Lecture Notes in Electrical Engineering 627

Sergio Saponara
Alessandro De Gloria *Editors*

Applications in Electronics Pervading Industry, Environment and Society

APPLEPIES 2019

Springer

Lecture Notes in Electrical Engineering

Volume 627

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Associate Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Executive Editor (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Web of Science and Springerlink ****

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

Sergio Saponara · Alessandro De Gloria
Editors

Applications in Electronics Pervading Industry, Environment and Society

APPLEPIES 2019

Springer

Editors
Sergio Saponara
DII
University of Pisa
Pisa, Italy

Alessandro De Gloria
DITEN
University of Genoa
Genoa, Italy

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-3-030-37276-7 ISBN 978-3-030-37277-4 (eBook)
https://doi.org/10.1007/978-3-030-37277-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The 2019 edition of the conference on *Applications in Electronics Pervading Industry, Environment and Society* was held in Pisa, Italy, on September 11–13, 2019, at the School of Engineering (Aula Magna U. Dini and Aula Magna A. Pacinotti).

During the three days, 110 registered participants, from 35 different entities (25 universities and 10 industries), discussed electronic applications in several domains, demonstrating how electronics has become pervasive and ever more embedded in everyday objects and processes.

The conference had the technical and/or financial support of University of Pisa, University of Genoa, SIE (Italian Association for Electronics), Giakova, and the H2020 European Processor Initiative.

After a strict blind-review selection process, 21 interactive posters and 43 lectures have been accepted (with co-authors from 14 different nations) in 11 sessions focused on circuits and electronic systems and their relevant applications in the following fields: wireless and IoT, health care, vehicles and robots (electrified and autonomous), power electronics and energy storage, cybersecurity, AI and data engineering.

More in details the interactive poster (IP) sessions involved contributions on IP1 *Vehicular, Robotic and Energy Electronic Systems,* IP2 *IoT and Integrated Circuits*, IP3 *Digital Circuits and Systems*, while the oral sessions involved contributions on O1 *Rad-Hard Electronics*, O2 *Internet of Things*, O3 *Processors and Memories,* O4 *VLSI and Signal Processing*, O5 *Digital Circuits and AI Data Processing*, O6 *Sensors and Sensing Electronic Systems,* O7 *Power and High Voltage Electronics,* O8 *Signal and Data Processing.*

There were also two special events:

– A round table on EuroHPC and the European Processor Initiative with contributions from E4, CINECA, STMicroelectronics, University of Bologna, University of Pisa

– A demo session of high-performance instrumentation and prototypes for battery management system, aerospace onboard data communication, high-speed drivers for optical modulators.

The proposed papers, collected in this book, and the talks and roundtables of the special events, prove that the computing, storage and networking capabilities of today electronic systems are such that their applications can fulfill the needs of humankind in terms of mobility, health care, connectivity, energy management, smart production, ambient intelligence, smart living, safety and security, education, entertainment, tourism, and cultural heritage.

To exploit such capabilities, multidisciplinary knowledge and expertise are needed to support a virtuous iterative cycle from user needs to the design, prototyping and testing of new products and services. The latter are more and more characterized by a digital core.

The design and testing cycles go through the whole system engineering process, which includes analysis of users' needs, specification definition, verification plan definition, software and hardware co-design, laboratory and user testing and verification, maintenance management, and lifecycle management of electronics applications. The design of electronics-enabled systems should provide key features such as innovation, high performance, real-time operations, implementations with low-cost and reduced budgets in terms of size, weight and power consumption. To succeed in this, one of the most important factors is the adoption of a suited design flow and relevant electronic design automation (EDA) tools. Platform-based design and meet in the middle between top-down and bottom-up design flows are needed to fulfill the time and cost-related challenges of nowadays' market scenarios.

All these challenging aspects call for the importance of the role of academia as a place where new generations of designers can learn and practice with cutting-the-edge technological tools and are stimulated to devise solutions for challenges coming from a variety of application domains.

The APPLEPIES 2019 conference aims at becoming a reference point in the field of electronics systems design and applications, trying to fill at scientific and technological R&D level a gap that the most farsighted industries have already indicated and are striving to cover.

Pisa, Italy — Sergio Saponara, General Chair

Genoa, Italy — Alessandro De Gloria, Honorary Chair

Contents

Part I Rad-Hard Electronics

1 Advanced Radiation Sensors VLSI Design in CMOS Technology for High Energy Physics Applications 3
Tommaso Croci, Arianna Morozzi, Pisana Placidi and Daniele Passeri

2 Design, Operation and BER Test of Multi-Gb/s Radiation-Hard Drivers in 65 nm Technology for Silicon Photonics Optical Modulators 11
G. Ciarpi, S. Cammarata, S. Faralli, P. Velha, G. Magazzù, F. Palla and Sergio Saponara

3 A Rad-Hard Bandgap Voltage Reference for High Energy Physics Experiments 19
G. Traversi, L. Gaioni, M. Manghisoni, M. Pezzoli, L. Ratti, V. Re, E. Riceputi and M. Sonzogni

4 Analysis and Comparison of Ring and LC-Tank Oscillators for 65 nm Integration of Rad-Hard VCO for SpaceFibre Applications 25
D. Monda, G. Ciarpi, G. Mangraviti, L. Berti and Sergio Saponara

5 A Compact Gated Integrator for Conditioning Pulsed Analog Signals 33
Sara Pettinato, Andrea Orsini, Maria Cristina Rossi, Diego Tagnani, Marco Girolami and Stefano Salvatori

Part II Internet of Things

6 Multivariate Microaggregation with Fixed Group Size Based on the Travelling Salesman Problem 43
Armando Maya López and Agusti Solanas

7 Modular Design of Electronic Appliances for Reliability Enhancement in a Circular Economy Perspective 51
Simone Orcioni, Cristiano Scavongelli and Massimo Conti

8 Pest Detection for Precision Agriculture Based on IoT Machine Learning 65
Andrea Albanese, Donato d'Acunto and Davide Brunelli

9 Statistical Flow Classification for the IoT 73
Gennaro Cirillo, Roberto Passerone, Antonio Posenato and Luca Rizzon

10 Using LPWAN Connectivity for Elderly Activity Monitoring in Smartcity Scenarios 81
D. Fernandes Carvalho, P. Ferrari, E. Sisinni, P. Bellitti, N. F. Lopomo and M. Serpelloni

Part III Processors and Memories

11 Characterization of a RISC-V Microcontroller Through Fault Injection 91
Dario Asciolla, Luigi Dilillo, Douglas Santos, Douglas Melo, Alessandra Menicucci and Marco Ottavi

12 Analyzing Machine Learning on Mainstream Microcontrollers . . . 103
Vincenzo Falbo, Tommaso Apicella, Daniele Aurioso, Luisa Danese, Francesco Bellotti, Riccardo Berta and Alessandro De Gloria

13 Quality Aware Selective ECC for Approximate DRAM 109
Giulia Stazi, Antonio Mastrandrea, Mauro Olivieri and Francesco Menichelli

14 Digital Random Number Generator Hardware Accelerator IP-Core for Security Applications 117
Luca Baldanzi, Luca Crocetti, Francesco Falaschi, Jacopo Belli, Luca Fanucci and Sergio Saponara

15 An Energy Optimized JPEG Encoder for Parallel Ultra-Low-Power Processing-Platforms 125
Tommaso Polonelli, Daniele Battistini, Manuele Rusci, Davide Brunelli and Luca Benini

Part IV VLSI & Signal Processing

16 VLSI Architectures for the Steerable-Discrete-Cosine-Transform (SDCT) 137
Luigi Sole, Riccardo Peloso, Maurizio Capra, Massimo Ruo Roch, Guido Masera and Maurizio Martina

17 Hardware Architecture for a Bit-Serial Odd-Even Transposition Sort Network with On-The-Fly Compare and Swap 145
Ghattas Akkad, Rafic Ayoubi, Ali Mansour and Bachar ElHassan

18 Variable-Rounded LMS Filter for Low-Power Applications 155
Gennaro Di Meo, Davide De Caro, Ettore Napoli, Nicola Petra and Antonio G. M. Strollo

19 A Simulink Model-Based Design of a Floating-Point Pipelined Accumulator with HDL Coder Compatibility for FPGA Implementation 163
Marco Bassoli, Valentina Bianchi and Ilaria De Munari

20 Bitmap Index: A Processing-in-Memory Reconfigurable Implementation 173
M. Andrighetti, G. Turvani, G. Santoro, M. Vacca, M. Ruo Roch, M. Graziano and M. Zamboni

Part V Digital Circuits and AI Data Processing

21 Digital Circuit for the Arbitrary Selection of Sample Rate in Digital Storage Oscilloscopes 183
M. D'Arco, E. Napoli and E. Zacharelos

22 An Intelligent Informative Totem Application Based on Deep CNN in Edge Regime 191
Paolo Giammatteo, Giacomo Valente and Alessandro D'Ortenzio

23 FPGA-Based Clock Phase Alignment Circuit for Frame Jitter Reduction 199
Dario Russo and Stefano Ricci

24 Real-Time Embedded System for Event-Driven sEMG Acquisition and Functional Electrical Stimulation Control 207
Fabio Rossi, Ricardo Maximiliano Rosales, Paolo Motto Ros and Danilo Demarchi

25 A Fast Approximation of the Hyperbolic Tangent When Using Posit Numbers and Its Application to Deep Neural Networks 213
Marco Cococcioni, Federico Rossi, Emanuele Ruffaldi and Sergio Saponara

Part VI Sensors and Sensing Electronic Systems

26 2-D Acoustic Particle Velocity Sensors Based on a Commercial Post-CMOS MEMS Technology 225
Andrea Ria, Massimo Piotto, Mattia Cicalini, Andrea Nannini and Paolo Bruschi

27 A High-SNR Distributed Acoustic Sensor Based on ϕ-OTDR Using a Scalable Phase Demodulation Scheme Without Phase Unwrapping 233
Yonas Muanenda, Stefano Faralli, Claudio J. Oton and Fabrizio Di Pasquale

28 Silicon Nanowires as Contact Between the Cell Membrane and CMOS Circuits 243
P. Piedimonte, D. A. M. Feyen, M. Mercola, E. Messina, M. Renzi and F. Palma

29 Ultra-Low Power Displacement Sensor 251
Alessandro Bertacchini, Marco Lasagni and Gabriele Sereni

30 Simulation of an Optical-to-Digital Converter for High Frequency FBG Interrogator 259
Vincenzo Romano Marrazzo, Francesco Fienga, Michele Riccio, Luca Maresca, Andrea Irace and Giovanni Breglio

31 Wireless Sensors for Intraoral Force Monitoring 267
M. Merenda, D. Laurendi, D. Iero, D. M. D'Addona and F. G. Della Corte

Part VII Power and High Voltage Electronics

32 Reinforced Galvanic Isolation: Integrated Approaches to Go Beyond 20-kV Surge Voltage (invited) 277
Egidio Ragonese, Nunzio Spina, Alessandro Parisi and Giuseppe Palmisano

33 Experimental Characterization of a Commercial Sodium-Nickel Chloride Battery for Telecom Applications 285
Federico Baronti, Roberto Di Rienzo, Roberto Roncella, Gianluca Simonte and Roberto Saletti

34 Design and Development of a Prototype of Flash Charge Systems for Public Transportation 293
Adriano Alessandrini, Riccardo Barbieri, Lorenzo Berzi, Fabio Cignini, Antonino Genovese, Fernando Ortenzi, Marco Pierini and Luca Pugi

35 Unsupervised Monitoring System for Predictive Maintenance of High Voltage Apparatus 301
Christian Gianoglio, Andrea Bruzzone, Edoardo Ragusa and Paolo Gastaldo

36 Control System Design for Cogging Torque Reduction Based on Sensor-Less Architecture 309
Dini Pierpaolo and Sergio Saponara

Part VIII Signal and Data Processing

37 Acoustic Emissions Detection and Ranging of Cracks in Metal Tanks Using Deep Learning 325
Gian Carlo Cardarilli, Luca Di Nunzio, Rocco Fazzolari, Daniele Giardino, Marco Matta, Marco Re and Sergio Spanò

38 Recognizing Breathing Rate and Movement While Sleeping in Home Environment 333
Maksym Gaiduk, Ralf Seepold, Natividad Martínez Madrid, Simone Orcioni and Massimo Conti

39 A Fast Face Recognition CNN Obtained by Distillation 341
Luca De Bortoli, Francesco Guzzi, Stefano Marsi, Sergio Carrato and Giovanni Ramponi

40 Fine-Grain Traffic Control for Smart Intersections 349
Jessica Bellitto, Valentina Schenone, Francesco Bellotti, Riccardo Berta and Alessandro De Gloria

41 A Graph Signal Processing Technique for Vibration Analysis with Clustered Sensor Networks 355
Federica Zonzini, Alberto Girolami, Davide Brunelli, Nicola Testoni, Alessandro Marzani and Luca De Marchi

42 Guided Waves Direction of Arrival Estimation Based on Calibrated Multiresolution Wavelet Analysis 363
Michelangelo Maria Malatesta, Nicola Testoni, Alessandro Marzani and Luca De Marchi

43 High-Frame-Rate Ultrasound Color Flow Imaging Based on an Open Scanner 371
Francesco Guidi, Enrico Boni, Alessandro Dallai, Valentino Meacci and Piero Tortoli

Part IX Vehicular, Robotic and Energy Electronic Systems

44 Empowering Deafblind Communication Capabilities by Means of AI-Based Body Parts Tracking and Remotely Controlled Robotic Arm for Sign Language Speakers 381
Silvia Panicacci, Gianluca Giuffrida, Luca Baldanzi, Luca Massari, Giuseppe Terruso, Martina Zalteri, Mariangela Filosa, Giovanni Tonietti, Calogero Maria Oddo and Luca Fanucci

45 Project VELA, Upgrades and Simulation Models of the UNIFI Autonomous Sail Drone 389
Enrico Boni, Marco Montagni and Luca Pugi

46 DC-Link Capacitor Sizing Method for a Wireless Power Transfer Circuit to Be Used in Drone Opportunity Charging 397
Andrea Carloni, Federico Baronti, Roberto Di Rienzo, Roberto Roncella and Roberto Saletti

47 Distributed Video Antifire Surveillance System Based on IoT Embedded Computing Nodes 405
Alessio Gagliardi and Sergio Saponara

48 Integrated Simulation Environment for Co-design/Verification of Mechanic, Electronic and Control of Automotive E-Drives: The Smart-Latch Case Study 413
Emanuele Abbatessa, Davide Dente and Sergio Saponara

49 Spice Model of Photovoltaic Panel for Electronic System Design 425
Mirco Muttillo, Tullio de Rubeis, Dario Ambrosini and Giuseppe Ferri

50 Exhaustive Modeling of Electric Vehicle Dynamics, Powertrain and Energy Storage/Conversion for Electrical Component Sizing and Diagnostic 433
Gaia Fiore, Lucian Mihet-Popa and Sergio Saponara

Part X IoT and Integrated Circuits

51 Analysis of 3-D MPPT for RF Harvesting 443
Michele Caselli and Andrea Boni

52 Analysis and Simulation of a PLL Architecture Towards a Fully Integrated 65 nm Solution for the New Spacefibre Standard 451
Marco Mestice, Bruno Neri and Sergio Saponara

53 Stability and Startup of Non Linear Loop Circuits 463
Francesca Cucchi, Stefano Di Pascoli and Giuseppe Iannaccone

54 IoT Ubiquitous Edge Engine Implementation on the Raspberry PI 469
Ahmad Kobeissi, Riccardo Berta, Francesco Bellotti and Alessandro De Gloria

55 Non-intrusive Load Monitoring on the Edge of the Network: A Smart Measurement Node 477
Hugo Wöhrl and Davide Brunelli

56 Design of a SpaceFibre High-Speed Satellite Interface ASIC 483
Pietro Nannipieri, Gianmarco Dinelli, Luca Dello Sterpaio, Antonino Marino and Luca Fanucci

57 An FPGA Realization for Real-Time Depth Estimation in Image Sequences 489
Stefano Marsi, Sergio Carrato, Luca De Bortoli, Paolo Gallina, Francesco Guzzi and Giovanni Ramponi

Part XI Digital Circuits and Systems

58 Integration of a SpaceFibre IP Core with the LEON3 Microprocessor Through an AMBA AHB Bus 499
Gianmarco Dinelli, Gabriele Meoni, Pietro Nannipieri, Luca Dello Sterpaio, Antonino Marino and Luca Fanucci

59 A RISC-V Fault-Tolerant Microcontroller Core Architecture Based on a Hardware Thread Full/Partial Protection and a Thread-Controlled Watch-Dog Timer 505
Luigi Blasi, Francesco Vigli, Abdallah Cheikh, Antonio Mastrandrea, Francesco Menichelli and Mauro Olivieri

60 Estimating the Downlink Data-Rate of a CCSDS File Delivery Protocol IP Core 513
Gabriele Meoni, Alberto Valverde, Giorgio Magistrati and Luca Fanucci

61 Automatic Detection of the Carotid Artery Position for Blind Echo-Doppler Blood Flow Investigation 521
Riccardo Matera and Stefano Ricci

62 Efficient Mathematical Accelerator Design Coupled with an Interleaved Multi-threading RISC-V Microprocessor 529
Abdallah Cheikh, Stefano Sordillo, Antonio Mastrandrea, Francesco Menichelli and Mauro Olivieri

63 AXI4LV: Design and Implementation of a Full-Speed AMBA AXI4-Burst DMA Interface for LabVIEW FPGA 541
Luca Dello Sterpaio, Antonino Marino, Pietro Nannipieri, Gianmarco Dinelli and Luca Fanucci

64 3D-HEVC Neighboring Block Based Disparity Vector (NBDV) Derivation Architecture: Complexity and Implementation Analysis 549
Waqar Ahmad, Naveed Khan Baloch, Fawad Hussain, Muhammad Asif Khan and Maurizio Martina

Author Index 559

Part I
Rad-Hard Electronics

Chapter 1
Advanced Radiation Sensors VLSI Design in CMOS Technology for High Energy Physics Applications

Tommaso Croci, Arianna Morozzi, Pisana Placidi and Daniele Passeri

Abstract In this paper we discuss some issues related to the design, implementation and test of a CMOS Active Pixel Sensor. Two different pixel layout have been proposed based on a standard architecture to investigate the suitability of a 110 nm standard technology for the realization of small pixels, high granularity detectors to be used in High-Energy Physics, medical and space applications, such as particle tracking or beam monitoring.

Keywords Active Pixel Sensor · CMOS · Radiation sensor · High energy physics applications

1.1 Introduction

The adoption of standard CMOS technology has been suggested as a viable option for the fabrication of particle detectors, integrating sensitive element and related read-out circuitry on the same substrate. The inherently lower detection efficiency of standard CMOS substrates can be compensated by the simultaneous integration of small capacitance detection nodes and signal conditioning and elaboration of circuitry [1]. This foster the realization of integrated detectors without the need of hybrid solutions, e.g. the very expensive bump-bonding between sensing nodes (pixels) and read-out circuitry or the adoption of dedicated, ad-hoc technology flavours and options (e.g. high-resistivity substrates, with thick epi-layers or multiple wells) [2, 3]. In this paper we discuss some design, implementation and test issues with respect to the development of conventional Active Pixel Sensor (APS) matrices in 110 nm LFoundry technology [4] conceived for CMOS Image Sensor (CIS) fabrication. The aim of this study is to investigate the suitability of such a technology for the realization

T. Croci (✉) · A. Morozzi · P. Placidi · D. Passeri
INFN-Section of Perugia, Perugia, Italy
e-mail: tommaso.croci@pg.infn.it

P. Placidi · D. Passeri
Department of Engineering, University of Perugia, Perugia, Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_1

of small pixels, high granularity detectors to be used in High-Energy Physics, medical and space applications, such as particle tracking or beam monitoring [5, 6].

1.2 System Architecture and Active Pixel Sensor

To evaluate the performance of the chosen technology for high energy physics applications, test structures based on single active pixels and on pixel arrays of limited dimensions have been designed. The structures are characterized by the use of the typical three-transistor pixel architecture (Fig. 1.1), with different geometries of the sensitive area. The designed chip houses also the interface circuits required for reading, addressing and interfacing the sensitive component.

The particle detection principle is based on a photodiode, a reverse biased pn junction used to detect the impinging radiation by converting in electrical charge the energy released into the material. In high energy physics the sensor requirements are typically very harsh, such as high efficiency and good spatial localization. A good tolerance to radiation damage is offered by modern submicrometric VLSI processes, guaranteeing the correct functionality of the sensor and a longer operating life.

To collect the maximum amount of charge inside the pixel, the chip substrate or the epitaxial layer, if available, tends to be used as the p-type region of the photodiode, whereas the n-type region is usually made by an n-well or an n+ implantation. In this work we explore the possibility of using a standard CMOS technology, provided that the layout of the sensitive element has been designed according to the technology itself for the specific particle to be detected.

The APS involves the use of a basic electronic signal processing inside the pixel, directly connected to the sensitive element. In this way it is possible to increase the reading speed and to reduce the noise due to the lower impact of the parasitic elements. The price to be paid, however, is the reduction of the fill factor (FF) due to the "blind" area dedicated to electronic circuits. Therefore, during the pixel design,

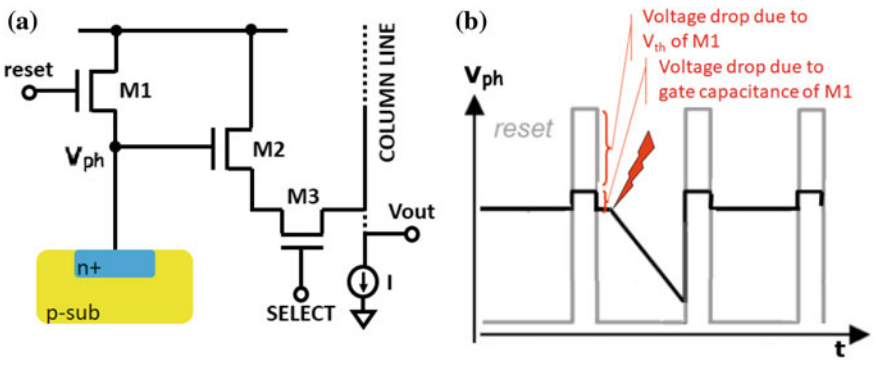


Fig. 1.1 **a** APS 3T circuit; **b** output voltage of the pixel

an effort has been made to limit the area occupancy of the front-end electronics and, at the same time, increasing the segmentation (pixel pitch) for better spatial resolution. The reading is the most critical operation because the photodiode has to be properly biased by connecting the cathode to the power supply through a NMOS (M1 in Fig. 1.1a) fixing its voltage to $(V_{DD} - V_{th})$. In Fig. 1.1b an additional voltage drop has been highlighted due to the capacitive coupling between gate and source of M1, when the transistor is turned off. The useful signal is represented by the voltage variation measured at the cathode of the photodiode with respect to this voltage and therefore this configuration limits the useful excursion of the signal. In addition, it should be avoid that the source follower (M2 in Fig. 1.1a) leaves the saturation region, otherwise it would further reduce the voltage swing.

The scaling of CMOS technology introduces significant advantages (for example in the reduction of area occupancy) but from the point of view of the sensitive element requires a greater attention in the design, creating new challenges. In fact the relationship between pixel dimensions and minimum channel length is not straightforward due to the different scaling. Consequently, beyond a certain level of technological integration, pixel scaling is no longer convenient, as the improvement in resolution is no longer sufficient to compensate for a bunch of new disadvantages. Indeed, while the decrease in the supply voltage tends to be proportional to the scaling, the threshold voltages do not decrease following the same trend, reducing the useful signal swing.

1.3 Simulation Results

The chip uses two different layout of 3 T pixels, called Small Pixel and Large Pixel (Fig. 1.2). They differ in the sensitive area dimensions, respectively 0.25 and 56 μm^2, while sharing the same square overall occupation, featuring 10 μm pixel pitch. Therefore, on a total pixel area of 100 μm^2, the FF of the Small Pixel is around 0.25% while the FF of the Large Pixel is 56%.

The sensing node (photodiode) is made by a n+ doped implantation, hosted in a deep p-well which is in turn realized on a standard, p-type substrate (Fig. 1.3). The metal interconnections have been shaped aiming at minimizing the antenna effects, at the same time aiming at multiple n+ contacts integration. Within the design flow, several parametric simulations have been carried out, aiming at exploring the different combinations of both reset and source follower transistors and photodiode node geometries and their impact on the pixel performance as a function of an external stimulus compatible with a MIP generation. As a general outcome, the Small Pixel exhibits better performance for low radiation intensity, as illustrated in the following.

In particular, in Table 1.1 the post-layout voltage drops (ΔV) on pixel output are reported, as a function of the sensitive node dimensions. A larger sensitive area would in principle collects more charge, with an upper limit corresponding to the Full Well Capacity (FWC). However, a larger area corresponds to a larger (parasitic) capacitance, thus reducing the charge to voltage conversion factor. Following these

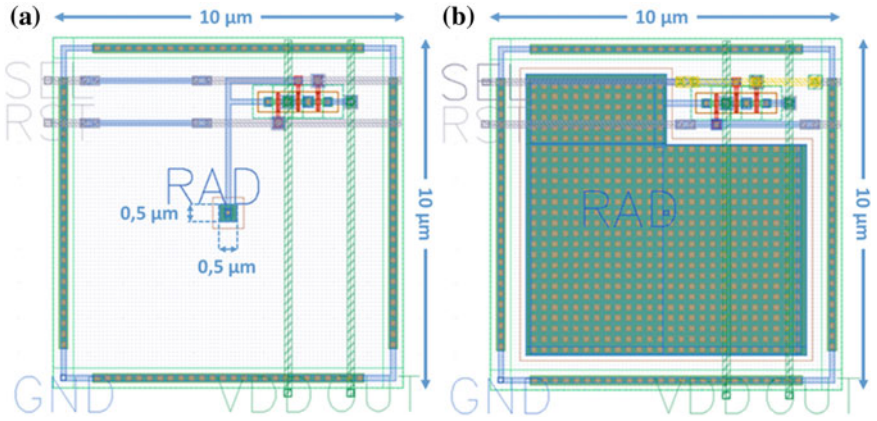


Fig. 1.2 **a** Small pixel and **b** Large pixel layouts

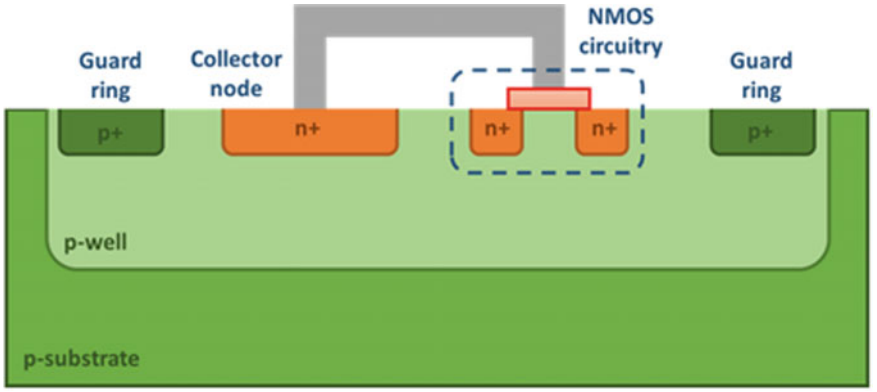


Fig. 1.3 Simplified cross section of the pixel

indications, we selected the Small Pixel with active area of $0.5 \times 0.5\ \mu m^2$, while for the Large Pixel we selected the option with the maximum area coverage (Fig. 1.2).

Considering the Small Pixel, Tables 1.2 and 1.3 show that the voltage drop in post-layout simulations tends to decrease at increasing transistor width (W) and length (L). This is due to the contribution of the reset and source follower transistors to the sensing node capacitance. According to this finding, the dimensions and aspect ratio of all the transistors within the pixel has been kept at the minimum value according to the design rules (150/110). Therefore, along the same line, the transistors within the Large Pixel have been kept at the minimum value according to the design rules as well, since the increase of their dimensions does not significantly affect the sensing node capacitance, being dominated by the large diode diffusion capacitance.

Eventually, in Table 1.4 are reported the post-layout voltage drops as a function of the radiative stimulus parameters, namely amplitude and duration of the resulting

Table 1.1 Voltage drops as a function of the photodiode sensitive area

Dimensions (μm^2)	ΔV (mV)
0.39×0.39	188.58
0.5×0.5	190.63
1×1	173.68
2×2	98.66
4×4	34.68
6×6	16.5
MAX	10.7

Table 1.2 Voltage drops versus width (W) and length (L) of M1 for the Small Pixel

W (nm)	ΔV (mV)
150	190.63
300	134.17
450	98.35
L (nm)	ΔV (mV)
110	190.63
220	153.75
330	149.35

Table 1.3 Voltage drops versus width (W) and length (L) of M2 for the Small Pixel

W (nm)	ΔV (mV)
150	190.63
300	187.18
450	179.73
L (nm)	ΔV (mV)
110	190.63
220	170.82
330	163.82

current pulse (used as input for circuit level simulation purposes). Data coming from device simulations were exploited to characterize a compact model of the sensing element: a junction diode was supplemented by a current generator describing a radiation-induced current pulse as predicted by device simulations. The quantitative effects of the increase of both pulse amplitude and width are reported in Fig. 1.4.

With reference to noise it should be underlined that the pixel-reset noise (N_{reset}) is determined by the thermal noise of the photodiode and is proportional to the inverse of the capacitance seen at the photodiode node. Charge-integration noise (N_{integr}) is instead due to dark current and is approximately proportional to the inverse of the

Table 1.4 Voltage drops as a function of the amplitude and duration of the radiative stimulus for the Large Pixel

Amplitude (A)	ΔV (mV)
600 n	10.7
1.2 µ	14.89
1.8 µ	19.11
2.4 µ	23.35
Duration (ns)	ΔV (mV)
2	23.35
4	37.36
6	51.21

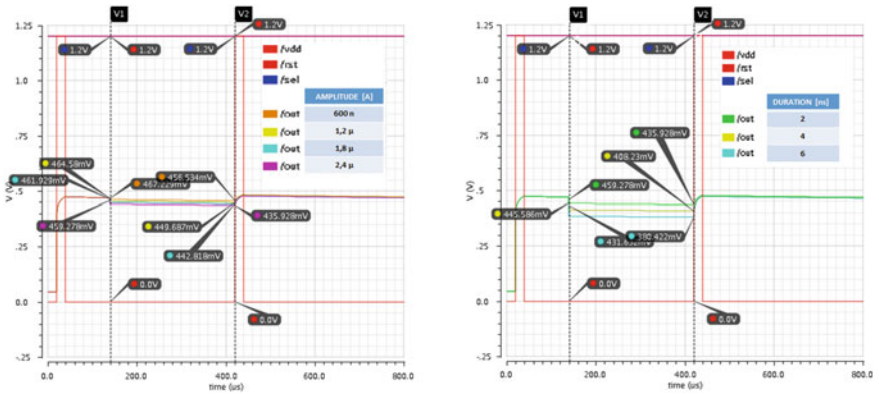


Fig. 1.4 Voltage drops as a function of the radiative stimulus parameters for the Large Pixel (amplitude on the left, duration on the right)

square of the capacitance [2]. Total pixel noise (obtained from the root mean square of reset and charge-integration noises) is expected to be in the order of a few mV.

1.4 Test Setup

A suitable test environment has been set up, due to the different features that have to be validated, ranging from stand-alone photodiode response to the test of small matrices. This results in a dedicated sequence of test signals to be generated and delivered to the chip which have been devised using a standard Arduino Due board based on a 32-bit ARM core microcontroller. A critical issue concerns the radiation source to be used for testing purposes. To allow for optical test, coverage of sensitive areas with metal layers has been avoided in the chip design. A dedicated PCB has

also been designed, accounting for size constraints coming from the optical setup. From the functional point of view, maximum flexibility has again been pursued, accounting for both manual and automatic test procedures. All the control and I/O signals can be generated either through on-board hardware circuitry, by means of routines driving the test board from a PC. Test-board assembly has currently been completed, and actual test is planned to be carried out in the next months.

1.5 Conclusion

This work aimed at the validation of basic performance of sensitive elements integrated in standard 110 nm LFoundry technology conceived for CMOS Image Sensor fabrication for particle detection application. The suitability of such an approach, in particular the adoption of a standard CMOS substrate with optimized pixel layout, has been verified. Results were very encouraging: a significant SNR, expressed in terms of output voltage drop, has been obtained in post-layout simulation. A dedicated PCB has also been designed and fabricated and test on actual chip are on-going.

Acknowledgements This work was supported by the *Department of Engineering* ("Ricerca di Base" 2017 and 2018) and by the *INFN* (SEED and ARCADIA projects).

References

1. Villani EG et al (2005) Simulation of a novel, radiation-resistant active pixel sensor in a standard 0.25 μm CMOS technology. IEEE Trans Nucl Sci 52(3):752–755
2. Passeri D et al (2004) Design, fabrication, and test of CMOS active-pixel radiation sensors. IEEE Trans Nucl Sci 51:1144–1149
3. Wang T et al (2017) Development of a depleted monolithic CMOS sensor in a 150 nm CMOS technology for the ATLAS inner tracker upgrade. JINST
4. LFoundry 'LFOUNDRY TECHNOLOGY 110nm' (Online). http://www.lfoundry.com/en/technology. Accessed 9 Sept 2019
5. Conti E et al (2013) Use of a CMOS image sensor for an active personal dosimeter in interventional radiology. IEEE Trans Instrum Meas 62(5):1065–1072
6. Gao W et al (2018) Total-ionization-dose radiation-induced noise modeling and analysis of a $2k \times 2k$ 4T CMOS active pixel sensor for space applications. IEEE Sens J 18(19):8053–8063

Chapter 2
Design, Operation and BER Test of Multi-Gb/s Radiation-Hard Drivers in 65 nm Technology for Silicon Photonics Optical Modulators

G. Ciarpi, S. Cammarata, S. Faralli, P. Velha, G. Magazzù, F. Palla and Sergio Saponara

Abstract The paper presents the design and the performance characterization, through system-level bit error rate (BER) tests, of a driver for silicon photonics Mach-Zehnder modulator (MZM) devices. Fabricated in TSMC 65 nm technology, the driver exploits a differential topology and a multi-stage current-mode logic architecture. It is designed to withstand radiation levels in compliance with the requirements for the on-detector systems in future particle physics experiments. The driver has been tested up to 800 Mrad showing about 30% degradation in voltage ratings. The BER test made on the stand-alone driver shows a capability of handling 5 Gb/s bit-rates with a quasi-error free BER of 10^{-11}. Electro-optical system-level BER tests carried out with an MZM wire-bonded to the designed driver showed an unexpected degradation in speed performances, which has been mainly attributed to packaging issues. Optimization and re-design activities, still working with 65 nm technology, are currently on-going to meet a data rate of 10 Gb/s for the same radiation hardness.

Keywords Silicon photonics · Mach-Zehnder modulator driver · Current-mode logic · Radiation hardness · High energy physics · BER characterization

2.1 Introduction

Silicon Photonics (SiPh) has become a viable technology for reducing the size, weight and energy consumption of optical devices for short-reach optical interconnects. All-

G. Ciarpi · S. Cammarata (✉) · G. Magazzù · S. Saponara
Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Via G. Caruso 16, Pisa, Italy
e-mail: cammarata.simone@yahoo.it

G. Ciarpi · S. Cammarata · S. Faralli · G. Magazzù · F. Palla
Istituto Nazionale di Fisica Nucleare – Sezione di Pisa, L. Pontecorvo 3, Pisa, Italy

S. Cammarata · S. Faralli · P. Velha
Scuola Superiore Sant'Anna – Istituto TeCIP, Via G. Moruzzi 1, Pisa, Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_2

silicon modulators are essential components for such communication links and are currently being evaluated at the Europen Organization for Nuclear Physics (CERN) in order to asses their suitability for use in high energy physics (HEP) experiments. Optical and electronic devices installed in the particle detection region have to ensure high reliability to radiation exposure. Custom-made SiPh Mach-Zehnder modulators (MZMs) have already been proved to tolerate radiation levels in line with those expected for future particle physics experiments [1]. In the context of CERN's large hadron collider (LHC) upgrade foreseen for 2026, the beam luminosity boosting will determine a significant increase in data traffic, on the order of dozens of tera-bits per second (Tb/s). The installation of optical transceivers with few Gb/s read-out capabilities will then be required [2]. It represents a data transfer speed roughly one order of magnitude higher than the throughputs currently achievable with state-of-the-art HEP front-end circuits, like those belonging the RD53 project [3].

Photonic devices easily reach operational bandwidths above 10 GHz, but the exploitation of these technologies in compact modules would be possible only after a careful design of the conditioning electronics which allows to encode a data stream onto an optical carrier. The aim of this work is to design a full-custom electronic integrated circuit (EIC) to operate with the MZM presented in [4], withstanding, at the same time, total ionizing doses (TID) up to 1 Grad and 1 MeV equivalent neutron fluences on the order of a few 10^{16} cm^{-2} regarding radiation damage from non-ionizing energy losses (NIEL).

Section 2.2 introduces the MZM driver (MZMD) core structure and the main circuital solutions which have been implemented to properly drive a traveling-wave MZM. A purely electrical characterization of the driver performances in terms of bandwidth, output voltage amplitude and bit error rate (BER) is detailedly reported in Sect. 2.3. The following section presents the overall system-level results and describes the electro-optical setup implemented to perform BER measurements of an hybrid transmitting unit made of an MZM driven by the developed MZMD. Conclusions are drawn in Sect. 2.5, mentioning the further activities that are currently ongoing towards the realization of a working prototype suitable for HEP environments.

