoked at unknown instants, the sliding window principle should be applied continuously to the received data. The sensor node transmits to the base station the speed signals every 100 ms. Therefore, every time a new frame is available, the relative samples are concatenated with the last 1.9 s. of the previous signals to create a window of 2 s and to have an overlap of 95 %. This large overlap is used to improve the performance of the system. In fact, a larger overlap involves a higher number of classifications such that a fall event will be considered over multiple windows, making the system much more immune to the noise that could generate a false positive in a single window classification. Finally, each segment is preprocessed and classified.

A custom parallelization technique involving the 8 cores of the TMS320C6678 DSP has been implemented to perform real-time classification. The strict requirement is that a segment classification must be performed before a new Zigbee communication occurs, that means in a time shorter than 100 ms. The pseudo-code of the classification is shown in Fig. 19.3. The core of the algorithm is to determine the Kernel array K , whose elements are calculated by the function $computeGAK(Test, Training)$, that is an off-the-shelf C code that computes the Global Alignment Kernel for the LS-SVM algorithm. Finally, the function $f(K)$ returns a value that, compared to a threshold, determines whether the segment contains fall or normal movement data. For each iteration of the loop, 8 elements of the Kernel matrix K are computed at the same time by the 8 cores of the DSP. Each core invokes the function $computeGAK()$ independently.


```

for (j = 0 to Ntrain - 1)
  core0 → Kj = computeGAK(Test, Trainingj);
  core1 → Kj+1 = computeGAK(Test, Trainingj+1);
  core2 → Kj+2 = computeGAK(Test, Trainingj+2);
  core3 → Kj+3 = computeGAK(Test, Trainingj+3);
  core4 → Kj+4 = computeGAK(Test, Trainingj+4);
  core5 → Kj+5 = computeGAK(Test, Trainingj+5);
  core6 → Kj+6 = computeGAK(Test, Trainingj+6);
  core7 → Kj+7 = computeGAK(Test, Trainingj+7);
  j = j+8;
end for;
val = f(K) = K·α + b;
if ( val < threshold )
  then fall event;
  else normal movement;

```

Fig. 19.3 Classification pseudo-code. N_{train} represents the number of matrices in the Training structure. Training and Test are three-dimensional matrices resulting respectively from the model estimation and the preprocessing in the testing phase. The vector α and the constant b are variables estimated in the training phase

19.4 Experimental Results

A training set containing 80 activities executed by a single test person is used to estimate the activity classification model. The models have consequently been tested for 3 h on 3 different test persons, for a total of 36 signals of 5 min each. Each of these signals was acquired with a single volunteer in the room at a time. He had not contributed to the training model. The subjects were allowed to mimic typical domestic situations without any restriction in their movements. Moreover, each signal contained only one fall event invoked in a random instant.

The success rate of the system was calculated as the percentage of detected falls. The results have indicated a success rate, in detecting fall incidents in real-time, of 100 %, without reporting any false positives. Figure 19.4 shows the result of the

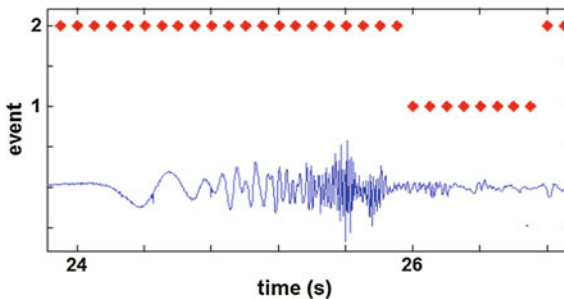


Fig. 19.4 Classification results of a small portion of a signal containing a fall event. In this example, the results of the classification are sent to Matlab for plotting. The fall is labelled as “1” while the normal movement as “2”

classification on a small portion of a signal containing fall event invoked at about 24 s. Each dot represents the class where a window of 2 s. of signal has been assigned. The event was classified as a fall for eight consecutive windows while the alarm was activated after the third. Since the time to classify a window is 16 ms, the total time to detect a fall event is about 316 ms.

19.5 Conclusion

In this paper, an embedded healthcare system for non-invasive fall detection in in-door environment has been described. It consists of a radar sensor and a base station for data processing. The implementation of the fall detection algorithm by means of a DSP platform has been also presented. Experimental results conducted with human subjects have shown a success rate, in detecting fall incidents in real-time, of 100 %, without reporting any false positives, with a maximum delay of about 316 ms.

The system is inline with the growing need for home health care applications and supervision technology for elderly people living at home. Next step is to integrate multiple sensors in a wireless sensor network in order to cover a whole room and also to perform indoor positioning.

Acknowledgments This work was supported by FWO-Flanders, KU Leuven GOA Project, and the Hercules Foundation.

References

1. Haentjens, P., Magaziner, J., et al.: Excess mortality after hip fracture among older women and men. *Ann. Intern. Med.* **152**, 380–390 (2010)
2. Lord, S.R., Sherrington, C., Menz, H.B.: *Falls in Older People: Risk Factors and Strategies for Prevention*. Cambridge University Press, Cambridge (2007)
3. Gurley, R.J., Lum, N., Sande, M., Lo, B., Katz, M.H.: Persons found in their homes helpless or dead. *N. Engl. J. Med.* **334**, 1710–1716 (1996)
4. Moran, C.G., Wenn, R.T., Sikand, M., Taylor, A.M.: Early mortality after hip fracture: is delay before surgery important. *J. Bone Joint Surg.* **87**, 483–489 (2005)
5. Yu, M., Naqvi, S.M., Chambers, J.: A robust fall detection system for elderly in a smart room. In: *International Conference on Acoustics Speech and Signal Processing*, pp. 1666–1669. Dallas, USA, March (2010)
6. Zigel, Y., Litvak, D., Gannot, I.: A method for automatic fall detection of elderly people using floor vibrations and sound—proof of concept on human mimicking doll falls. *Trans. Biomed. Eng.* **56**(12), 2858–2867 (2009)
7. Mercuri, M., Schreurs, D., Leroux, P.: Optimised waveform design for radar sensor aimed at contactless health monitoring. *IET Electron. Lett.* **48**(20), 1255–1257 (2012)

8. Mercuri, M., Soh, P., Pandey, G., Karsmakers, P., Vandenbosch, G.A.E., Leroux, P., Schreurs, D.: Analysis of an indoor biomedical radar-based system for health monitoring. *IEEE Trans. Microw. Theory Tech.* **61**(5), 2061–2068 (2013)
9. Karsmakers, P., Croonenborghs, T., Mercuri, M., Schreurs, D., Leroux, P.: Automatic in-door fall detection based on microwave radar measurements. In *Proceedings of the European Radar Conference, Amsterdam, The Netherlands, 2 Oct 31-Nov 2012*, pp. 202–205

Chapter 20

Analysis of Spread-Spectrum Clocking Modulations Under Synchronization Timing Constraint

Davide De Caro, Michele De Martino, Nicola Petra
and Antonio G.M. Strollo

Abstract Spread spectrum clocking slowly sweeps clock frequency of a digital system to reduce the Electromagnetic Interference (EMI). In a digital system-on-chip there can be subsystems where clock spreading is not allowed. This paper analyzes the performances achievable by spread spectrum clocking when a constraint is imposed to easily synchronize clock modulated and unmodulated subsystems. It is shown that the best modulation gain (7.7 dB) is achieved by using optimized discontinuous modulation.

Keywords Dithered clock · Electromagnetic interference (EMI) · Frequency modulation · Modulating waveform · Spread-spectrum clock

20.1 Introduction

Spread-Spectrum clocking [1] is an established approach to reduce the Electromagnetic Interference (EMI) in a number of applications [2–12]. In this technique the clock frequency is slowly swept (modulated) in a given frequency range with a predetermined modulation waveform to spread the energy associated to each clock harmonic over a certain bandwidth.

The peak power level reduction (modulation gain) that can be obtained by spread spectrum technique depends on modulation parameters (frequency deviation Δf , modulation frequency f_m) and on modulation waveform. According to Carson's rule the modulation gain mainly increases by increasing the ratio $\Delta f/f_m$. The EMI measurement standards [13–15] prescribe precise procedures to evaluate the peak power level of frequency modulated waveforms, requiring the analysis of the signal with a spectrum analyzer in swept frequency mode with a prescribed resolution

D. De Caro (✉) · M. De Martino · N. Petra · A.G.M. Strollo
Department of Electrical Engineering, Information Technology University of Napoli
“Federico II”, Via Claudio 21, I-80125 Naples, Italy
e-mail: dadeclaro@unina.it

bandwidth (RBW) and peak type detector. A consequence of this procedure is that the modulation gain depends in addition from the ratio f_m/RBW .

In [1] a simple Fourier series is used to study the spectrum of frequency modulated signal. In many cases this approach results in a remarkable difference with experimental data since Fourier analysis does not take into account the combined effect the spectrum analyzer RBW and peak type detector, required by the EMI standards. The most common modulation waveform is the triangular waveform. When analyzed by using Fourier series, this waveform provides a flat spectrum, therefore achieving the best modulation gain for given modulation frequency and frequency deviation. Unfortunately this property is lost when the effect of the spectrum analyzer RBW and detector is taken into account. A waveform providing a modulation gain higher than the triangular waveform, when the spectrum analyzer RBW and detector is considered, is obtained empirically in [16].

In [17, 18] Matsumoto et al. propose an approach which allows to study analytically the effect of spectrum analyzer RBW and peak-type detector. This work derives analytically the optimal modulation waveform that maximizes the modulation gain, for prescribed modulation parameters, under the hypothesis $f_m < \text{RBW}$. The obtained waveform results similar to the empirical solution proposed in [16].

The analysis of Matsumoto et al. is extended to discontinuous waveforms in [19]. Recent developments in all-digital spread spectrum clock generators [2–5] (SSCG) nowadays allow to implement SSCG where the output frequency can be varied without any limitation in speed by making very interesting the investigations on discontinuous modulations. The theoretical approach proposed in [19] shows that sawtooth modulation can achieve a modulation gain higher than continuous modulations like triangular and Matsumoto et al. [17, 18] modulations. The paper in addition derives the optimal discontinuous modulation profile which maximizes the modulation gain for prescribed modulation parameters, including spectrum analyzer effects.

In a digital system-on-chip there can be subsystems where clock spreading is not allowed (e.g. digital-to-analog and analog-to-digital converters, critical peripherals, ...). The paper of Zhou and Dehaene [20] face the problem of data interchange between two subsystems, one with frequency spreading and one clocked by a signal without spreading. As show in Fig. 20.1 the two subsystems can easily communicate through two registers if a constraint is imposed on modulated clock signal. As highlighted in the figure the modulated clock needs not to exceed a 2π phase difference with respect to unmodulated clock.

Unfortunately the paper [20] analyzes the achievable modulation gains by using a simple Fourier analysis, without considering spectrum analyzer effects. This paper analyzes the performances achievable by using both simple (Triangular, Sawtooth) and optimized (Matsumoto et al., optimized discontinuous) modulations under the above timing synchronization constraint by including spectrum analyzer effects.

The paper is organized as follows. The following Sect. 20.2 recalls simple modulations and formalize the synchronization constraint. The Sect. 20.3 presents simulation results.

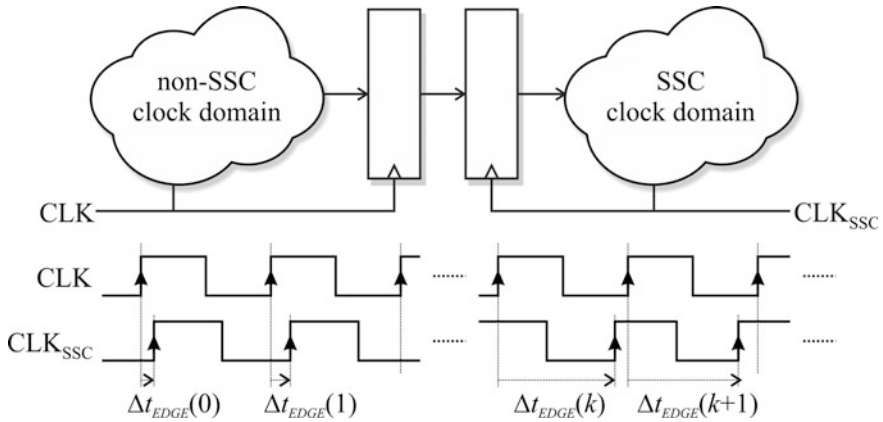


Fig. 20.1 Synchronization between two subsystems, one clocked by a clock signal without frequency spreading and one clocked by a frequency spreaded signal

20.2 Synchronization Constraint for Triangular and Sawtooth Modulations

The frequency-spreaded clock signal $u_s(t)$ can be written as:

$$u_s(t) = \sum_{k=-\infty}^{+\infty} \frac{1}{2} I_k(t) = \frac{I_{0,0}}{2} + \sum_{k=1}^{+\infty} \text{Re}[I_k(t)] \tag{20.1}$$

where:

$$I_k(t) = I_{0,k} \cdot \exp \left[j2\pi k f_0 \left(t + \frac{\delta}{f_m} \int_{-\infty}^{f_m t} V(\tau) d\tau \right) \right] \tag{20.2}$$

and the harmonics amplitudes $I_{0,k}$, for a square wave of amplitude A , are given by:

$$I_{0,k} = 2A \frac{\sin(k\pi/2)}{k\pi} \tag{20.3}$$

The instantaneous frequency $f(t)$ of this waveform can be easily computed as:

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \frac{d}{dt} \left[2\pi f_0 \left(t + \frac{\delta}{f_m} \int_{-\infty}^{f_m t} V(\tau) d\tau \right) \right] = \\ &= f_0 (1 + \delta V(f_m t)) \end{aligned} \tag{20.4}$$

This equation shows that f_m and δ correspond respectively to the modulation frequency and to the relative frequency deviation of the frequency-spreaded clock waveform:

$$\delta = \frac{\Delta f}{f_o} \quad (20.5)$$

In this paper we will focus on the spectrum of a the first frequency modulated harmonic $I_1(t)$. A similar reasoning will apply to other harmonics.

The modulation gain (*Gain*) is defined as the ratio between the amplitude of the un modulated harmonic ($I_{0,1}$) and the peak value of the spectrum $S(f_c)$ of a frequency modulated harmonic $I_1(t)$:

$$Gain \equiv \frac{I_{0,1}}{\max[S(f_c)]} \quad (20.6)$$

By looking to Eq. (20.2), we can write the time $t_{EDGE}(n)$ of the n th rising edge of the modulated clock waveform as the solution of the following equation:

$$2\pi f_o \left(t_{EDGE}(n) + \frac{\delta}{f_m} \int_{-\infty}^{f_m t_{EDGE}(n)} V(\tau) d\tau \right) = 2\pi n \quad (20.7)$$

that is:

$$t_{EDGE}(n) + \frac{\delta}{f_m} \int_{-\infty}^{f_m t_{EDGE}(n)} V(\tau) d\tau = nT_o \quad (20.8)$$

where $T_o = 1/f_o$ is the period of the unmodulated clock signal.

Therefore the timing difference between frequency-spreaded clock signal and unmodulated clock signal at frequency f_o , shown in Fig. 20.1, is given by:

$$\Delta t_{EDGE}(n) = t_{EDGE}(n) - nT_o = -\frac{\delta}{f_m} \int_{-\infty}^{f_m t_{EDGE}(n)} V(\tau) d\tau \quad (20.9)$$

The analysis of Fig. 20.1 shows that the synchronization between the two subsystems is correct when the peak-to-peak variation of Δt_{EDGE} ($\Delta t_{EDGE, pk}$) is lower than one period:

$$\frac{\Delta t_{EDGE, pk}}{T_o} \leq 1 \quad (20.10)$$

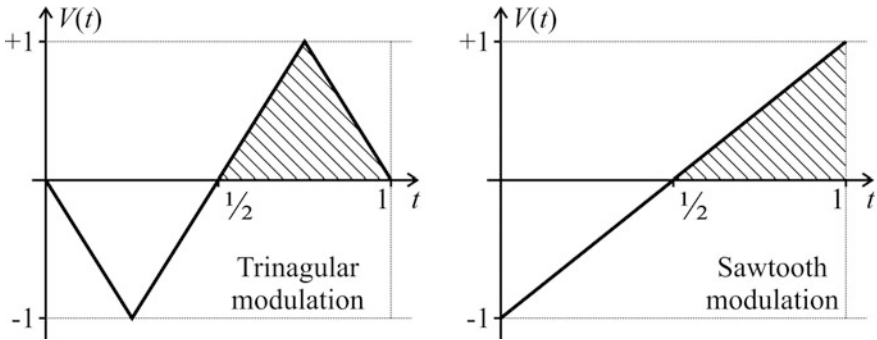


Fig. 20.2 Computation of the maximum value of the integral in (20.11) for triangular and sawtooth modulations

According to Eq. (20.9), $\Delta t_{EDGE\ pk}$ can be written as:

$$\Delta t_{EDGE\ pk} = \frac{\delta}{f_m} \left(\max_t \int_{-\infty}^t V(\tau) d\tau - \min_t \int_{-\infty}^t V(\tau) d\tau \right) \tag{20.11}$$

The Fig. 20.2 shows that, both for triangular modulation and for sawtooth modulation, the minimum value of the integral in (20.11) is 0 and the maximum value is 1/4.

Therefore, for this modulations:

$$\frac{\Delta t_{EDGE\ pk}}{T_o} = \frac{\Delta f}{4f_m} \tag{20.12}$$

As well known the modulation gain increases by increasing the ratio $\Delta f/f_m$. Equation (20.12) joined to inequality (20.10) show that this synchronization constraint impose a limitation on the maximum achievable modulation gain. The maximum modulation gain can be achieved in the limit case in which $\Delta t_{EDGE\ pk} = T_o$, which according to (20.12) corresponds to choose $\Delta f/f_m = 4$. Clearly the achieved performances will depend also on the modulation frequency, since, due to spectrum analyzer effects, the modulation gain depends on the ratio f_m/RBW (see [17–19]).

20.3 Simulation Results

In order to quantify the achievable modulation gains by the different modulations under the synchronization constraint (20.10) we performed a simulation by considering the filtering effect of the spectrum analyzer. The simulation results are shown in Fig. 20.3 as a function of the modulation frequency f_m .

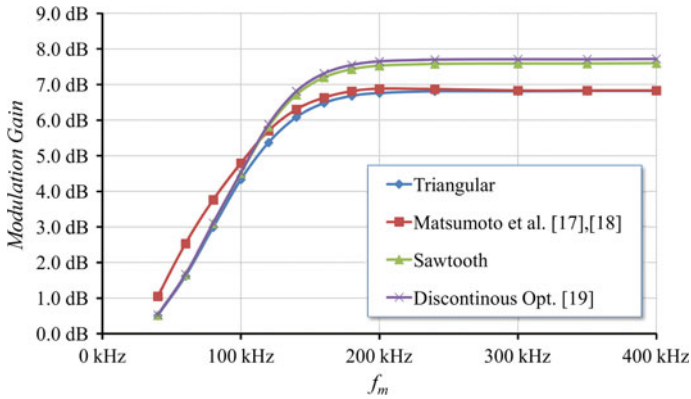


Fig. 20.3 Modulation gain simulated considering spectrum analyzer filtering (RBW = 100 kHz) under synchronization constraint (20.10)

The data reported in the figure show that, for each modulation, for $f_m \ll \text{RBW}$, the modulation gain increases by increasing the modulation frequency. In this region, in fact, the modulation gain is proportional to $\sqrt{\Delta f/f_m} \cdot f_m/\text{RBW}$ (see [17–19]) and, due to the second term of this expression, increases by increasing f_m for a constant $\Delta f/f_m$. When f_m becomes larger than RBW, we start entering into Fourier region where the modulation gain is proportional to only $\sqrt{\Delta f/f_m}$. In fact the curves saturate to a value independent from f_m . The data shows that the best modulation gain is achieved by using optimal discontinuous frequency modulation, and that this modulation allows to obtain of modulation gain of about 7.7 dB under the synchronization constraint considered in this paper.

References

1. Hardin, K.B., Fessler, J.T., Bush, D.R.: Spread spectrum clock generation for the reduction of radiated emissions. In: Proceedings of the IEEE International Symposium on Electromagnetic Compatibility, pp. 227–231 (1994)
2. Dampousse, S., Ouici, K., Rizki, A., Mallinson, M.: All-digital Spread Spectrum Clock Generator for EMI reduction. *IEEE J. Solid-State Cir.* **42**(1), 145–150 (2007)
3. Ebuchi, T., Komatsu, Y., Okamoto, T., Arima, Y., Yamada, Y., Sogawa, K., Okamoto, K., Morie, T., Hirata, T., Dosho, S., Yoshikawa, T.: A 125–1250 MHz process-independent adaptive bandwidth spread spectrum clock generator with digital controlled self-calibration. *IEEE J. Solid-State Circ.* **44**(3), 763–774 (2009)
4. De Caro, D.: Glitch-free NAND-based digitally controlled delay-lines. *IEEE Trans. Very Large Scale Integr. Syst.* **21**(1), 55–66 (2013)
5. De Caro, D., Romani, C.A., Petra, N., Strollo, A.G.M., Parrella, C.: A 1.27 GHz, all-digital spread spectrum clock generator/synthesizer in 65 nm CMOS. *IEEE J. Solid-State Circ.* **45**(5), 1048–1060 (2010)

6. Hwang, S., Song, M., Kwak, Y.H., Jung, I., Kim, C.: A 3.5 GHz spread spectrum clock generator with a memoryless Newton-Raphson modulation profile. *IEEE J. Solid-State Circ.* **47**(5), 1199–1208 (2012)
7. Lin, S.Y., Liu, S.I.: A 1.5 GHz all-digital spread-spectrum clock generator. *IEEE J. Solid-State Circ.* **44**(11), 3111–3119 (2009)
8. Pareschi, F., Setti, G., Rovatti, R.: A 3-GHz Serial ATA spread spectrum clock generator employing a chaotic PAM modulation. *IEEE Trans. Circ. Syst. I Regul. Pap.* **57**(10), 2577–2587 (2010)
9. Lin, F., Chen, D.Y.: Reduction of power supply EMI emission by switching frequency modulation. *IEEE Trans. Power Electron.* **9**(1), 132–137 (1994)
10. Tse, K.K., Chung, H.S.-H., Huo, S.Y., So, H.C.: Analysis and spectral characteristics of a spread spectrum technique for conducted EMI suppression. *IEEE Trans. Power Electron.* **15**(2), 399–410 (2000)
11. Balcells, J., Santolaria, A., Orlandi, A., Gonzalez, D., Gago, J.: EMI reduction in switched power converters using frequency modulation techniques. *IEEE Trans. Electromagn. Compat.* **47**(3), 569–576 (2005)
12. Tse, K.K., Ng, R.W.-M., Chung, H.S.-H., Hui S.Y.R.: An evaluation of the spectral characteristics of switching converters with chaotic carrier frequency modulation. In: *IEEE Trans. Industr. Electron.* **50**(1), 171–182 (2003)
13. FCC 47 CFR Part 15, Radio frequency devices (2008)
14. CISPR 22: Information technology equipment—radio disturbance characteristics—limits and methods of measurement (2003–2004)
15. ANSI C63.4: American national standard for methods of measurement of radio noise emissions from low voltage electrical and electronic equipment in the range of 9 kHz to 40 GHz (2003)
16. Hardin, K., Oglesbee, R.A., Fisher, F.: Investigation into the interference potential of spread-spectrum clock generation to broadband digital communications. *IEEE Trans. Electromagn. Compat.* **45**(1), 10–21 (2003)
17. Matsumoto, Y., Fujii, K., Sugiura, A.: An analytical method for determining the optimal modulating waveform for dithered clock generation. *IEEE Trans. Electromagn. Compat.* **47**(3), 577–584 (2005)
18. Matsumoto, Y., Fujii, K., Sugiura, A.: Estimating the amplitude reduction of clock harmonics due to frequency modulation. *IEEE Trans. Electromagn. Compat.* **48**(4), 734–741 (2006)
19. De Caro, D.: Optimal discontinuous frequency modulation for spread-spectrum clocking. *IEEE Trans. Electromagn. Compat.* **55**(5), 891–900 (2013)
20. Zhou, J., Dehaene, W.: A synchronization-free spread spectrum clock generation technique for automotive applications. *IEEE Trans. Electromagn. Compat.* **53**(1), 169–177 (2011)

Chapter 21

Towards a Frequency Domain Processor for Real-Time SIFT-based Filtering

Giorgio Lopez, Ettore Napoli and Antonio G.M. Strollo

Abstract The Scale Invariant Feature Transform (SIFT) extracts relevant features from images and video frames. The extracted features are robust against luminance variations, geometrical transformations, and image resolution. Due to its performances, the SIFT algorithm is of great importance in fields such as object recognition, content retrieval from image databases, robotic navigation, and gesture recognition. Main drawback of the SIFT algorithm is the high computational complexity. This paper presents the development of a hardware filtering accelerator for the implementation of SIFT-based visual search. The accelerator works in the frequency domain, operating on a block-by-block basis. This enables to work faithfully to the original Scale-Space theory, which employs non-separable Laplacian of Gaussian (LoG) filters. The targeted throughput is of ~ 20 fps, making the coprocessor suitable for real time processing.

Keywords SIFT · Computer Vision · FPGA

21.1 Introduction: SIFT, CVDS, and the TMuC

The extensive use of images and video streams in nowadays applications requires the development of algorithms that conduct automatic search of information in a frame. One of the most appreciated algorithm is the Scale Invariant Feature Transform (SIFT) proposed by D.G. Lowe [1].

G. Lopez (✉) · E. Napoli · A.G.M. Strollo
Department of Electrical and Information Technology Engineering,
University of Napoli Federico II, Napoli, Italy
e-mail: giorgio.lopez@unina.it

E. Napoli
e-mail: etnapoli@unina.it

A.G.M. Strollo
e-mail: astrollo@unina.it

The circuit proposed in this paper is part of a contribution to the CDVS standardization project (Compact Descriptors for Visual Search). This project wants to produce a standard for the extraction and interchange of image-extracted information. The process is helped by the adoption of an Evaluation Framework (the Test Model under Consideration—TMuC) which is used for simulating and validating proposals made by the members of the standardization committee: the algorithm that is now considered by the CDVS committee is based on SIFT.

The SIFT algorithm, [1] detects a number of point of interests (*keypoints*, *KP*) from an input image that are then described by means of a vector known as the SIFT descriptor. Such descriptor is based on a statistical characterization of the gradients of the luminance for the pixels surrounding the KP itself.

SIFT identifies the KPs building a band-pass pyramid of images and characterizes the KPs (i.e. builds the descriptor) using a low-pass pyramid of images. The generation of the image pyramids is accomplished, in [1], by filtering the input image with a set of Gaussian kernels to produce the low-pass images and, by subtracting low-pass images to obtain the Difference of Gaussian (DoG) images. These images, as proven by Lowe, constitute an approximation of the scale-normalized Laplacian of Gaussian (LoG). The LoG is an operator for obtaining the band-pass pyramid of an image, discussed in [2] and proven in [3] to produce the most stable features when compared to a range of other operators.

During the process, an iterative downsampling is performed and the whole cascade filtering and difference process is repeated to produce the next octave, until the complete pyramids are generated.

In the subsequent algorithm steps, local extrema are searched for along the scales leading to a first set of KP candidate locations. For these locations a stability analysis is then carried out. Next, a first statistical characterization of image gradients around the locations is performed to assign one or more intrinsic orientations to the keypoints; finally, each KP is normalized with respect to the orientations and its SIFT descriptor is calculated.

This process, while providing the best performances proposed to date, is very demanding in terms of computational resources: as a consequence, various hardware approaches to SIFT elaboration in real time scenarios have been proposed.

21.2 State of the Art Approaches to SIFT Elaboration

Current scientific literature provides several approaches to real time SIFT elaboration: some of them, [4], leverage on General Purpose GPU processing power; others exploit multi-processor and/or multi-core systems, [5].

For embedded systems applications, where limited general purpose processing capabilities and low power consumption are major concerns, typical solutions include use of specialized hardware, either in form of ASICs or FPGAs deployment. In [6], a mixed hardware/soft-processor environment is proposed, while in [7], a FPGA and a DSP processor are jointly used to speed up the overall processing.

Other recent works, [8], propose all-hardware solutions that operate in the space domain and adopt DoG filters. This paper presents a study towards the implementation of a SIFT keypoint detector that operates in the frequency domain and adopts LoG filters.

21.3 Frequency Based Approach and Block-Based LoG

In this study we explore the frequency domain approach to evaluate performance with respect to the TMuC standard (space domain) floating point implementation. Furthermore, this gives us the opportunity to utilize the LoG filters, which characterize the Scale Space theory image pyramid in its former formulation as given in [2], without incurring in the high computational costs related to a 2D convolution, which is necessary in space domain.

Processing the whole VGA image in the frequency domain would result in the need for large buffers (i.e. large enough to contain the whole Discrete Fourier Transform of the image, which would be composed by exactly 640×480 samples). This is a typical downside of operating in the frequency domain, which is less suitable for streaming applications, and a common solution is partitioning the frame into blocks. According to the approach proposed in [9], and taking into account the length of the finite impulse response of the filters which will be used, we choose the optimal block size.

This optimal choice is given by the minimization of the product of the time needed to perform analysis on a single block with the total number of blocks: if we denote the image width and height respectively by W and H , the block size by N and the maximum filter tail by L , the total number of blocks is:

$$N_{blocks} = \left\lceil \frac{W}{N - L + 1} \right\rceil \cdot \left\lceil \frac{H}{N - L + 1} \right\rceil. \quad (21.1)$$

The next point is determining the computational load for the elaboration of a single block. For each block we firstly perform N FFTs in the row direction and N FFTs in the column direction to obtain its frequency spectrum. Further, for each filter of the image pyramids, we perform N^2 multiplications. Finally, we perform N IFFTs in the row direction and N IFFTs in the column direction to return in the space domain. Since each FFT requires $N \log_2(N)$ operations, and the filter bank is composed by 8 filters coherently with the TMuC formulation of SIFT, we obtain:

$$N_{ops} = 9 \cdot (2N \cdot (N \log_2(N))) + 8N^2 \quad (21.2)$$

Limiting to block sizes that are powers of two, we obtain the computational loads shown in Table 21.1, which refer to VGA image resolution and maximum kernel size, L , equal to 33, as in the TMuC SIFT formulation. Table 21.1 indicates that operating on blocks of 128×128 pixels is the optimal choice.

Table 21.1 Computational loads for block sizes which are integer powers of two

Block size (pixels)	Total computational load (operations)
64	1.43×10^7
128	7.68×10^6
256	8.97×10^6
512	8.91×10^6

21.4 Proposed Filtering Processor Architecture

The proposed architecture is depicted in Fig. 21.1. The fundamental blocks of the processor are a 1D FFT unit which, iteratively used, transforms the block from the space domain to the frequency domain and vice versa, and a filter bank which allows for the calculation of both the LoG and the Gaussian-smoothed images needed by the SIFT algorithm to construct the keypoint descriptors.

The FFT unit is a mixed-radix dual-channel unit which operates following a decimation in frequency approach. Internal multiply operation are optimized in terms of speed and HW resource utilization exploiting the circuits and the design techniques proposed in [10–13]. The output from the FFT is not reordered inside the unit itself to avoid incurring in the associated resource utilization penalty: for this reason, a memory address generator issues to block buffers 1 and 2 the correct sequence of read/write addresses in order to allow the storage of the output data in the natural ordering. The memory address generator also allows to read and write the buffers in row/column order, which is needed to transform the rows and the columns separately via the 1D FFT unit.