2.2 Mach-Zehnder Modulator Driver Design

The full-custom MZM driver was designed in the commercial-grade TSMC 65 nm technology because of its recognized radiation hardness, mainly determined by its very thin gate oxide. Ionizing energy losses induce a build-up of positive trapped charges in oxide layers, causing threshold voltage shifts and current leakage in MOSFET devices. The thinner the oxide the less charges could be trapped and, in turn, the less detrimental will be the radiation effect on the electronic circuit. However, p-MOSFET devices are more sensitive to TID than their n-type counterpart, e.g. a minimum-sized diode-connected p-MOSFET loses the 100% of its on-current after being exposed to a TID of 1 Grad [5]. For this reason, the driver needed to be de-

veloped avoiding p-type MOSFETs [6]. A current-mode logic (CML) architecture, which exploits only n-MOSFETs and passive devices, has thus been adopted.

In order to meet speed and wide output swing constraints, the circuit was structured as sketched in Fig. 2.1. Five CML pre-driving stages, supplied at $V_{DD,L} = 1.2\,V$ and with gradually increasing sizes, forego an output stage with $V_{DD,H} = 2.4\,V$. The current drained from the $V_{DD,L}$ power supply (comprising the whole set of pre-driving stages) is around 35 mA while the output stage sinks approximately 60 mA, keeping the MZMD power consumption below 200 mW. Because of the radiation requirements only thin-oxide MOSFETs have been used, therefore the last stage exploits a cascode topology to share the wide voltage drop on two devices. Moreover, a bandwidth increment is obtained using inductive peaking techniques in the last two stages [7, 8].

2.3 Circuit-Level Electrical and Radiation Tolerance Testing

The electrical characterization of the driver was performed in terms of scattering parameters measurements, eye diagram plots and BER tests. S-parameters were carried out gluing the chip on a carrier board and contacting the chip pads with RF and DC probes, as shown in Fig. 2.2.

Figure 2.3 shows the S_{21} and S_{11} parameters of the driver. The 3-dB S_{21} bandwidth point is measured around 2.5 GHz, highlighting a potential application of the driver to bit-rates up to 5 Gb/s. The blue line shows that the input matching network of the driver works properly up to 4 GHz, whereupon the S_{11} parameter exceeds -10 dB.

Regarding eye diagrams and BER tests, the EIC was bonded on a custom-made printed circuit board (PCB). Standard SMA coaxial cables were used to connect the board to the instruments, while impedance-matched coplanar transmission lines convey the signals on the PCB. A 12.5 Gb/s pulse pattern generator (PPG) has been used to generate a pseudo random binary sequence (PRBS) following a PRBS-31 pattern, with voltage characteristics in compliance with standard CML levels. The eye diagrams obtained feeding the driver with this signal and measuring the output waveforms with a 23 GHz-bandwidth oscilloscope are shown in Fig. 2.2. The two eye diagrams present nearly the same amplitude, while higher noise and jitter appear at 5 Gb/s.

A BER tester (BERT) was then exploited to understand the impact of jitter-related penalties from a system-level viewpoint. Figure 2.3 reports the BER values for different bit-rates. A plateau at 10^{-11} is shown for data-rates up to 5 Gb/s, indicating that no error have been registered out of 1 Tb of transmitted data. This confirms a *quasi error-free* operation till a bit-rate of 5 Gb/s.

The circuit radiation resistance was investigated exposing the whole EIC to x-rays with a dose rate of 4.3 Mrad/h at the INFN-Padova facility. The normalized voltage amplitude degradation of the output signals with increasing dose level is

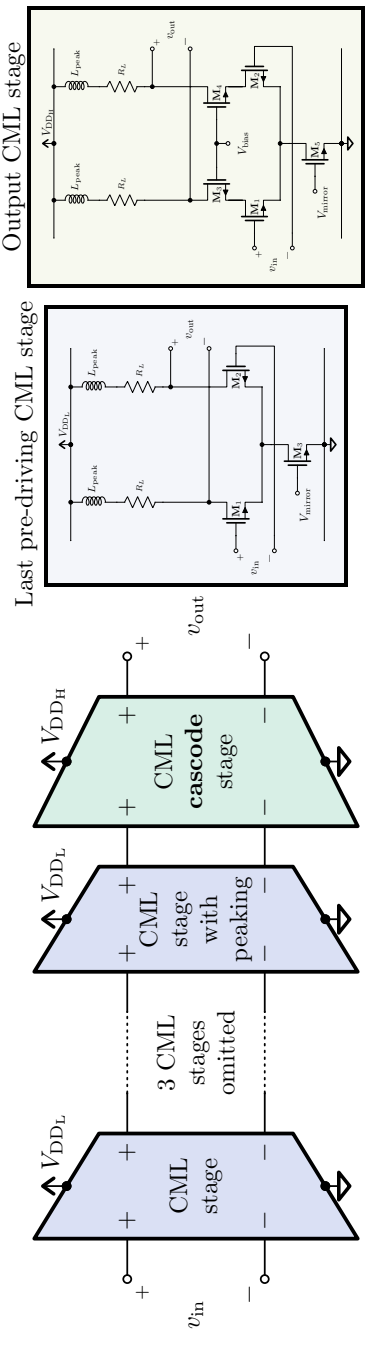


Fig. 2.1 Circuit architecture of the MZMD. Qualitative topologies of the last two stages are sketched on the right

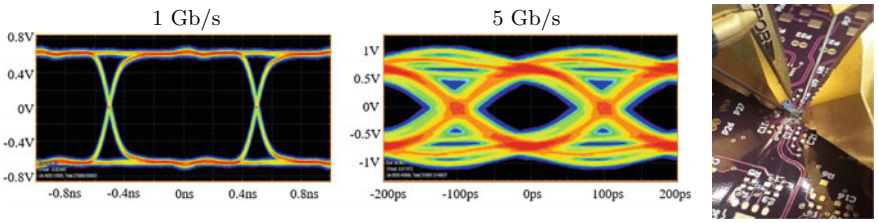


Fig. 2.2 Left: eye diagrams of driver output voltage at different bit-rates. Right: picture of the on-chip characterization setup

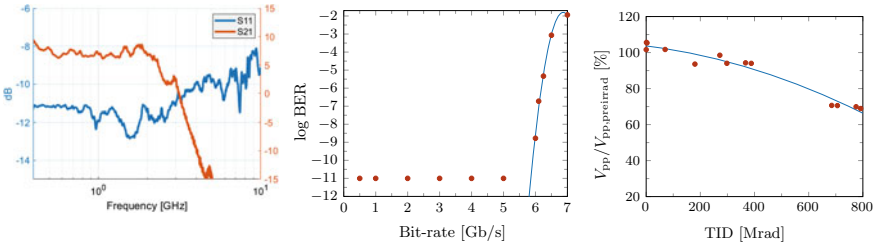


Fig. 2.3 MZMD circuit-level electrical characterization. Input-output S-parameters are reported on the left, while in the middle BER performances are shown. On the right, radiation-induced peak-to-peak voltage (V_{pp}) degradation is documented

reported in Fig. 2.3. At 800 Mrad, which was the highest dose level reached during the test because of limited testing time, the signal amplitude was reduced by 30% with respect to the pre-irradiation value.

2.4 System-Level Electo-Optical BER Testing

The fidelity of a data transmission system is ultimately quantified by the BER. An MZM fabricated in the Imec's isipp25g technology and the custom driver realized within this work have been hybridly integrated on a PCB. The only difference with respect to the testing scenario described in Sect. 2.3 is that the driver output pads are now wire-bonded with the MZM electrodes. The MZM under test is 1.5 mm-long and has no termination impedance. Measurements made on the same MZM with RF probes guarantee that its electro-optical modulation bandwidth remains above 5 GHz also with this load impedance mismatch, thus validating that the bandwidth-bottleneck remains in the electronic domain.

In the framework of fiber optic links, two types of characterization could be performed to carry out BER performances: optical noise loading and receiver sensitivity measurements. The former is an important metric for links which needs to be optically amplified while the latter is more suitable for non-amplified interconnects,

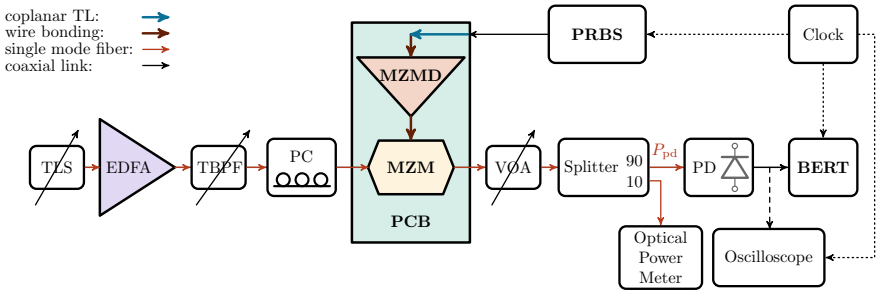


Fig. 2.4 Electro-optical setup for system-level characterization of an OOK link. Acronyms: TBPF (tunable band-pass filter), PC (polarization controller)

like those used within HEP experiments which cover 200 m at most. Hence, BER performances have been evaluated in function of bit-rate and received power on the photo-detector (PD).

As shown in Fig. 2.4, a standard on-off keying (OOK) transmitting system has been set up. The same PRBS-31 signal is applied at the driver input as before. A tunable laser source (TLS) was used to provide light in the C band near 1550 nm. The wavelength tuning allowed to set the MZM at the quadrature point. Light is coupled to the PIC using pigtailed fiber arrays and on-chip grating couplers. The modulated optical signal is attenuated with a variable optical attenuator (VOA) and then captured with a commercial PD, which is directly connected to the BER tester (BERT). Because of some issues encountered in the packaging procedure, which was performed manually, the fiber arrays resulted to be a little misaligned, causing an increase in optical insertion losses compared to similar devices realized in the same technology. Therefore, an erbium-doped fiber amplifier (EDFA) was required to perform BER tests. Even delivering the maximum rated optical power from the TLS the optical intensity at the MZM output was too low that an EDFA placed downstream the DUT failed to amplify the signal for photo-detection. The EDFA was then positioned *before* the MZM in the optical path, resulting in an injected power in the PIC of about 20 dBm, and an OSNR of 26 dB. Nevertheless, non-linear optical effects have not been captured throughout the measurement routines.

Optical eye diagrams and measured BERs as a function of input power P_{pd} on the photo-detector are shown for different data rates respectively in Figs. 2.5 and 2.6. The whole system is correctly working up to a bit-rate of 1.5 Gb/s while BER floors start to appear around 1.7 Gb/s, suggesting a systematic failure of the system. The eye diagrams at the PD output indeed report a sharp increase in jitter and inter-symbol interference (ISI) as the bit-rate reaches the 1.7 Gb/s level. Even if such poor speed achievements are in contrast with the previously presented BER performances of the stand-alone driver, these unexpected results could also be attributed to the non-optimum arrangement of wire bondings, as can be seen from Fig. 2.6.

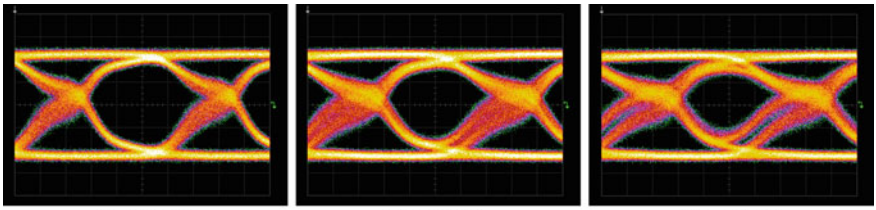


Fig. 2.5 Eye diagrams at the PD output for different data rates: **a** 1.5 Gb/s, **b** 1.7 Gb/s, **c** 1.75 Gb/s. All the plots have the same vertical scale of 20 mV/div

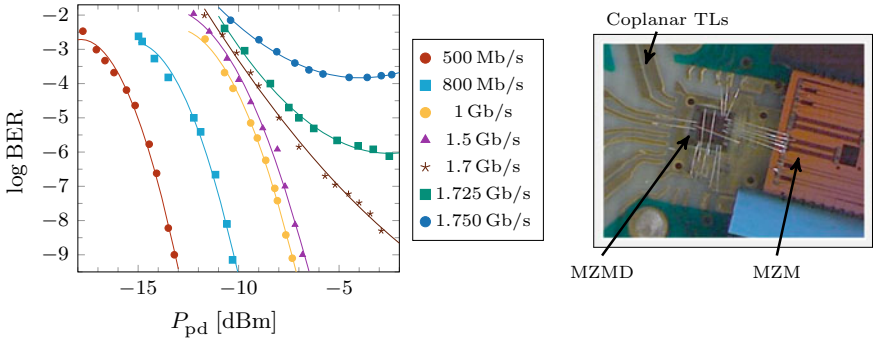


Fig. 2.6 BER performance of the data transmitting unit composed of the designed driver and a MZM bonded together. Acronyms: TL (transmission line)

2.5 Conclusions and Further Work

The design and the experimental characterization of a radiation-hard driver for a traveling-wave MZM have been reported at circuit-level as well as on a communication-link basis. Electrical measurements confirmed that the electronic driver is capable of withstanding data-rates up to 5 Gb/s as required by optical links specification in the HEP framework. A deviation from the expected speed capabilities has shown up during the system-level electro-optical BER test suggesting a leak in the packaging procedure. For this reason, further activities have already started to mitigate package-related parasitic effects and arrive to a working multi-Gb/s transmitter to be deployed in particle physics detectors. Also advanced solutions, such as flip-chip bump-bonding, are under investigation to avoid the usage of wire bondings between EICs and PICs when dealing with radio-frequency large signals as in this case study.

References

1. Kraxner A, Detraz S, Olantera L, Scarcella C, Sigaud C, Soos C, Troska J, Vasey F (2018) Investigation of the influence of temperature and annealing on the radiation hardness of silicon Mach-Zehnder modulators. IEEE Trans Nucl Sci 65(8):1624–1631
2. Colombo T, Amihalachioaei A, Arnaud K, Alessio F et al (2018) The LHCb online system in 2020: trigger-free read-out with (almost exclusively) off-the-shelf hardware. J Phys Conf Ser 1085:032041
3. Paternò A, Pacher L, Monteil E, Loddo F, Demaria N, Gaioni L, Canio FD, Traversi G, Re V, Ratti L, Rivetti A, Rolo MDR, Dellacasa G, Mazza G, Marzocca C, Licciulli F, Ciciriello F, Marconi S, Placidi P, Magazzù G, Stabile A, Mattiazzo S, Veri C (2017) A prototype of pixel readout ASIC in 65 nm CMOS technology for extreme hit rate detectors at HL-LHC. J Instrum 12(2):C02043
4. Zeiler M, El Nasr-Storey SS, Detraz S, Kraxner A, Olantera L, Scarcella C, Sigaud C, Soos C, Troska J, Vasey F (2017) Radiation damage in silicon photonic mach-zehnder modulators and photodiodes. IEEE Trans Nucl Sci 64(11):2794–2801
5. Faccio F, Borghello G, Lerario E, Fleetwood DM, Schrimpf RD, Gong H, Zhang EX, Wang P, Michelis S, Gerardin S, Paccagnella A, Bonaldo S (2018) Influence of ldd spacers and H^+transport on the total-ionizing-dose response of 65-nm MOSFETs irradiated to ultrahigh doses. IEEE Trans Nucl Sci 65(1):164–174
6. Ciarpi G, Saponara S, Magazzù G, Palla F (2019) Radiation hardness by design techniques for 1 grad tid rad-hard systems in 65 nm standard cmos technologies. In: Saponara S, De Gloria A (eds) Applications in electronics pervading industry, environment and society. Springer International Publishing, Cham, pp 269–276
7. Palla F, Ciarpi G, Magazzù G, Saponara S (2019) Design of a high radiation-hard driver for Mach–Zehnder modulators based high-speed links for hadron collider applications. Nucl Instrum Methods Phys Res Sect A Accelerators Spectrometers Detectors Assoc Equipment **936**:303–304 (Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors)
8. Ciarpi G, Magazzù G, Palla F, Saponara S (2018) Design of radiation-hard MZM drivers. In: 20th Italian national conference on photonic technologies (Fotonica 2018), May 2018, pp 1–4

Chapter 3
A Rad-Hard Bandgap Voltage Reference for High Energy Physics Experiments

G. Traversi, L. Gaioni, M. Manghisoni, M. Pezzoli, L. Ratti, V. Re, E. Riceputi and M. Sonzogni

Abstract This work is concerned with the characterization of a bandgap reference circuit, fabricated in a commercial 65 nm CMOS technology, designed for applications to HL-LHC experiments. Measurement results show a temperature coefficient of about 16 ppm/°C over a temperature range of 140 °C (from −40 to 100 °C) and a variation of 1.6% for V_{DD} from 1.08 to 1.32 V. The mean value of the bandgap output is about 400 mV, with a 5% maximum shift when exposed to a Total Ionizing Dose (TID) around 1 Grad (SiO_2). The power consumption is 165 μW at room temperature, with a core area of 0.02835 mm^2.

Keywords Bandgap voltage reference · Deep submicron · CMOS · Radiation effects · Total ionizing dose (TID)

3.1 Introduction

Voltage references, which provide precise, stable and temperature-insensitive DC voltages, are fundamental building blocks in mixed-mode circuits. The bandgap reference (BGR) is one of the most popular voltage reference that successfully achieves these requirements. It generates a voltage which is obtained from the sum of the voltage across a forward biased pn junction (inversely dependent on the absolute temperature) and a term directly proportional to the absolute temperature (PTAT). Unfortunately, this architecture is not suited for advanced CMOS technology where

G. Traversi (✉) · L. Gaioni · M. Manghisoni · V. Re · E. Riceputi · M. Sonzogni
Dipartimento di Ingegneria e Scienze Applicate, Università degli Studi di Bergamo, Via Marconi 5, 24044 Dalmine, BG, Italy
e-mail: gianluca.traversi@unibg.it

M. Pezzoli · L. Ratti
Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli Studi di Pavia, Via Ferrata 1, 27100 Pavia, Italy

G. Traversi · L. Gaioni · M. Manghisoni · M. Pezzoli · L. Ratti · V. Re · E. Riceputi · M. Sonzogni
Istituto Nazionale di Fisica Nucleare, Sezione di Pavia, Via Bassi 6, 27100 Pavia, Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_3

the supply voltage is 1.2 V or even lower. For this reason, in the last ten years, use of nonstandard devices in place of BJT or diodes has been proposed [1], but at the cost of a poor portability of the design and with the risks associated to the lack of accurate models for nonstandard devices. The resistive subdivision technique has been proposed to implement sub-1V BGR circuits [2], although this technique is not suitable for high-precision references working in a large temperature range. This paper discusses a BGR architecture based on a commercial 65 nm CMOS technology and capable of operating with 1.2 V supply. The proposed IP block has been designed for operation in the harsh radiation environment of the High Luminosity LHC. The 65 nm CMOS technology chosen for this prototype has been tested up to 1 Grad with promising results for CMOS transistors [3]. Nonetheless, other components of the BGR, namely bipolar devices, are affected by bulk damage effects. For this reason, in order to understand their behavior after irradiation, three different BGR versions (the first one based on parasitic PNP bipolar transistors, the second based on pn diodes and the third one based on enclosed-layout MOSFETs biased in weak inversion region) have been designed and submitted for fabrication in a prototype chip. These circuits have been fabricated and characterized before and after irradiation up to 225 Mrad(SiO_2) and the third design (the one based on MOSFETs) demonstrated the best performance in terms of radiation hardness [4]. Based on this work, a voltage reference circuit, designed in a commercial 65 nm CMOS technology and capable of operating in harsh radiation environments up to 1 Grad has been developed and its characterization is shown in this paper.

3.2 Operating Principle and Characterization Results

The bandgap circuit described in this paper and shown in Fig. 3.1, is based on a current mode approach [1]. Two currents, one (I_{2b}) proportional to absolute temperature (PTAT) and one (I_{2a}) complementary to absolute temperature (CTAT) are generated and summed in order to obtain a voltage insensitive to temperature. As already mentioned in the Introduction, with the purpose of increasing the radiation hardness of the circuit, only MOSFETs devices have been included in the circuit. In order to obtain a behavior similar to a bipolar transistor, they have been biased in the weak inversion region, where the I-V characteristic of the device is:

$$I_D = \frac{W}{L} I_0 \cdot exp\left(\frac{V_{GS} - V_{th}}{\eta V_T}\right) \cdot \left[1 - exp\left(-\frac{V_{DS}}{V_T}\right)\right] \tag{3.1}$$

where the V_{DS} dependence of the drain current can be neglected when $V_{DS} \geq 4V_T$. Being M1, M2 and M3 equally sized, the BGR output value is given by:

$$V_{REF} = \frac{R_3}{R_1}\left[V_{GS1} + \frac{R_1}{R_2}\Delta V_{GS}\right]. \tag{3.2}$$

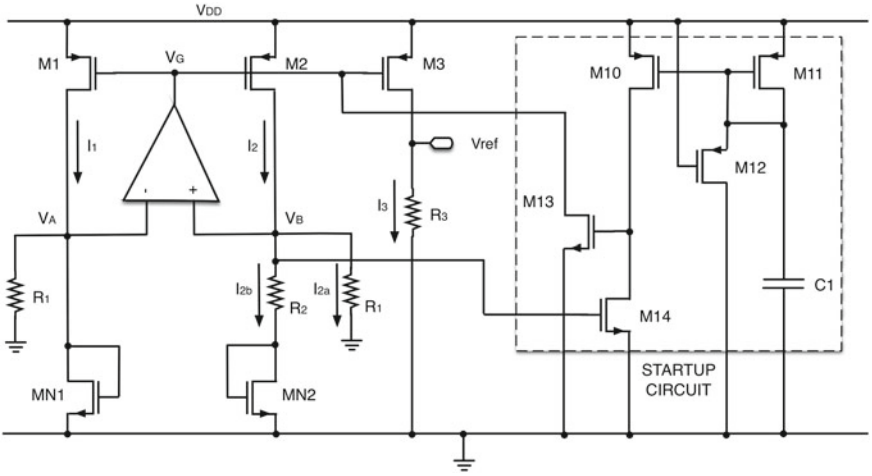


Fig. 3.1 Schematic of the bandgap reference together with the startup circuit

Since bandgap circuit has two stable operating points, it requires a start-up circuit to prevent operation in the undesired one. Figure 3.1 shows the startup circuit implemented [5]. It is based on a pull down capacitor. During the power on, a current starts to charge the capacitor C_1, the current is mirrored by M_{11} and M_{12} and it charges the gate of M_{13} thus turning the transistor on. M_{13} pulls down the gate of the bandgap current mirror injecting current into the bandgap. The power consumption of the startup circuit after power on is zero because, after startup, M_{14} is turned on and M_{13} is cutoff. Moreover, M_{12} discharges C_1 when power supply is switched off.

The proposed bandgap reference was fabricated in a commercial 65 nm CMOS technology. The chip microphotograph is presented in Fig. 3.2 (left). Extensive ex-

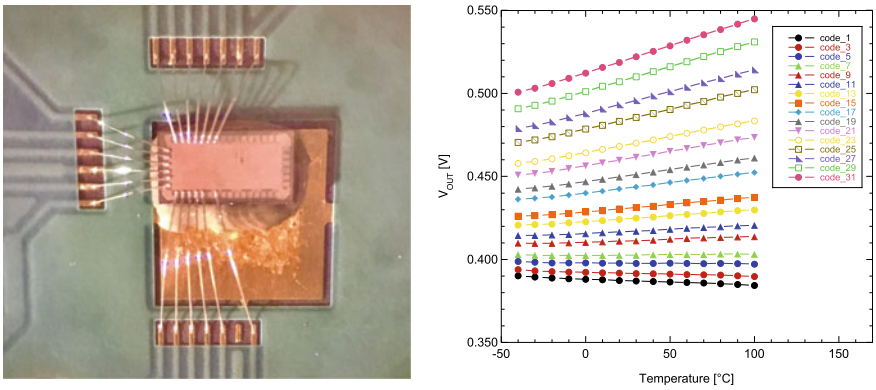


Fig. 3.2 (Left) Die microphotograph (2 mm × 1 mm); (right) measured temperature dependence of the bandgap reference voltage as a function of the temperature for different configuration bits of R_2

perimental measurements were performed in order to characterize the actual behavior of the proposed architecture. For example, the temperature behavior has been measured between −40 and 100 °C while the circuit will operate at −30 °C during the experiment and between about 20 and 40 °C during the operation without cooling. The measurements were performed using Keysight 34461A Digital Multimeter and GENVIRO-030LC Temperature Chamber. In order to be able to compensate for possible process and mismatch effects, the programmability of resistor R_2 (5 bits) has been included. Figure 3.2 (right) shows the measured output voltage as a function of the configuration word, while Table 3.1 summarizes the main characteristics of the bandgap circuits. The comparison shows that the proposed circuit provides the minimum variation of the reference voltage after irradiation. In addition, if needed, the line regulation of this work can be improved by adding a regulated cascode at the

Table 3.1 Performance summary of the proposed BGR circuit

	This work	[8]	[6]	[9]
Supply voltage (V)	1.2	1.2	1.2	1.2
Operating voltage range (V)	1.08–1.32	0.85–1.4	1.08–1.32	0.85–1.5
Nominal reference voltage (mV)	400	405	330	600
Line regulation (1.08–1.32 V) (%/V)	4	2.72	0.25	–
Temperature coefficient (ppm/°C)	16	30.5	130	15
Temperature range (°C)	−40 to 100	0–80	−40 to 80	−40 to 125
Power consumption @ 25 °C (μW)	165	–	240	60
Radiation induced ΔV_{REF}	5% @ 1 Grad	0.8% @ 45 Mrad	10% @ 800 Mrad	± 3% (5 samples) @ 450 Mrad
Layout Area (mm^2)	0.028	0.064	0.018	0.056
Technology	CMOS 65 nm	CMOS 130 nm	CMOS 65 nm	CMOS 130 nm

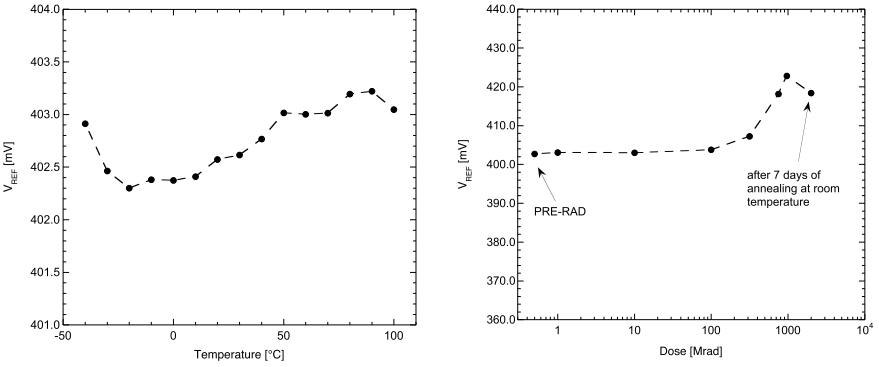


Fig. 3.3 Measured output voltage as a function of the temperature (left); measured output voltage of the bandgap as a function of the absorbed dose of 10 keV X-rays and after annealing (right)

output branch of the circuit, as implemented in [6]. The measured best temperature coefficient (TC) of the bandgap reference is 16 ppm/°C in a range of −40 to 100 °C, as shown in Fig. 3.3 (left).

Irradiation tests were carried out taking into account the unprecedented radiation tolerance requirements of demanding applications such as the HL-LHC [7]. To get an estimate of the performance of the bandgap circuit, we irradiated one device up to about 1 Grad(SiO_2) total dose of 10-keV X-rays. The irradiation was done at Laboratori Nazionali di Legnaro (Italy) with an X-ray machine at a dose rate of about 1 krad(SiO_2)/s. During irradiation, the bandgaps were biased as in the real application. Figure 3.3 (right) shows the variation of the output voltage as a function of the TID for the BGR with N-MOSFET. Annealing after one week at room temperature shows minor changes on the reference voltage with respect to the pre-irradiation value.

3.3 Conclusion

In this paper, a new radiation hard bandgap voltage reference circuit has been presented. The circuit has been characterized in a climatic chamber between −40 and +100 °C and irradiated up to 1 Grad(SiO_2), yielding up to 5% voltage change at the total ionizing dose. The BGR here proposed is able to face very high radiation doses, keeping a reasonable output accuracy, a relatively small area, and a simple architecture.

Acknowledgements The authors wish to thank Serena Mattiazzo and Devis Pantano (University of Padova) for providing the source for X-ray irradiation and for their constant support during the irradiation campaign, and Dr. Francesco De Canio for his contribution to the design and characterization activity. The authors are also in debt with Massimo Rossella (INFN Pavia) who have kindly made the climatic chamber available for the bandgap characterization.

References

1. Banba H et al (1999) A CMOS bandgap reference circuit with sub-1-V operation. IEEE J Solid State Circ 34:670
2. Neuteboom N, Kup BMJ, Janssens J (1997) A DSP-based hearing instrument IC. IEEE J Solid State Circ 32:1790–1806
3. Menouni M et al (2015) 1-Grad total dose evaluation of 65 nm CMOS technology for the HL-LHC upgrades. J Instrum 10(5), art. No. C05009
4. Traversi G et al (2016) Characterization of bandgap reference circuits designed for high energy physics applications. Nucl Instrum Methods A 824:371–373
5. Li W, Yao R, Guo L (2009) A low power CMOS bandgap voltage reference with enhanced power supply rejection. In: Proceedings of the 8th IEEE international conference on ASIC, pp 300–304
6. Vergine T, De Matteis M, Michelis S, Traversi G, De Canio F, Baschirotto A (2016) A 65 nm rad-hard bandgap voltage reference for LHC environment. IEEE Trans Nucl Sci 63(3):1762–1767

7. Garcia-Sciveres M, Christainsen J (2013) RD collaboration proposal: development of pixel readout integrated circuits for extreme rate and radiation. CERN-LHCC-2013-008, LHCC-P-006
8. Gromov V, Annema AJ, Kluit R, Visschers JL, Timmer P (2007) A radiation hard bandgap reference circuit in a standard 0.13 μm CMOS technology. IEEE Trans Nucl Sci 54(6):2727–2733
9. Cao Y, De Cock W, Steyaert M, Leroux P (2013) A 4.5 MGy TID-tolerant CMOS bandgap reference circuit using a dynamic base leakage compensation technique. IEEE Trans Nucl Sci 60(4):2819–2824

Chapter 4
Analysis and Comparison of Ring and LC-Tank Oscillators for 65 nm Integration of Rad-Hard VCO for SpaceFibre Applications

D. Monda, G. Ciarpi, G. Mangraviti, L. Berti and Sergio Saponara

Abstract The paper presents the comparison between two VCO (Voltage Controlled Oscillator) architectures designed in 65 nm CMOS for aerospace applications. In particular, the two VCOs have been designed targeting the 6.25 GHz frequency required in the SpaceFibre standard. The ring oscillator has been designed using three current mode logic stages connected in a loop. Although its performance in terms of low area occupation are attractive, the process variations simulations have demonstrated its inability to generate the target frequency in harsh operating conditions. Instead, the LC-Tank based oscillator, fixing the central frequency with the resonance of the L-C tank, has highlighted a lower influence through Process-Voltage-Temperature simulations on the oscillation frequency. Thanks to varactor-based voltage tuning control, it is able to cover the range from 5.18 to 6.41 GHz. Both architectures are biased with a supply voltage of 1.2 V. The complete layout of the last solution has been designed and its parasitic has been extracted for post-layout simulations. Achieved results are attractive to address the requirements of the new SpaceFibre aerospace standard.

Keywords Ring oscillator · LC-tank oscillator · SpaceFibre · Rad-hard circuit

4.1 Introduction

Current trends in satellites show a rapid increase in data traffic and digital processing. The throughput of next generation digital telecom satellites will exceed terabits per second of data, which have to be processed on board. For instance, the high-resolution cameras and synthetic aperture radars need high-speed communications between the instruments and storage [1]. The optical technology, thanks its high bandwidth-length product, the lightweight cabling and electromagnetic hardness, can potentially be the solution for data-rate increment in satellite. In this direction, the European

D. Monda · G. Ciarpi (✉) · S. Saponara
Department Information Engineering, University of Pisa, Pisa, Italy
e-mail: gabriele.ciarpi@ing.unipi.it

G. Mangraviti · L. Berti
IMEC, Louvain, Belgium

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_4

Space Agency (ESA) has recently released the new SpaceFibre standard for on-board satellite communication up to 6.25 Gbps [2, 3]. The communication performance is strongly related to the ability to synchronize the receiver and the transmitter and a key block for the synchronization is the Phase Locked Loop (PLL). The core system inside the PLL able to generate the suitable frequency is the Voltage Controlled Oscillator (VCO). It should be able to generate a tone at 6.25 GHz and be tolerant to SEE (Single Event Effects) and TID (Total Ionization Dose) up to 300 krad [4] as the whole PLL system. In literature, there are not examples of rad-hard VCOs able to work at 6.25 GHz. In [5] a comparison between Ring Oscillator (RO) and LC-Tank (LC) VCO for PLL were made for Large Hadron Collider's (LHC) for High Energy Physics (HEP) applications. Both were designed for a working frequency of 2.56 GHz and, after being exposed to irradiation, the LC oscillator showed a lower frequency shift than that of the RO solution and a jitter value one order of magnitude lower.

The goal of this work is to compare the performances of the widely used RO and LC circuits in radiation environments and to contribute with new approaches for exploiting the characteristics that have made these systems the most implemented. For a better comparison, both the VCOs were designed using the same 65 nm commercial-grade technology, which thanks its thin gate oxide is considered a radiation hard technology [6, 7]. The design of the VCO based on the ring oscillator and that based on the LC-Tank approach is presented in Sects. 2 and 3, respectively. Section 4 provides preliminary layout design and post-layout circuit performance results. Conclusions are drawn in Sect. 5.

4.2 Cascaded CML-Inverter Ring Oscillator in 65 nm Technology

A RO-VCO consists on a cascade of inverting amplifier in which the output of the last stage is connected to the first stage, as shown in the model of Fig. 4.1, where g_m and R are the transconductance and the equivalent output resistance, respectively of each stage, and C is the equivalent input capacitance of the following stages.

According with the Fig. 4.1 the open-loop gain of the system composed of N generic stages is expressed in Eq. 4.1.

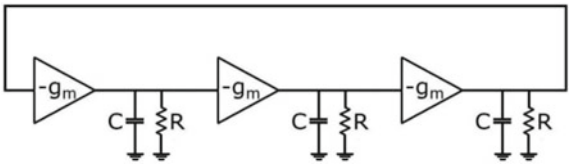


Fig. 4.1 Ring oscillator modalized using inverting stage amplifiers

$$H(j\omega) = \left(-\frac{g_m R}{1 + j\omega RC}\right)^N \tag{4.1}$$

For the Barkhausen oscillation criterion [8], the module of the transfer function has to be higher than one for the start-up condition and then equal to one to sustain the oscillation, while the transfer function phase has to be an integer multiple of 2π. Appling this criterion at the model in Fig. 4.1, we obtain the oscillation condition in term of design parameters, expressed in Eq. 4.2.

$$g_m R \geq \frac{1}{\cos\theta} \tag{4.2}$$

where θ is the phase shift introduced by each RC load, which for the Barkhausen criterion has to be an integer multiple of π/N.

In order to limiting the frequency variation due to process technology and to reduce area and power consumption, a number of three stages was chosen for the RO-VCO design. With this choice, in according with Eq. 4.2, the following condition (Eq. 4.3) is extracted as the main design guideline.

$$g_m R \geq 2 \tag{4.3}$$

The designed RO-VCO is composed by three CML (Current Mode Logic) stages, which thanks to their lower voltage swing and lower output impedance allow to reach higher frequency performance than the use of the standard CMOS approach [9]. Moreover, the use of a differential structure allows to obtain higher common mode disturb immunity than the use of a single ended structure, as CMOS circuits.

The single CML stage, shown in Fig. 4.2, is made by a differential pair amplifier with a resistive load.

The oscillation frequency of the RO-VCO is expressed by the relation $f_0 = 1/(2\pi RC)$. Where R is the parallel between the pull-up CML resistive load and the output MOSFET resistance, while C is the gate capacitance of the following stage. In order to make a control of the oscillation frequency a couple of varactors were added at the output of each stage. Accumulation n-MOSFETs devices were used to design varactors and increasing or decreasing their gate voltage, their capacitances change shifting the oscillation frequency.

The small length size n-MOSFETs allows to achieve high frequency performance, but on the other hand, this choice increase the deviation of the device's parameters from the typical condition. Although the use of varactors for frequency tuning, the frequency shift during the process corner simulations was so high that cannot be compensated using the control voltage.