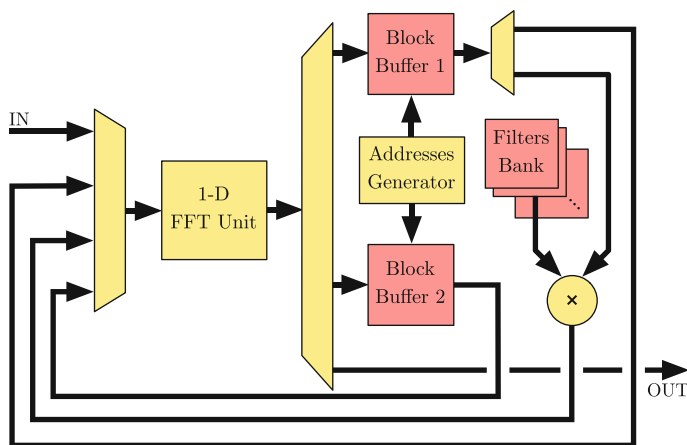


Fig. 21.1 Structure of the block processor: memory blocks are shown in red

The multiplexers at the sides of the FFT unit direct the data flow to allow an iterated use of the unit; we can distinguish 4 phases of operation, of which the last two are repeated for all the filters in the bank:

1. The block is fed from the circuit input to the FFT unit row-wise and the output is stored, half transformed, in Block Buffer 1.
2. The block is fetched column-wise from Buffer 1, transformed, and stored again in the buffer. This will now store the frequency spectrum of the block.
3. Samples of the frequency spectrum are multiplied with the corresponding samples of the filter, inversely transformed row-wise and then stored in Buffer 2, to keep the spectrum stored for processing by the other filters.
4. The final inverse FFT unit (in the column direction) is performed, and the result is fed to the circuit output.

21.5 Architecture Tuning and Experimental Results

Since the processor operates on fixed-point precision data, in order to obtain an acceptable fidelity with respect to floating point elaboration an accurate tuning of datapath and internal signal representation is needed, [14, 15]. In this case the aspects of the coprocessor that have been studied are mainly related to the intermediate data widths and lsb (least significant bit) weights of the input and output images. An exhaustive PSNR measurement with respect to the TMuC floating point elaboration has thus been conducted to find the best possible combination of resources allocations: considering that each FFT step produces an increment of $\log_2(N)$ bits in the dynamic range of the samples (in our case, being $N = 128$, we have an increment of 7 bits per FFT step) we have a set of constraints for our allocation problem. If we denote the datapath width by D , the weight of the lsb by l , the scaling factors after the first and second FFT by f and s , we have:

- input is given in 8 bit unsigned samples, which are extended to 9 bits to include sign information.
- it must be $9 + 2 \cdot 7 + l - (f + s) \leq D$ to avoid overflows or saturations, with the equality ensuring the maximum precision for the datapath width D .

From these constraints emerge a set of possible resources allocation schemes which result in the best combinations (for each of the given datapath widths) shown in Table 21.2.

Table 21.3 shows the implementation results for the 20 bit datapath version of the coprocessor on an Altera Stratix IV family FPGA.

Table 21.2 Best resources allocations for various datapath widths

Datapath Width (bits)	lsb Weight	1st Scaling Factor	2nd Scaling Factor	PSNR (dB)
14	2^{-5}	2^7	2^7	29.0
16	2^{-7}	2^7	2^7	38.3
18	2^{-9}	2^7	2^7	44.2
20	2^{-10}	2^6	2^7	45.6
22	2^{-11}	2^6	2^6	45.7
24	2^{-13}	2^6	2^6	45.7

Table 21.3 FPGA resources occupation for the 20 bit datapath version of the coprocessor

ALUTs	Registers	BRAM kbits	DSP Elements
5692	6004	2561	96

21.6 Conclusions

The proposed architecture constitutes an approach to SIFT elaboration in the frequency domain which obtains a PSNR of almost 50dB with respect to the TMuC's floating point elaboration with a small footprint with respect to FPGA deployment. The architecture is currently under development, with further stages to be introduced to complete the whole SIFT pipeline in a full-hardware environment.

References

1. Lowe, D.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **60**, 91–110 (2004)
2. Lindeberg, T.: Scale-space theory: a basic tool for analyzing structures at different scales. *J. Appl. Stat.* **21**, 225–270 (1994)
3. Mikołajczyk, K., Schmid, C.: A performance evaluation of local descriptors. In: *Proceedings. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol 2, pp. II-257. IEEE (2003)
4. Heymann, S., Frhlich, B., Medien, F., Mller, K., Wiegand, T.: SIFT implementation and optimization for general-purpose gpu. In: *WSCG 07 (2007)*
5. Zhang, Q., Chen, Y., Zhang, Y., Xu, Y.: SIFT implementation and optimization for multi-core systems. In: *IEEE International Symposium on Parallel and Distributed Processing, IPDPS 2008*, pp. 1–8. IEEE (2008)
6. Bonato, V., Marques, E., Constantinides, G.A.: A parallel hardware architecture for scale and rotation invariant feature detection. *IEEE Trans. Circuits Syst. Video Technol.* **18**, 1703–1712 (2008)
7. Zhong, S., Wang, J., Yan, L., Kang, L., Cao, Z.: A real-time embedded architecture for SIFT. *J. Syst. Archit.* **59**, 16–29 (2013)
8. Jiang, J., Li, X., Zhang, G.: SIFT hardware implementation for real-time image feature extraction. *IEEE Trans. Circuits Syst. Video Technol.* **24**, 1209–1220 (2014)

9. Hunt, B.: Minimizing the computation time for using the technique of sectioning for digital filtering of pictures. *IEEE Trans. Comput.* **100**, 1219–1222 (1972)
10. Garofalo, V., Petra, N., De Caro, D., Strollo, A., Napoli, E.: Low error truncated multipliers for DSP applications. *Proc. ICECS* **2008**, 29–32 (2008)
11. Garofalo, V., Coppola, M., De Caro, D., Napoli, E., Petra, N., Strollo, A.: A novel truncated squarer with linear compensation function. *ISCAS* **2010**, 4157–4160 (2010)
12. De Caro, D., Petra, N., Strollo, A., Tessitore, F., Napoli, E.: Fixed-width multipliers and multipliers-accumulators with min-max approximation error. *IEEE Trans. Circuits Syst. I Regul. Pap.* **60**, 2375–2388 (2013)
13. Petra, N., De Caro, D., Garofalo, V., Napoli, E., Strollo, A.: Truncated squarer with minimum mean-square error. *Microelectron. J.* **45**, 799–804 (2014)
14. Genovese, M., Napoli, E.: FPGA-based architecture for real time segmentation and denoising of HD video. *J. Real-Time Image Proc.* **8**, 389–401 (2013)
15. Genovese, M., Napoli, E.: ASIC and FPGA implementation of the gaussian mixture model algorithm for real-time segmentation of high definition video. *IEEE Trans. VLSI Syst.* **22**, 537–547 (2014)

Chapter 22

A Real-Time FPGA-based Solution for Binary Image Thinning

Daniele Davalle, Bernardino Carnevale, Sergio Saponara, Luca Fanucci and Pierangelo Terreni

Abstract This paper presents an optimized FPGA implementation for real-time binary image thinning algorithm. The reference thinning technique is based on iterated comparisons with a set of eight 3×3 binary masks. In the proposed architecture, the processing logic and the internal memory are implemented in a way that the mask matching on each 3×18 image segment can be done in parallel within a single clock cycle. This optimization entails a reduction of more than one order of magnitude in terms of execution cycles with respect to the original algorithm. The algorithm was implemented on an ALTERA Stratix II EP2S30 FPGA. The resource occupation of the thinning block and the dedicated memory controllers is 4% at 100 MHz clock frequency. The proposed solution produces the output in 0.03 s on a standard PAL 720×576 , allowing for further real-time processing.

Keywords Image thinning · Real-time video processing · FPGA

22.1 Introduction

Image thinning algorithms reduce objects to a simple set of lines representing approximately their medial axes [1]. A thinning algorithm can be thus a simple way to calculate binary image skeletons. The skeleton of an object A can be

D. Davalle (✉) · B. Carnevale · S. Saponara · L. Fanucci · P. Terreni
Department of Information Engineering, University of Pisa, Via Caruso 16, 56122 Pisa, Italy
e-mail: daniele.davalle@for.unipi.it

B. Carnevale
e-mail: berardino.carnevale@for.unipi.it

S. Saponara
e-mail: sergio.saponara@iet.unipi.it

L. Fanucci
e-mail: luca.fanucci@iet.unipi.it

P. Terreni
e-mail: pierangelo.terreni@iet.unipi.it

mathematically defined as the locus of centres of circles touching the boundary of A in more than one point.

There are several applications of this technique such as pattern recognition, biological shape analysis and motion tracking. Thinning algorithms are also well known to be computationally intensive and time consuming, therefore a hardware accelerator is often required to achieve real-time performance.

Several works are related to binary image thinning and are based on different algorithms.

Pattern-based solutions were presented in [2–4]. These algorithms are based on a sequence of raster scan iterations over the image, in which pattern matching computations are applied. The result of each thinning step is related to the previous one and the number of iterations is data dependent. The process stops its execution only when no more pixel can be removed. These techniques are also known as hit-and-miss thinning algorithms.

Another approach was introduced in [5, 6]. These algorithms are based on the Euclidean Distance Transform (EDT), which represents the first step to be executed before the image thinning algorithm. EDT assigns a number to each pixel of the object denoting the distance between the pixel itself and the background. Some examples of EDT algorithms can be found in [7–9]. After applying EDT the labelled image is eroded in order to obtain the thinned one.

Finally, a hardware implementation on FPGA was proposed in [10], which is slower and the elaborated image is smaller with respect to our solution.

This work is focused on hardware parallelization of a pattern-based thinning technique.

The paper is organized as follows. Section 22.2 describes the reference algorithm used as starting point of this work. The proposed hardware optimization for parallel processing is presented in Sect. 22.3. Results are given in Sect. 22.4 and conclusions are drawn in Sect. 22.5.

22.2 Algorithm Description

The proposed solution is functionally equivalent to a well-known literature approach [11]. We rearranged memory accesses and the data-flow in order to improve the execution time. The throughput increased by more than one order of magnitude with respect to the serial version of the algorithm.

Let us define a kernel of a given pixel as the pixel itself and its 3×3 neighbours. Thinning masks are applied to each image kernel in a raster fashion. 1-pixels are set to 0 if the mask matches the kernel. Figure 22.1 shows the thinning masks, ‘-’ sign is used for “don’t-care”.

The procedure described in [11] iterates until no more pixels can be removed, therefore the execution time is not known *a priori* and is data-dependent. In our HW implementation the number of iterations can be limited to have a known worst-case execution time, useful for real-time applications.

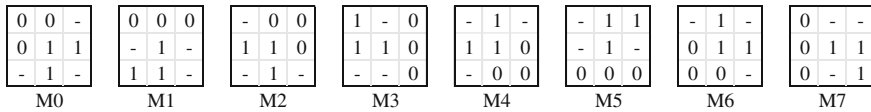


Fig. 22.1 Thinning algorithm masks

The pseudo-code is shown in Algorithm 22.1.

Algorithm 22.1 Image thinning

Input: B (binary image)

Output: S (thinned binary image)

```

1:  $H(r,c) = B(r,c) \forall r,c$ 
2: for  $n = 0$  to number of iterations - 1 do
3:   for  $i = 0$  to number of masks - 1 do
4:      $T(r,c) = 1 \forall r,c$ 
5:     for every pixel  $H(j,k)$  extracted in a raster fashion do
6:       if  $H(j,k) \neq 0$  then
7:         kernel = set of  $3 \times 3$  pixels from image S centred in  $j,k$ 
8:         if kernel =  $M_i$  then
9:            $T(j,k) = 0$ 
10:        end if
11:       end if
12:     end for
13:      $H(r,c) = H(r,c)$  and  $T(r,c) \forall r,c$ 
14:   end for
15: end for
16:  $S(r,c) = H(r,c) \forall r,c$ 

```

22.3 Hardware Architecture

The HW architecture of the thinning processing core is shown in Fig. 22.2.

A mask matching operation requires the 3×3 kernel of the pixel. In order to carry out the matching operation on P pixels in one clock cycle, P pixels and their 3×3 neighbourhood must be available in one clock cycle. To achieve this, the memory was organized in a way that P pixels of 3 contiguous rows are accessible in a single clock cycle. $P = 18$ was chosen as it is the maximum word size of M512 RAM blocks in Stratix II FPGA and leads to a good parallel trade-off.

The binary image is stored in 3 logically separated RAM blocks. The first memory block is used for storing rows satisfying $\{r \mid \text{mod}(r, 3) = 0\}$ where r is the row index. Similarly, $\{r \mid \text{mod}(r, 3) = 1\}$ are stored in the second memory block and

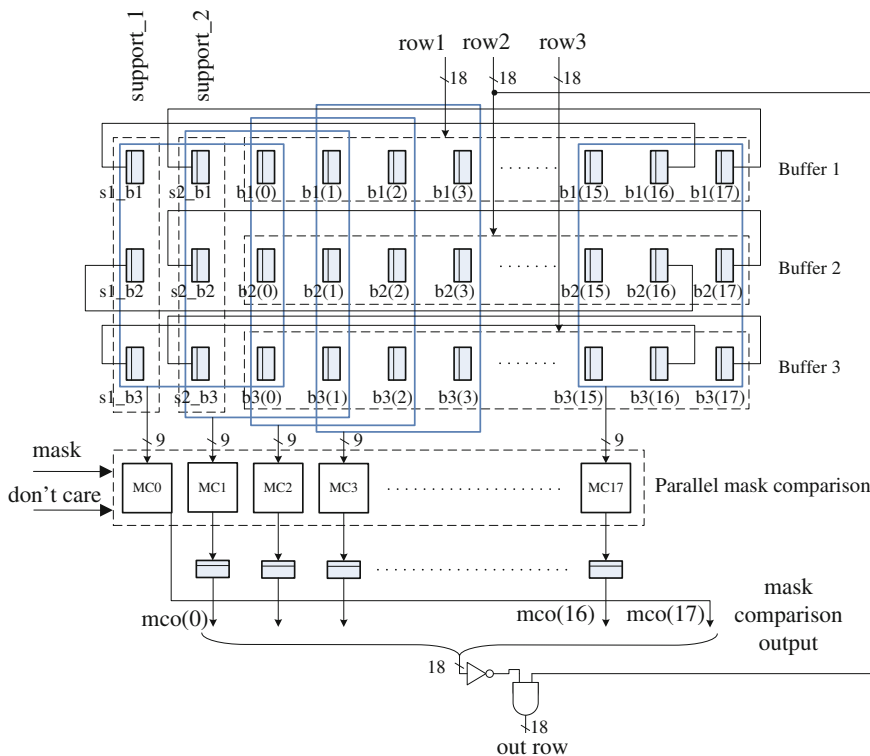


Fig. 22.2 HW architecture

$\{r \mid \text{mod}(r, 3) = 2\}$ in the third one. This “three-way interlaced” memory organization allows access to every 3×18 image segment in one clock cycle.

The mask matching operation is carried out by the Mask Comparison (MC) blocks in Fig. 22.2. Each 3×3 kernel is extracted from the buffered image segment and is compared with the current mask. Each mask is represented by two 9-bit words, one representing the mask with don't-care set to 0 and the other with just the don't-care locations set to 1. The latter is OR-ed with the mask comparison to implement don't-care conditions. The result of mask matching is a P -bit word, where 1's indicate that the mask is matched.

The image is scanned in a raster fashion, i.e., the first 3×18 segment has the top-left corner in $(0, 0)$, the second one in $(0, 18)$ and so forth. As the mask matching operation is carried out in parallel, connectivity problems arise on the left and right edges of the 3×18 image segments, as neighbours are not loaded for edges. To overcome this problem, the two support registers in Fig. 22.2 were used. Basically, the last two columns of the segment are stored for the next comparison operation. With reference to Fig. 22.2, the support register 2 is used to store the left neighbours of the first pixel in the new segment. The support register 1 is used to have the left neighbours of the last pixel in the old segment, whose right neighbours were not

available until the new segment was loaded. A pipeline stage is used to align the mask comparison output bits 1 to 17, because the last pixel comparison in the segment is available only with the new segment.

Two binary image memories are used, in order to maximize the throughput. While the first binary image memory is read, the result of the current iteration is stored in the second one. In the next iteration the roles of memories are inverted and so forth. The two-memories approach speeds up the execution and also avoids problems due to the overwriting of the source image during the mask matching operation. In fact, a temporary buffer is used in Algorithm 22.1 to store intermediate results without affecting the elaboration within the current iteration. Conversely, in the proposed solution the intermediate result is directly written to the memory used in the following iteration.

22.4 Results

In the following, the proposed solution will be analyzed in terms of execution cycles and will be compared to the equivalent serial implementation.

The whole image is iterated a maximum of $N = 16$ times. The algorithm works with $M = 8$ masks. In our solution, $P = 18$ pixels are elaborated in parallel. The total number of execution cycles is given by:

$$v_P = \frac{N \cdot M}{P} R \quad (22.1)$$

where R is the image resolution, $R = 720 \times 576$.

In the original serial solution, the elaboration time is data-dependent, since mask comparison is started only if the source image elaborated pixel is 1 and 3×3 mask matching can be shorter than 5 cycles. In fact, 5 cycles are necessary only if the mask is matched, otherwise the comparison is aborted at the first difference. Therefore, some assumptions have to be made in order to estimate the execution time. We assume that $\rho = 25\%$ image area is '1' and that on average $\chi = 2$ comparison operations are needed to evaluate mask matching. With these considerations, we obtain:

$$v_S = N \cdot M \cdot (1 + \chi \rho) \cdot R \quad (22.2)$$

The acceleration factor $\alpha \triangleq v_S/v_P$ attained with the proposed solution is given by using (22.1) and (22.2):

$$\alpha = P \cdot (1 + \chi \rho) = 27 \quad (22.3)$$

The implemented HW execution cycles are 27 times less than the serial solution.

The algorithm was implemented on the Altera Stratix II EP2S30 FPGA. The area occupation is 4% in terms of combinational logic cells and 1% of registers, including the thinning module plus the two memory controllers for optimized memory access. The clock frequency is $f_{ck} = 100$ MHz, that allows an execution time:

$$T_e = \frac{v_P}{f_{ck}} = 0.03 \text{ s}, \quad \tau_e = \frac{T_e}{T_{fr}} = 74 \% \quad (22.4)$$

where T_e is the absolute execution time while τ_e is normalized to the frame period $T_{fr} = 0.04 \text{ s}$.

The image thinning processing is carried out in the worst-case time of 74 % frame period.

22.5 Conclusion

In this paper, the HW acceleration of a thinning algorithm was treated.

A FPGA-based solution is proposed, achieving real time performance on a standard PAL 720×576 video at 25 frames-per-second. The thinned image is obtained in 0.03s which represents the 74 % of the frame period. The remaining frame time can be exploited for further real-time processing.

With reference to the state-of-the-art, the proposed solution appears as an appealing trade-off in terms of execution time, image resolution and FPGA resources occupation.

References

1. Jain, A.K.: Fundamentals of Digital Image Processing. Prentice Hall, Englewood Cliffs (1988)
2. Zhang, T.Y., Suen, C.Y.: A fast parallel algorithm for thinning digital patterns. *Commun. ACM* **27**(3), 236–239 (1984). Mar
3. Huang, L., Wan, G., Liu, C.: An improved parallel thinning algorithm. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition (2003)
4. Zhu, X., Zhang, S.: A shape-adaptive thinning method for binary images. In: International Conference on Cyberworlds (2008)
5. Arcelli, C., Di Baja, G.S.: A one-pass two-operation process to detect the skeletal pixels on the 4-distance transform. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(4), 411–414 (1989). April
6. Ranganathan, N., Doreswamy, K.: A VLSI chip for computing the medial axis transform of an image. Proceedings of Conference on Computer Architectures for Machine Perception (1995)
7. Breu, H., Gil, J., Kirkpatrick, D., Werman, M.: Linear time Euclidean distance transform algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(5), 529–533 (1995). May
8. Miyazawa, M., Zeng, P., Iso, N., Hirata, T.: A systolic algorithm for Euclidean distance transform. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(7), 1127–1134 (2006)
9. Wang, L., Zhang, Y., Feng, J.: On the Euclidean distance of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8), 1334–1339 (2005)
10. Hsiao, P.Y., Hua, C.H., Lin, C.C.: A novel FPGA architectural implementation of pipelined thinning algorithm. 2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512)
11. Arcelli, C., Cordella, L., Levialdi, S.: Parallel thinning of binary pictures. *Electron. Lett.* **11**(7), 148 (1975)

Chapter 23

Low Cost Electrical Current Sensors with Extremely Wide Measurement Range

N. Galioto, F. Lo Bue, L. Mistretta and C.G. Giaconia

Abstract A new electrical current measurement system is presented. It features the ability to dynamically and automatically change its measurement range to the sensed current amplitude without user action. It also exhibits galvanic isolation and near zero insertion loss characteristics.

Keywords Current sensor · Current transducer · Power meter

23.1 Introduction

The needs of accessing the power consumption information gained more and more importance as the technology to provide these information evolved. Nowadays, power metering capabilities became an important part of residential and industrial IT systems. The ability to collect voltage and current in these facilities allows indeed easier user appliances control and management, and could enable power distribution systems to effectively optimize their distribution networks based on the required power demand.

These measurement systems are referred to “Smart Sensors” and were already defined in the early 80’s. by the Institute of Electrical and Electronics Engineers (IEEE). They are usually composed by:

N. Galioto (✉) · F. Lo Bue · L. Mistretta · C.G. Giaconia
Viale Delle Scienze, Università Degli Studi di Palermo, DEIM, 90128 Palermo, Italy
e-mail: natale.galioto@unipa.it

F. Lo Bue
e-mail: francesco.lobue@unipa.it

L. Mistretta
e-mail: leonardo.mistretta@gmail.com

C.G. Giaconia
e-mail: costantino.giaconia@unipa.it

- A signal transducer;
- A signal conditioning/amplification block;
- A processing-capable block such as a microcontroller or a logic unit;

All of today's sensors include also a communication interface to be able to interoperate in complex sensors networks and complex control systems.

Smart sensors are already widely adopted in various applications scenarios: estimation and remote collection of power consumption in home buildings or in places where human access is restricted due to danger for life, estimation of tire pressures in automotive scenarios, e-health applications, Wireless Sensors Networks and many others fields.

23.2 System Overview

In this paper we present a new embedded Smart Current Sensor system which can read electrical currents as low as 100 mA, and up to 24 A, with dynamic and automatic range selection and galvanic isolation. It features two LEM FHS-40P/SP600 Hall effect sensors [1] in order to sample the same current signal with different amplitude, a Microchip MRF24J40MA IEEE 802.15.4 RF transceiver [2] to allow network operations, and an Atmel AT32UC3C2256C AVR 32-bit microcontroller [3] as the logic core. This device features 256 Kb of Flash memory, 64 Kb of SRAM, up to 45 I/O pins, up to 66 MHz clock, and 11 ADC channels multiplexed with 12-bit resolution up to 2Msps, two of which can be sampled in parallel, making it ideal for this metering application.

The system is composed by two main modules:

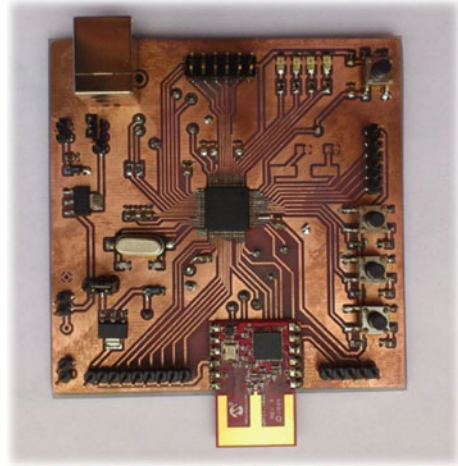
- The MCU board;
- The current sensor board;

This separation allowed us to switch between different current sensors and configurations without redesigning the entire system board during the various stages of the development.

23.2.1 MCU Board Overview

This board is equipped with the Atmel AT32UC3C2256C microcontroller [3] and the Microchip MRF24J40MA IEEE 802.15.4 RF transceiver [2]. It also features 4 status LEDs and an USB port, along with all the I/O pins routed to external expansion ports. The prototype of this board can be seen in Fig. 23.1. The microcontroller runs at 16MHz, and runs a predefined bootloader [4–6] which allows to program the device with a common PC or laptop directly connected to the onboard USB port without external programmers or tools.

Fig. 23.1 Prototype of the main system board. It features the Microchip MRF24J40MA IEEE 802.15.4 RF transceiver and the Atmel AT32UC3C2256C AVR 32-bit microcontroller



23.2.2 Current Measurement Overview

Four main techniques are generally used to sense electrical current, each with its strengths and weaknesses [7]:

- Rogowski coil;
- Current transformer;
- Low resistor shunt current monitor;
- Hall effect;

We decided to keep the system design as compact as possible, so Current Transformers and Rogowski coil were not taken into account since the beginning. Low resistor shunt current monitoring systems were analyzed but, since high voltage and high currents are involved, we abandoned the idea, as it is not possible to have a galvanic isolated design. So we had to find a solution with Hall effect sensors. Two sensors were analyzed: the Allegro ACS712T [8], and the LEM FHS-40P/SP600 [1].

23.2.2.1 Allegro ACS712T

The Allegro ACS712T is a fully integrated Hall effect sensor. The device consists of a linear Hall sensor circuit with a copper conductor path located inside the chip near the surface of the die. A current flowing *through* the sensor generates a magnetic field which is sensed by the integrated Hall sensor, and converted into a proportional output voltage.

The internal Hall sensor is located at fixed distance from the internal copper conductor, and has a factory-calibrated magnetic coupling of ± 1.2 mT/A. The output sensitivity of this device ranges from 66 to 186 mV/A thanks to a factory-programmed internal circuitry. The device is available in three different packages,

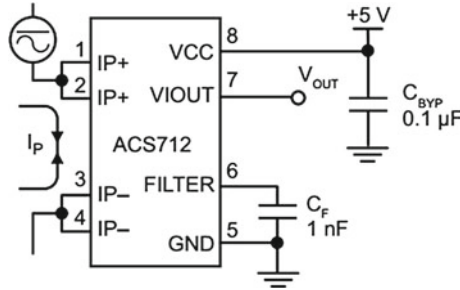


Fig. 23.2 Typical wiring diagram of the Allegro ACS712T Hall effect sensor. The ACS712T outputs an analog signal V_{OUT} that varies linearly with the uni- or bi-directional AC or DC primary sensed current I_p within the range specified. C_F is recommended for noise management with values that depend on the application

each optimized to maximize the full-scale output in one of the following ranges of sensed current: 5 A, 20 A, 30 A. Every configuration provide high voltage isolation ($2.1KV_{RMS}$) between high-side voltage and low-side voltage of the device. Authors have already successfully used this device in a Smart Power Meter application [9] and found it to be very effective. Figure 23.2 shows its typical wiring diagram.

23.2.2.2 LEM FHS-40P/SP600

The LEM FHS-40P/SP600 is a flat SMD open loop integrated circuit. It measures the magnetic field generated by the current flowing in a conductor placed *near* the die of the package, such as a PCB trace under (or over) the component, or a wire. The output voltage is proportional to the sensed magnetic field. Figure 23.3 shows its typical wiring diagram.

The magnetic flux density measuring range of this device is factory-calibrated at ± 3.3 mT with an input sensitivity of 600 mV/mT. The output voltage is proportional to the sensed magnetic field B :

$$V_s = G_B \cdot B \quad (23.1)$$

where G_B is the magnetic sensitivity of the device (600 mV/mT). A basic example with a current flowing in a long thin conductor is shown in Fig. 23.4. In this simple case, the generated flux density is:

$$B = \frac{\mu_0}{2\pi} \cdot \frac{I_p}{r} \text{ (T)} \quad (23.2)$$

where I_p is the current flowing the conductor (A), r is the distance from the center of the wire (m), and μ_0 is the permeability of the vacuum ($\mu_0 = 4\pi \cdot 10^{-7} \text{H/m}$). Replacing Eq. (23.2) in (23.1):

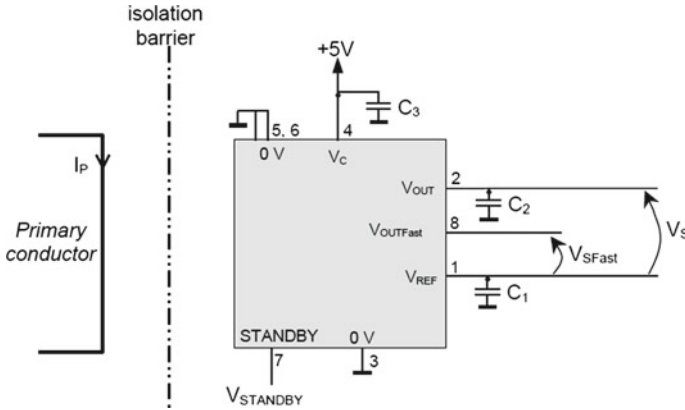
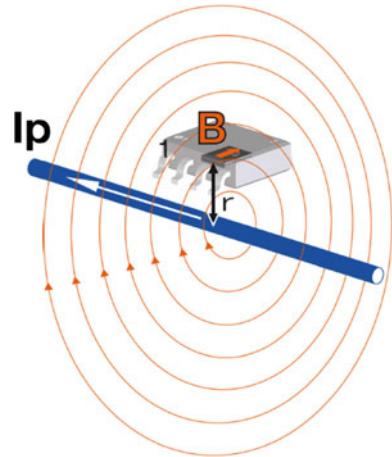


Fig. 23.3 Typical wiring diagram of the LEM FHS-40P/SP600 Hall effect sensor. It features galvanic isolation between the primary current conductor and the low voltage electronics, and hence no insertion loss

Fig. 23.4 Principle of operation of the LEM FHS-40P/SP600. Example of the orientation of the device to measure the magnetic field generated by a current along a conductor. Changing the distance between the sensor and the wire changes the sensed field and the sensitivity of the device



$$V_s = G_B \cdot \frac{\mu_0}{2\pi} \cdot \frac{I_p}{r} = 1.2 \cdot 10^{-4} \cdot \frac{I_p}{r} (V) \tag{23.3}$$

The output sensitivity is then defined by:

$$G = \frac{V_s}{I_p} = \frac{1.2 \cdot 10^{-4}}{r} (V/A) \tag{23.4}$$

Equation (23.4) clearly shows that the sensitivity of the device can be made arbitrary small by changing the distance between the device and the wire, and this represents a perfect condition for our metering system.