Table 4.1 lists the oscillation frequency and the tuning range values of the RO-VCO for the three corners process. The frequency values reported are extracted by schematic simulations performed with the minimum and the maximum values of the varactor tuning voltages. The oscillation frequency in the slow-slow corner case does not reach the 6.25 GHz frequency value required by the SpaceFibre standard,

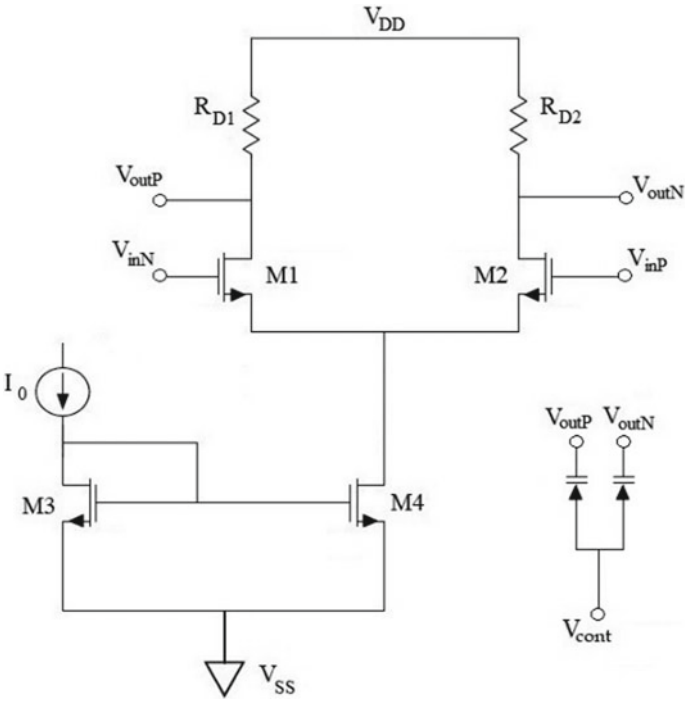


Fig. 4.2 Schematic of the single stage of the ring oscillator and the couple of varactors connected at the outputs

Table 4.1 Frequency range of the RO-VCO, expressed as function of the minimum and maximum control voltage value

Technology corner	Frequency (GHz)	Tuning range (GHz)
Fast-fast	7.61–9.11	1.50
Typical	5.65–6.70	1.05
Slow-slow	4.33–5.10	0.77

even using the maximum value of the control voltage. In the fast-fast corner case, the frequency is higher than the targeted frequency even with the minimum value of the control voltage. RO-VCO is strongly dependent on the device parameters making it not usable for this application.

4.3 LC-Tank Rad-Hard Oscillator in 65 nm Technology

In order to overcome the effects of the device parameters deviation on the oscillation frequency, a LC-Tank VCO architecture was designed to be compliance with the SpaceFibre protocol. This architecture bases its oscillation frequency on the filtering

effect of a L-C tank, leaving to active components only the role of setting the feedback gain [10] and compensate the loss of the inductor. Figure 4.3 shows the schematic of the LC-VCO designed to generate the target 6.25 GHz frequency.

A poly-silicon resistor is used to shift the output common mode level at VDD/2, preventing the damaging or lifetime reduction of the low-voltage MOSFETs used for the cross-coupled pair. This resistor is connected to the center tap of a symmetrical inductor chosen for its lower layout area than that of two separate inductors. In order to achieve the best frequency performance of this technology, the cross-coupled pair is sized using minimum length mosfets and a mosfet width of 3.6 μm to guarantee a cell gain of at least 6 dB for start-up condition. The design guideline to respect Barkhausen oscillation criterion should be $g_m > 1/R_p$, where g_m is the transconductance of the n-MOSFETs inside the cross-coupled cell and R_p is the parasitic resistance of the inductor [11]. The oscillation frequency of the LC-VCO is set by $f_0 = 1/\left(2\pi\sqrt{LC}\right)$ making possible to tune the central frequency with the use of two varactors connected at the LC output and using a control voltage in the range 0 V–V_{DD}.

In Fig. 4.4 is shown the frequency response of the VCO for the two extreme values of the control voltage, highlighting a tuning range of 1.23 GHz. Moreover, Fig. 4.4

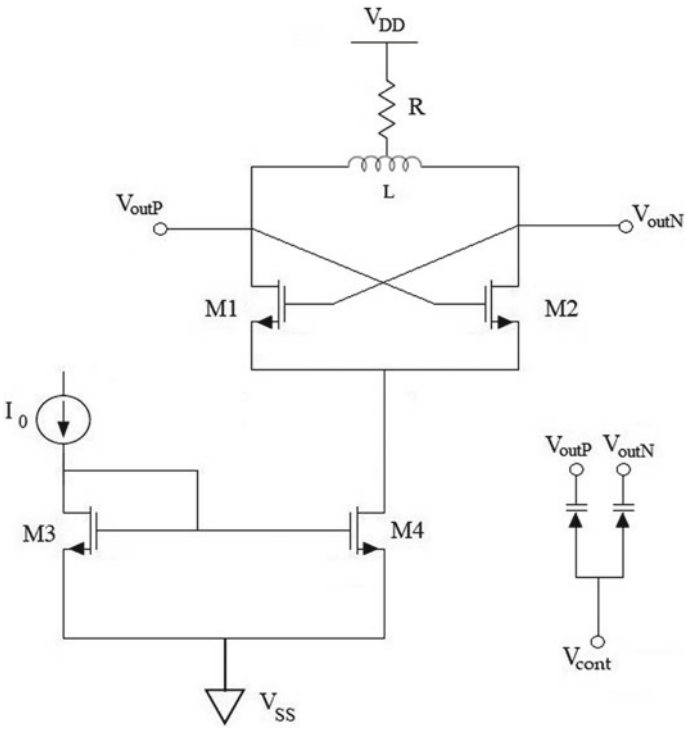


Fig. 4.3 Schematic of the LC-tank oscillator

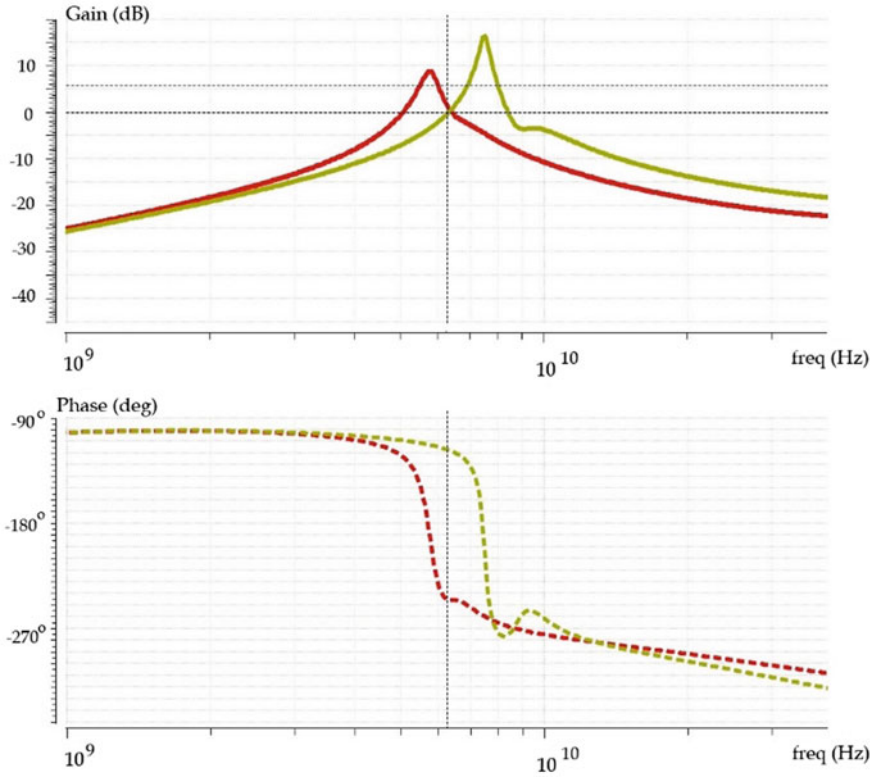


Fig. 4.4 Frequency response of the LC-VCO for control voltage equal to 0 V (red line) and for 1.2 V (yellow line); dot lines represent the phase for minimum and maximum value of the control voltage, respectively

shows a minimum cell gain of about 10 dB, for the minimum value of the control voltage, allowing to achieve a robust start-up condition for the oscillator.

Corner simulations were performed by changing the production process, temperature and supply voltage. The SpaceFibre standard requires to the system to properly work under harsh condition. In particular, the system was tested for temperature variations in the range -55 to 125 °C, fast-slow-typical process corners and for $\pm 10\%$ supply voltage and polarization current deviations.

4.4 LC-Tank Oscillator Layout

The layout for the VCO is shown in Fig. 4.5 where about the 85% of the total area is occupied by the inductor. For the design of this layout, all choices were made in order to reduce the parasitic resistance and to guarantee a good matching of simple current

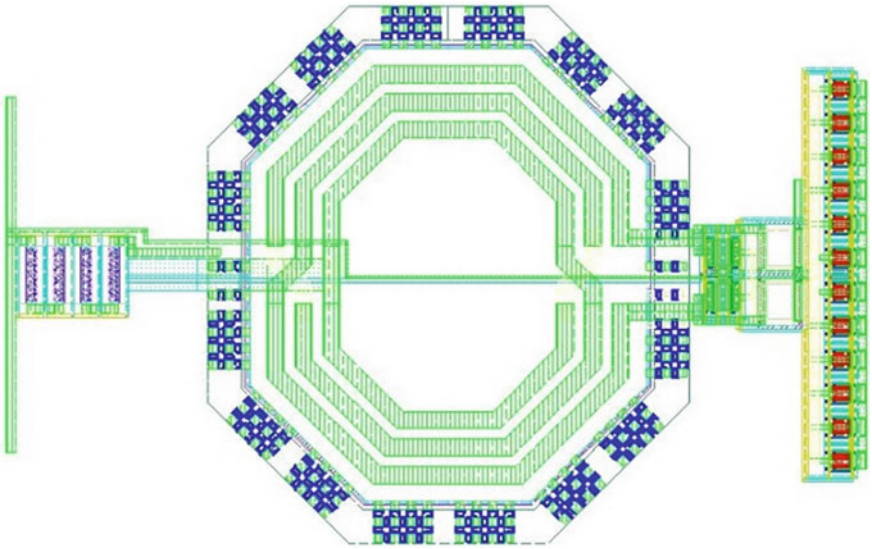


Fig. 4.5 Layout of the LC-VCO. From left to right there are the poly resistance, the inductor, the differential pair, varactors and the current tail mirror, respectively

mirror and cross coupled cell. A high parasitic resistance leads to a gain degradation and a weak start-up condition. For the simple current mirror, the two mosfets used to implement the diode MOSFETs were placed in the center of the other ten MOSFETs.

The space between the devices is the minimum allowed by technology and the Design Rule Check (DRC), helping to minimize the devices mismatch.

Post layout simulations show a tuning range of the LC-VCO from 5.18 to 6.41 GHz in the worst condition, highlighting the capability of this VCO to be used in the SpaceFibre communication protocol.

4.5 Conclusions and Future Work

In this work the comparison between two VCOs designed in 65 nm technology is made, targeting the SpaceFibre protocol applications. Although the RO-VCO is an appetible VCO configuration in terms of are occupancy, power consumption and tuning range than the other configuration, it is strongly dependent on the device parameters making it not usable for 6.25 GHz applications, as SpaceFibre protocol.

On the other side, the LC-VCO, despite its large area, mainly occupied by the inductor, presents promising performance in terms of frequency range, covering the 5.18–6.41 GHz range with a control voltage swing of 1.2 V.

The LC system has been integrated in a chip containing a 65 nm SERDES (Serializer-Deserializer) to test system level performance. The whole chip will be

electrically tested in standard condition and will be exposed to X-rays to achieve the 300 krad TID. and to heavy ions for SEE characterization.

References

1. Xie L, Wei L (2013) Research on vehicle detection in high resolution satellite images. In: IEEE fourth global congress on intelligent systems
2. ESA Requirements and Standards Division ESTEC, P.O. Box 299, 2200 AG Noordwijk The Netherlands. Space engineering, SpaceFibre—very high-speed serial link. European Space Agency for the members of ECSS, 2019
3. Parkers S, Ferrer A et al (2017) SpaceFibre specification draft K1. Copyright 2017, University of Dundee
4. Ciarpi G, Magazzù G et al (2018) Design of radiation-Hard MZM drivers. In: 20th Italian national conference on photonic technologies (Fotonica 2018), vol 26, pp 1–4
5. Prinzie J, Christiansen J et al (2017) Comparison of a 65 nm CMOS ring- and LC-oscillator based PLL in terms of TID and SEU sensitivity. IEEE Trans Nucl Sci 64(1):245–252
6. Ciarpi G, Saponara S et al (2019) Radiation hardness by design techniques for 1 grad TID rad-hard system in 65 nm standard CMOS technologies. In: Application in electronics pervading industry, environment and society, pp 269–276
7. Palla F, Ciarpi G et al (2019) Design of a high radiation-hard driver for Mach-Zehnder Modulators based high-speed links for hadron collider applications. Nucl Instrum Methods Phys Res Sect A 936:303–304
8. Voinigescu S (2013) High-frequency integrated circuits. Cambridge University Press
9. Heydari P (2003) Design and analysis of low-voltage current-mode logic buffers. In: Fourth international symposium on quality electronic design. IEEE
10. Razavi B (1996) A study of phase noise in CMOS oscillators. IEEE J Solid State Circ 31(3):331–343
11. Razavi B (1998) RF microelectronics, vol 1. Prentice Hall, Upper Saddle River, NJ

Chapter 5
A Compact Gated Integrator for Conditioning Pulsed Analog Signals

Sara Pettinato, Andrea Orsini, Maria Cristina Rossi, Diego Tagnani, Marco Girolami and Stefano Salvatori

Abstract An extremely compact gated integrator prototype has been realized and preliminarily characterized. Front-end section of the circuit is based on the high precision integrator IVC102, whereas the analog to digital conversion and data-acquisition, as well as the timing control, are performed by an LCP845 microcontroller. The system synchronizes signal detection with an external trigger generated in coincidence with the source pulse, i.e. the gated integrator amplifies the signal only when a pulse is generated, increasing significantly the signal-to-noise ratio. As a consequence, the proposed circuitry would represent an affordable, sensitive, and cost-effective alternative to the continuous-time regime measurement-technique largely adopted, for example, in radiation dosimetry.

5.1 Introduction

The development of increasingly sophisticated techniques for radiotherapy led in recent years to the requirement of dosimeters characterized by high sensitivity, accuracy, reliability and high spatial resolution to follow the dose gradient delivered to the patient [1]. However, especially when small fields are concerned (e.g. in IMRT, Intensity Modulated Radiation Therapy), small (<1 mm^3) diamond detectors [2, 3] can

S. Pettinato (✉) · A. Orsini · D. Tagnani · S. Salvatori
Engineering Department, Università degli Studi Niccolò Cusano,
via don Carlo Gnocchi 3, 00166 Rome, Italy
e-mail: sara.pettinato@unicusano.it

M. C. Rossi
Engineering Department, Università degli Studi Roma Tre,
via Vito Volterra 62, 00146 Rome, Italy

D. Tagnani
INFN, Sez. Roma Tre, via della Vasca Navale, 00146 Rome, Italy

M. Girolami
Istituto di Struttura della Materia (ISM), Consiglio Nazionale delle Ricerche (CNR),
Via Salaria, 00015 Monterotondo Scalo, Rome, Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_5

be used as valid solid-state alternatives to ionization chambers, due to their peculiar chemical-physical characteristics, such as tissue equivalence and radiation hardness. Diamond has also been shown to be an excellent material for the detection of UV [4], soft X-rays [5, 6], charged particles [7, 8] and neutrons [9]. Regardless of the type of detector, appropriate techniques are required to measure the photocurrent (which extends from a few pA to a few nA) or the charge collected by the dosimeter [10]. The typical measurement method is based on the use of an electrometer able to measure either currents or charges in a continuous-time regime in wide ranges with very high resolution [11, 12]. However, when fast and repetitive signals are concerned [13], continuous integration may imply significantly long periods of time in which noise is the only input of the front-end electronics, resulting in a non-optimized signal-to-noise ratio (SNR). Conversely, gated integrating technique represents a suitable approach to pulsed signal conditioning [13]. It is based on the synchronization between the signal detection and the pulse emission from the source, ensured by an external trigger generated in coincidence with the source. This implies that signal conditioning occurs only in a time interval around which a pulse is generated, thus leading both to a higher signal-to-noise ratio (SNR) and a better sensitivity in comparison to conventional continuous integration. In particular, SNR increases by a factor of $\sqrt{N}$ for a periodic signal, where N represents the number of averaged measurements [14, 15]. It appears clear that synchronous detection method would be particularly effective in case of X-ray pulses generated in a linear accelerator (LINAC) apparatus used in radiotherapy [16] and detected by a diamond dosimeter. Synchronous detection, therefore, would assure superior performances in terms of sensitivity, accuracy and system dynamics, in order to satisfy the necessary Quality Assurance (QA) requirements of modern RT treatment protocols. In this work, we introduce the prototype of a high precision gated integrator, specifically designed for detectors employed in dosimetric applications where weak charge pulses are concerned. Points of novelty of the prototype are its cost-effectiveness and compactness if compared to commercial devices. Indeed, the proposed solution is based on the low-cost high-precision switched integrator IVC102 [17] which represents an effective and commercially available solution for accurate charge/current measurements [10, 18]. An LPC845 microcontroller unit is used for signal acquisition and processing, as well as to generate all the internal control signals. Preliminary characterizations in the 0.1–10 pC range have been performed to verify the effectiveness of the realized circuit. The prototype showed excellent performance in terms of linearity and sensitivity, with values comparable to those reported for state-of-the art electrometers used for routine dosimetry [11].

5.2 Circuit Description and Preliminary Characterization

Figure 5.1 shows the schematic of the proposed gated-integrator circuitry. The front-end section is based on the commercially available switched integrator transimpedance amplifier IVC102 (by Texas Instruments) and an inverting amplifier stage

useful to establish, at reset ($V_O = 0$ V), an ADC input voltage around 0.7 V. The read-out section is based on the microcontroller LPC845 (by NXP) equipped with an ARM-Cortex M0+ processor. IVC102 chip integrates high quality metal/oxide capacitors characterized by low leakage, excellent dielectric characteristics (typical non-linearity of ±0.005%) and temperature stability (±25 ppm/°C) [15]. The IVC102 output voltage, which is proportional to the integrated input charge provided by the detector, is digitally converted by the 12-bit successive approximation A/D converter embedded in the LPC845 microcontroller.

The measurement cycle starts by resetting the integrator output at 0 V (closing the internal switch S2) and integration begins when S2 is open and the charge is transferred to the integration capacitor closing the S1 switch. A dual power supply voltage of $V_{CC} = \pm 15$ V was used for the IVC102, whereas microcontroller unit, hence its internal ADC also, is supplied at $V_{DD} = 3.3$ V. The timing control circuitry of the system uses the State Configurable Timer (SCT) integrated in the LPC845 microcontroller, which is used to generate the timing signals for IVC102 S1 and S2 MOS switches synchronized to the external *sync* signal. Figure 5.2 shows an example of S1 and S2 control signals generated by the realized prototype, as well as the voltage at the ADC input obtained by leaving float IVC102 input. The example reported in Fig. 5.2 highlights the case in which an integration period T_{INT} (S2 open, S1 closed) is located across the rising edge of the synchronism signal. It is worth to observe that, to null any error induced by charge transferred at the integrator input during switches commutations, signal acquisition is performed in two phases, before (pre-hold) and after (hold) the T_{INT} period. As shown in Fig. 5.2, two opposite voltage step ΔV_Q are found both at the start and at the end of the integration period. Therefore, the net contribution of offset charge injection becomes insignificant if the integration result is measured as the voltage difference $V_B - V_A$.

The SCT is a tool that can perform advanced timing and control operations with little or no CPU intervention. It allows comparing the timer-counter value with a match register content, as well as storing the current timer value in capture registers when certain conditions/events occur. Moreover, it supports distinct user-defined

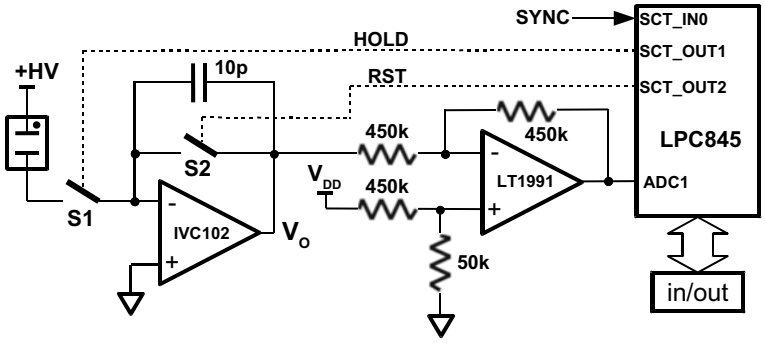


Fig. 5.1 Schematic of the proposed circuitry based on IVC102 integrator and LPC845 microcontroller

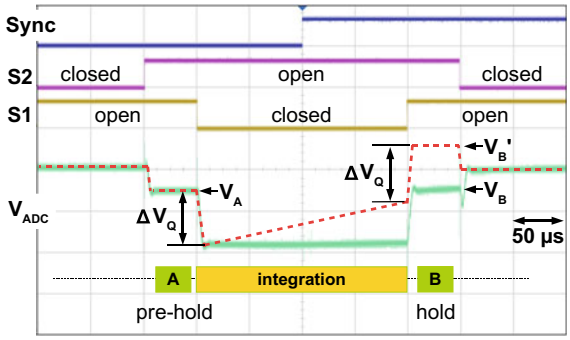


Fig. 5.2 Signal acquisition is performed in two phases, "pre-hold" and "hold", before and after the integration period, respectively, in order to null errors due to charge transfer during S1 switch commutations. V_{ADC} continuous (green) and dotted (red) lines refer to absence and presence of input current, respectively

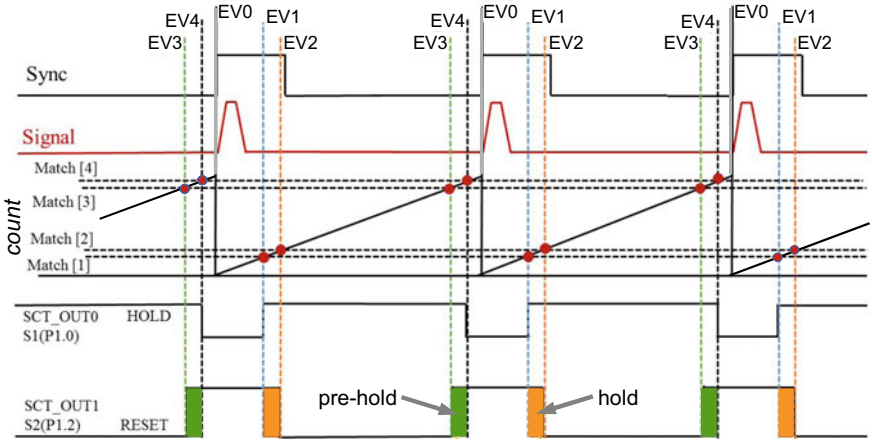


Fig. 5.3 Timing for S1 and S2 IVC102 switches performed by the SCT embedded in the LPC845 microcontroller

events based on a combination of parameters, including a match on one of the match registers. In our case the SCT was used as a 32-bit up-counter timer, with a clock frequency of 30 MHz, to manage five events, with one input (sync signal) and two outputs (S1 and S2 digital controls).

The time diagram reported in Fig. 5.3 refers to an input-pulse (red) generated in correspondence of the *sync* signal rising edge. In such a case the pre-hold periods have to occur *before* the sync signal arrival: by measuring the time period between two sync rising edges, the system will generate the pre-hold before the *next* pulse (pre-hold end time is calculated taking into account the possible jitter of pulse repetition rate). Hence, on the *sync* rising edge, event EV0 is generated: the timer counter value

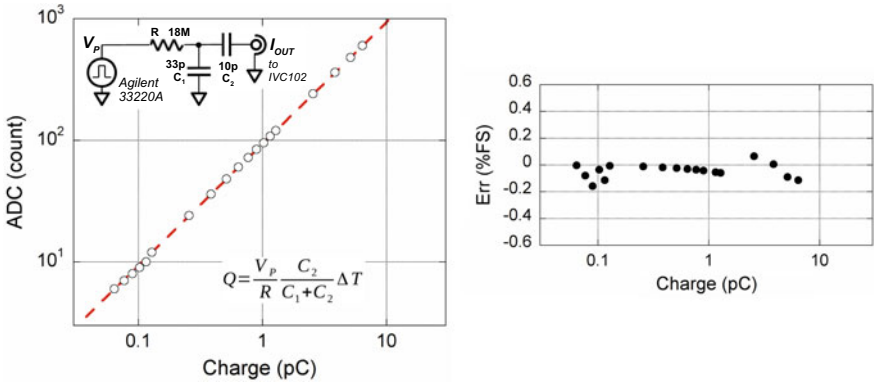


Fig. 5.4 ADC output as a function of injected charge packets (see formula) achieved by a pulsed voltage source (Agilent 33220A) coupled to the RC network reported in the inset. On the right, the error over the full scale for the investigated input-charge range

is captured and timer restarts. After restart condition, when the timer counter reaches the Match[1] (Match[2]) value, event EV1 (EV2) is generated. Match[1…4] values are user defined (in our case 100 and 150 μs, see Fig. 5.2). Events EV1 and EV2 are used for S1 and S2 control signals transitions and determine the hold-phase start and end times, respectively. During this period, the ADC acquires the V_B voltage amplitude.

Representing a measure of the time period T between two pulses, the captured timer value on EV0 event allows to calculate the match values Match[3] and Match[4]: the former, T—150 μs, represents the pre-hold start time; the latter, T—100 μs, the pre-hold end time (see Fig. 5.2). Obviously, such a pre-hold period will be used for V_A acquisition in correspondence of the *next* input pulse to calculate proper $V_B - V_A$ amplitude (here V_B represents the quantity acquired after the next EV0 event).

A preliminary characterization was performed in the lowest measurement range using the 10 pF internal capacitor of IVC102 in order to evaluate the circuit capability to acquire typical charge packets generated by a detector irradiated by a pulsed source. Data of Fig. 5.4 refer to mean values of $N = 512$ pulse acquisitions. Pulsed signals were emulated with an Agilent 33220A function generator, providing voltage pulses with amplitude in the 100 mV − 10 V range, 50 μs duration, and 500 Hz repetition rate. The function generator output was coupled to an RC network (see the inset of Fig. 5.4) to emulate charge packets in the 0.1 − 10 pC range generated by a detector having an equivalent 10 pF capacitance.

As can be seen from the best fit of experimental data shown in Fig. 5.4, the system shows excellent performance in terms of linearity in the investigated range of charge packets. The relative error, calculated with respect the nominal expected values, is lower than ±0.2%, and less than 0.04% for an input charge around 1 pC. Worth to mention a sensitivity of about 40 fC, estimated by the peak-to-peak output noise measured amplitude lower than 4 mV at IVC102 output.

5.3 Conclusions

The feasibility of a compact gated-integrator, implementing the precision switch-integrator transimpedance amplifier IVC102, has been demonstrated. Timing circuitry, based on the versatile State Configurable Timer embedded into the microcontroller used for signal acquisition and processing, synchronizes the integration period to the *sync* signal provided by the source. Moreover, the adopted two-phase differential measurements allow to null errors induced by charge transfer during integrator MOS-switch commutations. Experimental results demonstrate excellent linearity (with a relative error lower than $\pm 0.2\%$), and a sensitivity of 40 fC, which is a value comparable to those reported for state-of-the-art devices, all ranging from 10 to 30 fC [11]. More detailed analyses will be performed to evaluate measurement stability as well as its temperature dependence. Also, on-field measurements (clinical dosimetry) are planned for the next future. Finally, to null any offset induced by unavoidable asymmetric charge transfer during MOS switch commutation induced by the particular value of detector capacitance, a system upgrade will be implemented with the SCT, performing a two-phase real-time measurement of "zero-signal" in the midpoint time between two consecutive pulses.

Acknowledgements The authors would like to thank Marco Pacilli and Fabrizio Imperiali for fruitful discussions and technical support.

References

1. Bucciolini M et al (2003) Diamond detector versus silicon diode and ion chamber in photon beams of different energy and field size. Med Phys 30(8):2149–2154
2. Tromson D et al (2010) Single crystal CVD diamond detector for high resolution dose measurement for IMRT and novel radiation therapy needs. Diam Relat Mater 19:1012–1016
3. Marsolat F et al (2013) Diamond dosimeter for small beam stereotactic radiotherapy. Diam Relat Mater 33:63–70
4. Girolami M et al (2012) Diamond detectors for UV and X-ray source imaging. IEEE Electron Device Lett 33:224–226
5. Girolami M et al (2012) Optimization of X-ray beam profilers based on CVD diamond detectors. J Instrum 7:C11005
6. Conte G et al (2007) X-ray diamond detectors with energy resolution. Appl Phys Lett 91:183515
7. Salvatori S et al (2017) Nano-carbon pixels array for ionizing particles monitoring. Diamo Relat Mater 73:132–136
8. Pacilli M et al (2012) Polycrystalline CVD diamond pixel array detector for nuclear particles monitoring. J Instrum 8:C02043
9. Muraro A et al (2016) First neutron spectroscopy measurements with a pixelated diamond detector at JET. Rev Sci Instrum 87:11D833
10. D'Antonio E et al (2018) High precision integrator for CVD-diamond detectors for dosimetric applications. In: 2018 IEEE international symposium on medical measurements and applications (MeMeA), pp 1–6
11. See for example https://www.ptwdosimetry.com/en/products/unidos-webline/

12. Khan FM, Gibbons JP (2014) Khan's the physics of radiation therapy. Lippincott Williams & Wilkins, Philadelphia
13. Reichert J, Townsend J (1964) Gated integrator for repetitive signals. Rev Sci Instrum 35:1692–1697
14. Betts J (1970) Signal processing, modulation and noise. English Universities Press, London
15. Collier JL et al (1996) A low-cost gated integrator boxcar averager. Meas Sci Technol 7:1204
16. See for example ClinacR iX System, Varian Medical Systems, Inc., CA (USA). https://www.varian.com/oncology/products/treatment-delivery/clinac-ix-system
17. Precision switched integrator transimpedance amplifier, IVC102, datasheet, Texas Instruments. www.ti.com/lit/ds/symlink/ivc102.pdf
18. Salvatori S et al (2006) Compact front-end electronics for low-level current sensor measurements. Electron Lett 42:682–684

Part II
Internet of Things

Chapter 6
Multivariate Microaggregation with Fixed Group Size Based on the Travelling Salesman Problem

Armando Maya López and Agusti Solanas

Abstract Due to the growing use of IoT and 5G technologies, data are collected at an unprecedented pace. These data are used to improve decision-making processes. However, they could endanger individuals privacy, which is protected by international regulations. In this article, we propose a privacy-preserving microaggregation technique, inspired by the Travelling Salesman Problem, to protect individuals privacy through k-anonymity. We recall the basics on microaggregation and the TSP and, we describe the algorithm behind our approach. Also, we report experiments with real benchmark data sets showing that our approach outperforms current methods for low cardinality values.

Keywords Microaggregation · Travelling Salesman Problem · Privacy

6.1 Introduction

The massive use of information technologies, pervasive electronic devices, and telecommunications, in all areas of our society, has opened the door to the gathering of huge amounts of data. With the aim to obtain information and knowledge from these data [7], new disciplines focused on data analysis have been created, namely Data Science, Data and Process Mining [1], Big Data Analytics, Deep Learning, and so on. Although the collected data might include only small portions of personal and private data, they have to be protected. Otherwise, due to the capabilities of big-data-based technologies, sensible information, trends, patterns and behaviours could be revealed, thus, endangering people's privacy.

A. Maya López · A. Solanas (✉)
Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain
e-mail: agusti.solanas@urv.cat

A. Maya López
e-mail: armando.maya@estudiants.urv.cat

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_6

Recognizing the aforementioned risk, governments have reformed existing regulations to legally guarantee people's privacy. Consider for example, the General Data Protection Regulation (GDPR), which regulates the processing of personal data relating to individuals in the EU. Hence, once data are collected, they have to be anonymised prior to the application of any big data analysis technique. To do so, there exist well-known methods developed in the field of Statistical Disclosure Control (SDC) that aim to confer privacy properties to the data (e.g., k-anonymity, l-diversity, or t-closeness) to protect individuals privacy.

Amongst these data privacy-preserving techniques, microaggregation is one of the most consolidated and used ones, since it guarantees the k-anonymity property, this is, any record in the data set will be indistinguishable from other $k-1$ records. As a result, unambiguously identifying a single respondent/individual is impossible [6]. Microaggregation was proposed at *Eurostat* in the early nineties, and has since then been used by the British Office for National Statistics and other national agencies. Optimally solving the microaggregation problem is known to be NP-Hard, hence, heuristics and approximations are used.

In this article, we propose a new heuristic inspired in the well-known Travelling Salesman Problem (TSP) to solve the microaggregation problem. We show that our approach improves the results of other well-known approaches for small values of k. Also, we set the ground for further research in this field. The rest of the article is organised as follows: In Sect. 6.2 the basics of microaggregation and the TSP are recalled. Next, in Sect. 6.3 we describe our TSP-inspired method to solve the microaggregation problem, and we show some experimental results in Sect. 6.4. We conclude in Sect. 6.5 with some final remarks and future research directions.

6.2 Background

6.2.1 Basics of Microaggregation

Microdata refers to data belonging to individuals and they consist of several attributes with a diversity of features. Microaggregation is a family of perturbation-based statistical disclosure control (SDC) methods originally designed to protect continuous numerical microdata. Formally, microaggregation can be defined as follows:

Consider a microdata set D with p continuous numerical attributes and n records (i.e., the result of observing p attributes on n individuals). Groups (also called subsets) of D are formed with n_i records in the ith group ($n_i \geq k$ and $n = \sum_{i=1}^{g} n_i$), where g is the number of resulting groups, and k a cardinality constraint. Optimal microaggregation is defined as the one yielding a k-partition maximizing the within-groups homogeneity. The sum of squares criterion is commonly used for measuring the homogeneity in each group. In terms of sums of squares, maximising within-groups homogeneity is equivalent to finding a k-partition minimizing the within-groups sum of squares (SSE) [8] defined as:

$$SSE = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{i,j} - \hat{x}_i)(x_{i,j} - \hat{x}_i)', \tag{6.1}$$

where $x_{i,j}$ is the jth record in group i, and $\hat{x}_i$ is the average record of group i. The total sum of squares (SST), an upper bound on the partitioning information loss, is computed as if only a single group exists, as follows:

$$SST = \sum_{i=1}^{n} (x_i - \hat{x})(x_i - \hat{x})', \tag{6.2}$$

where x_i is the ith record in D and $\hat{x}_i$ is the average record of D. Note that all the above equations use vector notation, so x_i is a vector belonging to $\mathbb{R}^p$.

The microaggregation problem consists in finding a k-partition with minimum SSE, this is, the set of disjoint subsets of D so that $D = \bigcup_{m=1}^{g} s_m$, where s_m is the mth subset and g is the number of subsets, with minimum SSE (it is ease to see that the cardinality of the groups in the optimal k-partition must lie between k and $2k - 1$). A normalised measure $L = \left(\frac{SSE}{SST}\right)$ of information loss is typically used (i.e., $0 \leq L \leq 1$). Optimal microaggregation is an NP-hard problem [2] for multivariate data and it requires heuristic approaches, which can be divided in two big families:

- Fixed-size microaggregation: These heuristics yield k-partitions where all subsets/groups have size k, except perhaps one group which has size between k and $2k - 1$, when the total number of records is not divisible by k.
- Variable-size microaggregation: These heuristics yield k-partitions where all groups have sizes in $(k, 2k - 1)$. The challenge is how to enforce cardinality constraints on groups without substantially increasing SSE.

6.2.2 The Travelling Salesman Problem: Foundations

In this section, we briefly recall the TSP by summarizing two of its most important formulations [3, 5]. First, we describe the TSP as a permutation problem and, next, we formulate it as a graph theoretic problem.

- Combinatorial optimization formulation: Given a set of cities, the goal is to find the shortest tour that visits each city exactly once and then returns to the starting city. Formally, the TSP can be stated as follows: The distances between n cities are stored in a distance matrix D with elements $d_{i,j}$ where $i, j = 1, \ldots, n$ and the diagonal elements $d_{i,i}$ are zero. A *tour* can be represented by a cyclic permutation π of $\{1, 2, \ldots, n\}$ where $\pi(i)$ represents the city that follows city i on the tour. Therefore, the TSP is reduced to finding a permutation π that minimizes the *length of the tour* $L = \sum_{i=1}^{n} d_{i,\pi(i)}$. Following a brute-force approach, the tour length of $(n - 1)!$ permutation vectors have to be compared, and it is known to be an NP-complete problem [5].

– Graph theory formulation: In this case the problem is modelled by a graph $G = (V, E)$, where cities are the nodes set $V = \{1, 2, \ldots, n\}$ and each edge $e_{ij} \in E$ has an associated weight w_{ij} representing the distance between nodes i and j. The goal is to find a *Hamiltonian cycle*, i.e, a cycle which visits each node in the graph exactly once, with the least total weight. This formulation leads to procedures involving minimum spanning trees for tour construction or edge exchanges to improve existing tours.
  An alternative approach to the *Hamiltonian cycle* to solve the TSP is finding the *Shortest Hamiltonian path*. The problem of finding the shortest Hamiltonian path through a graph (i.e., a path which visits each node in the graph exactly once) can be transformed into the TSP with cities an distances representing graphs vertices and edge weights, respectively. Finding the shortest Hamiltonian path disregarding the endpoints can be achieved by inserting a "dummy city" with zero distance to all other cities.

Finding the exact solution to the TSP with n cities requires to check $(n - 1)!$ possible tours [5]. The problem is known to be NP-Hard. However, solving it is an important step in many areas including vehicle routing, computing wiring, machine sequencing and scheduling, frequency assignment in communication networks and, thus, many heuristic approaches have been suggested [4].

6.3 Our TSP-Inspired Microaggregation Approach

The microaggregation method proposed in this article is based on the well-known TSP [5]. As stated in Sect. 6.2, the TSP consists in finding the shortest possible tour for a given list of cities that visits each city, exactly once, and ends back to the starting city. The TSP can be used to obtain a clustering object [3]. The idea is that objects in clusters are visited in consecutive order. The innovation of our method lies in the representation as a graph theoretic problem to find clusters of records, in the data set. In other words, we suggest to use efficient heuristics that find good approximations of the Hamiltonian Cycle Problem and, from the obtained Hamiltonian Cycle, we create the clusters (subsets) of a k-partition that solves the microaggregation problem. In fact, we do not guarantee finding the optimal k-partition. However, our intuition is that this approach could lead to good results and this is what we explore in this article.