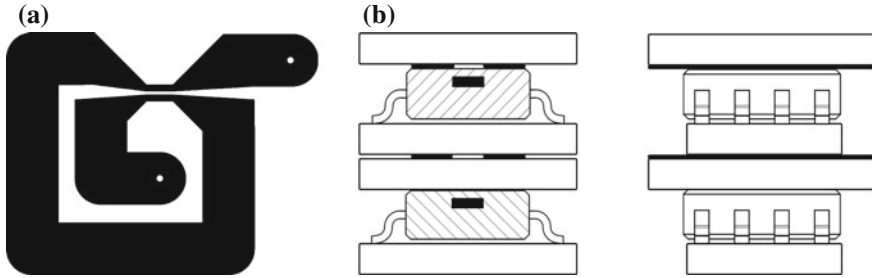


Fig. 23.5 PCB trace design and assembled sensors. *Top* sensor is placed inside two mirrored PCBs and gets two strong contributions from the traces of both the PCBs. *Bottom* sensor gets only the strong contribution of the traces of the *bottom* PCB. **a** PCB trace design. **b** Assembled sensors

23.3 System Design

The above example is limited in practical uses, and often a custom primary conductor design must be done. Custom non-round and non-long conductors can be used, but they change the overall sensor output response. In such situations, the Eq. (23.4) is no longer valid, and it becomes hard to predict the overall system response.

23.3.1 Current Conductor and Sensor Placement Design

In order to avoid the design of a very complex conductor, we took an existing and already tested design [10] and made some small modifications. We indeed prototyped two mirrored two-turns PCBs [10], with customized thickness and width of the traces in order to manage the heat dissipation produced by the maximum sustained current we would like to sense. We placed one LEM sensor inside these two PCBs, and placed another LEM sensor outside this “package”. Figure 23.5a shows our two-turns PCB design, and Fig. 23.5b shows the final assembly.

23.3.2 System Sensitivity

With our custom design, we can sense the current with two different sensitivities: the sensor mounted inside the two PCBs will get a strong contribution from the traces on both the PCBs, the sensor mounted outside will only get a strong contribution from the traces on the bottom PCB.

In order to gather information about the sensitivity of each sensor, we run the system with some known flowing current configurations, and annotated the output voltage of each sensor. Table 23.1 shows the obtained sensitivities at each flowing

Table 23.1 Sensitivity of the LEM sensors

I_{RMS} (A)	$V1_{pk-pk}$ (mV)	$V2_{pk-pk}$ (mV)	S1 (mV/A)	S2 (mV/A)
1.65	300	960	64	206
2.19	410	1330	67	215
3.39	580	1920	60	200
3.88	700	2260	64	206
5.10	830	2860	58	199
5.18	860	2900	58	198
5.60	930	3220	58	204
7.18	1210	4000	59	–
8.88	1450	4000	58	–
8.96	1480	4000	57	–
9.00	1520	4000	60	–

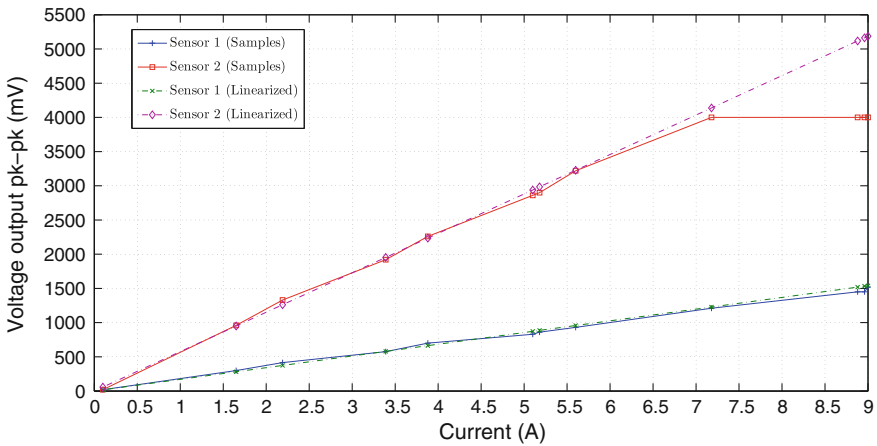


Fig. 23.6 Output voltage of the sensors. The linearized model fits the acquired data with good approximation

current configuration. Currents higher than 7 A saturate the output of the sensor with higher sensitivity, hence the constant reading of 4000 mV and the consequently missing sensitivities. For each sensor, all the reported sensitivities were averaged, in order to linearize the output response and to build an approximated linear model of the system. The obtained values, 61 mV/A for the sensor with one contribution only, and 204 mV/A for the sensor with two contributions, were used to code the linear model inside the MCU firmware. Figure 23.6 shows that the linearized model fits the acquired data with good approximation. Figure 23.7 shows the analog output voltage of the sensors with two different flowing current configurations.

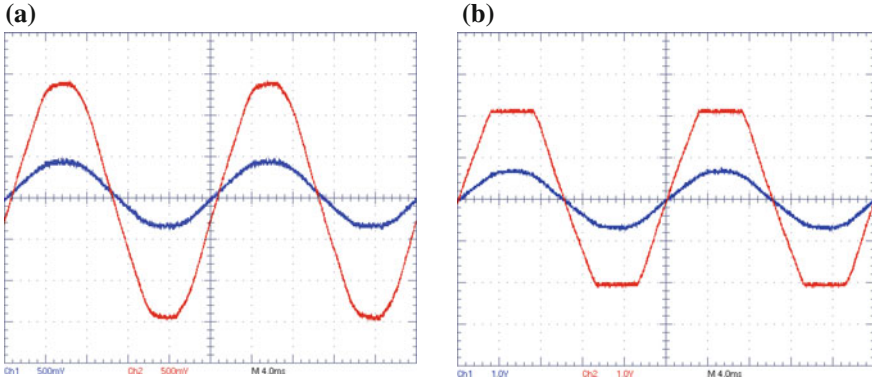


Fig. 23.7 Voltage output with different flowing current configurations viewed at the oscilloscope. With flowing currents higher than 7 A, the sensor with higher sensitivity saturates its output. **a** Voltage output with a owing current of 5.18 A. Both sensors are within their measurement range. **b** Voltage output with a owing current of 9 A. One of the sensors saturated its output

23.4 Firmware and Auto-Range Selection

Data acquisition is performed by the integrated ADC of the selected MCU. In particular, the analog output signals of the sensors are conditioned with a resistor network in order to meet the ADC voltage specifications.

Given the MCU capabilities, the output voltage of the two sensors is acquired at the same time and with a 12-bit resolution. The auto-range selection, then, is simply done by analyzing the values of the sampled signals. In particular, for each pair of sampled values (V_1 , V_2) we take the value V_2 , coming from the sensor with higher sensitivity, and compare it with a specific threshold. If this value is greater/lower than ($\pm 1.9V$), we replace it with the value coming from the other sensor V_1 :

$$V = \begin{cases} V_1 & \text{if } |V_2| > 1.9 V \\ V_2 & \text{if } |V_2| \leq 1.9 V \end{cases}$$

Once we have the “right” value V , we convert it to a known-scale floating-point number and perform the usual steps to compute the RMS value of the signal.

23.5 Conclusions

We have developed an new electrical current measurement system. It uses two low-cost current sensors, and with a custom PCB design of the primary current conductor, we were able to dynamically and automatically change the measurement range. This system exhibits galvanic isolation and near zero insertion loss characteristics.

References

1. LEM FHS-40P/SP600 datasheet <http://www.lem.com/>
2. Microchip MRF24J40MA 2.4 GHz IEEE Std. 802.15.4 RF Transceiver Module datasheet <http://www.microchip.com/>
3. Atmel AT32UC3C2256C datasheet <http://www.atmel.com>
4. AVR32806: AVR UC3 USB DFU Boot Loader, Version 1.1.0 and Higher <http://www.atmel.com/images/doc32166.pdf>
5. AVR32760: AVR32 UC3 USB DFU Bootloader Protocol <http://www.atmel.com/images/doc32131.pdf>
6. AVR4023: FLIP USB DFU Protocol <http://www.atmel.com/images/doc8457.pdf>
7. Ziegler, S., Woodward, R.C., Lu, H.H.-C., Borle, L.J.: Current sensing techniques: a review. *Sens. J. IEEE* **9**(4), 354–376 (2009)
8. Allegro MicroSystems ACS712T datasheet <http://www.allegromicro.com/>
9. Galioto, N., Lo Bue, F., Rizzo, D., Mistretta, L., Giaconia, C.G.: A novel wireless sensor network for electric power metering. In: *Applications in Electronics Pervading Industry, Environment and Society*. Springer (2014). ISBN 978-3-319-04369-2
10. LEM FHS-40P/SP600 design guide <http://www.lem.com/>

Chapter 24

Pathological Voice Analysis via Digital Signal Processing

Francesco Lo Bue, Natale Galioto and Costantino Giaconia

Abstract The interest in pathological voice analysis for specific neurological diseases is growing up aiming to offer more Health-care tele monitoring services since new high performing electronic devices are available for the end-user. In this article we show some parameters that can be digitally extracted and analyzed from pathological voices, in order to find a distinctive sign of the Parkinson disease. As a result, we will show a parameter that gives some information about the Parkinson disease characterization, particularly for male patients. We will also discuss about the needed computational cost related to parameters extraction and elaboration, aiming to target a possible tough yet portable hardware architecture capable to carry out the whole calculation or at least part of them locally.

Keywords Parkinson disease · Wavelet transform · Pathological voice · Mutual information

24.1 Introduction

The effects on vocal tracts, induced by the neurodegenerative diseases, are already known being the main topic of several research laboratories and medical test center since the last 30 years [1–7]. In particular some specific voice tests became the most used examinations of the medical anamnesis for the Parkinson disease classification and diagnosis. As example, a specific voice test is included within the UPDRS

F.L. Bue (✉) · N. Galioto · C. Giaconia

Department of Energy Information Technology and Mathematical Models (DEIM),
University of Palermo, Viale delle Scienze Bldg. 9, 90128 Palermo, Italy
e-mail: francesco.lobue@unipa.it

N. Galioto
e-mail: natale.galioto@unipa.it

C. Giaconia
e-mail: costantino.giaconia@unipa.it

(Unified Parkinson Disease Rating Scale) classification, one of the most used Parkinson Scale. Another traditional classification, used in the last 40 years, is the Hoehn–Yahr (H-Y) ‘staging’ that is actually the most popular because it requires an easier and empiric method for the anamnesis, so it’s still preferred by several neurologists.

Our work is intended to find other additional independent and unbiased parameters that can help neurologist to identify and classify the Parkinson disease. Previous researches contribute to develop several tools of voice and sounds analysis in order to study the speech signals [8–12]. They calculate some important parameters such as the fundamental frequency contour F0 (Pitch contour) and the voice formants. The Pitch values are strictly related to the glottal enclosure period, thus the F0 contour is one of the most interesting signal to analyze for the vocal tract modifications. The formants contours are instead related to the F0 but their positions and time trends give us a more complete spectral voice characterization.

In this paper we carried out an investigation on some specific formants by analyzing them with the wavelet transform and we envisaged how its coefficients can be linked to the presence of some kind of dysphonic disturbs, induced by the Parkinson disease. The studied parameters gave encouraging results, particularly in male patients, and could represent some additional features for the dysphonic voice analysis.

The described study was headed by a data collection campaign, in order to create a database of voice samples from both healthy and unhealthy people. During this acquisition phase we collected about 825 vocal samples from 165 patients among males and females. The vocal sampling consists on pronouncing the five sustained vowels (*a/e/i/o/u*) as long as the patient can do. The acquisition has been done using a digital voice recorder with a sampling rate of 44,100 Hz, set on PCM mode. The speech analysis was separately done for male and female samples, mainly because the Parkinson disease differently modifies their vocal tracts.

24.2 Methods

In this work several software and tools have been used to analyze voice files. Mainly we use PRAAT, Snack and Straight through the Matlab’s VoiceSauce interface. These tools allow to calculate several voice parameters like Shimmer (amplitude variations referred to a time period), Frequency jitter, Spectrograms and, above all, the pitch and the first five formants contour.

These software use different methods to calculate the pitch and formants, with different performance in terms of precision and computational cost [13, 14]. The trade off between optimal precision and calculus complexity, is an open issue for several research laboratories. Finding new and faster way for Pitch and formant tracking is another objective of our research but it goes beyond the work presented in this paper. The main target is to find parameters for pathological voice characterization with a low additional computational cost on top of the formant tracking

algorithm, in order to virtually implement it on a specific hardware such as a Smartphone or a DSP microcontroller.

To underline frequency variations and to find some typical voice markers of the Parkinson disease we applied the wavelet transform to the formants contour obtaining the whole wavelet decomposition on ten levels. We decided to use two basic mother wavelet Haar and Daubechies (type 2–5), because the their decomposition bank is lighter than others mother functions in term of computational cost. By the way we also tested other mother wavelets obtaining similar results. The overall calculation consist in these few steps:

- (1) Computing the Daubouchie wavelet transform;
- (2) Computing the variance and energy of Approximation and Detailed coefficients;
- (3) Compare the obtained value for each item.

24.3 Formant Analysis

The Parkinson disease causes psycho-motoric disturbs that induce effects on speech. Thus we can observe some phenomena related to a voice pathology like an enhanced amplitude variations (shimmer) within the time interval between two or more glottal periods, or an enhanced voice roughness and frequency jitter.

This paper focused onto the frequency components irregularities of the voice induced by the Parkinson disease otherwise known as dysphonia.

In the next figure we report the First three formants contour for an healthy and unhealthy person, in order to show some characteristic differences. The plotted data comes from the voices of two 70 years old males. On the left we reported the frequency contour of an unhealthy person and on the right the contour of a healthy one. The unhealthy one is affected by Parkinson's disease with H-Y staging of 2 (Fig. 24.1).

The two graphs have been plotted using the Straight Algorithm for the F0 and the snack Algorithm for the formants. We used the Straight algorithm implemented in the VoiceSauce Matlab tool. It is quite clear how the frequency contour for the unhealthy person is more irregular than the one referred to the healthy person, particularly at the higher formants. The analysis of the other samples in the database confirms this trend and justify a deeper analysis of the Formants' contour.

24.4 Introducing the Wavelet Transform

The wavelet decomposition allows to study the signals in a time-frequency multi scale domain, by introducing a separate scale and a translation factor. A wavelet transform using a generic mother wavelet can be described as the following equation

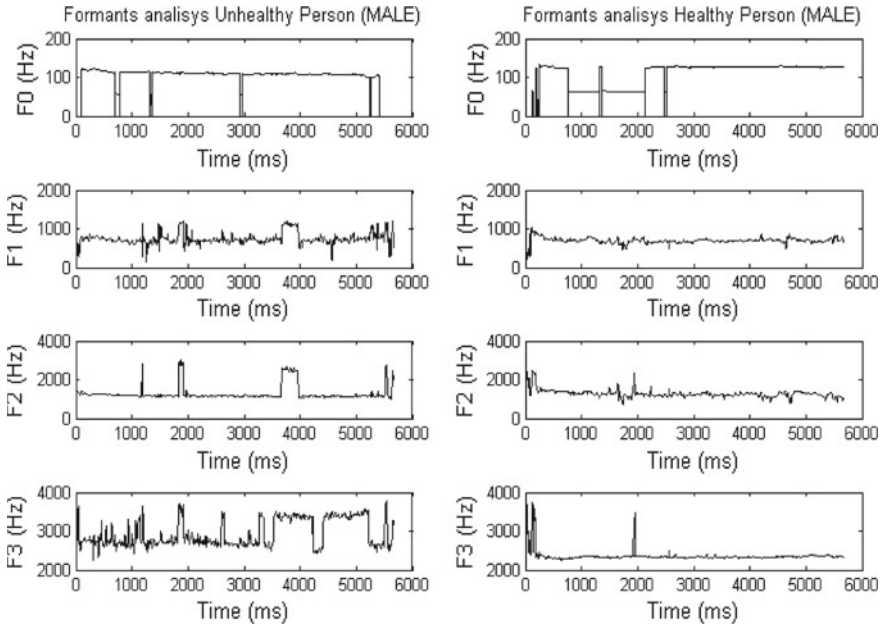


Fig. 24.1 Fundamental Frequency and formants contour of two 70 years old males. On the *left* the unhealthy one, on the *right* the healthy one

$$W_{\psi x} = \frac{1}{\sqrt{2}} \cdot \int_{-\infty}^{\infty} x(t) \cdot \psi^* \left(\frac{t-b}{a} \right) dt \quad (24.1)$$

By varying a and b parameters, the wavelet coefficients allows us to locally analyze the information content of $x(t)$ in the time-scaling factor domain. The transformation coefficients give information about the $x(t)$ perturbation at ‘frequency’ $1/a$ in a time interval around b . So the $x(t)$ variability is strictly related to the module of transformation coefficients and can be analyzed at different time resolutions depending on the frequency interval. Then this transformation allows to test the variability of non-stationary signals. From a mathematic point of view, this transformation appears to be quite complicated, but it can be implemented with a multi level digital filter bank. For each wavelet level the signal is simultaneously filtered with a Low-pass and High-pass filter. The resulting signals are reduced by two decimators. The decomposition can be schematized like the block diagram sketched in Fig. 24.2. For the type 5 Daubechies wavelet used in this work, $h(n)$ and $g(n)$ are a 10th order High-pass and Low-pass FIR filter respectively [14–17].

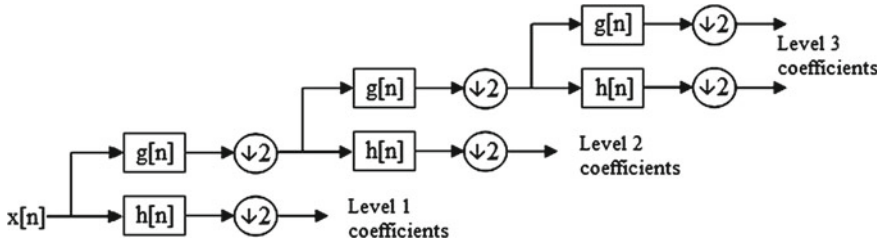


Fig. 24.2 Wavelet decomposition using filter banks. A level corresponds to an high-pass filter $h(n)$ and a low pass $g(n)$ filter bank

24.5 Data Analysis

For our tests, we used the data collected under a specific recordings protocol consisting of 59 recordings set from unhealthy subjects and 106 from healthy subjects. The voice data have been separated in males (78 patients) and females (87 patients). For each test we extracted the pitch and formants contour using PRAAT and Snack. For all formants we calculate the wavelet transform till the 10th level and finally computed both variance and energy for each approximation (CA_x) and Detailed (CD_x) coefficients. Since the obtained values are quite higher in patients affected by Parkinson’s disease, we carried out a complete analysis between the 1st and the 5th formant. As an example, the results related to the 3rd formant are presented in Fig. 24.3, where we can observe that there is a visible difference between healthy and unhealthy subjects for male patients. This results can be used to highlight the irregularities of the formants behaviour in the voice of the patients affected by the Parkinson disease. The main target of this kind of analysis is to shrink the voice characteristics in a few number of features in order to use them for an automatic PD staging classification method.

The differences between males and females led to separately study their voices characteristics. Moreover the complete voice analysis can’t be carried out by calculating one parameter only, but a deeper search within the data set is necessary in order to mine enough information and better classify healthy and unhealthy patients.

In order to investigate the information content of a data set, derived from formants evaluation, we used a theorem [18, 19], here reported, capable to analyse if a matrix of possible features represents a good set of attributes, a class in fact, that can be used to describe the and potentially classify the disease stage with different level of development.

Theorem If the mutual information between X and y ($I(X; y)$) is equal to the entropy of y ($H(y)$), then y is function of X .

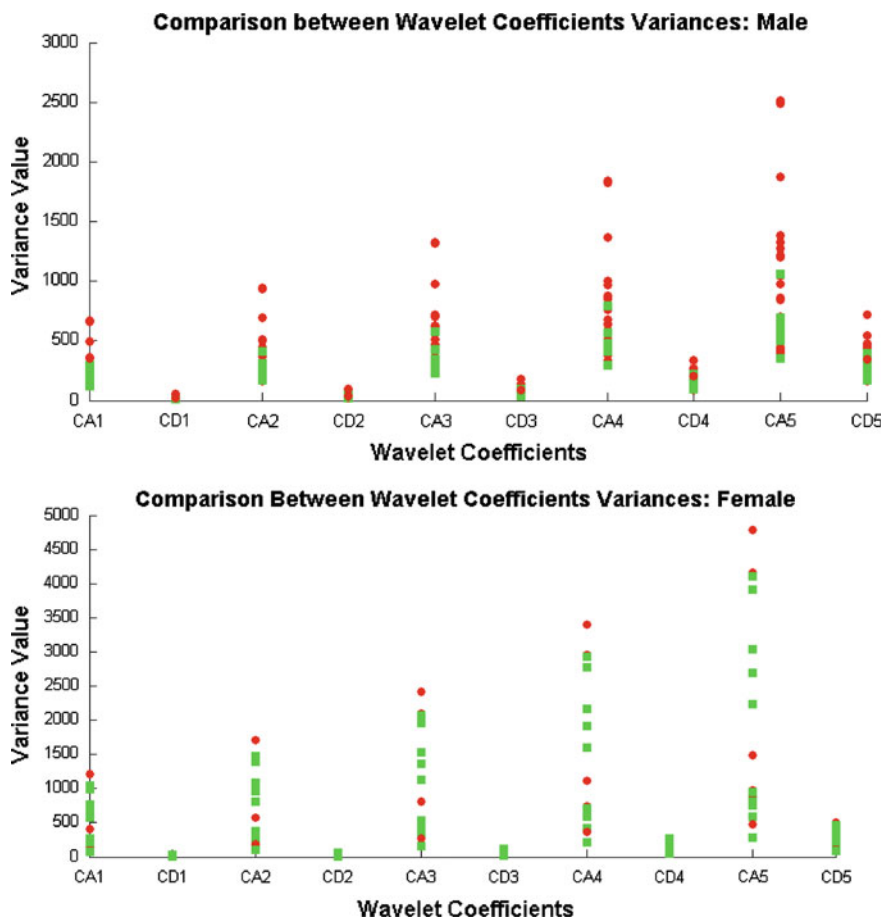


Fig. 24.3 Comparison between variances of the wavelet coefficients for both male and female patients. *Squared green dots* are used to represent data from healthy subjects, *circle red dots* for unhealthy subjects

where:

X is a $n \times m$ matrix representing a set of m parameters (features) derived from n patients;

Y is the vector of classes with the value 0 representing healthy persons and 1 representing the unhealthy ones;

$I(X; y)$ is the mutual information between the features and the classes vector

$H(y)$ is the entropy of vector y

$H(y|X)$ is the entropy of y for a given X

So, if the condition $I(X; y) \approx H(y)$, also corresponding to $H(y|X) \approx 0$, is respected, then X is a good candidate to be a feature subset for predicting y .

Starting from this theorem, the mutual information between a data set matrix X and the class attributes y can be found with the following formulas:

$$I(X; y) = H(y) - H(y|X) = H(X) - H(X|y) = H(y) + H(X) - H(X; y) \quad (24.2)$$

The results reported in Table 24.1 clearly shows that the parameters evaluated from the formants analysis can be used as a data set for a prediction model. As a second result, in the next section we show an application of this parameters to a simple linear classification model.

From the computational cost point of view, we analyzed the additional cost of the wavelet variance computation using Matlab. The tests have been done on a Intel Pentium E5200 based PC, using Windows XP operating system. The results are shown on Table 24.2, where computation time for each processing step is reported as a percentage of the overall processing time (126.86 s using PRAAT and 7.05 s with Snack). As we can see the wavelet analysis introduces a little extra computational cost compared with other processing phases.

24.6 Experimental Test and Results

By using the above described method, a study of a tailored model for the best classification has been carried out in order to obtain a validation test for a particular set of extracted features. In particular a simple linear model in order to evaluate the prediction performances is shown. The selected data set is obtained from the a collection of 200 features extracted from all 165 subjects and using the sustained vowel /a/only.

The test has been carried out with a basic pre-filtering of the data set using the minimum Redundancy Maximum Relevance (mRMR) [20] and a raw forward selection wrapping method, thus reducing the data set to the best 6–8 features.

Table 24.1 The full dataset derived from the formants analysis gives the results here resumed

Information content			
	I (X; y)	H(y)	H(y X)
Males	0.9995	0.9995	0
Females	0.7973	0.7973	0

(Information content of the dataset derived from formant analysis)

Table 24.2 Computation time resume

Computation resume			
	F0 contour (%)	Formants (F1–F4) contour (%)	Wavelet variance (%)
PRAAT ^a	85.5	14.17	0.33
Snack	54.73	43.93	1.34

Values are referred to the overall computation time

^aThe F0 contour was also evaluated by using the Straight algorithm and ending up to same results

Table 24.3 Test on a simple linear logistic regression model

Test on linear classification model			
	Training samples ^a	Selected features	Correct classifications ^b (%)
Males	62	8	72.5
Females	70	6	70.2

^aThe dataset has been divided in Training—Validation set with a ratio of 80–20 %

^bThe percentage is referred to the correct evaluations in the validation set

The test was reiterated about 50 times, keeping the mean error values on the validation set. The results are summarized in Table 24.3 and seem promising since with a subset containing a relatively low number of features, the percentage of correct classification is clearly beyond around 70 %.

The relative low number of samples doesn't allow to be completely sure to have a strongly validated model, but the information content of the extracted features and the little additional computation cost introduced, make these analysis very interesting for future applications.

24.7 Conclusions and Future Work

In this paper we show how the vocal formants' analysis, carried out by using wavelet based algorithms, can give good indications about the dysphonia induced by Parkinson disease. The obtained results was reached by introducing a low additional elaboration cost on top of the formants contour calculation. A deeper analysis of the overall tonal frequencies, instead of the vocal formants, could give some other features for helping the Parkinson disease classification [21]. The efficient voice components extraction is, by itself, an open issue in continuous development. The additional parameters and the relative features extraction method, developed in this work, will be furtherly exploited toward the implementation of an automatic classification method of Parkinson disease.

Acknowledgments This work has been developed within the Digital Electronics Systems Laboratory at the Electronics Engineering Department of the University of Palermo. Telecom Italia also supported the Ph.D. program of Francesco Lo Bue.

References

1. Dromey, C., Kumar, R., Lang, A.E., Lozano, A.M.: An investigation of the effects of subthalamic nucleus stimulation on acoustic measures of voice. *Mov. Disord.* **15**(6), 1132–1138 (2000)
2. Imaizumi, S.: Acoustic measurement of pathological voice qualities for medical purposes. In: *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Tokyo, (1986)

3. Henríquez, P., Alonso, J.B., Ferrer, M.A., Travieso, C.M., Godino-Llorente, J.I., Díaz-de-María, F.: Characterization of healthy and pathological voice through measures based on nonlinear dynamics. *IEEE Trans. Audio Speech Lang. Process.* **17**(6), 1186–1195 (2009)
4. Skodda, S., Schlegel, U.: Speech rate and rhythm in Parkinson's disease. *Mov. Disord.* **23**(7), 985–992 (2008). doi:[10.1002/mds.21996](https://doi.org/10.1002/mds.21996)
5. Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O.: Suitability of Dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomed. Eng.* **56**(4), 1015–1018 (2009)
6. Tsanas, A., Little, M.A., McSharry, P.E., Spielman, J., Ramig, L.O.: Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans. Biomed. Eng.* **59**(5), 1264–1271 (2012)
7. Tsanas A., Zañartu M., Little M.A., Fox C., Ramig L.O., Clifford, G.D.: Robust fundamental frequency estimation in sustained vowels: Detailed algorithmic comparisons and information fusion with adaptive Kalman filtering. *J. Acoust. Soc. Am.* **135**(5): 2885–2901 (2014)
8. Paul, B., David, W.: Praat: doing phonetics by computer [Computer program]. Version 5.3.71. <http://www.praat.org/> (2014). Accessed 9 April 2014
9. Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T.: Tandem-straight: a temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation. In: *ICASSP* (2008)
10. Kawahara, H., Morise, M.: Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework, *SADHANA Acad. Proc. Eng. Sci.* **36**(5), 713–722 (2011)
11. Sjölander, K.: The Snack Sound Toolkit (c), Department of Speech, Music and Hearing, Royal Institute of Technology Stockholm, Sweden, <http://www.speech.kth.se/snack/> (1997–2004)
12. Shue, Y.-L., Keating, P., Vicens, C.: A program for voice analysis. *J. Acoust. Soc. Am.* **126**, 2221–2221 (2009)
13. Jang, S.-J., Choi, S.-H., Kim, H.-M., Choi, H.-S., Yoon, Y.-R.: Evaluation of performance of several established pitch detection algorithms in pathological voices. In: *Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France, 23–26 August 2007*
14. Tsanas, A., Zañartu, M., Little, M.A., Fox, C., Ramig L.O., Clifford G.D.: Robust fundamental frequency estimation in sustained vowels: detailed algorithmic comparisons and information fusion with adaptive Kalman filtering. *J. Acoust. Soc. Am.* **135**, 2885–2901 (2014). doi:<http://dx.doi.org/10.1121/1.4870484>
15. Mallat, S.: *A Wavelet Tour of Signal Processing: The Sparse Way*, 1998, 3rd edn. Academic Press, Amsterdam (2009)
16. Vetterli, M., Herley, C.: Wavelets and filter banks: theory and design. In: *IEEE Transactions on Signal Processing*, Sep (1992). ISSN 1053-587X
17. Vetterli, M., Kovačević, J.: *Wavelets and Sub-band Coding*. Prentice Hall, Englewood Cliffs (1995)
18. Cover, T., Thomas, J.: *Elements of Information Theory*. Wiley, New York (1991)
19. McElice, R.: In *The Theory of Information and Coding: A Mathematical Framework for Communication*, ser. *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Company, USA, Reading, vol. 3 (1977)
20. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8) 1226–1238 (2005)
21. Lo Bue, L., Giaconia, C.G., Galio, N.: Pathological voice parameters extraction using digital signal processing. In: *AGIE Conference Proceedings, Cagliari, 18–20 June 2014*

Chapter 25

A Platform-Based Emulator for Mass-Storage Flash Cards Evaluation in Embedded Systems

Francesco Menichelli and Mauro Olivieri

Abstract In this work we present a simulation environment, built around the QEmu emulator, that allows the evaluation of mass-storage Flash-Card memories, specifically embedded Multimedia Cards (e-MMC). Flash card memories are internally complex systems containing, along with the memory array, an intelligent controller, running its own firmware. The controller is a critical unit, since its functions are not limited in providing a standard interface between the internal memory array and the user, but they are much more elaborate (e.g. buffering, erase sequences, garbage collection, flash memory wear leveling, etc.). It is then clear that the implementation of these functions can have a strong impact on performances. In this scenario, a simulation environment would be a valuable resource in the design flow, since it could allow the exploration of different internal architectures and firmware implementations, the verification and the estimation of performances of new devices during their design. Using QEmu as base environment, we have developed a fast emulator of a complete embedded system platform, containing a behavioral model of next-generation e-MMC devices, parametrized in order to be portable to future generations of e-MMCs. The whole emulator is fast enough to boot a complete Linux kernel and to launch applications, allowing the analysis of e-MMCs behavior on real use cases, based on actual file systems (e.g. ext2, FAT32, NTFS, etc.) and actual applications or benchmarks.

Keywords Embedded Systems emulation · Mass storage flash memory · MMC

F. Menichelli (✉) · M. Olivieri
Department of Information Engineering, Electronics and Telecommunications,
Sapienza University of Rome, via Eudossiana 18, 00184 Rome, Italy
e-mail: menichelli@diet.uniroma1.it

M. Olivieri
e-mail: olivieri@diet.uniroma1.it

25.1 Introduction

Embedded Systems, intended as computer system with a dedicated function, are rapidly evolving to new applications as long as technology allows rising computing power at low cost and low energy consumption. From small microcontrollers containing few Kbytes of memory and modest computing power, they have reached applications requiring much more performance and often including a complete operating system, such as communication/multimedia devices (PDA, tablets, smartphones), automotive infotainment systems, intelligent public terminals, etc.