For our initial investigation presented in this article, we consider a fixed-size microaggregation heuristic, in which all groups have k records (except the last formed grouped, that might have up to $2k - 1$ records). In our approach, a multivariate data set consisting of n records and p numerical attributes can be represented as n points $x_1, \ldots, x_n$ in $\mathbb{R}^p$. Inspired by the TSP, we consider that each record is represented by a city (located in a p-dimensional space) and, hence, it is a node in a connected graph. Our approach proceeds as follows:

1. For each starting city s, find a Hamiltonian path $H_{path}(s)$ traversing all n points in the dataset D with the minimum possible length, starting in city s. Let $\pi_{H_{path}(s)}$ be the permutation of $\{1, \ldots, n\}$ expressing the order in which the points are traversed by $H_{path}(s)$.
   - At the end of this iteration, we have a set of n Hamiltonian paths, each starting from each city (record) in the data set. This is, we have n permutations $\pi_{H_{path}(i)}, \forall i \in [1, n]$.
2. From the set $\pi_{H_{path}(i)}, \forall i \in [1, n]$ build a "neighbourhood matrix" (R) so that R is a squared matrix $n \times n$, whose elements r_{ij} represent the number of times node i and j have been found (in all permutations) at $k-1$ or less edges away from each other.
   - To build R we iterate a sliding window of k elements over each permutation position and for all permutations. Note that high values of r_{ij} indicate higher chances for i and j to be clustered together.
3. Given the aforementioned matrix R, generate clusters/groups of cities/records of size k: The group generation starts by finding the maximum value $r_{ij} \in R$, and assigning elements i and j to the first group. Next, the maximum value $max(r_{i,p}, r_{q,j}) \in R, \forall p, q \in [1, n] | (p \neq j, q \neq i)$ is found and element p or q, as appropriate, is added to the group. This procedure is repeated $(k-2)$ times to create each group. Groups are created following the same procedure until there remain no unassigned elements in D. As a result, a k-partition of D is obtained.
4. Finally, to obtain a microaggregated data set D' from D, compute the centroid (i.e., the average vector) of each group in the k-partition and replace each record x_i in D by the centroid $\hat{x}_g$ of the group g to which it belongs.

6.4 Experimental Results

With the aim to validate our intuition that TSP heuristics could be used to find good microaggregation solutions, we have compared our approach with two well-known and good-performing microaggregation algorithms (i.e., Maximum Distance to Average Vector (MDAV) and, Variable-MDAV [8]) over two real microdata sets that are frequently used in the literature as benchmarks (i.e., *Census* and *Tarragona* [2]). *Census* contains 1080 records with 13 numerical attributes and *Tarragona* has 834 records with 13 numerical attributes.

Our method is a fixed-size microaggregation heuristic. Therefore, to study the information loss for several group sizes, we have varied k in the range $[3, 4, 5, 10]$ – which are the typical values used for statistical agencies – , and we compared the results with those obtained by MDAV and V-MDAV, for the same values of k. The results are shown in Table 6.1.

Table 6.1 Information loss obtained by MDAV,V-MDAV and our method (MF-TSP)

Dataset	Method	$k = 3$	$k = 4$	$k = 5$	$k = 10$
Census	MDAV	5.66	7.51	9.01	14.07
	V-MDAV	5.69	7.52	8.98	14.07
	MF-TSP	5.30	8.47	10.01	17.01
Tarragona	MDAV	16.96	19.70	22.88	33.26
	V-MDAV	16.96	19.70	22.88	33.26
	MF-TSP	15.45	18.86	24.90	37.19

It can be observed that our approach performs better than MDAV and V-MDAV for $k = 3$ in *Census* and *Tarragona* and, for $k = 4$ for *Tarragona*. In a nutshell, we have an initial indication that our method could lead to better solutions for small values of k while it yields to worse results for larger cardinalities.

6.5 Conclusion

The deployment of IoT and 5G technologies opens the door to the collection of large amounts of data used to obtain information and make better decisions on business, healthcare [9], transportation, etc. Despite its utility, analysing huge amounts of data could jeopardise individuals privacy and current regulations mandate companies to put in place the right measures to guarantee individuals privacy. With this aim, we have proposed a new fixed-size multivariate microaggregation method inspired in the heuristic solutions of the Travelling Salesman Problem, that helps to guarantee individuals privacy through k-anonimity.

After introducing the basics on Microaggregation and the TSP, we have described our algorithm and we have empirically shown that it performs better than off-the-shelf, well-known microaggregation methods for low cardinalities over benchmark data sets frequently used in the literature. Our proposal represents the first step towards the creation of a more solid TSP-based microaggregation algorithm that would outperform current methods, not only for small cardinalities but for any k as well, and it opens the door to a fruitful research line in the field of SDC. As further work, we plan to improve our clustering algorithm over Hamiltonian paths permutations and test alternative TSP heuristics.

Acknowledgements The authors are supported by the Government of Catalonia (GC) with grant 2017-DI-002. A. Solanas is supported by the GC with project 2017-SGR-896, and by Fundació PuntCAT with the Vinton Cerf Distinction, and by the Spanish Ministry of Science & Technology with project RTI2018-095499-B-C32.

References

1. Batista E, Solanas A (2018) Process mining in healthcare: a systematic review. In: 9th International conference on information, intelligence, systems and applications. IEEE, pp 1–6
2. Domingo-Ferrer J, Sebé F, Solanas A (2008) A polynomial-time approximation to optimal multivariate microaggregation. Comput Math Appl 55(4):714–732
3. Hahsler M, Hornik K (2007) TSP—infrastructure for the traveling salesperson problem. J Stat Softw 23(2):1–21
4. Johnson O, Liu J (2006) A traveling salesman approach for predicting protein functions. Source Code Biol Med 1:3
5. Liao YF, Yau DH, Chen CL (2012) Evolutionary algorithm to traveling salesman problems. Comput Math Appl 64(5):788–797
6. Samarati P (2001) Protecting respondents identities in microdata release. IEEE Trans Knowl Data Eng 13(6):1010–1027
7. Solanas A, Casino F, Batista E, Rallo R (2017) Trends and challenges in smart healthcare research: a journey from data to wisdom. In: 3rd IEEE international forum on research and technologies for society and industry. Modena, Italy, pp 1–6
8. Solanas A, Martinez A (2006) VMDAV: a multivariate microaggregation with variable group size. In: 17th COMPSTAT symposium of the IASC, Rome, pp 917–925
9. Solanas A, Patsakis C, Conti M, Vlachos IS, Ramos V, Falcone F, Postolache O, Pérez-Martínez PA, Di Pietro R, Perrea DN, Martínez-Ballesté A (2014) Smart health: a context-aware health paradigm within smart cities. IEEE Commun Mag 52(8):74–81

Chapter 7
Modular Design of Electronic Appliances for Reliability Enhancement in a Circular Economy Perspective

Simone Orcioni, Cristiano Scavongelli and Massimo Conti

Abstract The design of electronic systems must consider the possibility of their repair, reuse and recycle, in order to reduce the waste. In this paper, we present a design methodology for modularization of electronic appliances which optimize its end of life cost. The optimization algorithm is based on the partitioning of electronic components by mean of simulated annealing optimization, and it has been applied to the design of a real industrial test case.

Keywords Reliability · Reuse · Recycle · WEEE · End-of-life

7.1 Introduction

Electronic devices keep spreading every day. The fundamental problem is that all these new electronic devices are usually not designed to last. When a phone breaks, probably the consumer is going to buy a new one, rather than repair the old one, and in that case the old phone simply becomes electronic waste, which has to be disposed. Therefore, more electronic products don't just mean more opportunities, but also more waste and this waste poses a serious environmental and economic issue [1].

The economic problem comes from the fact that the end-of-life (EoL) treatment of these devices is an expensive process; moreover, these electronic devices also contain precious materials. These problems are becoming so important that many countries introduced or are introducing specific directives and specifications about the WEEE recycling or disposal. The European Community first tackled the problem with the WEEE directive 2002/96/CE, in 2002, and today the WEEE disposal is ruled by the 2012/19/EU directive. Furthermore, the European Commission has launched an EU action plan for the Circular Economy which aims to support the transition towards an economy in which valuable materials, products and resources are maintained as long as possible, while reducing the generation of waste. Basically, the EoL industries

S. Orcioni · C. Scavongelli · M. Conti (✉)
Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy
e-mail: m.conti@univpm.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_7

have three possibilities: they can try to repair a broken device; they can dismantle it, in order to try to reuse part of it; they can recycle it, if the device is beyond repair, or cannot be dismantled, or it is simply too expensive to repair o dismantle it. These approaches go under the name of "3Rs": repair, reuse, recycle. The 3Rs approach is a possible solution to the reduction of the waste of electronic appliances, in opposition with the fact that some classes of products end their life even if they are working, mainly due to fashion choices.

In recent years, many researchers faced the aspect of the reuse of electronic components. Many authors tried to estimate the remaining useful life (RUL) using statistical models based on a real-world data fitting (see for example [2]). In all these works, the authors point out the importance of obtaining high-quality data related to the operating life of a device in order to get good RUL estimates. A few attempts to collect and manage lifecycle data for electronic equipment have also been done [3]. Other cloud-based approaches are discussed in [4–6].

If the RUL estimate suggests that can be economically convenient to reuse some parts of a device, an EoL industry can proceed to the disassembling phase. The disassembling of an equipment can either be economically feasible or not; it usually depends on the particular device assembly and layout, on the component's size, on the practical difficulty of accessing a particular component or device part. A strategy to optimize the disassembly sequence can be found in [7].

In order to make the disassembly process easier for the EoL industries, the disassembly problem has been tackled starting from the device's design phase. This approach leads to the so-called "design for disassembly". The authors of [8] propose a selective parallel planning method, which groups parts into modules and try to remove simultaneously grouped part from products.

The idea of grouping components with similar features into modules in order to speed up the disassembly process can be pushed further by dividing the appliance into modules. The idea of modularity has widely been used to improve the product reliability, scalability, feasibility of component change and maintenance, but not so much to improve disposal, ease of reuse, reduction of waste and recycling. If a modular device breaks, we can decide to change and waste only the module which broke, or we can decide to use the modules which keep working in new products.

In summary, there are just a handful of projects which try to find an optimal modular structure in the context of the 3Rs. In this paper, we present a design methodology which tries to find an optimal modularization for an appliance. The goal is to reduce the cost of the device considering the cost of the repair of the device in case of fault. The cost function considers the cost of a module, the cost of the interconnections between two modules, and a fault probability for each module. The optimum is found using the simulated annealing optimization algorithm. In Sect. 7.2 we present the design methodology. In Sect. 7.3 we briefly describe the optimization algorithm and the parameters required for the optimization. In Sect. 7.4 we'll present the results of the algorithm application to a real-world case.

7.2 Modular Design for Reliability

Many are the parameters used to define the probability that a system is correctly functioning. The failure probability density $f(t)$ is the probability that a failure occurs in the time interval $[t, t + \mathrm{d}t]$. The cumulative probability of failure $F(t)$ is the integral of the probability density

$$F(t) = \int_0^t f(\tau)\mathrm{d}\tau \tag{7.1}$$

with the normalization condition

$$F(\infty) = \int_0^\infty f(\tau)\mathrm{d}\tau = 1 \tag{7.2}$$

The cumulative probability of failure $F(t)$ of a system is the probability that a system is not correctly functioning at time t.

Failure rate $\lambda(t)$ is defined as the number of failures per unit time normalized to the number of systems that are still correctly functioning. The following relationship holds among failure rate and reliability (the complement of probability of failure) [9]

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{7.3}$$

The following relationships also holds [9]

$$F(t) = 1 - \mathrm{e}^{-\int_0^t \lambda(\tau)d\tau} \tag{7.4}$$

$$f(t) = \lambda(t)\mathrm{e}^{-\int_0^t \lambda(\tau)d\tau} \tag{7.5}$$

The failure rate of the equipment composed of N components with independent failure rate is [10]

$$\lambda_B(t) = \sum_{i=1}^{N} \lambda_i(t) \tag{7.6}$$

Therefore, the joint failure of the components becomes

$$F_B(t) = 1 - \mathrm{e}^{-\int_0^t \sum_{i=1}^{N} \lambda_i(\tau) d\tau} = 1 - \prod_{i=1}^{N} \mathrm{e}^{-\int_0^t \lambda_i(\tau) d\tau} = 1 - \prod_{i=1}^{N} (1 - F_i(t)) \quad (7.7)$$

Usually the companies are interested in a reduced failure rate for the first 5–15 years and they are not interested on the behavior in a longer term. Therefore, the parameter we consider in this work is the probability of failure of the device $F_B(t)$ in a fixed time t (for example 8 years).

The idea behind a modular design is to divide the whole equipment, consisting in N components, into M distinct, interconnected modules, the generic j-th module consists in N_j components. The following relationship holds

$$\mathrm{N} = \sum_{j=1}^{M} N_j \quad (7.8)$$

If a module breaks, it can be simply unplugged and replaced or repaired. Every module has a given economical cost, which can be essentially the sum of the costs of its internal components. If we are planning to repair or replace a broken module, we have to consider its fault probability, because the module cost has to be paid every time it breaks. Therefore, the overall cost of a module is given by the initial cost, plus the cost paid again every time the module fails. Hence

$$\mathrm{Cost}_j = C_{Bj} + C_{Bj} F_{Bj}(t) = C_{Bj}\big(1 + F_{Bj}(t)\big) \quad (7.9)$$

where C_{Bj} is the cost of the j-th module and F_{Bj} its fault probability that can be expressed as

$$C_{Bj} = \sum_{i=1}^{N_j} C_{ij} \quad F_{Bj}(t) = 1 - \prod_{i=1}^{N_j} \big(1 - F_{ij}(t)\big) \quad (7.10)$$

In (7.10) C_{ij} and F_{ij} are the costs and the fault probabilities of the i-th component of the j-th module, respectively. The total cost of the equipment is therefore

$$C_{TOT} = \sum_{j=1}^{M} \big[C_{Bj}\big(1 + F_{Bj}(t)\big)\big] + \sum_{j=1}^{M} \sum_{k=1}^{M} C_{conj,k} \quad (7.11)$$

where $C_{conj,k}$ is the additional cost term, which considers the cost of the interconnections between module j and module k $C_{conj,k}$ takes into account the cost of the connectors and cabling among the modules. If the components that are electrically connected are in the same module, they do not contribute to the connection cost. The number of modules M and the way the components are placed in the different modules are design parameter, that depends mainly on the modularization feasibility: the more the modules, the more complex the connections among them, the more

difficult to actually implement the design. In addition, the connection cost takes into account the disassembling and reassembling costs.

In this work, the optimization goal is to decide which component goes into which module, in order to minimize this cost function C_{TOT}. For example, we might expect that coupling a high cost component with a low fault probability component is going to reduce the overall cost, but the increase in the interconnections cost might frustrate this reduction. So, we have an enormous number of combinations which have to be searched to find the optimum components placement, and the simulated annealing is the algorithm we chose to perform this search.

7.3 Optimization Algorithm

Simulated annealing (SA) belongs to the class of random perturbation algorithms, and it has been widely used in many optimization problems. At the heart of SA there is the simple random perturbation algorithm: for any given cost function, we generate an initial random solution and then we start changing the parameters the cost function depends on. For every new value for the parameters, we evaluate the cost function. If the cost function decreases, we keep the new solution and we go on trying another solution. The SA overcomes the problem of local minima by sometime accepting a solution which increases the cost function. This approach is called "hill climbing".

Accepting a solution which increases the cost function allows to better explore the solutions space, but this cannot be done at the same rate both at the beginning and at the end of the search. The SA defines an additional parameter, called "temperature" for a physical analogy with the real annealing process in the context of material science. This parameter is set to a high value at the beginning of the search, and it gets decreased while the algorithm goes on, reaching a low value near the end of the search. The hill-climbing solutions are accepted with greater probability when this temperature is high, and with lower probability when this temperature is low. After a few iterations, we decrease the temperature and we keep iterating. The algorithm stops when the system has "frozen", i.e. the temperature has reached a very low value and the cost function has stopped decreasing over the last few iterations.

To minimize the cost function in (7.11), the algorithm needs as input: the maximum number of modules we are willing to generate the list of components, their interconnections, their cost and the fault probability for the specified time. The information on the components is given by the netlist file with the format shown in Table 7.1. In this example, we have 7 components. It follows a list of component/nodes association items. Each item starts with a component ID, the component type, and the component name. For example, the first component is a *serial_inputs_comp* that we have called *serial_inputs*. The first number after the names specifies the number of nodes the component is connected to, followed by the ID names of those nodes. For example, *serial_inputs* is connected to 33 nodes.

The last two numbers represent the economic cost of the component and its cumulative probability of failure in 8 years $F(t = 8\,\text{years})$, respectively. For example,

Table 7.1 Example of netlist

n	Type	Name	Node #	ID node1	…	Cost €	Failure prob
1	serial_inputs_comp	serial_inputs	33			6.351	0.00559
2	power_supply_comp	power_supply	2			1.202	0.00303
3	MCU_comp	MCU	28			4.842	0.00195
4	input2_comp	input2	24			0.113	0.00056
5	input1_comp	input1	20			0.111	0.00019
6	connector_comp	connector	35			0.725	0.00059
7	amplifier_comp	amplifier	15			1.775	0.00406

serial_inputs costs 6.351 euro, and its fault probability is 0.00559. Starting from this netlist, the software derives a connectivity matrix. An example is shown in Table 7.2. This matrix contains a row and a column for each component. Each cell in the matrix contains a number which specifies the number of connections between each couple of components. For example, the *MCU* and the *serial_inputs* share 14 connections, while the *amplifier* and the *power_supply* share two connections. This matrix is used to evaluate the interconnection cost between the modules $C_{conj,k}$ in the cost function in (7.11). The interconnections cost is evaluated by multiplying the cost of a single connector by the number of nodes shared between two components in two different modules.

In general, the SA algorithm must perform a huge number of iterations before reaching the "frozen" state, but its speed usually depends mainly on the dimensions of the solutions space. In common electronic equipment, there could be a few hundreds of components, and the SA should move around all these components. Nevertheless, the designer can force some components to be together in the same module. This allows us to reason in term of "macroblocks" rather than "components", to speed up the optimization algorithm.

Table 7.2 Example of connectivity matrix

	serial_inputs	power_supply	MCU	input2	input1	connector	amplifier
serial_inputs		1	14	7	10	5	3
power_supply	1		1	1	1	1	2
MCU	14	1		3		7	9
input2	7	1	3		1	14	
input1	10	1		1		9	
connector	5	1	7	14	9		7
amplifier	3	2	9			7	

7.4 Results

After developing a C++ implementation for the optimization algorithm, we applied the program to a real application of a board of the Vega s.r.l company. The appliance we used is an electronic board, called SVN400 and reported in Fig. 7.1, used to control an elevator. The board can be divided into the 7 macroblocks, shown in the schematic of Fig. 7.2. The macroblocks are described in the following:

- *MCU*: block relative to the microprocessor, consisting of 33 components: one 16 bit PIC microcontroller, resistors, capacitors, inductors, diodes, connectors;

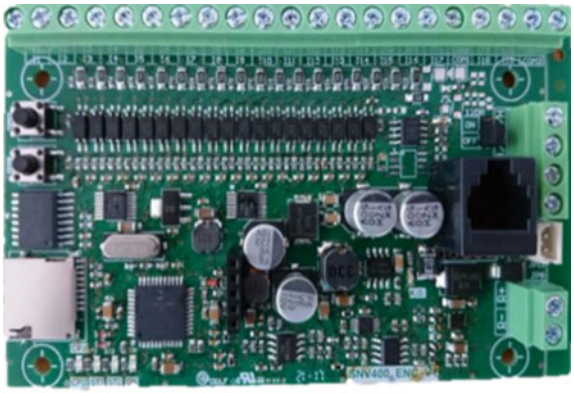


Fig. 7.1 Board SNV400 without partitioning

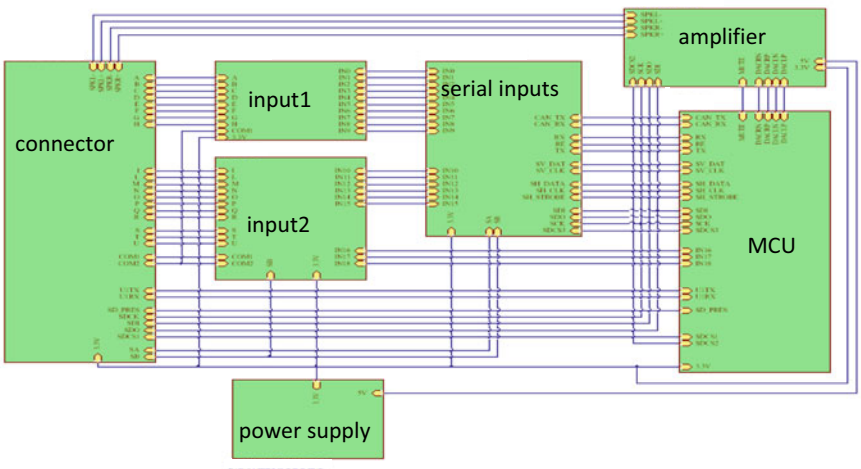


Fig. 7.2 Functional blocks of the SNV400 board

- *power_supply*: the power supply section which supplies the required voltage to all the components (5 and 3.3 V), consisting of 28 components;
- *amplifier*: audio amplification section that includes the operational, the digital rheostat that establishes the gain and the final amplifier that sends the signal to the speakers, consisting of 59 components;
- *connectors*: includes all the connectors that allow the board to interface with the display, speaker etc., consisting of 16 components;
- *serial_input*: part of serial communication including RS485, can_bus and VEGA_serial, consisting of 27 components;
- *input1* and *input2*: opto-isolated inputs that carry signals from the connectors to the micro through transceivers, consisting of 60 and 64 components, respectively.

The total number of single components are 287. The costs and the fault probabilities for these macroblocks are given in Table 7.1, while their connection matrix is shown in Table 7.2. Costs and fault probabilities during the first 8 years of life have been obtained analyzing the fault history of similar electronic boards. We used a mixture model for the cumulative probability of failure of the i-th component of the j-th macroblocks

$$F_{ij}(t) = a_\Gamma F_{\Gamma i,\mathrm{j}}(t) + a_\lambda F_{\lambda i,j}(t) + a_G F_{Gi,j}(t) \tag{7.12}$$

with the normalization condition $1 = a_\Gamma + a_\lambda + a_G$. We used the gamma function F_Γ to the infant mortality, the lambda function F_λ for constant failure rate and the gaussian function F_G for ageing.

Table 7.3 reports the partition, the cost and the total failure probability. The algorithm groups all the macroblocks into one module, which means that no modularization is performed or suggested. The cost function tells us that this choice is the searched optimum, because there are no further improvements we can make on the cost.

The reason for this result is that the failure probabilities for the macroblocks are so low that no good comes from separating the macroblocks into modules and allowing the extra interconnections cost. We can demonstrate this claim by running the algorithm without interconnections cost. The results are reported in Table 7.4. Now the algorithm is able to find a solution with 4 partitions that has a total cost lightly reduced with respect to the case of no partition.

To better understand the effect of the modularization in the cost function we considered a simplified example with 8 blocks. Each couple of blocks share only one connection; hence, the connectivity matrix will contain "0" in the diagonal and "1" for all the other terms. We added to the optimization algorithm the constraint that each module must have the same number of blocks. Therefore, we found the best solution with 1 module with 8 blocks, 2 modules with 4 blocks each, 4 modules

Table 7.3 Partition results

		Cost €	Failure prob
Partition 1	All macroblocks	15.3604	0.0160

Table 7.4 Partition results without interconnection cost

		Cost	Failure prob
Partition 1	input2, amplifier	1.888	0.0046
Partition 2	MCU	4.842	0.0020
Partition 3	power supply, input1, connector	2.038	0.0038
Partition 4	serial inputs	6.351	0.0056
TOT		15.119	0.0160

with 2 blocks, and 8 modules with 1 block. With the above defined constraints, the number of blocks per module in (7.10–7.11) is $N_j = p$, $j = 1..M$ and $M = N/p$. Therefore (7.11) becomes:

$$C_{TOT} = \sum_{j=1}^{N/p}\left[\sum_{i=1}^{p} C_{ij}\right]\left(2 - \prod_{i=1}^{p}(1 - F_{ij})\right) + C_{con}N(N - p) \tag{7.13}$$

We considered 4 tests cases with different costs and failure probabilities.

In test 1 the failure probability is the same for all the blocks (chosen equal to 1/8 = 0.133), blocks 1–4 have higher normalized cost (0.24) with respect to blocks 5–8 (0.01). The normalized connection cost is 0.008. With a constant failure probability $F_{ij} = F$, (7.13) can be further simplified

$$C_{TOT} = \sum_{i=1}^{N} C_i\left[2 - (1 - F)^p\right] + C_{con}N(N - p) \tag{7.14}$$

In test 2 the normalized cost is the same for all the blocks (chosen equal to 1/8 = 0.133), blocks 1-4 have higher failure probability (F = 0.2) with respect to blocks 5–8 (F = 0.05). The normalized connection cost is 0.008. In tests 3 and 4 the 8 blocks have different normalized cost and failure probability. The normalized connection cost is 0.012 and 0.008, respectively for test 3 and 4.

The configuration of the best solutions and the total cost defined in (7.13) are reported in Table 7.5. Table 7.6 reports, for the different tests, the normalized total cost including connection cost of the best solution as a function of the number of modules $M = N/p = 1, 2, 4, 8$ for each module A-H identified in Table 7.5.

For the selected value of connection cost, the best solution of test 1 and 2 is obtained using 8 modules. For the test 2 case, the best solution groups together the blocks with the highest fault probability. For example, in the case of two modules, all the high fault probability blocks are placed in the same module.

In test 3, with high connection cost, the best solution is obtained grouping all the blocks in a single module. In test 4, with medium connection cost, the best solution is obtained grouping the blocks in 4 modules, 2 blocks for each module. High fault probability and high cost blocks are placed together to low fault probability and low-cost blocks to reduce the total cost.

Table 7.5 Optimal configuration the tests

	Test 1				Test 2				Test 3				Test 4			
N/p	1	2	4	8	1	2	4	8	1	2	4	8	1	2	4	8
Module A	All	1, 2, 5, 6	1, 5	1	All	1, 2, 3, 4	1, 2	1	All	1, 3, 4, 7	1, 7	5	All	1, 3, 4, 7	1, 7	5
B		2, 3, 7, 8	2, 6	2		5, 6, 7, 8	3, 4	2		2, 5, 6, 8	2, 8	6		2, 5, 6, 8	2, 8	6
C			3, 7	3			5.6	5			3, 4	1			3, 4	1
D			4, 8	4			7.8	6			5, 6	2			5, 6	2
E				5				3				3				3
F				6				4				4				4
G				7				7				7				7
H				8				8				8				8

Table 7.6 Normalized total cost including connection cost of the best solution

	Test 1				Test 2				Test 3				Test 4			
N/p	1	2	4	8	1	2	4	8	1	2	4	8	1	2	4	8
Module A	1.66	0.83	0.4	0.33	1.67	0.92	0.44	0.21	1.67	1.15	0.67	0.34	1.67	1.09	0.62	0.31
B		0.83	0.4	0.33		0.72	0.44	0.21		0.6	0.45	0.34		0.54	0.41	0.31
C			0.4	0.33			0.37	0.21			0.45	0.37			0.41	0.34
D			0.4	0.33			0.37	0.21			0.17	0.37			0.12	0.34
E				0.07				0.19				0.1				0.07
F				0.07				0.19				0.1				0.07
G				0.07				0.19				0.09				0.07
H				0.07				0.19				0.09				0.07
Tot cost	1.656	1.670	1.618	1.573	1.666	1.644	1.613	1.573	1.666	1.752	1.750	1.797	1.666	1.624	1.558	1.573

In general, the best solution is found grouping as much as possible the blocks, when the connection cost is high. Each block is placed in a separated module, when the connection cost is zero, as can be seen in (13).

The optimum block partitioning of the real application of a board of the Vega s.r.l. company is one single module, that could be seen as an obvious solution. The simplified test cases have been useful to draw some considerations. In general, we can see how the optimizer produces modules with high cost and low fault probability, and modules with low cost and high fault probability.

7.5 Conclusions

The idea of grouping components in modules has been used to speed up disassembly in the recycle or reuse phase of EoL. The idea of modularity has considered in this work to improve the product reliability. In this paper, we present a design methodology to find an optimal modularization for an appliance, with the goal of reducing the cost of the device considering the cost of the repair of the device in case of fault. The methodology and software developed have been applied in a real test case of an electronic board for elevator control. The results show that the partition of the device into modules should keep high cost block separated to high fault probability blocks. The cost of the partitioning is taken into account in the cost function by the cost of the connectors.

Acknowledgements The authors would like to thank Andrea Vesprini and Andrea Medori of VEGA company. The work presented is part of a regional RAEEcovery project supported by EU funding (https://www.raeecovery.com).

References

1. Rifer W, Brody-Heine P, Peters A, Linnell J (2009) Closing the loop—electronics design to enhance reuse/recycling value. The Green Electronic Council in collaboration with the National Center for Electronics Recycling and Resource Recyclings
2. Park J, Bae S (2010) Direct prediction methods on lifetime distribution of organic light-emitting diodes from accelerated degradation test. IEEE Trans Reliab 59(1):74–90
3. Rudiger R, Hohaus C, Uriarte A, Ibanez N, Guarde D, Marquinez I, Manjon D, Kovacs P (2012) Towards efficient end-of-life processes of electrical and electronic waste with passive RF communication. IEEE Electron Goes Green
4. Xia K, Gao L, Chao K-M, Wang L (2015) A cloud-based disassembly planning approach towards sustainable management of WEEE. In: IEEE 12th international conference on e-business engineering
5. Capecci S, Cassisi E, Granatiero G, Scavongelli C, Orcioni S, Conti M (2017) Cloud-based system for waste electrical and electronic equipment. In: 2017 13th workshop on intelligent solutions in embedded systems, WISES 2017, pp 41–46
6. Conti M, Orcioni S (2019) Cloud-based sustainable management of electrical and electronic equipment from production to end-of-life. Int J Qual Reliab Manag 36(1):98–119

7. Smith S, Smith G, Chen W (2012) Disassembly sequence structure graphs: an optimal approach for multiple-target selective disassembly sequence planning. Adv Eng Inf 26:306–316
8. Smith S, Hung P (2012) A parallel disassembly method for green product design. In: Proceedings of IEEE conference 2012 electronics goes green 2012+, 9–12 Sept 2012, Berlin, Germany
9. Kumar V, Singh L, Tripathi AK (2017) Reliability analysis of safety-critical and control systems: a state-of-the-art review. IET Softw 12(1):1–18
10. Rausand M, Hoyland A (2004) System reliability theory: models, statistical methods and applications. Wiley

Chapter 8
Pest Detection for Precision Agriculture Based on IoT Machine Learning

Andrea Albanese, Donato d'Acunto and Davide Brunelli

Abstract Apple orchards are widely expanding in many countries of the world, and one of the major threats of these fruit crops is the attack of dangerous parasites such as the Codling Moth. IoT devices capable of executing machine learning applications in-situ offer nowadays the possibility of featuring immediate data analysis and anomaly detection in the orchard. In this paper, we present an embedded electronic system that automatically detects the Codling Moths from pictures taken by a camera on top of the insects-trap. Image pre-processing, cropping, and classification are done on a low-power platform that can be easily powered by a solar panel energy harvester.

Keywords Internet of Things · Machine learning · Precision agriculture

8.1 Introduction

Electronics and ICT technologies are gaining momentum in agriculture services. Precision farming is developing new solutions for pest detection [1], water management, treatments optimization nowadays; since the goal of precision agriculture is to get the most healthy product sustainably. Most of these applications use smart sensors which are managed from low cost and low power embedded systems [2, 3]. Usually, after sensing the surrounding environment, the system does not take any decision about the acquired data, and it is transmitted to remote servers for supports. The main drawback of this approach is a large amount of data to be transmitted that hampers scalability of such a distributed paradigm. The key idea is to shift processing

A. Albanese · D. d'Acunto · D. Brunelli (✉)
Department of Industrial Engineering, University of Trento, Via Sommarive 9, 38123 Povo, TN, Italy
e-mail: davide.brunelli@unitn.it

A. Albanese
e-mail: andrea.albanese@studenti.unitn.it

D. d'Acunto
e-mail: donato.dacunto@studenti.unitn.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_8

near the sensors and finally transmit a report of a few bytes, thanks to compression methods [4]. Moreover, machine learning can improve the performance of a precision agriculture application because this type of algorithms can quickly detect and classify parasites, diseases, and weeds.

This paper focuses on a smart application that detects automatically dangerous parasites for apple orchards, the *Codling Moth*. This insect looks like a butterfly, and it is a major problem for apple orchards. Thanks to an insect glue trap it is possible to take a picture and classify if there are any *Codling Moth* and finally send a notification to the farmer. The classification is done near sensor thanks to a specific low cost and low power hardware, and an energy-efficient solution is proposed to sustain the system as long as possible.

8.1.1 IoT Architecture

The system consists of a trap that looks like a little hive as shown in Fig. 8.1, where a pheromone bait and a glue layer capture the attracted insects even at low-density presence. The farmer usually takes periodic inspections of the traps or mount a wireless camera that sends the captured pictures wirelessly for remote evaluation. This process is expensive and time consuming for the farmer. The proposed work detects the presence of the parasites thanks to a machine learning approach that sends only notifications of threats and their position to the farmer.

The workflow of the proposed application is summarized in Fig. 8.2. A camera takes pictures inside the trap periodically, the board detects and crops new insects not yet analyzed for the classification. Eventually, a notification is transmitted to the farmer about the detection of parasites.

For this purpose, the hardware is based on a Raspberry Pi3 with a Pi Camera. It is in charge of image pre-processing and cropping, whereas a Movidius Neural Compute

(a) Commercial trap.

(b) Prototype of the IoT neural network Codling Moth smart trap.

Fig. 8.1 Codling Moth traps

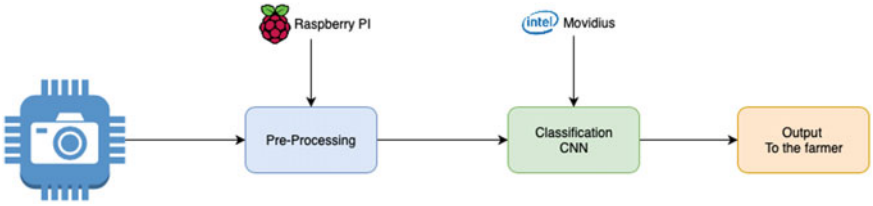


Fig. 8.2 Flowchart of the system application

Stick (NCS), which features the Intel Myriad X neural accelerator, completes the classification stage. Classification is done by a machine learning algorithm that uses a Convolutional Neural Network (CNN) model tailored for the NCS. The uncommon feature of this IoT application is that the classification stage is elaborated in-situ (near the camera). The processing results, consisting of few bytes after the classification, are transmitted using long-range and a low power communication like LoRaWAN [5–7]. Thanks to the technical features of this standard, the end nodes can transmit data in a range of 15 km [8]; additionally, LoRaWAN guarantees the integrity of the transmitted data because its protocol also defines security encryption [9].

8.1.2 Image Pre-processing and Deep Learning

Deep learning is a class of algorithms widely used in machine learning. The network implemented in this project is, in particular, a CNN. This type of networks are widely used in image classification and object recognition problems. Before the training stage of the Deep Neural Network (DNN), a clear and quite large dataset of pictures is necessary to build up the network in an optimal way. The dataset generation stage is fundamental for supervised methods, and each image used for training and validation stages is known and labeled a priori. It implies that a good dataset for the pictures used during training is crucial for global performance. The dataset generation session started with a small set of row pictures, as shown in Fig. 8.3a (approximately 300) that has been incremented when more insects have been trapped during the experiments. The dataset is divided into two classes: *codling moth* and *general insects*. For this specific task a VGG16 model, developed by the Oxford University, is used [10] training all the layers of the network. Then the model is converted to a graph model used to perform the classification on the Vision Processing Unit (VPU).

The camera captures the floor of the insect trap, as shown in Fig. 8.3, pictures may contain a high number of insects to classify. Thus, the images are processed with OpenCV functions to extract each insect in sub tiles from the original taken picture. The task is developed to extract easily features like the color (a dark subject on white background) and the shape of the insects through a Blob Extraction algorithm. The process for image crop is all developed through OpenCV functions, and it consists in:

(a) Raw picture.

(b) Cropped Codling Moth.

(c) Cropped general insect.

Fig. 8.3 Examples of pre-processed images

- Conversion of the frame from RGB to GRAY scale;
- Smoothing (or blurring) of the frame with a Gaussian filter with a size of 9;
- Edge extraction through Canny operator with 20 as the minimum value and 100 as the maximum value. This value represents the aperture size that is the size of the kernel used to find the image gradients;
- one closing and two expansions, all the operators are used to enhance the blobs, using a rectangular structure element which is the shape of the structure of the filter.

After the application of these morphological operators, the blobs are detected through the OpenCV blob detector. The blobs extracted are collected individually in a vector as a rectangle and, from the original frame, each of the corresponding region of interest (RoI) is cropped. All the new pictures are finally saved for the neural network. They are not of the same size, but all the pictures are square, in this way the CNN can take the image and resize to 52×52. The whole procedure is repeated only for the cropped images that contain more than one blob, in fact in one blob it is possible to find more than one insect, especially in the regions where insects are really crowded.

In this way, the iteration of the algorithm is useful for achieving better images for training and evaluation sessions, and also to extend the data-set.

8.2 Training, Validation and Test

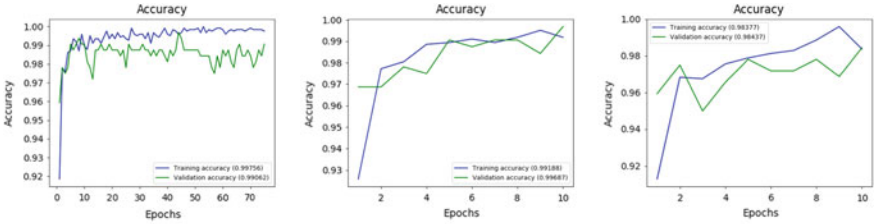
For the training stage, we use the effort of the rapid development of neural networks for image classification based on TensorFlow library [11].
This step is an offline process that is executed in a host computer (like a cluster), and it aims to optimize the neural network through a large dataset of labeled images. Therefore the system can learn from the category assigned the images. The basic element of a DNN is the neuron (or node). It is multiplied by a so-called weight value only when the input is ready. For example, if a neuron has four inputs, it has four weight values which can be adjusted during the training time. A DNN could be improved through many parameters involved in the process. In our case, the most important parameters, which affect the performance in a significant way, are the number of epochs and the image size. The first determines how many times the entire set of training vectors is used to update the weights; at the end of each epoch, a validation step is computed to evaluate the ongoing training process. The image size, instead, is obtained by scaling each picture that feeds the DNN. So the objective is to find the optimal tradeoff for the two parameters to complete the training stage while meeting the hardware constraints. In our application, the following three different configurations were used:

- 75 epochs, image size 224×224;
- 10 epochs, image size 112×112;
- 10 epochs, image size 52×52.

The results obtained in the training tests are shown in Fig. 8.4.

Notice that training and validation accuracy using 75 epochs (default parameter) is going to be saturated. This means that the network does not provide enough accuracy during the test stage and is not able to generalize as good as required.

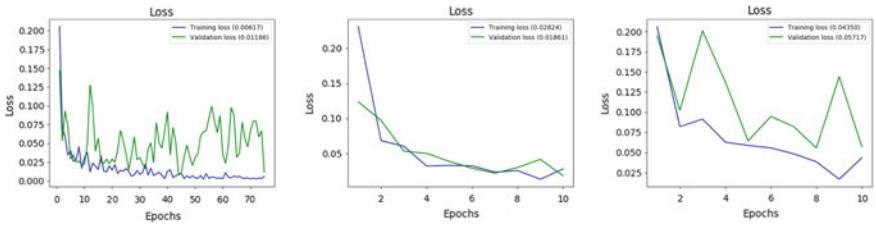
Thus, the epochs can be decreased to achieve better results: as shown in the graphs 10 epochs are enough for excellent accuracy. Moreover, in order to avoid possible overflow and to save memory on the Raspberry Pi 3, the image size is decreased to work with a simpler model and to meet the hardware constraints. Image size of 112×112 and 52×52 have been tested and used. The chosen image size shows worse performance with respect to the one obtained using a bigger image size. Nevertheless, the measured accuracy is 98% which satisfies the requirements of this class of parasites monitoring systems. After the training and the validation stage, the neural network model file is ready. It is possible to test the performance of the DNN model through a new set of data (a subset of the original dataset), which was never used by the DNN. This step helps to assess the performance and the generalization of the network, and it is crucial to confirm the accuracy computed during validation.



(a) 75 epochs, image size 224x224.

(b) 10 epochs, image size 112x112.

(c) 10 epochs, image size 52x52.



(d) 75 epochs, image size 224x224.

(e) 10 epochs, image size 112x112.

(f) 10 epochs, image size 52x52.

Fig. 8.4 Training and validation accuracy and loss function

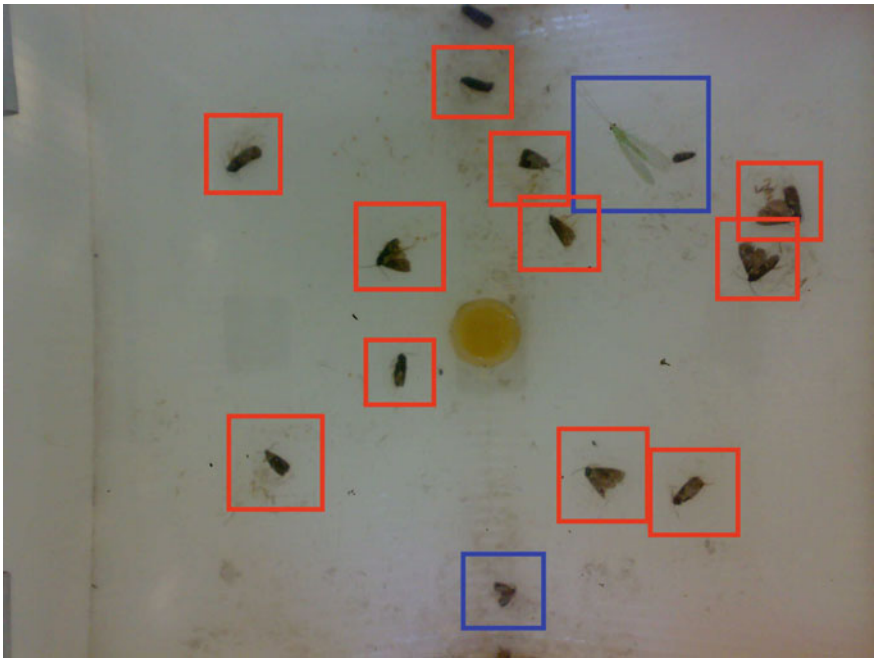


Fig. 8.5 Example of codling moth detection (red boxes) and general insects (blue boxes)

An example of the output from the classification stage is presented in Fig. 8.5. Our DNN provides a measure of the accuracy, which indicates how the detected insect is more similar to a general insect or to a *Codling Moth*. The tests were done in an apple orchard for 12 weeks, with the insect glue trap shown in Fig. 8.1, where 62 insects were captured. The 70% of them were *Codling Moth*, while the remaining 30% were general insects. In this case, the tested pictures are of different sizes.

Classification results are summarized as follows:

- 80.6% was classified correctly;
- 4.8% was false positives;
- 6.4% was false negatives;
- 8.1% was uncertain.

8.3 Conclusions

This paper presents a machine learning-based smart camera tailored for precision agriculture services. The camera detects automatically if dangerous parasites are trapped by the commercial pheromone boxes, in apple orchards and sends an alarm to the farmer. Future work will investigate the performance improvement in terms of classification accuracy and energy consumption, by developing a custom DNN and by extending the training dataset for addition pest types. Moreover, we will include an energy harvester capable of self-sustaining the energy consumption of the smart trap, to permit an unattended activity indefinitely.

Acknowledgements This research was supported by the IoT Rapid-Proto Labs projects, funded by Erasmus+ Knowledge Alliances program of the European Union (588386-EPP-1-2017-FI-EPPKA2-KA).

References

1. Ding W, Taylor G (2016) Automatic moth detection from trap images for pest management. Comput Electron Agric 123(C):17–28
2. Magno M, Tombari F, Brunelli D, Di Stefano L, Benini L (2013) Multimodal video analysis on self-powered resource-limited wireless smart camera. IEEE J Emerg Sel Top Circ Syst 3(2):223–235
3. Magno M, Tombari F, Brunelli D, Di Stefano L, Benini L (2009) Multimodal abandoned/removed object detection for low power video surveillance systems. In: 2009 Sixth IEEE international conference on advanced video and signal based surveillance, pp 188–193
4. Brunelli D, Caione C (2015) Sparse recovery optimization in wireless sensor networks with a sub-nyquist sampling rate. Sensors 15(7):16654–16673
5. Polonelli T, Brunelli D, Benini L (2018) Slotted ALOHA overlay on LoRaWAN—a distributed synchronization approach. In: 2018 IEEE 16th international conference on embedded and ubiquitous computing (EUC), Oct 2018, pp 129–132

6. Adelantado F, Vilajosana X, Tuset-Peiro P, Martinez B, Melia-Segui J, Watteyne T (2017) Understanding the limits of LoRaWAN. IEEE Commun Mag 55(9):34–40
7. Polonelli T, Brunelli D, Girolami A, Demmi GN, Benini L (2019) A multi-protocol system for configurable data streaming on IoT healthcare devices. In: 2019 IEEE 8th international workshop on advances in sensors and interfaces (IWASI), June 2019, pp 112–117
8. Polonelli T, Brunelli D, Marzocchi A, Benini L (2019) Slotted aloha on LoRaWAN-design, analysis, and deployment. Sensors 19(4):838
9. Tessaro L, Raffaldi C, Rossi M, Brunelli D (2018) Lightweight synchronization algorithm with self-calibration for industrial LoRa sensor networks. In: 2018 Workshop on metrology for industry 4.0 and IoT, Apr 2018, pp 259–263
10. Liu S, Deng W (2015) Very deep convolutional neural network based image classification using small training sample size. In: 2015 3rd IAPR Asian conference on pattern recognition (ACPR), Nov 2015, pp 730–734
11. NeuralNetworks. https://github.com/frank1789/NeuralNetworks (Online). Accessed 25 May 2019

Chapter 9
Statistical Flow Classification for the IoT

Gennaro Cirillo, Roberto Passerone, Antonio Posenato and Luca Rizzon

Abstract The objective of this work is to analyze packet flows and classify them as traffic that belongs to IoT devices or to traditional non-IoT communication. We employ two methods: a clustering approach, which learns directly from the structure of the dataset, and a classification tree, trained with the collected data and evaluated using 10-fold cross validation. The results show that classification trees outperform clustering on all datasets, and achieve high accuracy on both homogeneous simulated and real deployment traffic data.

Keywords IoT · Traffic classification · Clustering · J48

9.1 Introduction

Protocol and packet classification is at the basis of several services that can be offered by network operators and by device manufacturers. For instance, differentiated services require that the kind of communication be recognized, in order to provide customized quality of service or to analyze the performance of the network. In our specific case, we are interested in distinguishing between traditional user traffic, such as e-mail, web surfing and media streaming, from traffic originating from independent devices, such as sensors, remote controls, fleet tracking and environmental monitoring. This last category of devices constitutes what is known as the *Internet of Things*

G. Cirillo · R. Passerone (✉)
DISI, University of Trento, Via Sommarive 9, 38123 Povo, TN, Italy
e-mail: roberto.passerone@unitn.it

G. Cirillo
e-mail: gennaro.cirillo@studenti.unitn.it

A. Posenato · L. Rizzon
Microtel Innovation Srl, via Armentera 8, 38051 Borgo Valsugana, TN, Italy
e-mail: antonio.posenato@microtelinnovation.it

L. Rizzon
e-mail: luca.rizzon@microtelinnovation.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_9

(IoT). In this paper we employ a statistical flow classification, distinguishing between traditional and IoT related communication. To construct a classification procedure we use machine learning algorithms that can estimate the parameters for recognition from available labeled data. In our case, we have used several different techniques for generating flows of packets, from real IoT deployments to simulated systems, to have variety. For the classification algorithm we use the Weka framework [1], which provides the most popular machine learning methods. In particular, we rely on the J48 decision tree algorithm, a Java implementation of C4.5, which has been shown to perform well for this class of problems [2]. We also explore clustering methods, using the k-means algorithm, to determine whether there is intrinsic structure in the data that can set apart the behavior of IoT devices from traditional communication. Our results show that this is only partly the case, and that the decision tree can provide better performance when devices of different kinds are employed.

9.2 Dataset Preparation and Attribute Selection

The dataset consists of information collected using Tstat [2] on approximately 77 thousand flows. Most of the non-IoT flows (around 54,000) are obtained from an online repository in Japan [3]. The vast majority of IoT flows (around 12,500) of the initial dataset are obtained by capturing the simulated IoT systems using an IoT software simulator. An additional set of 15,081 IoT flows was obtained from an online repository of real IoT traffic [4]. Traffic is captured in a domestic environment from a real deployment in Australia, where a house was instrumented with several devices interconnected through a wireless network. We first report the results obtained with the initial dataset, and then analyze how these change with the addition of more IoT flows. This highlights the difference in clustering and classification accuracy between a simulated and a real environment.

Attribute Selection. The flow analysis with Tstat provides a large number of features, all of which are not necessarily relevant to our objectives. We assume that certain parameters, such as the IP addresses and the port numbers, are not visible to the application. Instead, we focus on the "behavioral" parameters, which are more independent of the protocols and robust to encryption. We rely on the features identified by a recent study on the behavior of Machine-to-Machine (M2M) communication [5], i.e., the *packet rate*, the *packet size* and the *round trip time*, distinguishing between client and server. Some of the attributes of interest are not directly generated by the Tstat default distribution. The software was therefore modified to compute the *fraction of ACK packets* over the total packets, the *fraction of uplink and downlink packets*, and the *fraction of uplink and downlink bytes* exchanged in the communication, over the total.

9.3 Results

We have followed two methodologies to develop a flow classification method. The first is based on clustering, while the second is based on a classification tree.

9.3.1 Clustering

The first method that we have used for classification is a semi-supervised clustering approach, based on the SeLeCT self learning classifier proposed by Grimaudo et al. [6]. We proceed as follows. We select in Weka the SimpleKMeans algorithm, and partition the dataset into a number of disjoint classes. We instruct the algorithm to ignore the IoT/NON_IoT label given to each flow, making the approach unsupervised. In other words, the algorithm will try to determine classes irrespective of the label that was assigned in the first place. Clustering is then run several times, progressively increasing the number of clusters. Ideally, two clusters would be sufficient, but naturally the unsupervised method is unable to aggregate IoT and non-IoT flows so that they are completely separated. With several clusters, instead, we might find smaller aggregates which are mostly IoT or mostly non-IoT. To make the determination, for each cluster, we inspect the number of actual IoT and NON_IoT flows that belong to the cluster. Clusters which have a majority of IoT flows are then labeled as IoT, while the others are labeled as NON_IoT. This is the supervised step of the approach: while clusters are identified based solely on the flow features, the destination of the cluster is determined based on the previous knowledge of the flow classification.

The first set of experiments makes use of the initial dataset, comprising mostly the simulated IoT flows. Table 9.1 shows in detail the results of clustering, obtained through 10-fold cross validation. The first column reports the number of clusters. The second and third columns report the confusion matrix: for each class (shown in the last column), the table shows the number of flows that were included in a cluster which was labeled as IoT or NON_IoT, respectively. The following four columns give a summary of the performance: we compute the True Positive (TP) and the False Positive (FP) rates, as well as the Precision and Recall measures for both IoT and NON_IoT flows. As the number of cluster increases, we get a better Recall for the IoT flows, reaching a maximum of 96.6% for the division in 50 clusters. As we increase the number of clusters, the overall performance slightly increases, although we are less accurate on the IoT flow.

We observe that the number of IoT flows correctly categorized as IoT flows increases up to 50 clusters. Increasing the number of clusters gives no improvement, in fact the number slightly decreases. The number of NON_IoT flows incorrectly categorized as IoT, on the other hand, steadily decreases as the number of clusters increases. A division in 50 clusters seems to provide the best trade off. Figure 9.1, left, shows in dark color the four clusters labeled as IoT traffic for the 50-cluster case, in terms of acknowledge rate from client and server.

Table 9.1 Clustering results

Clusters	IoT	NON_IoT	TP rate (%)	FP rate (%)	Precision (%)	Recall (%)	MCC (%)	Class
10 Simul.	6113	1649	78.8	1.6	84.4	78.8	79.5	IoT
	1134	68,223	98.4	21.2	97.6	98.4	79.5	NON_IoT
			96.4	19.3	96.3	96.4	79.5	Average
50 Simul.	7496	266	96.6	1.3	89.2	96.6	92.0	IoT
	904	68,453	98.7	3.4	99.6	98.7	92.0	NON_IoT
			98.5	3.2	98.6	98.5	92.0	Average
100 Simul.	7481	281	96.4	1.1	90.5	96.4	92.7	IoT
	781	68,576	98.9	3.6	99.6	98.9	92.7	NON_IoT
			98.6	3.4	98.7	98.6	92.7	Average
50 Compl.	13,266	9577	58.1	5.4	77.9	58.1	58.6	IoT
	3767	65,590	94.6	41.9	87.3	94.6	58.6	NON_IoT
			90.9	38.3	86.3	90.9	58.6	Average
100 Compl.	15,001	7842	65.7	3.4	86.4	65.7	68.8	IoT
	2353	67,004	96.6	34.3	89.5	96.6	68.8	NON_IoT
			93.5	31.2	89.2	93.5	68.8	Average

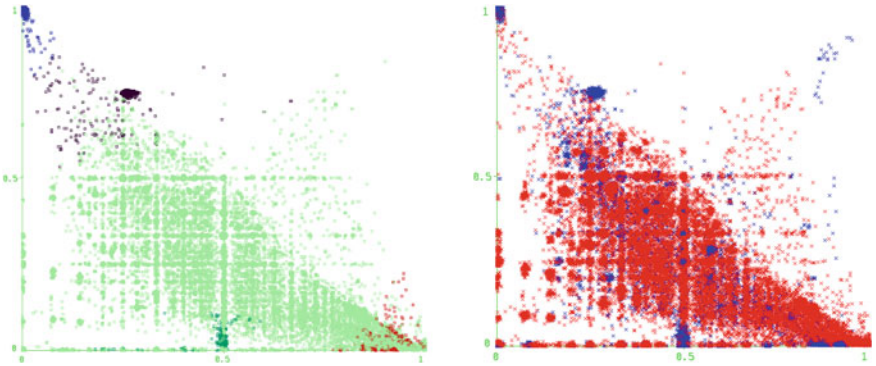


Fig. 9.1 Acknowledge rate of the client (X axis) and of the server (Y axis). IoT labeled clusters shown in dark color, non-IoT clusters in light green and red. Left: simulated IOT flows. Right: captured IoT flows

We have conducted the same analysis including the 15,000 flows from the Australian deployment. The expectation is that the results will be somewhat less satisfactory, because of the increased diversity of the devices in use. While there still are areas which are clearly identifiable, overall the distribution of IoT flows (blue dots) shown in Fig. 9.1, right, is much more dispersed. The confusion matrix is therefore far from ideal, as shown in the second part of Table 9.1. The situation slightly improves when using 100 clusters, however the precision is still fairly low, and the computational complexity of determining cluster membership increases. One of the reason why clustering does not provide good performance is that the different fea-

tures contribute symmetrically to the Euclidean distance from the cluster centroid. This is less of a concern with more homogeneous features, but induces confusion when traffic has a higher degree of overlapping.

9.3.2 J48 Classification Tree

Classification trees have been shown to perform well in protocol recognition [2]. We have generated several classification tree, for the initial simulated dataset and for the complete dataset. We have also analyzed the influence of the different parameters on both performance and tree size. As usual, the accuracy is evaluated through the resulting confusion matrix using 10-fold cross validation. Our first experiment deals with the simulated and the complete dataset using the full set of attributes. The trees performs particularly well, as shown in Table 9.2 where the confusion matrix highlights that only a few of the flows are misclassified.

In particular, the performance is superior to many other methods that we have analyzed (including SVM, Naïve Bayes and 3-level perceptron [7]), with an average precision and recall that exceed 99% for both datasets. Table 9.3 shows the tree information in terms of computational complexity.

The first column reports the total size of the tree (number of nodes), while the second column counts the number of leaves in the tree. The size of the tree gives an estimate of the amount of memory required to store the tree information. The following three columns provide information regarding the depth of the tree: the minimum and the maximum depth to reach a leaf, as well as the average, where the

Table 9.2 Accuracy of the classification trees with all attributes

Config	IoT	NON_IoT	TP rate (%)	FP rate (%)	Precision (%)	Recall (%)	MCC (%)	Class
J48 simulated	7686	76	99.0	0.1	99.4	99.0	99.1	IoT
	47	69,310	99.9	1.0	99.9	99.9	99.1	NON_IoT
			99.8	0.9	99.8	99.8	99.1	Average
J48 complete	22,449	394	98.3	0.5	98.5	98.3	97.8	IoT
	345	69,012	99.5	1.7	99.4	99.5	97.8	NON_IoT
			99.2	1.4	99.2	99.2	97.8	Average

Table 9.3 Complexity of the classification trees with all attributes

Config	Size	Leaves	Min depth	Max depth	Avg depth
J48 simulated	125	63	1	11	4.4
J48 complete	635	318	2	23	7.4

Table 9.4 Classification tree performance, Weka selected attributes

Config	IoT	NON_IoT	TP rate (%)	FP rate (%)	Precision (%)	Recall (%)	MCC (%)	Class
J48 simulated	7701	61	99.2	0.1	99.5	99.2	99.3	IoT
	36	69,321	99.9	0.8	99.9	99.9	99.3	NON_IoT
			99.9	0.7	99.9	99.9	99.3	Average
J48 complete	22,475	368	98.4	0.5	98.6	98.4	98.0	IoT
	315	69,042	99.5	1.6	99.5	99.5	98.0	NON_IoT
			99.3	1.3	99.3	99.3	98.0	Average

depth is weighted by the number of flows in the training set that are associated with each particular leaf. The data shows that the lower variability associated with the simulated flows results in a much smaller and shallower tree for classification.

It is interesting to study the influence of each attribute on the classification accuracy. This could be useful, for instance, to select only a subset of the attributes that provide most of the performance. To choose the most relevant attributes, we proceed in two ways. The first is a greedy search, whereby we evaluate the classification performance using progressively more attributes. Hence, we start by evaluating the performance of all trees that use only one attribute, and keep the attribute that provides the best performance. Then, we evaluate all trees with two attributes, having fixed the first in the previous step. The results show that performance increases quickly with the addition of more attributes. In both the simulated and the complete dataset, three specific attributes are selected among the first four. These correspond to the fraction of acknowledge from the client to the server, the fraction of bytes from client to server, and the client minimum round trip time. In the simulated case, the fraction of packets from client to server completes the set, whereas for the complete case the minimum server round trip time is used. The second mechanism for attribute selection makes use of the facility provided by the Weka framework. We perform a Wrapper Subset Evaluation, which is a scheme similar to the one employed above, using a Greedy Step-wise incremental search. In all cases, we select the J48 algorithm for evaluation. In both the simulated and complete case, Weka selects eight attributes out of the available 14, including the ones that we have determined using the manual procedure above. The results of generating the classification tree are shown in Table 9.4. The accuracy is slightly better than that of the tree that uses all the attributes together. This may be an indication that there is some degree of "overfitting", i.e., that there are too many parameters to choose from.

9.4 Conclusions

In this paper, we have discussed the statistical classification methods to discriminate between IoT and non-IoT traffic. We have shown that semi-supervised clustering works reasonably well in the case of homogeneous traffic. We have then considered supervised methods, such as classification trees. The results of 10-fold cross validation show that classification trees provide the best results, with performance in excess of 99% accuracy. Attribute selection is used to narrow down the set of attributes and to avoid overfitting.

Our future work is moving in two directions. The first is to explore different attributes that can be extracted from the data. Regularity in the time interval transmission has been highlighted as one peculiar feature of IoT traffic [5]. The use of frequency analysis could therefore help with flow characterization, although the method could suffer from a computational complexity point of view. Another direction includes a form of dynamic learning to follow the evolution of the behavior of devices [6]. We could evaluate the degree of confidence in classification by looking at the distance of the flows from the edge of the hypercubes identified by the tree in the attribute space. When the system observes that the overall classification confidence has decreased significantly, a new round of supervised learning could be employed to restore the lost accuracy.

References

1. Witten IH, Frank E, Hall MA, Pal CJ (2016) Data mining: practical machine learning tools and techniques. Morgan Kaufmann, Cambridge
2. Grimaudo L, Mellia M, Baralis E (2012) Hierarchical learning for fine grained internet traffic classification. In: Proceedings of IWCMC, Aug 2012
3. Fontugne R et al (2010) Mawilab: Combining diverse anomaly detectors for automated anomaly labeling and performance benchmarking. In: ACM CoNEXT10, Dec 2010
4. Sivanathan A, Habibi Gharakheili H, Loi F, Radford A, Wijenayake C, Vishwanath A, Sivaraman V (2018) Classifying IoT devices in smart environments using network traffic characteristics. IEEE Trans Mob Comput
5. Shafiq MZ, Ji L, Liu AX, Pang J, Wang J (2013) Large-scale measurement and characterization of cellular machine-to-machine traffic. IEEE/ACM Trans Netw 21(6):1960–1973
6. Grimaudo L, Mellia M, Baralis E, Keralapura R (2014) SeLeCT: self-learning classifier for internet traffic. IEEE Trans Netw Serv Manage 11(2):144–157
7. Pant V, Passerone R, Welponer M, Rizzon L, Lavagnolo R (2017) Efficient neural computation on network processors for IoT protocol classification. In: Proceedings of the first new generation of circuits and systems conference, NGCAS 2017, Genova, Italy, 7–9 Sept 2017

Chapter 10
Using LPWAN Connectivity for Elderly Activity Monitoring in Smartcity Scenarios

D. Fernandes Carvalho, P. Ferrari, E. Sisinni, P. Bellitti, N. F. Lopomo and M. Serpelloni

Abstract Home care is an increasing research area; an example is the interest in daily activity and mobility tracking, known to be a strong indicator for people's health. In particular, the digital mobility assessment of elderly can anticipate and prevent hard clinical events such as falls, that could result in hospitalizations and deaths. In this work, the use of LoRaWAN is verified in a real-world scenario as an effective communication infrastructure for transmitting activity level information to a supervisor structure like a clinic or a hospital. An experimental setup has been purposely implemented to evaluate the feasibility; in particular, the activity level inferred analyzing accelerometers data can be notified with an average delay in the order of 500 ms.

10.1 Introduction and Motivation

The Internet of Things paradigm has already affected the way healthcare services are provided [1]. In non-urban areas, in mountain areas, in smaller islands, or in any case characterized by a sparse population, in which the use of single clinical sites is not conceivable, it is necessary to promote the use of telemonitoring, teleassistance and, more in general, telehealth solutions. In this perspective, the use of ICT applications in home care results to be an increasing research area, with a huge set of ICT solutions that can be used to enhance accessibility to home care [2]. For instance, daily activity and mobility result to be a strong indicator for people's health [3]. Additionally, permanent digital monitoring would allow earlier diagnosis and faster response times, providing new digital biomarkers able to anticipate and prevent hard clinical endpoint such as falls. Here relies the importance of monitoring the activities of elderly people and chronic patients in the home ecology.

D. Fernandes Carvalho · P. Ferrari · E. Sisinni (✉) · P. Bellitti · N. F. Lopomo · M. Serpelloni
Department of Information Engineering, University of Brescia, Via Branze, 38-23123 Brescia, Italy
e-mail: emiliano.sisinni@unibs.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_10

In light of these considerations, this work suggests using LoRaWAN, a wireless communication solution belonging to the LPWAN category, for telehealth. Purposely designed for addressing many IoT-related applications requiring wide coverage and sporadic transmissions, LoRaWAN allows to implement cellular networks without the support of a third-party provider. Thus, limitations of mobile-based solutions are avoided. Additionally, the high sensitivity offered by the LoRa radios ensure good coverage, overcoming WiFi-based solutions when hybrid indoor/outdoor scenarios are taken into account. Security aspects are considered as well; encryption on both the network and application level is implemented. Consequently, LoRaWAN has been already proposed as a viable solution for e-health monitoring by many researchers, as demonstrated by the available literature [4–9]. In this paper, differently from other works, several innovative wearable devices, including a LoRaWAN modem complemented by an accelerometer-based monitoring system, are deployed and tested in a real-world public infrastructure. Each device allows to track body movements, offering minimum invasiveness. A parameter is locally evaluated, assessing the physical activity, and periodically sent to a supervisory center. Results about the communication delays confirmed the suitability of the proposed solution not only in monitoring the activity in elderly in a daily-life scenario, but for fall detection as well.

10.2 The Proposed Wearable System for Tracking Elderly Activity

As stated in the introduction, falls are ones of the leading cause of injuries [10] in geriatric population, and a sedentary lifestyle leads to a lower quality life [11]. Figure 10.1 shows the proposed tracking system application scenario, where a self-sufficient elderly person can carry out normal daily activities wearing the device. The data, collected by local LoRaWAN gateway(s), are tunneled through an Internet

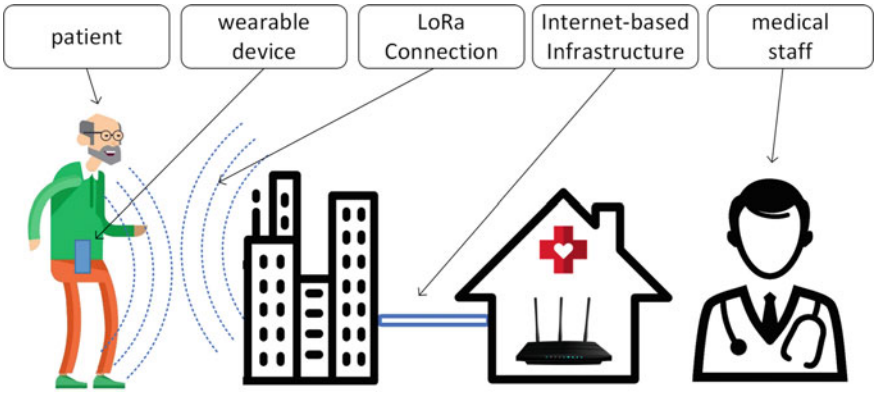


Fig. 10.1 Typical scenario application of the proposed wearable device

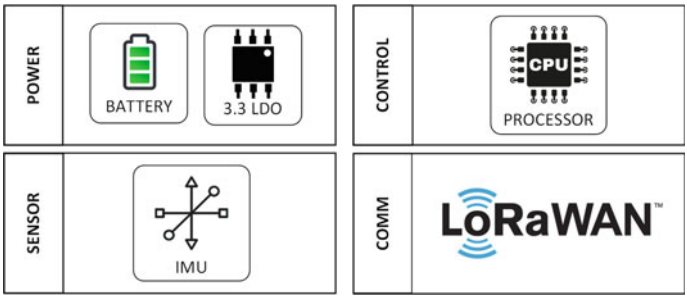


Fig. 10.2 Wearable device block diagram

connection to a supervisory structure, like a clinic or a hospital, where medical staff can infer the patient's health status by analyzing the data. Additionally, in case of a detected fall, emergency services can be provided promptly.

The proposed system is composed by a small and lightweight wireless wearable module that can track the motion thanks to its sensors. The main components are represented in Fig. 10.2.

All the measuring and transmitting operation are coordinated by an ATmega328P microcontroller unit. The motion data are retrieved from the accelerometer section of an inertial motion unit (IMU), LSM9DS1 from STMicroelectronics. The wireless low-power data transmission is provided by a LoRaWAN modem (an RN2483 from Microchip). The overall board is supplied by a small-size LiPo battery (20 mm × 11 mm × 3 mm) which guarantees proper functioning for about two days when the accelerometer data are sampled at 20 Hz and the activity level parameter is transmitted once per hour (and excluding event-based transmissions due to fall events recognition). If needed, the processor can reduce the measuring and transmitting frequency to save energy. Both electronic board and battery are closed inside a box fabricated with addictive manufacturing technique (3D printing), box size: 36 mm × 26 mm × 10 mm. The overall device weight is about 15 g to increase its wearable characteristic.

10.3 LPWAN for Smartcities: The LoRaWAN Solution

The LoRaWAN is a network with star-of-stars topology. The vast majority of information is transferred with "uplink" transactions: they are started by the end nodes and directed to the backend servers. Wireless messages are collected by gateways, which run the "packet forwarder" software, that tunnels messages over the air into the wired backhaul network (and vice versa, when reversed transactions—"downlink"—are needed). Regarding security aspects, messages are encrypted on a session base by means of application keys, while authentication at the network level is provided by network keys; another backend server is generally in charge of managing the

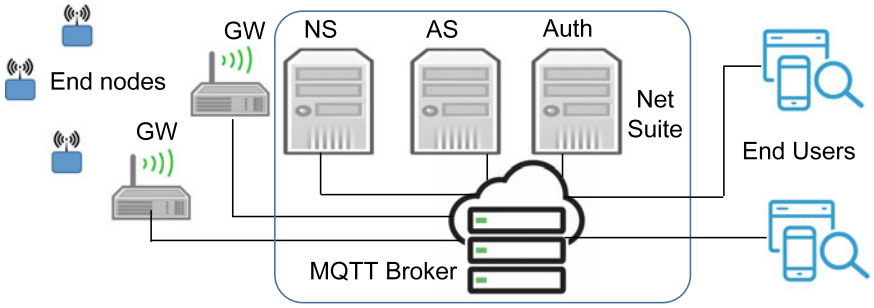


Fig. 10.3 Architecture of the LoRaWAN solution used in the BSL project

keys depending on the activation procedure. An example of possibilities offered by LoRaWAN for smartcity applications is given by the "Brescia Smart Living" (BSL) project. The Patavina NetSuite solution, provided by A2A Smart City, is used as the LoRaWAN backend (see the block diagram of the implemented architecture shown in Fig. 10.3).

It implements the Network, the Application and the Authentication Servers (NS, AS and Auth in Fig. 10.3), for managing the network, allowing end user application integration through a MQTT Broker and handling keys. Each end node uplink is published by the Broker as an MQTT topic, which can be subscribed by the end users interested in the information. It has to be pointed out that more than 100 LoRaWAN gateways are currently used to cover all urban areas of the city of Brescia, making BSL one of the wider LoRaWAN project across the world.

10.4 Experimental Validation

In this section the capabilities of the proposed wearable device are detailed. In particular, first it is shown how the system can collect information about physical activity and then the delays in transmitting such information are evaluated.

10.4.1 Activity Monitoring

In Fig. 10.4 an example is reported, regarding the data obtained from the analysis of two movements. In the left part (Fig. 10.4a) there are the acceleration components measured during a walk at a normal rate. The system is able to compute an activity level related parameter which is periodically sent to the healthcare physician for helping him in deciding if the patient has a sufficiently active lifestyle. In Fig. 10.4b

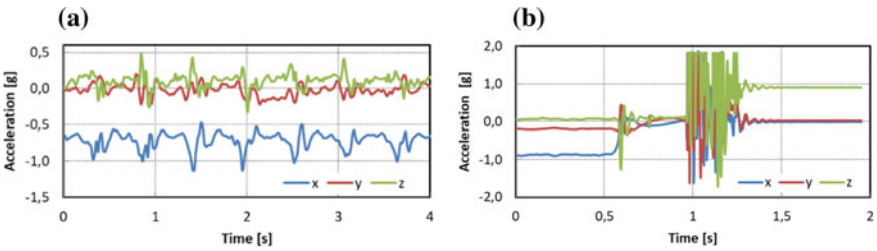


Fig. 10.4 Acceleration component retrieved by the system: **a** normal walking, **b** ahead fall ending face downward

we can observe an ahead fall ending with face downward. In this case, the device can send an automatic help request.

10.4.2 Application Delay of the LoRaWAN Network in BSL

In order to measure the application delay [12] inside the Patavina NetSuite infrastructure, the experimental setup of Fig. 10.5 has been built; it consists of one single node (located in the University laboratory and based on a PC connected to the LoRaWAN modem RN2483) sending information via uplink to several user end points (implemented by IOT2040 platforms; EP1 is connected to the Internet via the University reliable and fast access; EP2, located in Brescia and EP3, located in Milan, leverage on ADSL links).

In this way, timestamp T1 is registered when a LoRaWAN uplink transmission initiates. Each EP_n is a MQTT subscriber of the topic "event of interest" in the MQTT

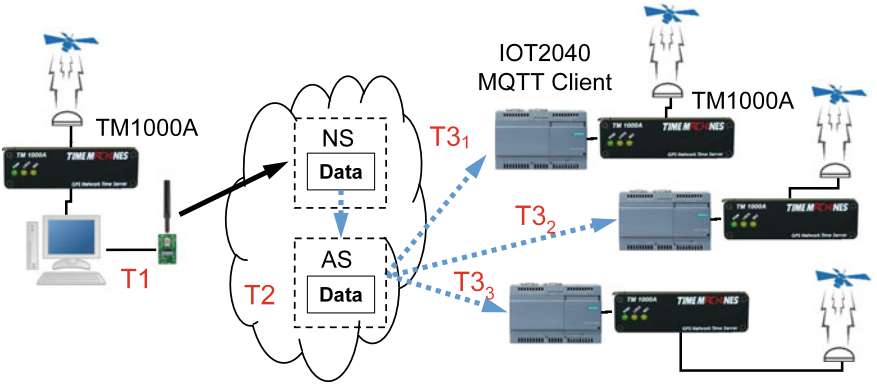


Fig. 10.5 Experimental setup with different end points (EP1 is connected to the Internet via the University reliable and fast access; EP2, located in Brescia and EP3, located in Milan, leverage on ADSL links)

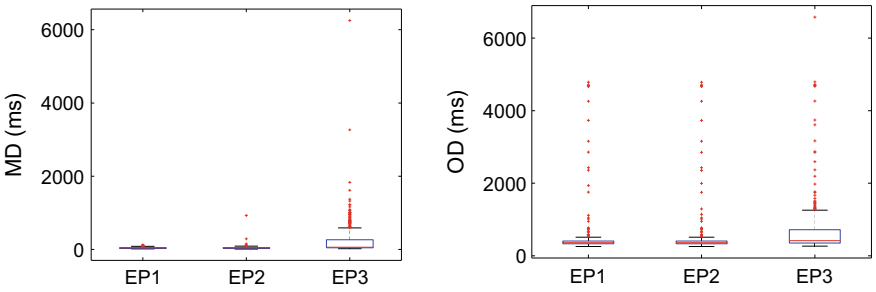


Fig. 10.6 Boxplot of the overall end-to-end delay OD for the three considered endpoints

Broker; when a new message is received, the message is timestamp tagged as $T3_n$. Moreover, the AS is in charge of registering the timestamp T2 when the "event of interest" arrives. The following metrics are calculated based on these timestamps: the LoraWAN backbone delay is ND = T2 − T1; the MQTT broker delay is MD_n = $T3_n$ − T2; and the overall end-to-end application delay is $OD_n = T3_n - T1$. Time dissemination is performed by means of TM1000A NTP time servers, each one UTC-synchronized via a GPS receiver. The NetSuite is natively UTC-synchronized.