Along with the evolution of computing power, other requirements, such as mass-storage capacity, are rapidly increasing in terms of size, speed, reliability. In this work we focus on a particular family of mass-storage devices, embedded multimedia cards (e-MMC). e-MMCs belong (along with, for example, SD cards, Compact Flash cards) to the group of managed NAND flash memories, meaning that the details on memory read and write addressing and timings are hidden by an integrated controller which exposes to the user a standard interface and protocol, based on high level commands. e-MMCs are specifically built to meet strict requirements of embedded systems, regarding both physical parameters (e.g. extended temperature range) and functionality (e.g. minimum guaranteed read/write speed/latency, reliability, endurance). Most of these functions are provided by the firmware running on the e-MMC controller, sometimes trying to balance the tradeoff between opposite requirements.

e-MMCs interface timings and supported operations are standardized by the JEDEC association [1]. While devices based on JEDEC specifications for e-MMC version 4.41 and 4.5 are now in production, next generation e-MMC, based on JEDEC spec. 5.0 are currently under design. The functions implemented in the e-MMC internal controller are not limited to *read* and *write* commands, but the standard requires a series of more complex commands and a number of internal operations as *garbage collection*, to collect free blocks of memory into bigger chunks, and *memory wear leveling*, to mitigate wearing effects intrinsic to flash memory technology. The implementation of these functions strongly impacts the characteristics of the device in opposite directions, to the point that sometimes ad-hoc firmwares are tailored to custom requirements, depending if the application is stressed toward read/write speed or reliability.

25.2 State of the Art

System emulation at architectural level is an established requirement for every new digital design [2] and in the years many environments have been presented to satisfy this request [3–6]. In general these environments can be classified depending on the level of abstraction at which they model the architecture, starting from cycle-accurate RTL level, through TLM, instruction level and higher. While being very accurate and

fundamental for many design verification steps, cycle-accurate simulators are the slowest in the group, to the point of being unpractical when simulations involve one or more high performance CPUs booting and running a complete operating system, accessing peripherals as mass-storage devices, USB hosts, network controllers, etc. In this scenario, an emulator as QEmu [4] can be a good choice for its speed and flexibility, even at the expense of reduced timing estimation accuracy.

The target of our work are high performance embedded systems, where mass-storage devices are used to store large amount of data (e.g. multimedia files, interactive maps) or programs (navigation applications, user installable apps, etc.). Given the characteristics of the environment they will generally be employed, as limited amount of available energy or large temperature operating range, traditional mass-storage devices, as hard disks, are not a good solution, while solid-state devices, as flash memory cards, are a much more desirable choice.

We concentrate on modeling e-MMC cards, whose electrical requirements, protocol interface and commands set are standardized by the JEDEC organization. At the moment future generation of e-MMC requirements are described in the JEDEC 5.0 standard [7].

25.3 Emulator and Platform Architecture

Our work is based on the QEmu emulator for building a complete virtual embedded platform, introducing a new and configurable e-MMC model, compliant to JEDEC e-MMC standard command set, protocol and interface. Model parameters are defined to configure some characteristics as individual command delays, buffers size, interface speed, in order to meet the specifications of different implementations of e-MMC internal hardware and firmware functions, allowing the exploration of design solutions and trade-offs between requirements.

Among with the e-MMC functional model we inserted a statistical delay model, meaning that each command and operation is annotated with an execution latency described by a statistical distribution. Data for delay annotations are extracted from instruction level simulation and profiling of e-MMC firmware, which is not covered by this paper.

The virtual platform architecture can be chosen between those available in the QEmu emulator. For the scope of our work we centered our attention on building a reference embedded architecture closely resembling the ARM Versatile Express family of boards [8]. The whole emulated architecture, as shown in Fig. 25.1 is composed of an ARM Cortex-A9 core, an e-MMC host controller linked to our e-MMC model and a group of standard peripheral, not detailed in the figure since not central to this work. They are present since they are required in order to run an operating system on the virtual platform. QEmu version used is 1.5.1.

A custom Linux OS runs on the emulated platform. We obtained the OS using Buildroot software [9] and customizing the kernel, based on version 3.12.1, in order to include non-mainline patches and extensions that support the latest e-MMC standard JEDEC version [7].

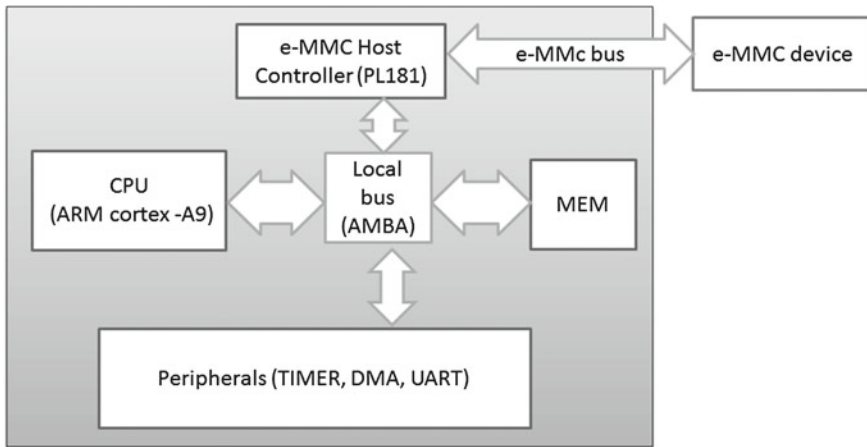


Fig. 25.1 Emulated hardware architecture

25.4 Simulation Setup and Results

The simulation setup is built in order to boot a complete embedded Linux OS, with an e-MMC connected and seen as an additional mass-storage device. Commands directed to the e-MMC are executed considering delays and latency provided by the statistical model. The e-MMC delays are integrated in QEmu internal time count, so that they can be accounted within total execution time.

In order to show the features of the emulator, we formatted and mounted the virtual e-MMC device using several filesystems for mass-storage devices (i.e. ext2, ext3, FAT32, NTFS, REISERFS), using custom application and the IOZone [10] suite for benchmarking.

As for the e-MMC delay model configuration, we extracted data from four different devices, that correspond to four different generations of e-MMCs. We report briefly in Table 25.1 some of the data, showing the differences between them for two typical operations *block read* (N_{rd}) and *block write* (N_{wr}). Data are expressed in number of cycles required to complete an operation, we specifically report the minimum and maximum number, as resume of the complete statistical distribution which is instead used inside the e-MMC model. The data are extracted from datasheets and measurements on actual e-MMC devices. As we can see, MMC4 is the newest

Table 25.1 Execution latency (clock cycles) for two typical commands

	MMC4	MMC3	MMC2	MMC1
$N_{rd}min$	2	2	2	2
$N_{rd}max$	5000	7500	12000	30000
$N_{wr}min$	2	2	2	2
$N_{wr}max$	5000	7500	12000	30000

Table 25.2 Average time (seconds) to create or read a file

Operation	MMC4	MMC3	MMC2	MMC1	SIZE(MB)
<i>EXT2</i> create	0.642	0.768	0.963	1.872	1
	1.257	1.508	1.931	3.630	2
	2.534	3.081	3.850	7.308	4
	5.106	5.981	7.725	14.613	8
<i>EXT2</i> read	0.291	0.356	0.447	0.861	1
	0.571	0.698	0.877	1.650	2
	1.151	1.427	1.750	3.322	4
	2.321	2.772	3.511	6.642	8
<i>FAT32</i> create	0.625	0.749	0.948	1.829	1
	1.283	1.509	1.925	3.639	2
	2.372	2.962	3.835	7.215	4
	4.792	5.920	7.646	14.478	8
<i>FAT32</i> read	0.215	0.256	0.324	0.609	1
	0.442	0.520	0.641	1.213	2
	0.818	1.021	1.278	2.405	4
	1.652	2.041	2.548	4.825	8

and most performing device, while the others are devices belonging to progressively older technology generations. In the same way, every command accepted by e-MMCs and every internal operation has been annotated in the model.

Table 25.2 reports average time required to create (write) and to read a file, for increasing file size. The simulation are repeated for two different filesystems, namely *ext2* and *FAT32*, using custom benchmark programs. The table allows to compare data in several ways. Considering a fixed file size, actual performance increase between e-MMC generations can be extracted, resulting in about a 3x speedup between the oldest and the newest one. Moreover, read and write time for the same e-MMC device can be compared. As expected for mass-storage devices based on FLASH memory technology, read time is substantially less than write time, resulting in about 2.5x read speed with respect to write speed. This results are obtained even if in

Table 25.1 maximum and minimum execution latencies are the same for read and write commands, because of the different statistical distributions of the two commands.

Eventually, we can compare performances between the two filesystems under test. We can see that, irrespective of e-MMC device generation, *FAT32* expresses a slightly better read and write speed than *ext2*.

Considering now just the newest and most performing device (MMC4), Table 25.3 presents file write and read average time, obtained using the IOzone benchmark suite [10], along with directory creation time. In this case we repeated the simulation for several filesystems (*ext2*, *ext3*, *ext4*, *FAT32*, *reiserfs*) and different file size. The table can be used to compare performances between different filesystems

Table 25.3 Average time (seconds) to create or read a file using IOzone and to create empty directories

Operation	EXT2	EXT3	EXT4	FAT32	REISER	SIZE (MB)
Create file	0.496	0.626	0.567	0.661	0.804	1
	0.976	1.221	1.096	1.312	1.602	2
	1.953	3.202	2.195	2.625	3.202	4
	3.936	6.404	4.390	5.258	6.404	8
	31.788	39.577	35.223	42.047	51.208	64
Read file	0.226	0.226	0.232	0.226	0.226	1
	0.453	0.453	0.458	0.453	0.453	2
	0.906	0.910	0.910	0.905	0.905	4
	1.812	1.816	1.815	1.811	1.810	8
	14.496	14.500	14.500	14.497	14.495	64
Create 10000 directories	1.417	4.552	4.392	2.468	1.102	

in the specific case of an e-MMC used as mass-storage device. We note that, comparing results in Tables 25.2 and 25.3 for the same e-MMC generation (MMC4) and the same filesystem, an increase in performance is clearly visible when passing from our custom benchmarks to the IOzone suite benchmarks. This observation, valid for both file read and file write operations and for both *ext2* and *FAT32* filesystems, can be attributed to optimized buffering strategies used in the IOzone benchmarks. We can also see that, using optimized read/write strategies as in IOzone, *ext2* speed is above *FAT32* speed, contrarily from what we deduced from tests in Table 25.2.

25.5 Conclusions

In this work we presented an emulation environment for mass-storage e-MMC devices, based on QEmu, that allows the emulation of a complete hardware platform, in order to test e-MMC performances under real stress conditions. The latency of e-MMC operations is accounted using a behavioral statistical model, in order to minimize its impact on QEmu speed. In fact, the emulator is fast to the point of being able to boot a complete linux OS and run standard or custom benchmarks, as well as real applications.

The data that are obtained can be used to deduce how a specific e-MMC device behaves in real case scenarios, in contrast with theoretical numbers that can be extracted from specifications. The main application of the emulator is during the design of new devices, when decisions must be taken at firmware level in order to optimize performances on a specific direction or to tailor trade-offs according to customer needs.

Future works regard the extension of the e-MMC model to cover completely the JEDEC 5.0 (or later) standard, as well as to improve the behavioral statistical delay model, which at the moment requires a characterization extracted from real devices or instruction-level simulations of the e-MMC firmware.

References

1. JEDEC Solid State Technology Association. <http://www.jedec.org>
2. Menichelli, F., Olivieri, M., Benini, L., Donno, M.C., Bisdounis, L.: A simulation-based power-aware architecture exploration of a multiprocessor system-on-chip design. In: Proceedings of the Conference on Design, Automation and Test in Europe, vol. 3. DATE '04, Washington, DC, USA, p. 30312. IEEE Computer Society (2004)
3. Benini, L., Bertozzi, D., Bogliolo, A., Menichelli, F., Olivieri, M.: Mparm: Exploring the multi-processor soc design space with systemc. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **41**(2), 169–182 (2005)
4. Bellard, F.: QEmu, a fast and portable dynamic translator. In: USENIX Annual Technical Conference, FREENIX Track, pp. 41–46 (2005)
5. Ferrari, A., Carloni, M., Mignogna, A., Menichelli, F., Ginsberg, D., Scholte, E., Nguyen, D.: Scalable virtual prototyping of distributed embedded control in a modern elevator system. In: 2012 7th IEEE International Symposium on Industrial Embedded Systems (SIES), IEEE, pp. 267–270 (2012)
6. Fraboulet, A., Risset, T., Scherrer, A.: Cycle accurate simulation model generation for soc prototyping. In: Computer Systems: Architectures, Modeling, and Simulation. Springer, New York, pp. 453–462 (2004)
7. JEDEC Solid State Technology Association: EMBEDDED MULTIMEDIACARD (e-MMC) e-MMC/CARD PRODUCT STANDARD (V5.0). <http://www.jedec.org/standards-documents/results/jesd84-b50>
8. ARM Inc.: ARM Versatile Express board. <http://www.arm.com/products/tools/development-boards/versatile-express/index.php>
9. Andersen, E.: Buildroot: Making Embedded Linux Easy. <http://buildroot.uclibc.org>
10. Norcott, W.D., Capps, D.: IOzone filesystem benchmark. www.iozone.org

Chapter 26

A Model-Based Methodology to Generate Code for Timer Units

Marco Marazza, Francesco Menichelli, Mauro Olivieri,
Orlando Ferrante and Alberto Ferrari

Abstract In this paper we present a model-based methodology and a tool-chain supporting pseudo-automated code generation for different Timer Units, which represent a new approach in this field. Programmable Timer Units are timing co-processors used to elaborate complex high-resolution timing functions subject to hard real-time constraints. Verification at the different design stages, as required per safety standards' certification, is becoming a major concern for Timer Units code development life-cycle. Enabling correct-by-construction code generation, our methodology supports code development, integration and testing across all design phases. We show how high-level functional models derived from functional requirements can be mapped onto the target architecture and how architecture-specific code can be generated. Our methodology is then applied to an automotive reference example.

Keywords Embedded Systems · Code generation · Timer unit

M. Marazza (✉) · F. Menichelli · M. Olivieri
Department of Information Engineering, Electronics and Telecommunications (DIET),
Sapienza University of Rome, Via Eudossiana, 18, 00184 Roma, Italy
e-mail: marazza@diet.uniroma1.it; marco.marazza@utsce.utc.com

F. Menichelli
e-mail: menichelli@diet.uniroma1.it

M. Olivieri
e-mail: olivieri@diet.uniroma1.it

M. Marazza · O. Ferrante · A. Ferrari
Advanced Laboratory on Embedded Systems (ALES), Via Barberini, 50, 00187 Roma, Italy

O. Ferrante
e-mail: orlando.ferrante@utsce.utc.com

A. Ferrari
e-mail: alberto.ferrari@utsce.utc.com

26.1 Introduction

An increasing number of today's industrial applications demand accurate control of timing synchronization. Typical examples come from the automotive domain: cylinder spark timing, fuel injection timing and fuel mixture control must be precisely controlled to achieve the highest gain in terms of fuel economy, unwanted emissions and engine performance, while guaranteeing low energy consumption [9]. In these cases even the use of low-latency software interrupts might not achieve the required high time resolution. Moreover, the great number and the concurrent nature of these timing functions exaggeratedly increases the workload of the Electronic Control Unit's (ECU) processor. To help delivering hard real-time functions, programmable Timer Units can be integrated into the ECU architecture. These co-processors are provided with custom hardware and software to reduce the amount of I/O processing on single- or multi-core CPUs [8]. Examples of programmable Timer Units can be found in [4, 6]. Each programmable Timer Unit is provided with its own programming language: while the ETPU [4] is provided with a customized high-level assembly programming model, the GTM [6] is provided with a C-like compiler prototyped in LLVM [7]. Despite high-level languages have been developed, still Timer Units' programming models differ significantly and require a great amount of time to develop and debug the code. In this paper we propose a methodology to add a model-based programming layer. The benefits of such an approach are manifold: (1) the programmer would be provided with a single programming environment for many Timer Units; (2) a model-based environment typically allows simulating the models to check whether they fulfill functional requirements; (3) many tools in the market supporting model-based design provide automatic code generation, thus saving code development time considerably and enable automatic test generation [3]; (4) many formal analysis techniques (e.g. [3, 10]) can be successfully addressed on functional models, thus reducing the risk of generating unspecified or bugged source code. To the best of the author's knowledge this approach has not been addressed yet for Timer Units and represents an opportunity to improve code development and verification for Timer Units.