The experiments last for one day, summing a total number of 1440 messages transmitted every 60 s. Without losing generality, the user message length is 30 B and includes the transmission timestamp and a sequence number for sorting, totalizing the time on air of about 226 ms (Spreading Factor = 7 and Coding Rate = 4/5). Regarding the network delay, the average delay is $ND_{AVE} = 438$ ms and the standard deviation is $ND_{STD} = 592$ ms; however, it is interesting to highlight that some outliers exist, leading to a maximum value $ND_{MAX} = 4738$ ms. The distribution of the MD and OD metrics are reported in Fig. 10.6a and b, respectively. The three endpoints (EP1, EP2 and EP3) have an average OD delay of about 500 ms, enough for long-term monitoring and possible fall detection and notification. As expected, the EP3 has the worst performance, due to the poor performance of the available internet connection.

10.5 Conclusions

In this work a wearable system for continuously tracking the physical activity of elderly has been proposed and described. Patient movements are collected by means of a MEMS accelerometer and used to compute resuming activity-related parameters by the local microcontroller. The device is complemented by a LoRaWAN modem, which exploits the LoRaWAN infrastructure to update periodically several supervisory center (e.g. hospital) or patient relatives. Doctors can then estimate if the patient is doing enough activity or not. Accelerometer data are used to detect falls as well; in such a case, a notification is promptly sent.

References

1. Depari A, Carvalho DF, Bellagente P, Ferrari P, Sisinni E, Flammini A, Padovani A (2019) An IoT based architecture for enhancing the effectiveness of prototype medical instruments applied to neurodegenerative disease diagnosis. Sensors (Switzerland) 19(7), art. no. 1564
2. Lindberg B, Nilsson C, Zotterman D, Söderberg S, Skär L (2013) Using information and communication technology in home care for communication between patients, family members, and healthcare professionals: a systematic review. Int J Telemed Appl 2013, Article ID 461829, 31 pages
3. Fritz et al. (2009) White paper: 'walking speed: the sixth vital sign'. Journal of Geriatric Physical Therapy, 32
4. Koekuza, H.: Utilization of loT in the Long-term Care Field in Japan, in 2nd International Conference on Cloud Computing and Internet of Things (CCIOT), Dalian, China (2016)
5. Wang H, Fapojuwo A (2018) Performance evaluation of LoRaWAN in North America urban scenario 2. In: IEEE 88th vehicular technology conference (VTC-Fall), Chicago, USA (2018)
6. Hossain T, Inoue S (2019) Sensor-based daily activity understanding in Caregiving Center. In: Sensor-based daily activity understanding in Caregiving Center, Kyoto, Japan
7. Yang G, Liang H (2018 Nov 15) A smart wireless paging sensor network for elderly care application using LoRaWAN. IEEE Sens J 18(22)
8. Mdhaffar A, Chaari T, Larbi K, Jmaiel M, Bernd F (2017) IoT-based health monitoring via LoRaWAN. In: IEEE EUROCON—17th international conference on smart technologies, Ohrid, Macedonia, July, 2017
9. Catherwood P, Rafferty J, McComb S, McLaughlin J (2018) LPWAN Wearable intelligent healthcare monitoring for heart failure prevention. In: 20th international conference on human computer interaction, Las Vegas, Nevada, EUA
10. Kamińska M, Brodowski J, Karakiewicz B (2015) Fall risk factors in community-dwelling elderly depending on their physical function, cognitive status and symptoms of depression. Int J Environ Res Public Health 12(4):3406–3416
11. Cvecka J, Tirpakova V, Sedliak M, Kern H, Mayr W, Hamar D (2015) Physical activity in elderly. Eur J Transl Myol 25(4):249. Available: https://doi.org/10.4081/ejtm.2015.5280
12. Fernandes Carvalho D, Ferrari P, Sisinni E, Depari A, Rinaldi S, Pasetti M, Silva D (2019) A test methodology for evaluating architectural delays of LoRaWAN implementations. Pervasive Mob Comput 56:1–17

Part III
Processors and Memories

Chapter 11
Characterization of a RISC-V Microcontroller Through Fault Injection

Dario Asciolla, Luigi Dilillo, Douglas Santos, Douglas Melo, Alessandra Menicucci and Marco Ottavi

Abstract This article reports the results of fault injection on a microcontroller based on the RISC-V (Riscy) architecture. The fault injection approach uses fault simulation based on Modelsim and targets a set of 1000 fault injected per microcontroller block and per benchmarck. The chosen benchmarks are the Dhrystone and CoreMark that may represent generic workloads. The results show certain block are more prone to fault than others, as also confirmed by a vulnerability analysis that correlates the number of observed faults and the rate of access to the blocks.

Keywords RISC-V · Fault injection · Micronontroller · Simulation

D. Asciolla
LIRMM, University of Montpellier, Montpellier, France
e-mail: dasciolla@lirmm.fr

L. Dilillo
LIRMM, CNRS, University of Montpellier, Montpellier, France
e-mail: dilillo@lirmm.fr

D. Santos · D. Melo
Laboratory of Embedded and Distributed Systems, University of Vale do Itajaí, Itajaí, Brazil
e-mail: douglasas@edu.univali.br

D. Melo
e-mail: drm@univali.br

A. Menicucci
Department of Space Engineering, Delft University of Technology, Delft, Netherlands
e-mail: A.Menicucci@tudelft.nl

M. Ottavi (✉)
Department of Electronics Engineering, University of Roma Tor Vergata, Rome, Italy
e-mail: ottavi@ing.uniroma2.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_11

11.1 Introduction

The space environment interaction with electronics represents an important challenge for satellite missions. Ionizing particles and electromagnetic radiations affect electronic devices by inducing faults in specific circuit areas that may lead to system failures. These failures can be temporary, with the occurrence of the so-called Soft Errors, or permanent with the occurrence of Hard Errors [1]. Moreover, the exposition of electronics to radiation induces to premature aging, with the deterioration of performance and/or functionality of the systems. For this reason, the space industry resorts to hardening techniques that are generally based on hardware and software redundancy, with the generation of custom devices. This type of solutions, while effective, are energy and hardware greedy and leads to costs several times higher than for conventional COTS (Commercial Off The Shelf) electronics.

Completely programmable hardware platforms such as FPGA make the development of a system extremely flexible but, at the same time, require ad-hoc designing and therefore they are not available to a broad programming community. On the other side, the use of standard processors ISA architectures allows many developers to design applications, but the closeness of the architectures does not allow the designers to make easy and cheap modifications to the underlying hardware.

RISC-V allows developers to combine the advantages of both worlds [2], providing flexibility to both hardware and software. On one side we can modify the architecture to obtain specific applications, while on the other side we can open to applications made by programmers, who are unaware of the underlying hardware. RISC-V is an open ISA born from both the academia and research environment [3]. The RISC-V ISA was originally developed in the Computer Science Division of the EECS Department at the University of California, Berkeley [3]. RISC-V represents a promising platform to experiment different techniques and to design new architectures.

The purpose of this paper is characterizing a RISC-V core, through an extensive simulation-based fault injection campaign with the target of identifying the most critical modules within the core. For this purpose, the study of sequential modules is fundamental, especially for COTS components, because they can suffer from bit flips [4]. The data stored in the memory element can be corrupted, with Single Event Upsets (SEUs), after the interaction with ionizing particles and electromagnetic radiations in general. This paper analyzes the effects of SEU (Single Event Upset) in a RISC-V core. This characterization is useful in perspective to design a fault-tolerant version of a RISC-V core for space applications by applying targeted hardening techniques, shaped on the sensitivity of the different blocks composing the system. The final target is the use of this kind of low cost hardened processor within nanosatellites, like Cubesats, and in other systems where high reliability, flexibility and low cost are required.

The rest of the paper is organized as follows. Sections 11.2 and 11.3 detail the RISC-V platform and fault injection methodology, respectively. Section 11.4 introduces the chosen algorithmic benchmarks. Sections 11.5 and 11.6 present the simulation results as well as an analysis of the microcontroller vulnerability. Conclusions are given in Sect. 11.7.

11.2 Platform

The chosen RISC-V platform is the Parallel Ultra Low Power (PULP) Platform. It is been designed from Integrated Systems Laboratory (IIS) of ETH Zrich and Energy-efficient Embedded Systems (EEES) group of the University of Bologna [5]. It is an open-source platform and very useful for our purposes, because it is possible to access all parts of the system. In this specific case, from this platform, the Pulpino microcontroller has been chosen. It is built for RISC-V Riscy and zero-riscy core [6]. Pulpino offers a separate memory for instructions and data. It uses AXI (Advanced eXtensible Interface) interface as its main interconnect and a bridge to APB (Advanced Peripheral Bus) for simple peripherals [6]. All architectural details about Pulpino platform [6] are shown in Fig. 11.1.

The choice of Riscy core for this study is based on the following reasons. Firstly, it is a four-stage RISC-V core and it can run most of the typical workloads. It is a 32-bit core and for the chosen configuration, it can manage only integer numbers. It implements the RV32I instruction set. All software that runs over this core has been compiled using the GNU RISC-V Toolchain [7] with an optimization level equal to 3.

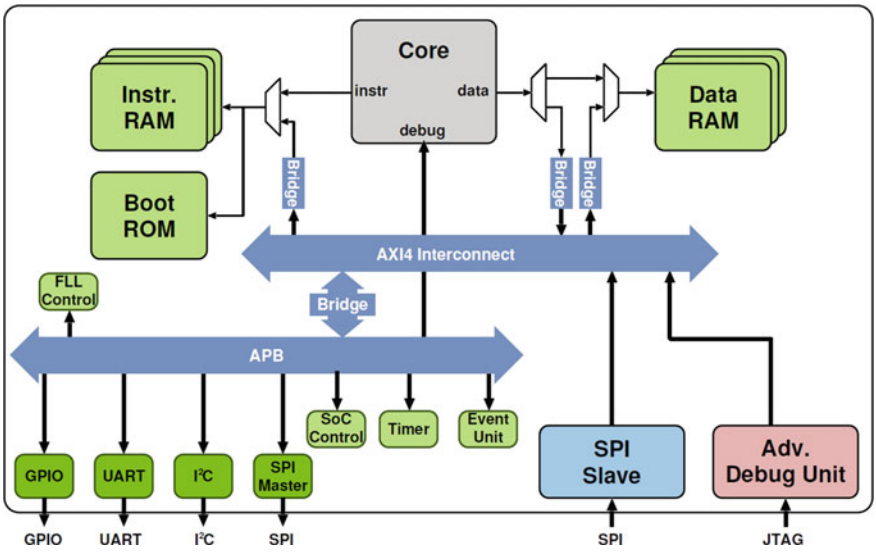


Fig. 11.1 RISC-V PULPino platform architecture overview

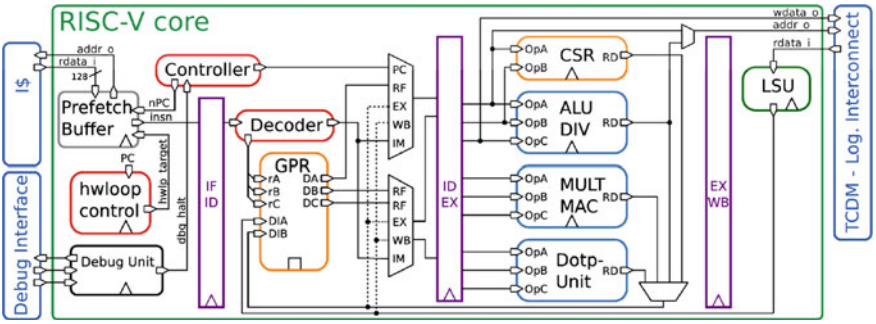


Fig. 11.2 RISC-V Riscy core architecture overview

Riscy core architecture overview [6] is shown in Fig. 11.2. For the characterization campaign, we are focusing over all sequential modules inside the core. In Riscy core there are four stages: Instruction Fetch, Instruction Decode, Execution and the Write Back. All stages are separated by an interface register. In the Instruction Fetch stage, sequential parts are inside the prefetch buffer, the hardware loop control module and inside the instruction fetch top-level registers. In the Instruction Decode stage, sequential modules are inside the register file and the controller unit. In the Execution stage, the control-state registers and the multiplier are both sequential modules. In the write-back stage, the load-store unit is the only sequential module. The only architectural modification that was made consists of the redefinition of state machines inside sequential modules by using the binary codification instead of labels. This modification was introduced to simplify the fault injection procedure.

11.3 Fault Injection Environment

For replicating the typical effects of space radiation environment on electronics [8], a simulation-based fault injection technique was chosen. This technique allows full access to the entire processor without any architectural modifications. One of the main disadvantages of this technique is that, being simulation based, required a long time to run compared to execution on hardware emulators such as the FPGA based ones [9]. This simulation-based strategy is based on TCL (Tool Command Language) scripts, that allow the manipulation of signals for fault injection and observe fault effects. The HDL (Hardware description language) simulator used for the system simulation and to run TCL scripts was the Modelsim [10] from Mentor Graphics.

11.3.1 Fault Model

The used fault model is based on SEU occurrence in sequential logic blocks. In each simulation, a single fault is injected to cause a bit flip inside the chosen sequential

block. Other effects could be Multiple Bit Upset (MBUs), Multi Cell Upsets Single Event Latchups, but they are out of the scope of this study.

11.3.2 Simulation Procedure

The first step of the procedure, a golden simulation, with no injected fault , is performed to obtain the reference data to be used for the detection of mismatches caused by fault injections. The Riscy core fault injection is performed for all sequential subsystems. For each sequential subsystem, 1000 simulations are performed, 1 fault injected per simulation run. The following steps [11] summarize the tasks executed for each simulation:

(1) Selection of a flip-flop, in a certain sequential subsystem, where the fault will be injected. This is done selecting, in a random way, from a list that contains all signals, in the VHDL code, which implement registers. Each signal corresponds to each bit of the register.
(2) Selection of a random instant when the fault will be injected. In order to avoid a fault during the logging process, the fault is injected before the reporting process.
(3) Simulation runs until the chosen injection instant.
(4) Injection of the fault by forcing a bit flip in the target sequential element.
(5) Simulation runs until the end of the algorithmic benchmark.
(6) Making a copy of the register file content.
(7) Storing the print out of the program results.

If exceptions are generated during the execution, they are stored in a file and whether the core doesn't respond after a threshold time a relative log is generated.

11.3.3 Fault Effects Classification

Data obtained during the simulation campaign are used to classify fault effects that can be summarized in five categories [11] that are listed below:

- *No Effect*—The simulation finishes obtaining the correct result from the program and the content of the register file is equal to the reference one.
- *Latent*—The simulation finishes obtaining the right result, but the content of the register file is not equal to the reference.
- *Wrong result*—The system has a failure and the simulation finishes obtaining the wrong program result.
- *Timed out*—The simulation takes an abnormal amount of time to finish the program execution compared to the reference.
- *Exceptions*—The core generates exceptions during the simulation.

Latent errors potentially can propagate and lead to a system failure in the future, but these errors may also be masked by the normal core functioning.

11.4 Chosen Benchmarks

Benchmarks usually are used to evaluate performances of microcontrollers, microprocessors and computing systems in general. In this characterization campaign, they are used because they offer a generic workload that can cover almost all operations that a core can execute. Benchmarks perform a large number of different operations such as logic, numeric and string operations. The chosen benchmarks, for this campaign, are Dhrystone and CoreMark, described below.

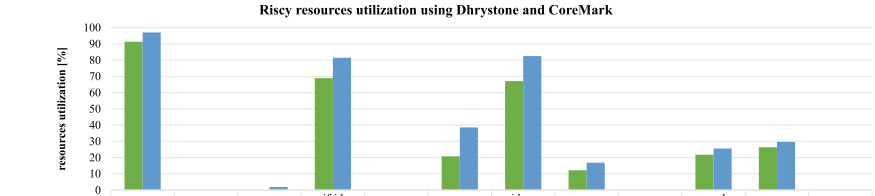
- *Dhrystone benchmark* provides a measure of integer performance and no floating point instructions. Here, it has been used the 2.1 C version that avoids over optimization problems [12] encountered with the first version. Dhrystone benchmark workload [13] can be categorized in: ALU operations for 42% of the instructions; 20% load instructions; 15% store instructions; 21% branch instructions; 2% shifting instructions.
- *CoreMark benchmark* from EEMBC [14] (Embedded Microprocessor Benchmark Consortium) is specifically designed for embedded systems and it can be used to measure microcontrollers and microprocessors performances. It is considered the next version of Dhrystone [14]. It implements numbers of algorithms like find, sort, matrix manipulation, state machine and crc. The crc is used both to provide a typical workload for an embedded application and to check the results of the operations. This benchmark is designed to be independent of the compiler optimization options and this is one of the improvements respect to Dhrystone [14] .

11.5 Simulation Results

This section presents and discusses the results of the fault injection campaign and the measure of the utilization of sequential modules. These simulations are useful for vulnerability estimation.

11.5.1 Resources Utilization Using Dhrystone and Coremark

The resources utilization has been measured using a Modelsim simulation running a TCL script. Coremark is configured to perform 1 cycle while Dhrystone 1000 cycles. In this study, the focus is over the sequential parts inside modules that are the target



	prefetch unit	if top level registers	hw loop	if-id pipeline register	controller	register file	id-ex pipeline register	control-state registers	multiplier registers	ex-wb pipeline register	load store unit	interrupt controller
r.u. - Dhrystone [%]	91.31	0.00013	0.00	68.89	0.00150	20.77	67.04	12.18	0.00	21.70	26.31	0.00
r.u. - CoreMark [%]	97.01	0.09246	1.88	81.45	0.00179	38.56	82.49	16.81	0.00	25.58	29.67	0.00

Fig. 11.3 RISC-V Riscy sequential resources utilization

of the characterization through fault injection. A simulation both for Dhrystone and CoreMark is performed obtaining the resources utilization for the entire simulation time. In this simulation, the measure of the resource utilization is performed counting how many times the value, stored in a given register, changes from the beginning to the end of the simulation. This was made using a TCL script that runs in Modelsim.

The workload is similar for both simulations, as shown in Fig. 11.3, for Dhrystone and CoreMark. The most used modules are the prefetch unit, the instruction fetch-instruction decode pipeline register and the instruction decode-execute pipeline register. There are modules that are never used during the benchmark execution like the hardware loop for Dhrystone and the multiplier registers and the interrupt controller for both benchmarks (in the run simulations).

11.5.2 Characterization Through Fault Injection Using Dhrystone Benchmark as Workload

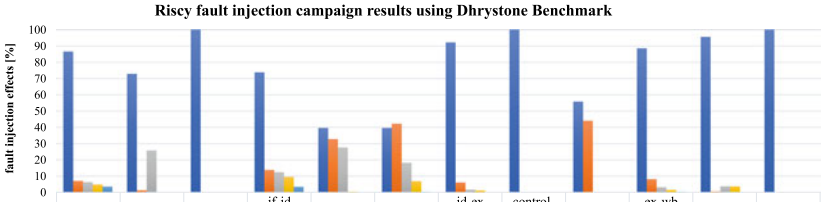
In Fig. 11.4 are shown the results of the simulation campaign. As mentioned above, for each microcontroller block, 1000 simulations were performed, with a fault has been injected for each run.

This procedure is repeated for each block, obtaining the results showed in the plot.

The graph shows that the most critical sequential modules are the controller and the register file, with injected faults that cause a large number of latent errors and wrong results. Despite the fact that the controller is used with a lower frequency than the register file, it causes a large number of failures when it undergoes to fault injection.

Exceptions are generated from modules inside the instruction fetch stage, in the instruction decoder stage and in the execution-write back pipeline register. In this core, exceptions are used to report a wrong instruction operation code.

The hardware loop module, the interrupt controller and the control-state registers don't cause any failure when faults are injected.



	prefetch unit	if top level registers	hw loop	if-id pipeline register	controller	register file	id-ex pipeline register	control-state registers	multiplier registers	ex-wb pipeline register	load store unit	interrupt controller
no effect [%]	86.6	72.9	100.0	73.8	39.6	39.6	92.2	100.0	55.8	88.6	95.6	100
latent [%]	7.1	1.3	0.0	13.8	32.8	42.2	6.1	0.0	44.1	8.2	0.6	0
wrong result [%]	6.3	25.8	0.0	12.4	27.6	18.2	1.7	0.0	0.1	3.2	3.8	0
timed out [%]	4.8	0.0	0.0	9.5	0.5	6.9	1.2	0.0	0.0	1.5	3.7	0
exceptions [%]	3.6	0.0	0.0	3.5	0.0	0.4	0.0	0.0	0.0	0.4	0.0	0

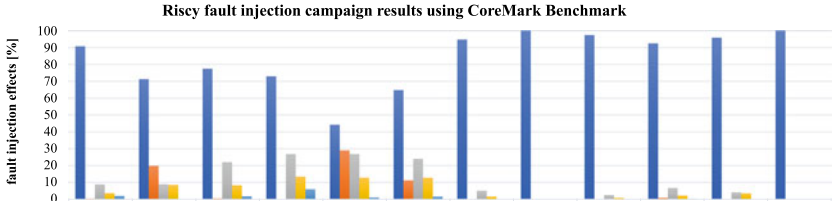
Fig. 11.4 Fault injection campaign results using Dhrystone Benchmark

A particular behavior to be noticed is about the multiplier module. It is never used during the program execution but it causes failures when faults are injected, because due to its implementation, faults can propagate in other modules.

11.5.3 Characterization Through Fault Injection Using CoreMark Benchmark as Workload

Figure 11.5 shows the results of the simulation campaign with the same procedure used above.

Like in the analysis concerning the other benchmark, it can be noticed that the most critical modules are the controller and the register file, which present a large number of latent errors and wrong results. The controller is again accessed with lower frequency w.r.t. the register file, but it displays high vulnerability.



	prefetch unit	if top level registers	hw loop	if-id pipeline register	controller	register file	id-ex pipeline register	control-state registers	multiplier registers	ex-wb pipeline register	load store unit	interrupt controller
no effect [%]	90.8	71.3	77.5	73.0	44.2	64.8	94.8	100.0	97.5	92.5	95.9	100
latent [%]	0.4	19.8	0.5	0.2	29.0	11.2	0.1	0.0	0.0	0.8	0.0	0
wrong result [%]	8.8	8.9	22.0	26.8	26.8	24.0	5.1	0.0	2.5	6.7	4.1	0
timed out [%]	3.6	8.5	8.3	13.4	12.8	12.8	1.6	0.0	0.8	2.2	3.5	0
exceptions [%]	2.0	0.0	1.8	6.0	1.1	1.6	0.1	0.0	0.0	0.3	0.0	0

Fig. 11.5 Fault injection campaign results using CoreMark Benchmark

Exceptions are generated from modules inside the instruction fetch stage, in the instruction decoder stage and in the execution-write back pipeline register. In this core, exceptions are used to report a wrong operation code of the instruction.

The interrupt controller and the control-state registers don't cause any failure when faults are injected. In this case, CoreMark stimulates the usage of the hardware loop module and we noticed system failures caused by faults injected inside this module.

In this case, the multiplier module shows less latent errors respect the Dhrystone workload.

11.6 Vulnerability Estimation

Results obtained from the fault injection campaigns present similar trends for the two workloads. In this section, we try to calculate the vulnerability of each block of the RISC-V Riscy core, by introducing the following equation:

$$v = \begin{cases} \frac{f}{u} \times c, & \text{if } u > 0 \\ 0, & \text{otherwise} \end{cases} \tag{11.1}$$

- v resource vulnerability.
- f failure rate. It is equal to the number of wrong results normalized to the number of simulations;
- u resource utilization. For each module, it is equal to the number of clock cycles of activity over the total number of clock cycles.
- c normalization constant equal to 100.

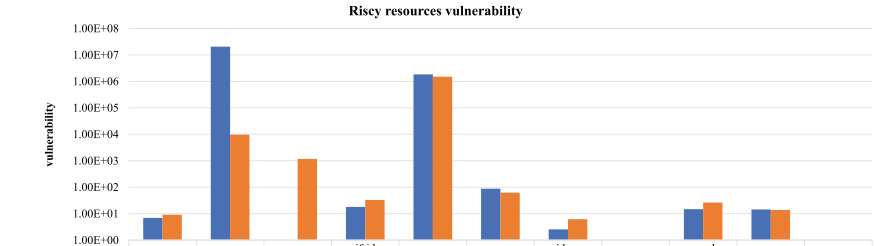
This approach allows to extrapolate a general evaluation of the block vulnerability that is independent of the used benchmark algorithm. Since it is based on the correlation between the amount of detected failures and the actual use of the blocks of the microcontroller. Figure 11.6 shows the results of the vulnerability associated with each block.

The plot uses a logarithmic axis to easily visualize the results.

It can be noticed that there is a correlation between the vulnerability calculated from both campaigns. The subsystems that show a high vulnerability are the instruction fetch registers, the controller and the register file.

Lower but significant vulnerability magnitude is showed for the prefetch unit, instruction fetch-instruction decode pipeline register, the instruction decode-execution pipeline register, ex-wb pipeline register and the load store unit.

Vulnerability is normalized to the resource utilization for the chosen benchmark. There is information about the vulnerability for the hardware loop only from the campaign using CoreMark. For Dhrystone this module wasn't used and it didn't generate system failures.



	prefetch unit	if top level registers	hw loop	if-id pipeline register	controller	register file	id-ex pipeline register	control-state registers	ex-wb pipeline register	load store unit	interrupt controller
vulnerability-Dhrystone	6.9	20624081.4	0.0	18.0	1838580.9	87.6	2.5	0.0	14.7	14.4	0.0
vulnerability-CoreMark	9.1	9625.9	1173.2	32.9	1500829.8	62.2	6.2	0.0	26.2	13.8	0.0

Fig. 11.6 RISC-V Riscy sequential resources vulnerability

The limitation of the adopted vulnerability model is due to the occurrence of system failures that are observed also in blocks that are not supposed to be used by the algorithm. This is the case for the multiplier registers that generate system failures during the campaign, whether this block is supposed to be never used. For this reason, the multiplier module vulnerability is not shown in Fig. 11.6. These occurrences, which are treated as exceptions, are caused by the propagation of a fault injected in other blocks. These events depend on how the module is designed and can be avoided with modifications in the system design.

11.7 Conclusion

This paper introduces a detailed analysis of the SEU effects in the RISC-V Riscy core. The results are based on data obtained from the fault injection campaigns based on simulation-based injection technique. The workload is similar both for Dhrystone and Coremark benchmarks as representative of generic applications.

From simulation results, showed in Fig. 11.4 for Dhrystone workload and in Fig. 11.5 for CoreMark workload, the most critical sequential modules are the controller and the register file. Despite the fact that the controller is used with lower frequency than the register file, it causes a large number of failures when it undergoes to fault injection. Exceptions are caused by faults injected in modules inside the instruction fetch stage, in the instruction decoder stage and in the execution-write back pipeline register. CoreMark stimulates the usage of the hardware loop module and we noticed a relevant system failures caused by faults injected inside this module.

The interrupt controller and the control-state registers don't cause any failure when faults are injected.

The most used resources are the prefetch unit, the instruction fetch-instruction decode pipeline register and the instruction decode-execute pipeline register. There are modules that are never used during the benchmark execution like the hardware

loop for Dhrystone and the multiplier registers and the interrupt controller for both benchmarks (in the run simulations).

From the study of vulnerability, showed in Fig. 11.6, the most critical module result to be the controller module.

The Vulnerability can be used to estimate, for each module, the system failure rate when executing other software. This can be done simply making the product between the tabled vulnerability values and the utilization value measured for the given application. These represents an important information for the design of fault-tolerant Risc-V core, since it can be used to evaluate the best redundancy techniques in terms of time usage and hardening impact for each composing block, on the base of its vulnerability.

References

1. Calligaro C, Gatti U (2018) Rad-hard semiconductor memories. Series in Electronic materials and devices
2. Di Mascio S, Menicucci A, Furano G, Monteleone C, Ottavi M (2019) The case for RISC-V in space. In: Saponara S, De Gloria A (eds) Applications in electronics pervading industry, environment and society. Springer International Publishing, Cham, pp 319–325
3. "About the RISC-V Foundation," [Online]. Available: https://riscv.org/risc-v-foundation/. Accessed May 24, 2019
4. Dilillo L, Tsiligiannis G, Gupta V, Bosser A, Saign F, Wrobel F (2016) Soft errors in commercial off-the-shelf static random access memories. J Semicond Sci Technol 32
5. Pulp-Platform (June 2017) "Project info," [Online]. Available: https://pulp-platform.org/projectinfo.html
6. Pulp-Platform (2017 June) "PULPino: Datasheet," PULPino: Datasheet
7. "GNU RISC-V Toolchain," [Online]. Available: https://github.com/riscv/riscv-gnu-toolchain. Accessed May 09, 2019
8. Gupta V, Bosser A, Wrobel F, Saigne F, Dusseau L, Zadeh A, Dilillo L (2016) MTCube project: SEE ground-test results and in-orbit error rate prediction. The 4S symposium, small satellites systems and services symposium, Valletta, Malta
9. Cho H (2018) Impact of microarchitectural differences of RISC-V processor cores on soft error effects. IEEE Access, 6:41302–41313
10. "Mentor Graphics," [Online]. Available: https://www.mentor.com/company/higher_ed/modelsim-student-edition. Accessed May 09, 2019
11. Travessini R, Villa PRC, Vargas FL, Bezerra EA (2018) Processor core profiling for SEU effect analysis. In: Test symposium (LATS)
12. "Roy Longbottom's PC Benchmark collection," [Online]. Available: http://www.roylongbottom.org.uk/dhrystoneresults.htm. Accessed May 09, 2019
13. Price WJ (1989) A benchmark tutorial. IEEE Micro, pp 28–43
14. "Embedded microprocessor benchmark consortium," [Online]. Available: https://www.eembc.org/coremark/. Accessed May 16, 2019

Chapter 12
Analyzing Machine Learning on Mainstream Microcontrollers

Vincenzo Falbo, Tommaso Apicella, Daniele Aurioso, Luisa Danese, Francesco Bellotti, Riccardo Berta and Alessandro De Gloria

Abstract Machine learning in embedded systems has become a reality, with the first tools for neural network firmware development already being made available for ARM microcontroller developers. This paper explores the use of one of such tools, namely the STM X-Cube-AI, on mainstream ARM Cortex-M microcontrollers, analyzing their performance, and comparing support and performance of other two common supervised ML algorithms, namely Support Vector Machines (SVM) and k-Nearest Neighbours (k-NN). Results on three datasets show that X-Cube-AI provides quite constant good performance even with the limitations of the embedded platform. The workflow is well integrated with mainstream desktop tools, such as Tensorflow and Keras.

Keywords Edge computing · Machine learning · Artificial neural networks · Microcontrollers · X-Cube-AI

12.1 Introduction

Internet of Things (IoT) technologies are enabling a variety of new applications directly in the field. The huge quantity of data being generated by IoT sensors is ever more being processed near to the source, on the edge, which typically reduces latencies, bandwidth, overhead of the cloud and of remote units, and should limit privacy issues [1]. This is the edge computing paradigm, which is complementing the well known cloud computing model.

Artificial Intelligence, and particularly Machine Learning (ML), has started to play an important role also in this context. ARM has recently released the Project

V. Falbo · T. Apicella · D. Aurioso · L. Danese · F. Bellotti · R. Berta (✉) · A. De Gloria
DITEN, Università degli Studi di Genova, Via Opera Pia 11/a, 16145 Genoa, Italy
e-mail: Berta@elios.unige.it

F. Bellotti
e-mail: Franz@elios.unige.it

A. De Gloria
e-mail: ADG@elios.unige.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_12

Trillium ML platform, an IP designed for ML and object detection, typically targeting "super smart" phones [2]. But a wide diffusion is likely to take place also on already widespread platforms. On the firmware side, in fact, ARM released in 2018 CMSIS-NN, an open-source library of optimized kernels that maximize Neural Network (NN) performance on Cortex-M processors, which are the most common platforms deployed in the field [3]. Google released TensorFlow Lite for ARM 64 microcontrollers, again with a focus on NNs [4]. Similarly, STM recently released STM X-Cube-AI expansion package, for 32 bit microcontrollers [5].

Still, the number of articles about experiences with ML on the edge is by far not comparable with that concerning desktop/cloud computing. This is exacerbated by the very limited availability of freely available IoT datasets, which does not favor development of research works.

The goal of this paper is to explore the use of one of the above mentioned NN libraries, namely X-Cube-AI, on mainstream ARM Cortex-M microcontrollers, analyzing their performance, and considering also other two common supervised ML algorithms, namely Support Vector Machine (SVM) and k-Nearest Neighbours (k-NN). As a starting point, we have limited our analysis to classification, since the training phase is quite heavier to perform [6] and typically requires human supervision, which is easier to do in the cloud.

12.2 Related Work

A lot of research is ongoing to embed ANNs in autonomous devices, tackling issues of energy efficiency, resource usage and accuracy. Andrade et al. [7] provides a comprehensive analysis of the efforts recently made in this area.

Lai and Suda [8] discusses the challenges of deploying neural networks on microcontrollers with limited memory, computation resources and power budgets. The authors introduce CMSIS-NN, a library of optimized software kernels to enable deployment of NNs on Cortex-M cores. They also present techniques for NN algorithm exploration to develop light-weight models suitable for resource constrained systems, using keyword spotting as an example.

Cerutti et al. [9] presents a new system that merges a low resolution thermal camera with advanced feature extraction techniques such as Convolutional Neural Networks. The paper demonstrates the possibility of adapting the classification execution to a resource-constrained platform without significant loss of performance, by processing data on a 32-bit low power microcontroller. They achieve a 77% accuracy, using 6 kB of RAM.

[10] is a C++ library explicitly declared as dedicated to the embedded world. It implements a NN for the classification, and other algorithms such as Genetic and Reinforcement Learning.

To the best of our knowledge, there is no paper in literature describing the utilization of the STM X-Cube-AI expansion package.

12.3 Machine Learning Implementation

As seen above, NNs have gained momentum also in the embedded system field. Our analysis focuses in particular on one of the above mentioned recently released libraries, namely the STM X-Cube-AI expansion package, which is usable within the STM32CubeMX configuration tool. The package provides automatic conversion of pre-trained Neural Network and integration of generated optimized library into the user's project. The workflow we accustomed to consists in developing a NN on a PC in python, using the Tensorflow library and Keras as wrapper. We normalize the vectors, in order to reduce the convergence time. Once the developer finds a NN configuration providing acceptable accuracy according to tests on the PC, its model is saved in a .HDF5 file, which is imported by CubeMX. The CubeMX "Analyze" function then estimates the memory footprint (Flash and RAM) and suggests a list of possible target microcontrollers, accordingly. Once the target is decided (or the developers has checked suitability of the target at hand), a new project can be started, including the "AI-Application" and "X-CUBE-AI" packs. CubeMX allows then performing a validation both on desktop, which estimates complexity through the Multiply and Accumulate Operation (MACC) figure, and on target. Writing the C program for the target, exploiting the "network" library, can be done in few lines of code that configure the network from the recorded weights, set the input and output tensors and then execute the prediction.

As a term of comparison, we employed also the following two algorithms:

- Support Vector Machine (SVM). We used the sklearn python framework for training the SVM on the PC, with linear kernel and the model obtained through cross-validation. sklearn does not support the gpu acceleration, and the svm method is not able to exploit multi-core architectures. This is a limitation of our approach, as the long training times prevented us from a full exploration of the alternatives (e.g., for more complex kernels). The implementation on the target is as simple as executing the y = w*x + b prediction, where x and y are the inputs and output, w the support vectors and b the bias.
- k-nearest neighbours (k-NN). In k-NN, no model is learned, and all the training set is recorded. We implemented the algorithm in C from scratch, using the Euclidean distance criterion and majority voting.

12.4 Experimental Analysis

We conducted the experimental analysis using two well established ARM microcontrollers produced by STM, namely an F401RE and an F746. The former belongs to the mainstream Cortex-M4 family, the latter to the high performance M7. Results are generally reported in Tables 12.1, 12.2, 12.3 and 12.4 for the F4 case only, while F7 is explicitly considered in Table 12.3. In all cases, we first developed the classifiers

Table 12.1 Results for the Sonar dataset

Classifier	60 features		
	Acc (%)	Time	Flash
K-NN	81	25 ms	46 kB
SVM	85.7	< ms	250 B
NN	90.5	< ms	50 kB

Table 12.2 Results for the heart diseases dataset

Classifier	13 features			6 features		
	Acc (%)	Time	Flash	Acc (%)	Time	Flash
K-NN	63	38 ms	15 kB	93	36 ms	8 kB
SVM	87	< ms	30 B	93	< ms	30 B
NN	93.5	< ms	2.9 kB	87	< ms	2 kB

Table 12.3 Results for the Viruses dataset

Classifier	F401			F746		
	Acc (%)	Time	Flash	Acc (%)	Time	Flash
NN	87	9 ms	65 kB	87	1 ms	524 kB
SVM	71	< ms	60 B	71	< ms	60 B

Table 12.4 Results for a reduced version of the Viruses dataset

Classifier	94% reduced			90% reduced + Feature selection		
	Acc	Time	Flash	Acc	Time	Flash
K-NN	85%	986 ms	250 kB	97%	2.8 s	75 kB

on a PC, and then deployed on the target, with the needed adjustments, especially in terms of performance.