26.2 Reference Example

As a reference example we consider a function controlling the cylinder ignition timing of an internal combustion engine [5]. This function controls the generation of spark pulses that feed a spark plug actuator. Spark pulses must be synchronized to specific angular positions of the rotating shaft of the engine. Our reference ignition function consists of generating a main spark pulse, followed by a sequence of multi-spark pulses at each complete engine cycle (720°). Generally, the properties of these I/O functions (i.e. the angular value at which the main pulse must start for a specific cylinder, etc.) are parametrized and can be changed at run-time by the appli-

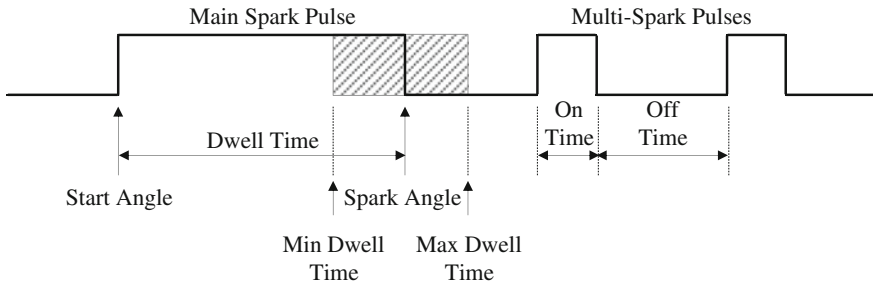


Fig. 26.1 Specification of the ignition function

cation running on the ECU. Figure 26.1 represents the specification of the ignition function. The *Dwell Time* is a function parameter set by the application running on the ECU and indicating the required active time of the main spark pulse. *Start Angle* is the required shaft's angular position at which the main spark pulse must switch to active and *End Angle* is the required angle for the opposite transition of the main spark pulse. Since the ignition function must guarantee that the spark pulse ends at the correct engine angle irrespective to engine acceleration or deceleration, it also has two additional parameters: *Minimum Dwell* (time) and *Maximum Dwell* (time). To ensure that the ignition coil has been charged sufficiently to generate a reliable spark after the pulse ends the spark pulse must remain active for at least a *Minimum Dwell* time length. Conversely, the spark pulse must be shorter than *Maximum Dwell* to ensure that the ignition coil being charged is not damaged by too much current and heat. After the main spark pulse has been generated, a set of equidistant short spark pulses can follow. This sequence is characterized by only three parameters: *Number of Multi-spark Pulses*, *Off Time* and *On Time*, indicating the number, and inactive time and active time lengths respectively.

26.3 Methodology

We propose a model-based methodology to automatically generate code for Timer Units. The main steps are summarized in Fig. 26.2; in this Section we briefly review each of the activities.

26.3.1 Functional Model

In our model-based methodology we start from the functional model M which is a formal representation of a function of the system, implementing the specified functional requirements. The benefit of implementing requirements by means of a formal

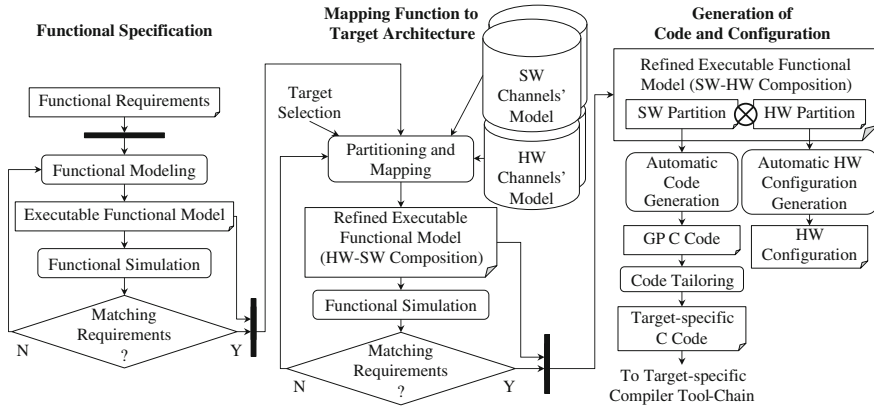


Fig. 26.2 Illustration of the proposed methodology

model rely on the possibility of executing and refining it. The refinement level of the model depends both on the designer’s expertise and on the purpose of the model in the work-flow. As shown in the left-hand side of Fig. 26.2, the functional modeling phase is often an iterative process, converging to the needs of the designer. The formal nature of the functional model also opens the way to a set of formal verification activities that can be leveraged (1) to verify the correctness of the model against its (possibly formalized) requirements and (2) to generate test scenarios [3] that can be applied on the design at subsequent phases of the development life-cycle (e.g. for back-to-back testing).

26.3.2 Mapping Functions on the Target Architecture

The partitioning and mapping activity (central portion of Fig. 26.2) aims at the *decomposition* of the function into a set of functional components and *allocation* of each functional component to the proper architectural component(s) in the target Timer Unit. The inputs of this activity are: the executable functional model provided by the preceding modeling activity, the selection of the target Timer Unit, a library of hardware channels’ models per each Timer Unit, and a library of channels’ control software models per each Timer Unit. Timer Units’ configurable hardware channels [4, 6], can be thought as a set of hardware-implemented services provided to the software. Each library model is still a functional (possibly hierarchical) model, but enriched with the hardware and software peculiarities of the specific Timer Unit. Model libraries can be designed by the end user or could be provided by third parties, e.g. by the Timer Unit vendor. To match the functional behavior, architectural components are picked up from the hardware and software libraries pertaining to the selected Timer Unit. Different Timer Units provide differing sets of services to

their control software. Hence, mapping the same functional model on different architectures can be reduced to a graph covering problem. The result of this activity is a model $C = H \otimes S$ resulting from the composition (\otimes) of a hardware library model H and a software library model S and equivalent (\equiv) to the input functional model M , ($C \equiv M$).

26.3.3 Automatic Generation of Target Code and Configuration

Referring to the right-hand part of Fig. 26.2, the enriched model $C = H \otimes S$ is used to generate both the configuration and the source code controlling the behavior of each I/O hardware channel. The I/O hardware channel configuration is derived straightforwardly from the hardware partition model H . In fact, H models the function performed by the hardware channel, which only depends on its configuration. On the other side, the software partition model S is used to automatically generate correct-by-construction source code. The generated target source code, along with header files, have to conform to the various Timer Units programming languages. This code has to be compiled in a later stage of the work-flow in order to be executed in the target Timer Unit. This approach subsumes that a C or high level assembly compiler is available to the developer, so that the automatically generated source code can be effectively translated into the executable machine-code.

26.4 Application to the Reference Example

We applied our methodology by using the Matlab/Simulink/Stateflow tool, along with the Embedded Coder Simulink Toolbox to automatically generate standard C code from our models. Such code has then been modified by hand and compiled with the specific Timer Unit's compiler [1, 7]. This Section gives a short description about how we accomplished the different phases of our methodology. The ignition function has been modeled as a Simulink/Stateflow Extended Finite State Machine (EFSM) by starting from the function specification. The formal model depicted in Fig. 26.3 is a function $F[\mathbf{u}, \mathbf{x}, \mathbf{f}, k]$ that at each discrete time k maps the inputs vector $\mathbf{u}[k]$ and the current state vector $\mathbf{x}[k]$ to a vector of outputs $\mathbf{y}[k]$ and next-state values $\mathbf{x}[k + 1]$. All the input parameters are generated by a subsystem external to the EFSM in Fig. 26.3 and can change at every time, as required by the application. Figures 26.4 and 26.5 represent the execution of the ignition function in two corner cases: in Fig. 26.4 *End Angle* arrives before *Max Dwell*, whereas in the example in Fig. 26.5 *Max Dwell* occurs before *End Angle*. The waveform at the bottom of the two figures indicates the time between *Min Dwell* and *Max Dwell*, where the *End Angle* is expected to arrive. The function in Fig. 26.3 has been refined for the two

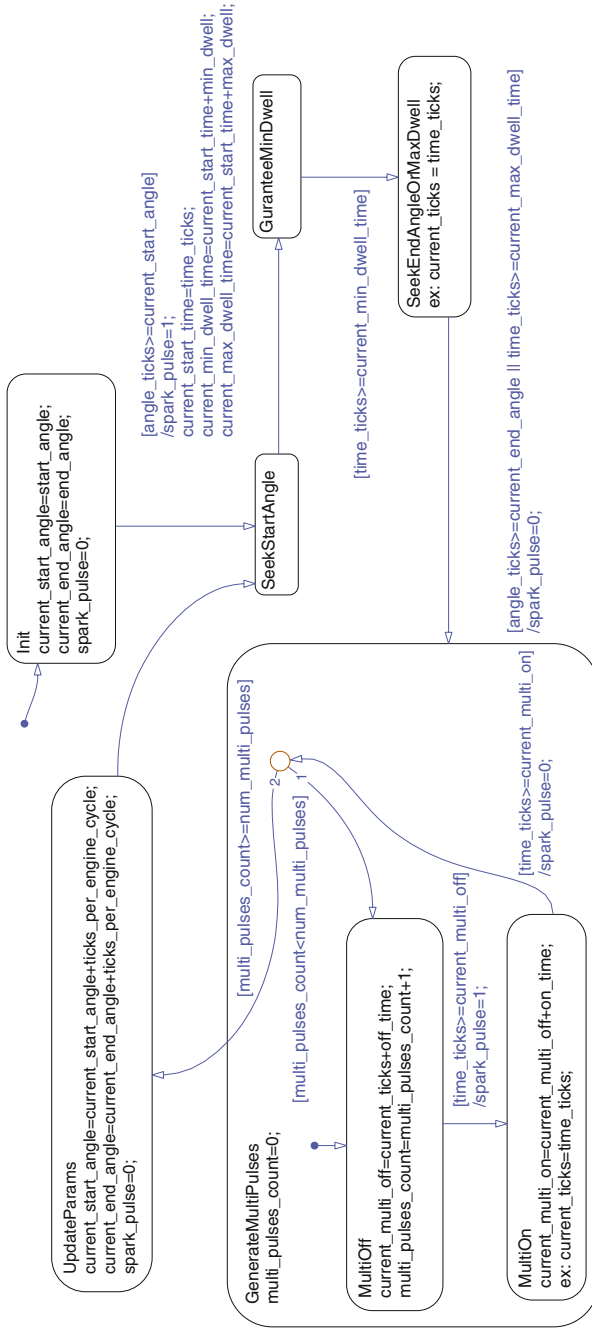


Fig. 26.3 Ignition function EFSM modeled in simulink/stateflow

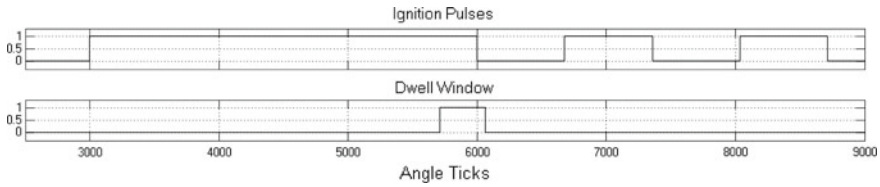


Fig. 26.4 Main spark pulse terminating at *MaxAngle*

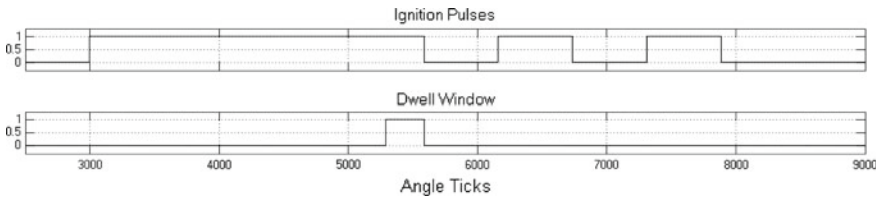


Fig. 26.5 Main spark pulse terminating at *MaxDwell*

architectures in [4, 6]. The hardware channels and the related control software have been modeled through the EFSM formalism. The composition of both the hardware and the software machines represents the ignition function as implemented on the two distinct architectures. The simulations of the refinements for the two architectures give the same results as shown in Figs. 26.4 and 26.5. As defined in Sect. 26.3, we exploited the software partition of the refined EFSM to generate the code specific to the target Timer Unit. In this preliminary work we generated code for a standard x86 platform and then manually tailored the resulting C code to match the programming model of the specific Timer Unit [1, 7]. This step helped us filling a set of tables of correspondences between I/O channel modes of the two Timer Units. The correspondences can be used to map particular “patterns” in the functional model to the corresponding library models of the channel mode or software code specific to the target architecture.

26.5 Conclusion

In this paper we presented a model-based methodology along with the supporting tool-chain for pseudo-automated code generation for different Timer Units, which represents a new approach in this field. The benefits are manifold: code developers can spend their effort on the modeling of the desired function, independently of the target Timer Unit; source code and hardware configuration can be generated automatically from the model and model-based automatic test generation and verification techniques can be exploited to test the design across its development phases. Future work will be devoted to automation of those phases that still require manual intervention.

References

1. ASH WARE: Compiler Reference Manual. v.2.01, 12 2011
2. Ferrante, O., Ferrari, A., Marazza, M.: Automatic Generation of Failure Scenarios for SoC. ERTS, Toulouse, France, 5–7 Feb 2014
3. Ferrante, O., Ferrari, A., Marazza, M.: Model based generation of high coverage test suites for embedded systems. In: Proceedings of the IEEE European Test Symposium, Paderborn, Germany, 26–30 May 2014
4. Freescale: Enhanced Time Processing Unit (eTPU) Reference Manual. 05 2004
5. Freescale: Using the eTPU Spark Function. Application Note. 07 2009. http://www.freescale.com/files/32bit/doc/app_note/AN3771.pdf
6. GTM-IP Specification revision. 06 2013. http://www.bosch-semiconductors.de/media/en/pdf_1/ipmodules_1/timer/GTM-IP_Specification_v1551.pdf
7. Marazza, M., Cremona, F., Ceraolo Spurio, D., Demuth, C., Nastasi, C., Ferrari, A.: Towards a Programming and Analysis Framework for Timer Units. In: JRWRTC, Sophia Antipolis, France, 16–18 October 2013
8. Menichelli, F., Olivieri, M., Benini, L., Donno, M., Bisdounis, L.: A Simulation-Based Power-Aware Architecture Exploration of a Multiprocessor System-on-Chip Design. DATE, pp. 312–317 (2004)
9. Menichelli, F., Olivieri, M.: Static minimization of total energy consumption in memory subsystem for scratchpad-based systems-on-chips. IEEE Trans. VLSI Syst. **17**(2), 161–171 (2009)
10. Rodrigues, C.: A Case Study for Formal Verification of a Timing Coprocessor. IEEE, LATW (2009)