We used three binary classification datasets: Sonar (209 samples $\times$ 60 features) [11], the UCI Heart diseases available on Kaggle (303 $\times$ 13) [12], and Viruses (24,736 $\times$ 13), a data traffic analysis dataset developed by the University of Genova. All the datasets are cast to float32, according to the target execution platform.

For the Sonar dataset (Table 12.1), we report data for a NN with two hidden dense layers (40 and 30 tanh neurons each, after an initial ReLU dense layer with 100 nodes, and an output sigmoid node). With a more complex network (5 wider layers, 300 ReLU input), we get a Flash footprint of 253 kB, and a lower accuracy, of 86%. For k-NN, the best k is 1. For all classifiers, accuracy is the same as on an i7 core PC.

For the Heart disease dataset (Table 12.2), feature selection (implemented through the Orthogonal Matching Pursuit (OMP) algorithm) was necessary to improve the

performance of k-NN (k = 17 for 13 features; k = 5 for 6 features). We used a 3-layer NN, with 30, 10, 1 nodes, tanh nonlinearity for all nodes but the output (sigmoid). For all classifiers, accuracy is the same as on an i7 core PC, apart from the SVM, despite using the same code and dataset.

The Viruses dataset is characterized by a much higher number of samples. On a NN on a PC we could achieve an 88% accuracy, with a 5-layer NN, with 400, 250, 100, 30, 1 nodes, tanh nonlinearity for all nodes but the output (sigmoid). For F401 target, validation on desktop fails, but the NN could anyway be loaded on the microcontroller, achieving an 87% accuracy, with a latency of about 10 ms (Table 12.3). Validation on desktop was successful on the F746, and the accuracy on target was 87%, with a latency of about 1 ms. The flash occupation is of 65 kB on the F401 and much higher on the F746. As an alternative, we achieved 85% accuracy with a smaller 3-layer (40, 30 and 1 nodes) NN, with quite a smaller footprint. The linear SVM could not reach convergency during training, and the average accuracy is 71%.

For the k-NN (Table 12.4), the memory footprint for the Viruses dataset was by far too large, so we exploited a new dataset obtained by randomly decimating the original one (reducing the number of samples by 94%), and we achieved a very high accuracy with a slightly lower sample size reduction, but applying the OMP feature selection. However, execution times grew up to the order of the second.

12.5 Conclusions and Future Work

ML in embedded systems has become a reality, with the first tools for NN firmware development already being available for developers. Analyzing three different algorithms with three different datasets, we saw that the NNs implemented by the STM X-Cube-AI package provides quite constant good performance even with the limitations of the embedded platform. The workflow is well integrated with mainstream desktop tools, such as Tensorflow and Keras. Also SVM performs quite well, with a small footprint. But its development is less well supported by tools compared to NNs. For k-NN, it is known that performance tends to worsen as the training set size increases [13].

Research still lacks publicly available IoT datasets, that would facilitate the experience by scholars and practitioners, in different application domains.

For future work, we are interested in a more detailed analysis (particularly on the space-time tradeoff), with different types of NNs and more relevant datasets. Moreover, it will be interesting to study performance and application of unsupervised learning algorithms, that look even more suited for field deployment, as they do not need human data processing for the training phase. Finally, given the limited facilities of the edge, distributing embedded ML computation is likely to become a major architectural challenge for the upcoming years.

References

1. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) Edge computing: vision and challenges. IEEE Internet Things J 3(5):637–646
2. https://www.arm.com/products/silicon-ip-cpu/machine-learning/project-trillium
3. https://pages.arm.com/machine-learning-on-arm-cortex-m-microcontroller.html
4. https://www.tensorflow.org/lite/guide/build_arm64
5. https://www.st.com/en/embedded-software/x-cube-ai.html
6. Parodi A, Bellotti F, Berta R, De Gloria A (2018) Developing a machine learning library for microcontrollers. In: Saponara S, De Gloria A (eds) Applications in electronics pervading industry, environment and society. ApplePies 2018. Lecture Notes in Electrical Engineering, vol 550. Springer, Berlin
7. Andrade L, Prost-Boucle A, Pétrot F (2018) Overview of the state of the art in embedded machine learning. In: 2018 Design, automation & test in Europe conference & exhibition (DATE), Dresden, pp 1033–1038
8. Lai L, Suda N (2018) Enabling deep learning at the LoT Edge. In: 2018 IEEE/ACM international conference on computer-aided design (ICCAD), San Diego, CA, pp 1–6
9. Cerutti G, Prasad R, Farella E (2019) Convolutional neural network on embedded platform for people presence detection in low resolution thermal images. In: ICASSP 2019—2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), Brighton, United Kingdom, pp 7610–7614
10. http://fidoproject.github.io/
11. http://fizyka.umk.pl/kis-old/projects/datasets.html#Sonar
12. https://www.kaggle.com/ronitf/heart-disease-uci
13. Islam MJ, Wu QMJ, Ahmadi M, Sid-Ahmed MA (2007) Investigating the performance of Naive-Bayes classifiers and K-nearest neighbor classifiers. In: International conference on convergence information technology (ICCIT 2007), Gyeongju, pp 1541–1546

Chapter 13
Quality Aware Selective ECC for Approximate DRAM

Giulia Stazi, Antonio Mastrandrea, Mauro Olivieri and Francesco Menichelli

Abstract Approximate DRAMs are DRAM memories where energy saving techniques have been implemented by trading off bit-cell error rate with power consumption. They are considered part of the building blocks in the larger area of approximate computing. Relaxing refresh rate has been proposed as an interesting solution to achieve better efficiency at the expense of rising error rate. However, some works have demonstrated that much better results are achieved if at word-level some bits are retained without errors (i.e. their cells are refreshed at nominal rate), resulting in architectures using multiple refresh rates. In this paper we present a technique that can be applied to approximate DRAMs under reduced refresh rate. It allows to trim error rate at word-level, while still performing the refresh operation at the same rate for all cells. The number of bits that are protected is configurable and depends on output quality degradation that can be accepted by the application.

Keywords Approximate memory · Transprecision computing · ECC memory

13.1 Introduction and Previous Works

Approximate computing is a design paradigm for low power systems that proposes to expand the degrees of freedom in digital system design by allowing inaccurate or approximate operations in circuits. The idea at the base of approximate comput-

G. Stazi · A. Mastrandrea · M. Olivieri · F. Menichelli (✉)
Department of Information Engineering, Electronics and Telecommunications (DIET) Sapienza, University of Rome, Via Eudossiana, 18, 00184 Roma, Italy
e-mail: francesco.menichelli@uniroma1.it

G. Stazi
e-mail: g.stazi@uniroma1.it

A. Mastrandrea
e-mail: antonio.mastrandrea@uniroma1.it

M. Olivieri
e-mail: mauro.olivieri@uniroma1.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_13

ing is the fact that many real-world applications do not require exact mathematical computations, since their input and output data are inherently affected by noise and errors. Approximate memories are part of the building blocks of this approach and are intended as memory circuits that do not store data exactly and indefinitely, but are affected by errors during read/write operations or tend to spontaneously forget data with the passage of time [1, 2].

Depending on their technology, circuits for approximate memory have been proposed by scaling V_{dd} for SRAMs [3] and by reducing refresh rate under the nominal value for DRAMs [4]. These circuit-level proposals lay the groundwork for practical implementations that can be used in programmable architectures as main approximate memory [5]. Applications that can tolerate a certain amount of errors can then allocate their data structures and buffers in these memories. These application, called ETAs (Error Tolerant Applications), will produce an output with degraded quality as the effect using approximate memories. The final assumption of approximate computing is that the amount of approximation (i.e. errors) can be tailored on the specific problem, trading off energy savings up to the limit of acceptable output quality.

13.2 Approximate Memories in Real Applications

The first approach to approximate memories relies on allowing errors uniformly distributed on the array of bit cells (i.e. all cells are subject to the same voltage scaling or the same refresh rate). The validity of the choice is based mostly on the simplification that it involves at circuit level, since it does not require to modify the array internal circuit, but signals and power supply at the interfaces. However, considering uniform error distribution means to not take into account the exponential relation between different bit weights in a data word, which is instead an important characteristic that should be considered by approximate memory circuit, even at the expense of increasing circuit complexity.

13.2.1 Exact MSBs in an Approximate Data Word

The first and intuitive approach is to design the memory array in order to save MSBs in exact bit-cells. Considering DRAMs, [6] proposes using two different refresh rates, one at nominal rate and one at reduced rate. Cell arrays are rearranged in a way that the nominal refresh rate is applied to bit cells for MSBs (exact MSBs), while the reduced refresh rate is applied to bit cells for LSBs (approximate LSBs). The number of exact MSBs and approximate LSBs depends on applications, for example, for 32 bit words a number from 1 to 8 exact MSBs and, respectively, 31–24 approximate LSBs have been found to be of interest [7]. Requiring exact cells in an approximate data word has direct impact on the following characteristics:

– it raises output quality under the same error rate in LSB cells or, conversely, allows for higher error rate in LSBs while meeting the required output quality;
– it reduces overall energy saving, since a portion of the cells is working at nominal conditions;
– it increases circuit complexity requiring dual refresh rate in DRAMs.

13.2.2 Bit Dropping for LSBs and Bit Reuse

The second approach results from exploration of the relation between output quality and BER on the LSBs [7]. LSBs in a data word can be dropped and set to a constant value (i.e. 0) with a marginal impact on output quality degradation. It is a technique that is proposed since it achieves energy savings with a simple circuit implementation (bit cells are powered off or even omitted).

Previous works have proposed to use selective ECC in SRAM to reduce errors in MSB (1) by enlarging memory words as in classical ECC memory systems (i.e. 32 bit memory word are expanded to 36 bit, introducing 4 bit ECC) [8] (2) by reusing LSB dropped bits [9]. The contribute of our work is (1) to design selective ECC specific for approximate DRAM memory systems (2) to allow tailoring selective ECC to the specific application, by first analyzing its output quality degradation related to bit error rate, looseness level and dropped bits.

13.3 Quality Aware Selective ECC

The idea of quality aware selective ECC consists in a two step process. First, an application is analyzed in order to find the desired tradeoff between output quality and approximate memory parameters (i.e. error rate, level of approximation [1], dropped bit); then an error correcting code is chosen in order to reduce error rate in a specific portion of data bits. In order to avoid increasing memory requirements with additional ECC bit, bit dropping and reuse is always considered for the additional check bits required by ECC.

13.3.1 ECC Codes for Approximate Memories

In order to reduce hardware complexity, (n,k) SEC (single error correcting) Hamming codes were considered. In this notation, k indicates the number of protected bits (data bits), while n is the code length, including additional check bits. We note that SEC codes can provide also error detection (e.g. double error detection typically), but for our scope error detection is not used: in case of detected errors, program execution continues as for undetected errors, in approximate memory. Table 13.1 summarizes

Table 13.1 List of Hamming codes

#Check bits (n-k)	#Total bits (n)	#Data bits (k)	Name	Rate
2	3	1	Hamming(3,1)	1/3
3	7	4	Hamming(7,4)	4/7
4	15	11	Hamming(15,11)	11/15
5	31	26	Hamming(31,26)	26/31
6	63	57	Hamming(63,57)	57/63

the most common Hamming codes. We note that, as a general rule, increasing the number of data bits k produces more efficient codes, since the rate k/n increases. However, larger k are effective at very small error rates (as is common in exact memories). In approximate memories typical error rates are much larger (i.e. from 10^{-4} to 10^{-2} errors/(bit $\times$ s) [1]) and, as consequence, shorter codes are desirable since enlarging n increases the probability of multiple errors within the same word, which cannot be corrected.

13.3.2 Looseness Level

With Looseness Level we intend the concept, introduced in [1], of having a certain number of exact MSBs in an approximate data words. As an example, Table 13.2 reports results obtained on a 32 bit integer FIR filter, showing how Looseness level (i.e. the number of exact MSBs) can impact output SNR.

Instead of using exact DRAM cells for MSBs, the idea is to use a single, and slower, refresh rate for all cells, while using SEC ECC in order to reduce error rate

Table 13.2 FIR, output SNR [dB]

Looseness level	Fault rate [errors/(bit $\times$ s)]			
	10^{-1}	10^{-2}	10^{-3}	10^{-4}
12 MSBs	70.5	83.4	93.5	104.2
8 MSBs	46.5	59.6	69.3	80.3
4 MSBs	22.6	35.3	45.5	56.4
1 MSB	4.6	17.2	27.6	38.2

Table 13.3 # of dropped bits

4 LSBs	8 LSBs	12 LSBs	16 LSBs	
134.7	122.4	106.1	82.2	

in MSBs. In this way, MSBs are still affected by errors, but their error rate is reduced with respect to LSB cells.

13.3.3 Impact of Bit Dropping and Bit Reuse

Table 13.3 reports results obtained on the same 32 bit integer FIR filter, showing how bit dropping (i.e. powering them off and reading them as '0') impacts output SNR. As already confirmed in literature, output SNR is only slightly dependent on LSBs. Instead of powering them off, these LSBs can be effectively reused as checkbits for the MSBs, without requiring additional bits.

13.4 Implementation and Results

Given the list of Hamming codes in Table 13.1, it appears that the most suitable for our application are Hamming (3,1),(7,4) and (15,11). This choice depends on two factors, first we assume to protect single 32 bit words in memory, in order to not impact read/write speed; infact, protecting with a single code larger data size would require to multiple read/write on the entire data. Secondly, given the relatively high bit error rate of approximate memories, longer SEC codes tend to fail due to the rising probability of multiple errors.

Figure 13.1 shows the formats considered for 32 bit data, where k MSBs are protected by SEC ECC, $32 - n$ bits are left unprotected and $n - k$ dropped and reused as checkbits. Assuming a uniform error probability p_e for each bit, expressed as errors/(bit $\times$ s), the probability of having i errors in a set of n bits is:

$$P_e(n, i) = \binom{n}{i} p_e^i (1 - p_e)^{n-i};$$

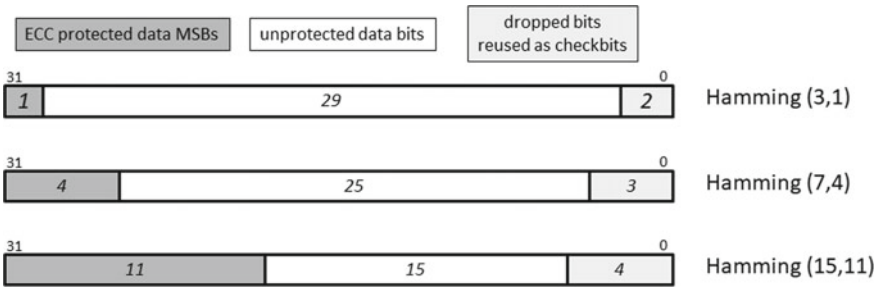


Fig. 13.1 32 bit ECC data format in approximate memory

Considering the SEC ECC code, protected bits will contain errors for $i \geq 2$; hence:

$$Pecc_e(n) = \sum_{i=2}^{n} P_e(n,i) = \sum_{i=2}^{n} \binom{n}{i} p_e^i (1-p_e)^{n-i};$$

In order to get a measure of the improvement, we can find the equivalent error rate peq_e, considered as the error rate that n bits (without ECC) should have to produce the same $Pecc_e(n)$.

$$Peq_e(n) = \sum_{i=1}^{n} P_e(n,i) = 1 - \sum_{i=0}^{0} P_e(n,i) = 1 - (1-peq_e)^n;$$

Equivalent bit error rate peq_e for ECC protected bits can be obtained with $Peq_e(n) = Pecc_e(n)$:

$$peq_e = 1 - \sqrt[n]{1 - Pecc_e(n)};$$

Table 13.4 BER for 32 bit data in approximate memory

Hamming (3,1) word		
ECC prot. 1 bit	Unprot. 29 bit	Drop 2 bit
9.42E−03	1.00E−01	–
9.93E−05	1.00E−02	–
9.99E−07	1.00E−03	–
1.00E−08	1.00E−04	–
1.00E−10	1.00E−05	–
Hamming (7,4) word		
ECC prot. 4 bit	Unprot. 25 bit	Drop 3 bit
2.29E−02	1.00E−01	–
2.90E−04	1.00E−02	–
2.99E−06	1.00E−03	–
3.00E−08	1.00E−04	–
3.00E−10	1.00E−05	–
Hamming (15,11) word		
ECC prot. 11 bit	Unprot. 15 bit	Drop 4 bit
3.92E−02	1.00E−01	–
6.45E−04	1.00E−02	–
6.94E−06	1.00E−03	–
6.99E−08	1.00E−04	–
7.00E−10	1.00E−05	–

Assuming 32 bit data stored in approximate memory, Fig. 13.1 resumes how selective ECC could be applied using Hamming (3,1), (7,4) and (15,11) codes. The most appropriate choice depends on the application; for example, according to Tables 13.2 and 13.3, a range from 8 to 12 protected MSBs results in an output SNR between 60 and 93 dB, while dropping 4 LSBs does not significantly impact SNR. In this case Hamming (15,11) seems the most suitable choice.

Table 13.4 reports the results that can be obtained on typical target application considering the previous Hamming codes. It shows that MSBs protected by SEC codes expose and equivalent BER significantly lower than unprotected bits. Considering the previous example, Hamming (15,11) and a BER of 10^{-3} on cells produces an equivalent BER of 6.94×10^{-6} on MSBs.

13.5 Conclusion

In this paper we proposed the use of selective ECC in approximate DRAM memory tailored to quality requirements of applications. We started from the consideration that in many works and use cases it has been demonstrated the effectiveness of limiting approximate cells to LSBs while leaving a portion of MSBs exact. However, this approach requires higher complexity in memory circuits and circuits surrounding the cell array. For DRAMs, it requires to produce and distribute multiple refresh rates in the array.

Due to the relatively high error rates in approximate memories, SEC codes reduce but do not eliminate errors. This is completely acceptable and we demonstrated that for typical error rates in the order of 10^{-3} to 10^{-4}, Hamming codes (7,4) and (15,11) can reduce error rate on MSBs of factor between $1/100$ and $1/1000$. Future works will implement the technique in simulation models and apply it to error tolerant applications, allowing the characterization and the comparison with respect to previous techniques.

References

1. Stazi G, Mastrandrea A, Olivieri M, Menichelli F (2019) Full system emulation of approximate memory platforms with appropinquo. J Low Power Electron 15(1):30–39
2. Menichelli F, Stazi G, Mastrandrea A, Olivieri M (2016) An emulator for approximate memory platforms based on QEmu. In: International conference on applications in electronics pervading industry, environment and society. Springer, Berlin, pp 153–159
3. Frustaci F, Blaauw D, Sylvester D, Alioto M (2016) Approximate SRAMs with dynamic energy-quality management. IEEE Trans Very Large Scale Integr (VLSI) Syst 24(6):2128–2141
4. Raha A, Sutar S, Jayakumar H, Raghunathan V (2017) Quality configurable approximate dram. IEEE Trans Comput 66(7):1172–1187

5. Stazi G, Menichelli F, Mastrandrea A, Olivieri M (2017) Introducing approximate memory support in linux kernel. In: 2017 13th conference on Ph.D. research in microelectronics and electronics (PRIME). IEEE, New York, pp 97–100
6. Lucas J, Alvarez-Mesa M, Andersch M, Juurlink B (2014) Sparkk: quality-scalable approximate storage in dram. In: The memory forum, pp 1–9
7. Stazi G, Adani L, Mastrandrea A, Olivieri M, Menichelli F (2018) Impact of approximate memory data allocation on a h.264 software video encoder. In: International workshop on Approximate and Transprecision Computing on Emerging technologies (ATCET). Springer, Berlin
8. Lee I, Kwon J, Park J, Park J (2013) Priority based error correction code (ecc) for the embedded sram memories in h. 264 system. J Signal Process Syst 73(2):123–136
9. Frustaci F, Blaauw D, Sylvester D, Alioto M (2015) Better-than-voltage scaling energy reduction in approximate SRAMs via bit dropping and bit reuse. In: (2015) 25th international workshop on power and timing modeling, optimization and simulation (PATMOS). IEEE, 132–139

Chapter 14
Digital Random Number Generator Hardware Accelerator IP-Core for Security Applications

Luca Baldanzi, Luca Crocetti, Francesco Falaschi, Jacopo Belli, Luca Fanucci and Sergio Saponara

Abstract Random numbers are widely employed in cryptography and security applications, and they represent one of the main aspects to take care of along a security chain. They are employed for creation of encryption keys, and if generation process is weak, the whole chain can be compromised: weaknesses could be exploited to retrieve the key, thus breaking even the strongest cipher. This paper presents the architecture of a digital Random Number Generator (RNG) IP-core to be employed as hardware accelerator for cryptographically secure applications. Such design has been developed starting from specifications based on literature and standards, and in order to assess the randomness degree of generated output, it has been successfully validated through the official NIST Statistical Test Suite. Finally the RNG IP-core has been characterized on Field Programmable Gate Array (FPGA) and ASIC standard-cell technologies: on Intel Stratix IV FPGA it offers a throughput of 720 Mbps requiring up to 6000 Adaptive Logic Modules, while on 45 nm it reaches a throughput of 4 Gbps with a complexity of 119 kGE.

L. Baldanzi · L. Crocetti · F. Falaschi (✉) · J. Belli · L. Fanucci · S. Saponara
Department of Information Engineering, University of Pisa, Via G. Caruso, 16, 56122 Pisa, Italy
e-mail: francesco.falaschi@phd.unipi.it

L. Baldanzi
e-mail: luca.baldanzi@ing.unipi.it

L. Crocetti
e-mail: luca.crocetti@phd.unipi.it

J. Belli
e-mail: jacopo.belli23@gmail.com

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

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

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_14

14.1 Introduction

In modern cryptography one of the fundamental primitives to be employed is the Random Number Generator (RNG), the component in charge of generation of arbitrary length random bit sequences. It represents the core part for several security applications which are required to ensure authentication, confidentiality and message integrity for a broad range of activities, such as payments, on-line authentication, instant messaging and operating systems updates [4]. The creation of cryptographic keys requires a high degree of randomness so that an attacker is unable to derive the secret key of a cipher thus compromising the whole chain, authentication protocols nonces represent a valid countermeasure against replay attacks, in digital signature random numbers prevent attackers to derive private keys [3].

During the last decades, several circuits have been proposed to cope with generation of RNG sequence, in particular the True Random Number (or Bit) Generators (TRNGs) which are based on analog noise as physical source to generate random bits [1–7]. Such devices have a high-quality output, but they are affected by significant drawbacks, because they typically offer low throughput or require high power consumption. Moreover, they can be unreliable for long term use due to unexpected behaviors caused by changes in the device operating conditions. These are strong limitations especially considering the target to be employed in high performances and high complexity digital integrated systems such as hardware accelerators.

The limitations of TRNG devices can be worked around by implementing RNGs as Deterministic Random Bit Generators (DRBGs): in this case the output sequences are generated by means of deterministic algorithms instead of random processes, therefore in order to guarantee the expected level of randomness it is required to periodically give a new seed to such DRBG mechanisms (i.e., *reseed* operation, high entropy content is given to the deterministic algorithm to restart the sequence generation). This allow to pursue the requirement of indistinguishability between the output bit sequence and truly random sequence.

The reminder of this paper is organized as it follows: Sect. 14.2 presents the trade-off analysis among the different algorithms suitable for DRBG module, Sect. 14.3 describes the DRBG design architecture, Sect. 14.4 collects the characterization results, and Sect. 14.5 discusses about conclusions of this work.

14.2 DRBG Algorithms Trade-Off Analysis

As already mentioned, NIST has approved a certain number of DRBG mechanisms [2]: those mechanisms are based on Hash functions (SHA, Secure Hash Algorithm), keyed-Hash Message Authentication Code (HMAC), and Counter (CTR) mode of Advanced Encryption Standard (AES) and Triple Data Encryption Standard (TDES), and they are briefly presented, focusing on performance evaluation in terms of security strength and hardware implementation.

Hash DRBG family is based on SHA1 and SHA2 functions, but only SHA2 cryptographic primitives are taken into exam since SHA1 offers low security strength and it is considered outdated. The parameters related to a DRBG mechanism based on SHA2 Hash function are reported in Table 14.1.

CTR[1] DRBG mechanism is based onto a block cipher core used in *counter mode*. The parameters of this mechanism are listed in Table 14.2.

Concerning Hash DRBG, the characteristics of available SHA2 IP core are listed in Table 14.3. SHA-224 and SHA-384 are discarded from the options, since they offer a shorter output block keeping area and latency equal to respectively SHA-256 and SHA-512. The two remaining functions show some differences:

- SHA-256 has lower latency per block than SHA-512 but the latter offers a higher throughput since it provides 512 bit every 80 clock cycles;

Table 14.1 Hash DRBG mechanisms parameters (SHA2 only) [2]

	SHA algorithm			
	SHA-224	SHA-256	SHA-384	SHA-512
Highest security strength	192	256	256	256
Output block length (*outlen*) (bits)	224	256	384	512
Min. entropy for *Instance* and *Reseed* (bits)	192	256	256	256
Seed length (*seedlen*) bits	440	440	888	888
Max. num. of bit per request	2^{19}	2^{19}	2^{19}	2^{19}
Max. num. of requests between *Reseeds*	2^{48}	2^{48}	2^{48}	2^{48}

Table 14.2 CTR DRBG mechanisms parameters

	AES Algorithm			
	3Key TDEA	AES-128	AES-192	AES-256
Highest security strength	112	128	192	256
Input/output block length (*blocklen*) (bits)	64	128	128	128
Key length (*keylen*)	168	128	192	256
Counter field length (*ctr_len*)	$4 \leq ctr\_len \leq blocklen$			
Min. entropy for *Instance* and *Reseed* (bits)	112	128	192	256
Seed length (*seedlen*) (bits)	232	256	320	384
Max. num. of bit per request	min(B, 2^{13})	min(B, 2^{19})	min(B, 2^{19})	min(B, 2^{19})
Max. num. of requests between *Reseeds*	2^{48}	2^{48}	2^{48}	2^{48}

$B = (2^{ctrl\_len} - 4)\ blocklen$ [2]

[1]CTR is an abbreviation for *Counter*.

Table 14.3 SHA2 IP core specifications

SHA2 algorithm	Area (kGE)	Latency per block (clock cycles)	Output block size (bits)
SHA-224	15	64	224
SHA-256	15	64	256
SHA-384	30	80	384
SHA-512	30	80	512

Table 14.4 AES IP core specifications

AES algorithm	Area (kGE)	Latency per block (clock cycles)	Output block size (bits)
AES-128	11	11	128
AES-256	12.5	15	128

- comparing the areas, SHA-256 results to be more compact and this reflects also on internal state registers area footprint: as it can be seen in Table 14.1, the variable *seedlen* is 440 for SHA-256 and 888 for SHA-512; this implies that the internal state requires around 900 registers for the former and 1800 for the latter.

Now, the expected throughput of these two hash functions during generation phase in a Hash DRBG implementation can be calculated:

$$T_{SHA-256} = 256/64 \cdot f_{clk} \cdot n_{parallel\_core} = 4 \cdot f_{clk} \cdot n_{parallel\_core}\,\text{bit/s} \quad (1)$$

$$T_{SHA-512} = 512/80 \cdot f_{clk} \cdot n_{parallel\_core} = 6.4 \cdot f_{clk} \cdot n_{parallel\_core}\,\text{bit/s} \quad (2)$$

CTR DRBG proved to be best in class for both area and throughput. The characteristics of available AES IP core are presented in Table 14.4 for AES-128 and AES-256.

Since our focus is on highest level security strength implementations, only AES-256 is to be considered for the trade-off. As shown in the table, area is lower than SHA-256 and throughput is higher than SHA-512:

$$T_{AES-256} = 128/15 \cdot f_{clk} \cdot n_{parallel\_core} = 8.53 \cdot f_{clk} \cdot n_{parallel\_core}\,\text{bit/s} \quad (3)$$

Despite all these considerations, CTR DRBG has not been chosen to be implemented. The reason lays in the doubts about the effective capability of this mechanism to reach maximum security strength. In [8], the author claims that, while Hash-based DBRGs satisfy security requirements, block cipher-based ones should be avoided since the pseudo-random permutation inside each AES round coupled with the counter mode outputs a sequence which is indeed distinguishable from a random source. The choice ultimately fell on Hash DRBG, implemented with SHA-256

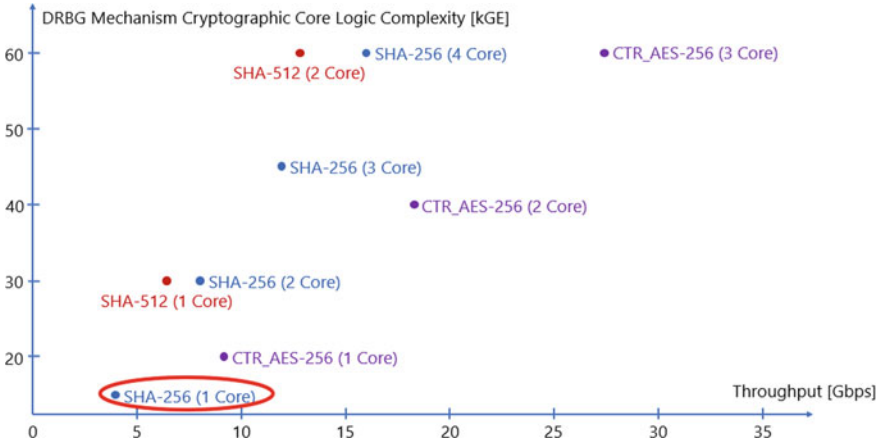


Fig. 14.1 Comparison between NIST approved DRBG mechanisms based on logic complexity in kGE and throughput

core. This ensures a compact implementation for the mechanism and the possibility to extend the design for supporting multiple cores to increase the throughput. Figure 14.1 reports the characteristics in terms of logic complexity and throughput of several DRBG implementations, relying on the available IP cores (SHA and AES) as primitives, their features when synthesizing on 45 nm standard-cell technology [9] and methods to construct DRBG using such primitives [2].

14.3 Hash DRBG Design Architecture

The design architecture of Hash DRBG with SHA-256 core is shown in Fig. 14.2, and it makes use of the following blocks:

- state registers for *V*, *C* and *Reseed counter*, with length respectively of 440, 440 and 20 bits, a 128-bit register to store an optional personalization string, for internal state randomization, and a 512-bit entropy register to store the input entropy content;
- a SHA-256 core with 512-bit input and 256-bit output, with a latency of 64 clock cycles;
- a serial adder with 440-bit inputs and modulo 440-bit output, which works in parallel with the SHA-256 core and stores the result of the addition into one of the its input registers, as shown in Fig. 14.2, in order to minimize area occupation;
- multiplexer network to address all data in internal state and from the previous operation to the inputs of the SHA-256 core and adder;
- a Finite State Machine (FSM), which controls the flow of operations, i.e., *instance*, *reseed* and *generate*;

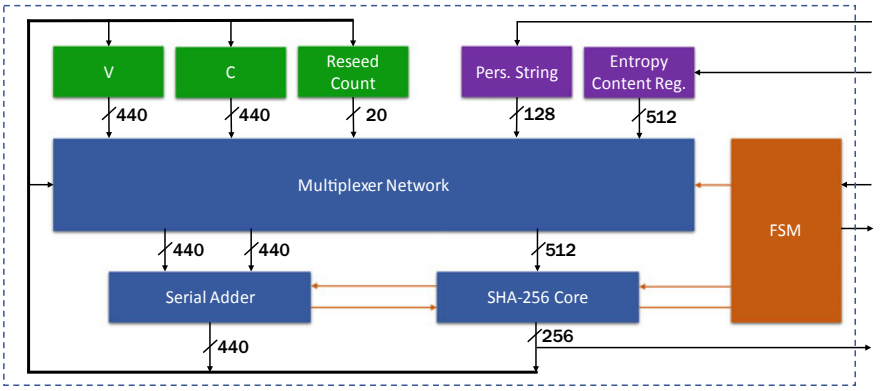


Fig. 14.2 Hash DRBG design architecture developed

- a DRBG self-test module (not present in Fig. 14.2), in order to diagnose possible failures inside the circuitry.

14.4 Results

For the Hash DRBG IP-core characterization, two different technologies have been identified as representative of potential targets for implementations of such hardware accelerator for security applications: Intel Stratix IV FPGA and Silvaco PDK 45 nm Open Cell Library [7] (i.e., ASIC standard-cell technology). In both cases different implementation effort corners were tested, in order to evaluate the trade-off between throughput and area. Concerning the Intel Stratix IV FPGA technology, the synthesis and layout flow performed with high performance constraints gives a maximum operative frequency of 180 MHz, meaning a throughput of 720 Mbps considering the single core instance, for an overall occupation of 5949 ALMs (Adaptive Logic Modules). The implementation on Silvaco ASIC standard-cell is able to reach a throughput even up to 4 Gbps, since the maximum frequency is equal to 1 GHz still for single core version of the IP-core, for a logic complexity of 118.98 kGE corresponding to an area of approximately 0.094 mm^2.

14.5 Conclusions

This paper presented the IP-core design related to a digital Random Number Generator (RNG), one of the most significant part required to implement algorithms for authentication, confidentiality, message integrity and security applications in general.

The proposed architecture is based on one of the Deterministic Random Bit Generators (DRBGs) approved by NIST according to trade-off analysis between throughput, area and security strength. Hash DRBG with SHA-256 as cryptographic core proved to be the most efficient solution in terms of throughput per logic complexity, among the solutions offering maximum security strength (i.e., 256 bits).

The RNG IP-core obtained has been tested by means of NIST Statistical Test Suite, thus stating that the sequences of bits generated cannot be distinguished from a true random sequence of numbers, and therefore validating its use for cryptographic applications. It has been also implemented on FPGA and ASIC standard-cell technologies for characterization. The implementation on Intel Stratix IV FPGA reported a throughput of 720 Mbps at 180 MHz with a maximum occupation of about 6000 ALMs, while the synthesis on Silvaco 45 nm ASIC standard-cell [7] reported a throughput of 4 Gbps at 1 GHz with a maximum logic complexity of about 119 kGE.

References

1. Barker E, Kelsey J (2016) Recommendation for Random Bit Generator (RBG) constructions. Special Publication 800-90C, NIST
2. Barker E, Kelsey J (2015) Recommendation for random number generation using deterministic random bit generators. Special Publication 800-90A, NIST
3. Lo Bello L, Mariani R, Mubeen S, Saponara S (2019) Recent advances and trends in on-board embedded and networked automotive systems. IEEE Trans Ind Inf 15:1038–1051
4. Pelzl J, Paar C (2011) Understanding cryptography. Springer, Berlin
5. Dang QH (2015) Secure hash standard. Technical report, NIST
6. Dichtl M, Golić JD (2007) High speed true random number generation with logic gates only. In: Cryptographic hardware and embedded systems—CHES 2007. Lecture Notes in Computer Science, vol 4727. Springer, Berlin, 45–62
7. Vasyltsov I, Hambardzumyan E, KimBohdan Y-S, Karpinskyy B (2008) Fast digital TRNG based on metastable ring oscillator. In: Cryptographic hardware and embedded systems—CHES 2008. Lecture Notes in Computer Science, vol 5154. Springer, Berlin, 164–180
8. Schmid M (2015) ECDSA—Application and implementation failures
9. Silvaco PDK 45nm Open Cell Library. https://www.silvaco.com/products/nangate/FreePDK45_Open_Cell_Library/index.html

Chapter 15
An Energy Optimized JPEG Encoder for Parallel Ultra-Low-Power Processing-Platforms

Tommaso Polonelli, Daniele Battistini, Manuele Rusci, Davide Brunelli and Luca Benini

Abstract The energy autonomy and the lifetime of battery-operated sensors are primary concerns in industrial, healthcare and IoT applications, in particular when a high amount of data needs to be sent wirelessly such as in Wireless Camera Sensors (WCS). Onboard real-time image compression is the appropriate solution to decrease the system's energy. This paper proposes an optimized algorithm implementation tailored for PULP (Parallel Ultra Low Power) processors, that permits to shrink the image size and the data to transmit. Our optimized JPEG encoder based on a Fast-Discrete Cosine Transform (DCT) function is designed to achieve the best trade-off between energy consumption and image distortion. The parallel software implementation requires only 0.495 mJ per frame and can support up to 80 fps satisfying the most stringent requirements in WCSs applications without requiring a dedicated hardware accelerator.

T. Polonelli (✉) · D. Battistini · M. Rusci · L. Benini
University of Bologna, Bologna, Italy
e-mail: tommaso.polonelli2@unibo.it

D. Battistini
e-mail: daniele.battistini3@studio.unibo.it

M. Rusci
e-mail: manuele.rusci@unibo.it

L. Benini
e-mail: luca.benini@unibo.it; luca.benini@iis.ee.ethz.ch

D. Brunelli
University of Trento, Trento, Italy
e-mail: davide.brunelli@unitn.it

L. Benini
ETH Zurich, Zürich, Switzerland

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_15

15.1 Introduction

The energy autonomy and the lifetime of battery-operated sensors are primary concerns in industrial, healthcare, and IoT applications, in particular when a high amount of data needs to be sent wirelessly. In this scenario, Wireless Camera Sensors are usually left in the environment to acquire and transmit visual data [1, 2]. From a system-level viewpoint, the energy consumption is dominated by the radio subsystem and is proportional to the number of bytes to transfer [3–5]. Concerning WCSs, on-board real-time image compression is the appropriate solution to decrease the system's energy [6, 7]. In fact, bringing the intelligence close to the sensor enables the reduction of transmission costs thanks to the compression of the data dimensionality [8].

Executing computationally heavy tasks, such as an image compression pipeline, without assuming a dedicated hardware acceleration engine (which may not be available or affordable for cost reasons) typically requires adequate computing capabilities and a large memory footprint. However, because of the available energy supply resources (i.e., small batteries or inefficient energy harvesters), [9] WCSs usually includes low-power MCUs (e.g., ARM Cortex-M or RISC-V PULP), which presents limited resources that can prevent executing data filtering tasks under real-time constraints [10]. To address this challenge, we propose an optimized image compression algorithm implementation tailored for a RISC-V multi-core processor, that permits to shrink the image size and the data to transmit. We developed an optimized JPEG (Joint Photographic Experts Group) encoder based on Fast-DCT (FDCT) image compression algorithm, with an adaptive trade-off between energy consumption and image distortion. Our software solution is tailored for a parallel fixed-point computing hardware and exploits the DSP-oriented instructions included into the RISC-V extended ISA (Instruction Set Architecture) of PULP. When compared with a JPEG implementation on ARM Cortex-M4, our solution achieves a frame rate of 22 fps and is eight times more energy-efficient, if running on the GAP-8 processor, an eight cores embodiment of the PULP architecture.

15.1.1 Related Works

Several hardware accelerators are available as standalone chips or add-on-IP blocks for system-on-chip integration [11]. However, the extra cost (in silicon area and/or bill of materials) for a hardware JPEG encoder may not be affordable in many application scenarios that require software JPEG compression. Since the '90s, FDCT algorithms for image compression have been intensively studied in the literature [12] to reduce the number of CPU instruction needed to operate on a standard block, an 8×8 matrix of pixels. Indeed, image compression function based on the 2-D 8-point DCT is prevalent, which is typically the most computationally intensive. Among the various fast DCT algorithm proposed [13], the following four are the most common.

Table 15.1 Number of cycles required to execute the JPEG algorithm on different implementations

Functions	Cycles 1	Cycles 2	Cycles 3	Parallel
DCT + Zig-Zag (8 × 8 block)	107,500	1947	1873	1611
Quantization (8 × 8 block)	2539	2539	2220	2368
Huffman (8 × 8 block)	984	984	984	802
Total (all image)	130,147,237	13,072,581	6,092,954	2,307,672
MSE	53	100	100	100
PSNR (dB)	31	29	29	29
Speedup	–	10	21	56

The first fast DCT was proposed by Chen [14], which has an excellent regular structure, but it requires as many as 16 multiplications for each 8-point block. Hou [15] proposed a recursive algorithm, with 12 multiplications and 29 additions. Although the number of operations is the same as other fast algorithms, it has the advantage of the smaller number of variables necessary for the execution. The function proposed by Loffler [16] involves 11 multiplications and 29 additions. Additionally, the authors proposed a parallel solution that simultaneously executes three multiplications. Finally, the algorithm proposed by Arai [17] features a simplification of the DCT processing. It requires only 5 multiplications and 29 additions. Moreover, it can be easily implemented with fixed-point operations, speeding up the code execution in the absence of a Floating Point Unit (FPU). The aforementioned works make clear that using an optimized DCT algorithm heavily decreases the number of operations required by the JPEG encoder and that a parallelized execution can be applied.

In this work, we based our development on Noritsuna, a JPEG encoder optimized for Cortex-M4 [18]. This implementation supports floating-point operation at low memory impact, but it is not tailored for real-time compression since it is based on a non-fast DCT algorithm (Table 15.1—Cycles 1). To overcome this issue, we replace the DCT algorithm with the Arai [17] FDCT implementation. However, the Noritsuna's algorithm implementation applies to individual 8 × 8 image blocks, hence demanding low L1 memory footprint and favoring a block-wise parallelization scheme for multi-core implementation. After an in-depth study, we selected the application described in [19] as a comparison for this paper; indeed, it needs only 10 Mcycles (220 ms @ 48 MHz) to compress a QVGA grayscale frame, about 8 Kcycles/block, one of the best performance with a low-power ARM Cortex-M4. Similarly to our solution, this implementation exploits fast DCT, but it is optimized on Cortex-M4 architecture featuring an L1 scratchpad memory of 80kB (with QVGA resolution), greater than the GAP-8 cluster memory. Among other solutions, the authors in [12] describe an optimized firmware that needs 22–26 Mcycles to compress a 752 × 480 pixel in RGB format (~9 Kcycles/block), whereas the paper in [6] requires 300 Kcycles to process a single 8 × 8 block, with an average execution time of 9207 ms on a Texas Instruments MSP430. The deployment in [6] uses up to 29 mJ to encode a single 128 × 128 picture. These latter implementations feature higher energy consumption than our solution.

15.2 GAP-8

In this work, we use GAP-8 SoC; a RISC-V ISA multi-core processor based on the PULP open-source computing platform [20]. It integrates a state-of-the-art RISC-V microcontroller core with a rich set of peripherals, and a powerful programmable parallel processing engine for flexible multi-sensor (image, audio, inertial) data analysis. These two subsystems are shown in Fig. 15.2c and are respectively called Fabric Controller (FC) and Cluster. The FC is an advanced MCU based on a RISC-V single-core. It features an extended ISA for energy-efficient digital signal processing, and it is equipped with a fast access-time data memory (L1). The 512 KB L2 memory is used for storing the code and most of the volatile variables. The cluster, residing on a dedicated frequency and voltage domain, is turned on when applications need computation-intensive functions. It contains 8 RISC-V cores identical to the FC, allowing the SoC to execute the same code on either the fabric controller or the cluster. This 8-core cluster is served by a shared L1 data memory (64 kB). The shared L1 can serve all memory requests from the cores in the cluster with single-cycle access latency and low average contention rate (<10% on data-intensive kernels).

Maximizing the power efficiency is an essential factor in low power devices; hence, GAP-8 contains an internal DC/DC directly connected to an external battery or energy harvester sources. It provides voltage in 1.0–1.2 V range when the circuit is active.

15.3 JPEG Algorithm: Implementation and Optimization

The original version of the firmware [18] is composed of the following steps: (i) generation of the header file; (ii) image decomposition into 8×8 pixel blocks, and if the overall dimensions are not multiple integers of 8, the missing blocks are padded with values calculated from the average value on the edges, then the level shifting is executed; (iii) application of the DCT to every block, followed by the quantization, and zigzag operations; (iv) Huffman; (v) writing back the compressed data into the L2 memory.

Since the GAP-8 architecture is not equipped with an FPU, all the operations are implemented with a fixed-point representation. For this data type, we must select in advance the number of bits dedicated to the integer and the fractional parts and, depending on this choice, the JPEG encoder can achieve higher precision (increasing the number of fractional bits) or a broader dynamic range. We individuate the best trade-off by selecting 15 bits for the fractional part and 16 bits for the integer part (16Q15). To quantitatively evaluate the differences between both representations, we adopt as mean metrics the Peak Signal to Noise Ratio (PSNR) and the Mean Squared Error (MSE) since they are widely used in the scientific community as evaluation indexes in the field of image processing [21]. The 16Q15 representation covers the

dynamical range required by the algorithm and increases the PSNR of 0.3%, and the MSE is practically unchanged concerning the floating-point original code.

Table 15.1—Cycles1 reports the GAP-8 performance metric to run the JPEG implementation on a QVGA image (324 × 240). This initial version requires more than 130 M cycles, at 50 MHz the frame conversion time is approximately 2.5 s. The latency breakdown individuates the DCT routine as the most onerous part from a computational point of view, as we expected from the description of the firmware in [18]. The two-dimensional DCT on an area of 8 × 8 pixels has been described previously and, following the formula given in [18], we need 3136 additions and 8192 multiplications, meaning a considerable load on processors, especially RISC, where multiplications require greater use of resources. The optimization usually focuses on reducing the number of arithmetic operations to be performed during the DCT. Like most of the fast algorithms, also the one proposed by Arai, Agui, and Nakajima [17] exploits the separability of the two-dimensional DCT and reduces it to the calculation of a one-dimensional DCT on eight elements for all the rows and subsequently for the columns. This algorithm is considered the fastest: it requires 29 additions and 5 multiplications for the DCT 1D and 464 additions and 144 multiplications for the 2D DCT on the 8 × 8 block. The JPEG encoder performance with AAN (Arai Arui Nakajama) DCT is presented in Table 15.1—Cycles 2. With this change, the major improvement in performance was achieved, dropping the total number of cycles by 89%, mainly due to the relative reduction by 98% in the execution of the DCT. On the other hand, since the AAN algorithm is an approximation of a standard DCT, it has an impact on the quality of the output image, increasing the MSE of about 86%. However, as shown in Fig. 15.1, the image quality difference perceived from a human eye is negligible despite the MSE and PSNR indexes drop; hence the AAN DCT can be considered a suitable replacement in our JPEG encoder.

The first (Cycles 1) implementation (Noritsuna [18]), written following the theoretical definition of the 2D DCT, presents a complexity $O(n^4)$; the AAN instead reduced the complexity to O(n log2 n) motivating the notable latency reduction.

Fig. 15.1 Image quality comparison between both FAST-DCT (AAN) and DCT algorithms

A third optimization step for sequential execution is performed using the hardware features available on GAP-8 SoC, such as the DSP-oriented extended-ISA instructions (built-in) and the single cycles access memory (L1). The built-in functions are extensions of the RISC-V instruction set, developed to speed up some computationally heavy operations. Among the most commonly used, we exploit the Multiply Accumulate (MAC) instructions, which multiply two variables and accumulate the partial sums, and the FIXED_MUL, which multiplies two fixed-point variables in one single cycle. The final number of cycles required for an in-line execution is presented in Table 15.1—Cycles 3.

To run the JPEG encoder on the GAP-8 cluster, the algorithm steps are executed by making use of the available 8 RISC-V cores. The initial section of the JPEG file header can be performed only once at the beginning of the program since it is fixed (Fig. 15.2a—Header Writing). The rest of the JPEG algorithm workload is distributed among the cluster by letting any core operates on different image 8 × 8 block (Fig. 15.2a—Multi-core functions). Indeed, during the compression of the pictures, it is sufficient writing to the output file (L2) the bytes containing only the information concerning the actual image starting from the byte following the last of the header (Fig. 15.2a—Footer writing). The image blocks reading function can be easily performed in parallel, similarly to level shifting, discrete transform of cosines, zigzag reordering, and quantization tasks. Instead, the Huffman task operates on data produced by previous steps. Hence it is executed as a sequential task on a single core. In addition to this, the Huffman encoding does not have a predefined number of bits needed to encode a symbol, but the output is of variable length. For this reason, it was considered necessary to separate this last step from parallel execution

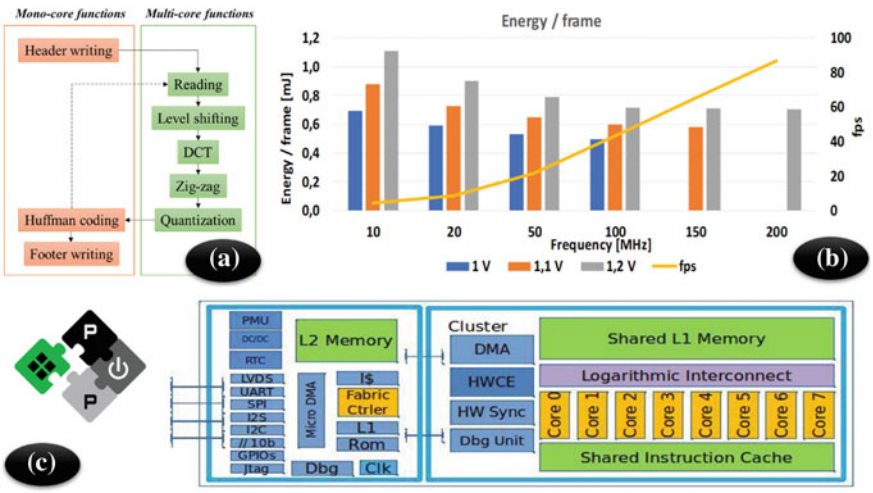


Fig. 15.2 **a** JPEG sub-functions, the multi-core algorithms can be parallelized. Instead, the mono-core function must be executed sequentially; **b** Energy per frame and maximum fps compared to the cluster frequency and voltage; **c** GAP-8 overview

by executing it sequentially from a single cluster core. With the parallel execution of the firmware, we reach 2,307,672 cycles (Table 15.1—Parallel) and conversion time of about 45 ms @ 50 MHz, which corresponds to 22 frames per second. The speedup reached with a parallel execution is 2.64 with eight cores because the Huffman is executed sequentially.

15.4 Estimation of the System Energy

In the case of multi-core execution, the highest energy efficiency is achieved at 1 V at the maximum frequency of 100 MHz (Fig. 15.2b). At this operative point, the GAP-8 compresses a frame with 0.495 mJ, while at 50 MHz the energy consumption results to be 0.532 mJ. With a supply voltage of 1.2 V, we can reach 200 MHz, compressing around 86 images per second, but the energy required for compression reaches 0.7 mJ per frame.

We analyzed the obtained performance metrics with respect to a low-power MCU device, such as STM32L476G from STMicroelectronics, running the JPEG implementation of [18]. The reference MCU is an ultra-low-power platform based on a 32-bit ARM Cortex-M4 core capable of operating at a frequency up to 80 MHz. The STM32L476G, in RUN mode @ 48 MHz, consumes 18.29 mW. The obtained number of cycles is equal to 10,528,330 with a single QVGA image conversion. With this information, we computed the compression latency and the energy consumption as 0.22 s and 4.011 mJ per frame. In the same scenario, our JPEG implementation, in conjunction with GAP-8, reaches an execution time ~5× faster than an STM32L476, with an average energy consumption 8 times lower.

One of the most power-consuming tasks in WCS applications is to transfer the images acquired by the camera either to cloud servers or to personal gateways (e.g., a mobile phone) for low-latency feedback [6]. Consequently, wireless communication is an essential feature, although it is often the bottleneck both for the throughput and for the power budget of the entire system, considering that applications might need to stream images and videos continuously. In our previous papers [22, 23], we studied the joint challenge of communication energy minimization and maximization of the communication flexibility under several different connectivity scenarios. The article [22] shows that to stream raw images, the Wi-Fi requires an average of 30 nJ/bit. Our QVGA sensor generates an 80 kB/frame that a 20 fps produces up to 12.8 Mbps data, which needs 384 mJ to send 20 frames. On the other hand, using our JPEG encoder, the compressed image uses only 3.8 kB, generating 608 kbps. The GAP-8 needs 9.9 mJ, but the energy used by the Wi-Fi decreases to 18 mJ with overall consumption of 27.9 mJ, which is an improvement of 14× in system energy efficiency.

15.5 Conclusions

In this paper, we present an optimized JPEG encoder based on the FDCT, which is parallel executed on GAP-8, a multi-core RISC-V SoC.

The encoder can reach up to 86 fps @ 200 MHz, but at 100 MHz the MCU requires only 0.495 mJ to compress a frame, reaching the best trade-off between the compression rate (46 fps) and the energy consumption. When compared with a JPEG implementation on ARM Cortex-M4 (48 MHz), our solution (@ 50 MHz) achieves a frame rate 4.8× higher with and requires 8 times less energy to encode a single image. Instead, if compared to Noritsuna [18], our solution features 56× lower number of clock cycles. Lastly, we exploit the JPEG encoder in a real deployment, a QVGA sensor with a Wi-Fi module. In this application, our solution can reduce the system energy up to 14× at 20 fps with respect to stream raw images through a Wi-Fi connection.

References

1. Magno M et al (2013 June) Multimodal video analysis on self-powered resource-limited wireless smart camera. IEEE J Emerg Sel Top Circuits Syst 3(2):223–235
2. Magno M et al (2009 Sept) Multimodal abandoned/removed object detection for low power video surveillance systems. In: 2009 Sixth IEEE international conference on advanced video and signal based surveillance, Genova, pp 188–193
3. Polonelli T et al (2019 June) A multi-protocol system for configurable data streaming on IoT healthcare devices. In: 2019 IEEE 8th international workshop on advances in sensors and interfaces (IWASI), Otranto, Italy, pp 112–117
4. Negri L et al (2004 Aug) FSM-based power modeling of wireless protocols: the case of Bluetooth. In Proceedings of the 2004 international symposium on low power electronics and design (IEEE Cat. No.04TH8758), Newport Beach, CA, USA, pp 369–374
5. Ballerini M et al (2019 July) Experimental evaluation on NB-IoT and LoRaWAN for industrial and IoT applications. In: 2019 IEEE 19th international conference on industrial informatics (INDIN), Helsinki, 2019
6. Makkaoui L et al (2010 July) Fast zonal DCT-based image compression for wireless camera sensor networks. In: 2010 2nd international conference on image processing theory, tools and applications. IEEE, pp 126–129
7. Rusci M et al (2016) An event-driven ultra-low-power smart visual sensor. IEEE Sens J 16(13):5344–5353
8. Chen S et al (2011) A 64 × 64 Pixels UWB wireless temporal-difference digital image sensor. IEEE Trans Very Large Scale Integr (VLSI) Syst 20(12):2232–2240
9. Torfs T et al (2012) Low power wireless sensor network for building monitoring. IEEE Sens J 13(3):909–915
10. Rossi D et al (2015 Oct) A −1.8 V to 0.9 V body bias, 60 GOPS/W 4-core cluster in low-power 28 nm UTBB FD-SOI technology. In: 2015 IEEE SOI-3D-subthreshold microelectronics technology unified conference (S3S). IEEE, pp 1–3
11. Osman H et al (2007 Nov) JPEG encoder for low-cost FPGAs. In: 2007 international conference on computer engineering & systems. IEEE, pp 406–411
12. Sakamoto T et al (1998) Software JPEG for a 32-bit MCU with dual issue. IEEE Trans Consum Electron 44(4):1334–1341

13. Rao K et al (2014). Discrete cosine transform: algorithms, advantages, applications. Academic Press
14. Chen W et al (1977) A fast computational algorithm for the discrete cosine transform. IEEE Trans Commun 25(9):1004–1009
15. Hou H (1987) A fast recursive algorithm for computing the discrete cosine transform. IEEE Trans Acoust Speech Signal Process 35(10):1455–1461
16. Loeffler C et al (1989 May) Practical fast 1-D DCT algorithms with 11 multiplications. In: International conference on acoustics, speech, and signal processing. IEEE, pp 988–991
17. Arai Y et al (1988) A fast DCT-SQ scheme for images. IEICE Trans (1976–1990) 71(11):1095–1097
18. Noritsuna 2019, https://github.com/noritsuna/JPEGEncoder4Cortex-M. Available online: July 2019
19. Moodstocks 2016, https://github.com/Moodstocks/jpec. Available online: July 2019
20. Flamand E et al (2018 July) GAP-8: a RISC-V SoC for AI at the edge of the IoT. In: 2018 IEEE 29th international conference on application-specific systems, architectures and processors (ASAP). IEEE, pp 1–4
21. Hore A et al (2010 Aug) Image quality metrics: PSNR vs. SSIM. In: 2010 20th international conference on pattern recognition. IEEE, pp 2366–2369
22. Polonelli T et al (2018 Oct) Slotted ALOHA overlay on LoRaWAN-A distributed synchronization approach. In: 2018 IEEE 16th international conference on embedded and ubiquitous computing (EUC). IEEE, pp 129–132
23. Polonelli T et al (2019 Feb) Slotted ALOHA on LoRaWAN-design, analysis, and deployment. In: Sensors (Switzerland), 19(4)

Part IV
VLSI & Signal Processing

Chapter 16
VLSI Architectures for the Steerable-Discrete-Cosine-Transform (SDCT)

Luigi Sole, Riccardo Peloso, Maurizio Capra, Massimo Ruo Roch, Guido Masera and Maurizio Martina

Abstract Since frame resolution of modern video streams is rapidly growing, the need for more complex and efficient video compression methods arises. H.265/HEVC represents the state of the art in video coding standard. Its architecture is however not completely standardized, as many parts are only described at software level to allow the designer to implement new compression techniques. This paper presents an innovative hardware architecture for the Steerable Discrete Cosine Transform (SDCT), which has been recently embedded into the HEVC standard, providing better compression ratios. Such technique exploits directional DCT using basis having different orientation angles, leading to a sparser representation which translates to an improved coding efficiency. The final design is able to work at a frequency of 188 MHZ, reaching a throughput of 3.00 GSample/s. In particular, this architecture supports 8k UltraHigh Definition (UHD) (7680 × 4320) with a frame rate of 60 Hz, which is one of the best resolutions supported by HEVC.

Keywords Video coding · Discrete Cosine Transform · Directional transform · VLSI

16.1 Introduction

In recent years, a large effort has been devoted to the field of video compression to cope with the increasing demand of high resolution multimedia contents. The latest standard proposed by ITU-T and ISO/IEC groups is the H.265/HEVC compression algorithm [8]. It extensively employs inter-frame and intra-frame prediction to exploit the temporal and the spatial redundancies present in video streams. H.265/HEVC requires computational load to detect and process intra mode, so many efforts have been done in order to lower the complexity [6] of the detection phase. The difference between the predicted block and the actual block of pixels is called *residual block* and

L. Sole · R. Peloso · M. Capra (✉) · M. Ruo Roch · G. Masera · M. Martina
Politecnico di Torino, Turin, Italy
e-mail: maurizio.capra@polito.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_16

it is lossly coded taking advantage of transforms (Discrete Sine Transform, DST, and Discrete Cosine Transform, DCT) and quantization. While the DST is used only for the smallest block size, namely 4×4 pixels, the DCT is used for all the other sizes, typically up to 32×32. Chen et al. [1] has shown how to reduce the complexity of the Integer Cosine Transform enabling solution up to 64×64. Since DCT is increasing in complexity and computational load, faster and low-power architectural solutions such as [7, 9] are required. Recently, Fracastoro et al. [3] proposed a directional DCT, called Steerable DCT (SDCT), which is better suited than DCT to compress directional data. The SDCT is based on the work of Zeng et al. [10] and makes possible to divide the directional cosine transform into a traditional DCT followed by a geometrical rotation. The kernels used for the SDCT are different from the DCT ones as they depend on the steering angle, with the limit case of 0 degrees rotation for which the SDCT coincides with the DCT. This paper presents a low power hardware accelerator for SDCT able to reach the throughput required by HEVC for the 8k UltraHigh Definition of 7680×4320 pixels. At first the architecture is analysed in Sect. 16.2 and then Sect. 16.3 will present the obtained results for the basic SDCT accelerator and some implementations stemming from it.

16.2 Architectural Implementation

While the 2D-DCT employed in HEVC is an inherently separable operation, the SDCT must be computed all at once. The complexity of a transform that is not separable is far greater than a separable one, so this may be a big drawback for the implementation. However, the complexity can be decreased drastically by splitting the SDCT in two parts, namely a separable 2D DCT followed by some rotations, and then by computing the separable transform before applying rotations, as reported in [4]:

$$\tilde{\boldsymbol{x}} = T(\boldsymbol{\theta})\boldsymbol{x} = R(\boldsymbol{\theta})T\boldsymbol{x} = R(\boldsymbol{\theta})\hat{\boldsymbol{x}} \tag{16.1}$$

where $\boldsymbol{x}$ are the input samples, $\hat{\boldsymbol{x}}$ are the results obtained by applying the T transform matrix, $R(\boldsymbol{\theta})$ is the rotation matrix, while $\tilde{\boldsymbol{x}}$ is the result of the SDCT. The SDCT can be thus implemented as a DCT followed by a steering transformation. The DCT part can be implemented as suggested in the literature, for example using a folded architecture [5], and then applying rotations when all the samples returned by the 2D-DCT are available. This means that the steering part of the architecture, which handles the rotations, has to work faster than the DCT. This issue has been addressed in this work and one of the possible solution is to define two clock regimes, one for the 2D-DCT and one, faster, for the steering part, in order to comply with the throughput offered by the 2D-DCT transform block. A FIFO memory between the two parts acts as a buffer memory. The whole structure is depicted in Fig. 16.1. The 2D-DCT block is based on the architecture proposed in [5] by Meher et al., which is very efficient, especially in the folded fashion, and scalable to transforms of size 4, 8, 16 and 32. The steerable part is shown in Fig. 16.2. It is composed by an input memory

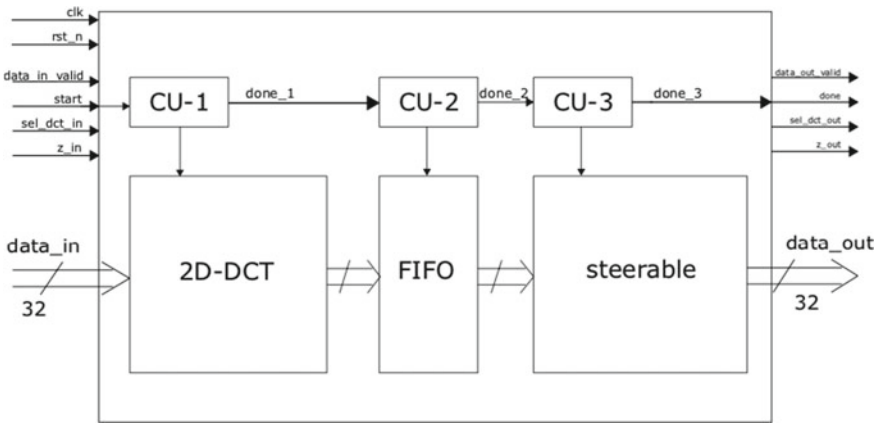


Fig. 16.1 Whole SDCT structure

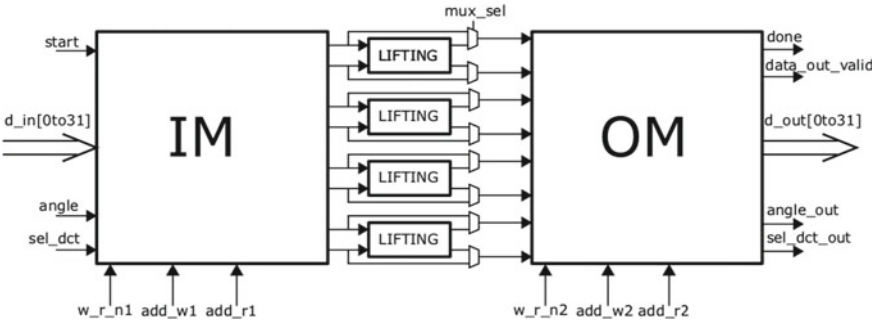


Fig. 16.2 Steerable block structure

(IM), an output memory (OM) and the lifting blocks that perform the rotation [2]. Some multiplexers are used to bypass the lifting blocks for the case of no rotation, returning directly the result given by the DCT. The IM is required also to reorder the samples as the steering process is computed on the custom zig-zag order, given in Fig. 16.3, that is different from the classic zig-zag ordering, as the vectors are rotated in pairs with respect to the diagonal elements. Rotation by lifting scheme:

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-\cos\theta}{\sin\theta} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\sin\theta & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{1-\cos\theta}{\sin\theta} \\ 0 & 1 \end{pmatrix} \tag{16.2}$$

The rotation matrix is decomposed in the multiplication of other three rotation matrices, in such a way the resulting structure, shown in Fig. 16.4, presents a lower complexity. Indeed, this implementation requires only three multipliers, while the original rotation matrix would need four multipliers to achieve the same result. In

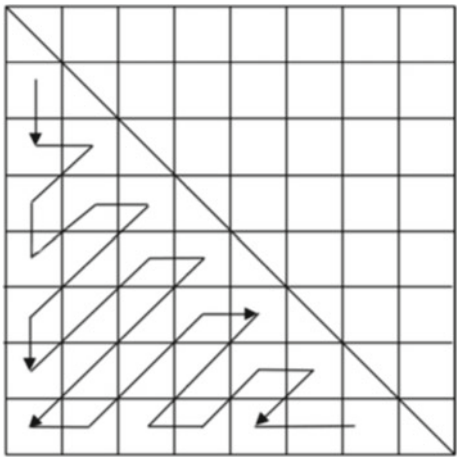


Fig. 16.3 Zig-zag scanning order

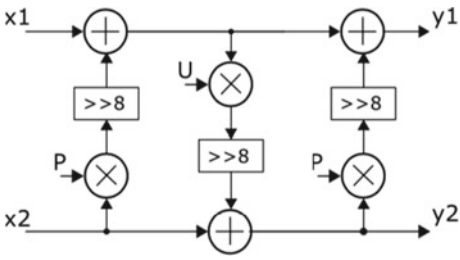


Fig. 16.4 Lifting-based rotation

order to further simplify the architecture, the multiplication for P and U coefficients from Eq. 16.2

$$P = \frac{1 - \cos\theta}{\sin\theta} \tag{16.3}$$

$$U = -\sin\theta \tag{16.4}$$

In Fig. 16.4 is implemented as shift and add, as the number of possible rotation angles have been fixed to 8 (from 0, no rotation, to 7), as reported as optimum in [4] by Masera et al. The steerable block thus introduces $2 \times N$ clock cycles of latency for the reordering stage plus 4 clock cycles due to the internal pipeline. Therefore, in the event that all the SDCT have a length $N = 32$, the latency is equal to 68 clock cycles, which corresponds to the worst case.

16.2.1 Reduced SDCT Architectures

The unit presented so far is able to compute SDCT of lengths 4, 8, 16 and 32. This type of structure has been designed to be implemented inside the HEVC standard.

Anyway, this algorithm could be also used for video compression standards with lower constraints and for image compression standard, such as JPEG. Therefore, two reduced SDCT unit have also been developed. The first is able to compute SDCT of length 4, 8 and 16, named SDCT-16, while the second is capable of computing SDCT of length 4 and 8, named SDCT-8. These two units have a reduced throughput of 50% and 75% respectively, so they have a parallelism of 16 or 8 data instead of 32, reducing the size of all the memories. In particular the length of both rows and columns of all memories is halved in the SDCT-16 unit, while is four time lower in the SDCT-8 unit with respect to SDCT-32. As a result the area occupation of these units is much lower than the SDCT-32 one. Moreover, just one clock domain has been used for both DCT and steerable block.

16.3 Results

In order to satisfy the HEVC speed requirements for a video resolution of 7680 × 4320 and a frame rate of 60 fps, the proposed structure needs a throughput of almost 3 GSample/s. As discussed in Sect. 16.2, the folded version presented in [5] has been implemented since this approach guarantees the required throughput. This structure has a processing rate of 16 pixels per cycle, therefore the architecture needs a frequency of at least 187 MHz (2.99 × 109/16 MHz). Clock gating has been enabled for the synthesis, leading to a smaller area and lower power consumption. The technology employed for the synthesis is the UMC 65 nm. The following architectures have been considered and synthesized:

– two-dimensional DCT
– SDCT
– reduced SDCT-16
– reduced SDCT-8.

For the SDCT implementation, several clocks have been tested for the steering part, namely 1×, 2×, 4× and 8×. By increasing the Steerable unit frequency it is possible to decrease the parallelism and consequently the number of input/output ports of the buffers (Table 16.1).

It can be noticed that by reducing the data parallelism of the Steerable unit, the size of the input memory (IM) and output memory (OM) decreases considerably, while the size of all the other sub-blocks slightly increases (Table 16.2).

In literature there are no other SDCT hardware architectures, so it is not possible to make comparisons. However, Table 16.3 presents an overview of the obtained results. As it can be noticed, the area and power results of the SDCT-16 are around 60% smaller than the complete SDCT. On the other hand, the SDCT-8 area is around 75% smaller than the SDCT-16 and 90% smaller than the complete SDCT while the throughputs are reduced respectively by 50% and 75%. Finally, comparing the DCT and the SDCT architecture we can observe that the hardware overhead to support up to N = 32 is very large. However, removing the hardware support for the steering

Table 16.1 SDCT area occupation for different clock regimes

Cell	1× total area (μm^2)	2× total area (μm^2)	4× total area (μm^2)	8× total area (μm^2)
SDCT	4,337,744	3,042,226	1,608,759	1,301,522
2D-DCT	438,866	601,970	455,150	474,167
IM	1,401,523	820,032	495,856	335,932
OM	2,377,837	1,418,162	482,048	319,037
FIFO	86,542	110,594	113,008	110,604
ROM	5895	22,228	13,227	33,223

Table 16.2 Estimated power consumption at 188 MHz

Power	Internal (mW)	Switching (mW)	Total dynamic (mW)	Leakage (mW)
Basic DCT	36.55	17.72	54.47	33
Clock gated DCT	21	12.52	33.52	30
Basic SDCT	290.47	60.33	350.88	106
Clock gated SDCT	88.71	59.85	48.67	94
Clock gated SDCT-16	27.86	28.97	56.85	27
Clock gated SDCT-8	6.56	7.20	14.17	7

Table 16.3 Overview of the obtained architectures

Architecture	DCT	SDCT	SDCT-16	SDCT-8
Technology (nm)	65	65	65	65
Frequency (MHz)	188	188	188	188
Power (mW)	33.52	148.67	56.85	14.17
Throughput	$2.992G$	$2.992G$	$1.496G$	$0.748G$
Area (mm^2)	0.321	1.427	0.444	0.110

part with N = 32 (SDCT-16), the area becomes comparable with the one of the DCT. As a consequence, this solution can be of interest to increase the rate-distortion performance [4].

16.4 Conclusion

This paper provides an efficient and compact hardware architecture accelerator for the SDCT algorithm to be used in the HEVC algorithm. Many of the design choices explained above present an optimized approach, such as the lifting-based approach, in which the hardware resources are reduced to a minimum. Moreover, the flexibility showed by this architecture makes it appealing for a wide range of applications, being able to work with different coding formats. The proposed SDCT framework is able to cope with 8k UltraHigh Definition (UHD) ($7680 \times 4320 pixels$) with a frame rate of 60 Hz for the 4:2:0 YUV format, which is one of the highest resolution supported by HEVC. The steerable DCT is a viable solution to improve compression efficiency, as reported in [4]. Further work will cover the integration of the proposed accelerator in a complete HEVC framework to validate the performances in a real case scenario.

References

1. Chen Z, Han Q, Cham W (2018) Low-complexity order-64 integer cosine transform design and its application in hevc. IEEE Trans Circ Syst Video Technol 28(9):2407–2412
2. Daubechies I, Sweldens W (1998) Factoring wavelet transforms into lifting steps. J Fourier Anal Appl 4(3):247–269. https://doi.org/10.1007/BF02476026
3. Fracastoro G, Fosson SM, Magli E (2017) Steerable discrete cosine transform. IEEE Trans Image Process 26(1):303–314
4. Masera M, Fracastoro G, Martina M, Magli E (2019) A novel framework for designing directional linear transforms with application to video compression. In: ICASSP 2019—2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 1812–1816
5. Meher PK, Park SY, Mohanty BK, Lim KS, Yeo C (2014) Efficient integer DCT architectures for HEVC. IEEE Trans Circ Syst Video Technol 24(1):168–178
6. Ogata J, Ichige K (2018) Fast intra mode decision method based on outliers of DCT coefficients and neighboring block information for h.265/hevc. In: 2018 IEEE international symposium on circuits and systems (ISCAS), pp 1–5
7. Oliveira RS, Cintra RJ, Bayer FM, da Silveira TLT, Madanayake A, Leite A (2019) Low-complexity 8-point dct approximation based on angle similarity for image and video coding. Multidimension Syst Signal Process 30(3):1363–1394. https://doi.org/10.1007/s11045-018-0601-5
8. Sullivan GJ, Ohm J, Han W, Wiegand T (2012) Overview of the high efficiency video coding (HEVC) standard. IEEE Trans Circ Syst Video Technol 22(12):1649–1668
9. Sun H, Cheng Z, Gharehbaghi AM, Kimura S, Fujita M (2019) Approximate DCT design for video encoding based on novel truncation scheme. IEEE Trans Circ Syst I Regul Pap 66(4):1517–1530
10. Zeng B, Fu J (2008) Directional discrete cosine transforms–a new framework for image coding. IEEE Trans Circ Syst Video Technol 18(3):305–313

Chapter 17
Hardware Architecture for a Bit-Serial Odd-Even Transposition Sort Network with On-The-Fly Compare and Swap

Ghattas Akkad, Rafic Ayoubi, Ali Mansour and Bachar ElHassan

Abstract Sorting algorithms are computationally expensive routines frequently executed on modern computers and embedded systems. Implementing sorting algorithms on dedicated hardware can contribute significantly to the overall execution time of the processes and applications embodying them. However, such algorithms are known to suffer from a trade off between convergence time and computational complexity. Consequently, this causes performance degradation i.e. bottleneck, when implemented on dedicated hardware with limited resources. In this respect, this paper proposes a novel sequential hardware architecture for a bit-serial Odd-Even transposition sorting network with on-the-fly compare and swap, on field programmable gate array (FPGA). In contrast to the classical parallel-data architecture, which operates on N data bits, this implementation significantly minimizes resource utilization while offering higher clock frequency, on the fly compare and swap and preserving $O(N)$ performance complexity. Simulation and synthesis results demonstrates that the proposed architecture is parallel, minimal in size, can operate on much larger arrays for a reference area size, can be easily expanded, and can achieve higher operating frequency.

Keywords Sorting · Odd-Even transposition · Hardware architecture · FPGA · Low latency · Embedded systems · Bit-serial · Median filter

G. Akkad (✉) · R. Ayoubi
Department of Computer Engineering, University of Balamand, Koura, Lebanon
e-mail: Ghattas.Akkad@balamand.edu.lb; Ghattas.akkad@ensta-bretagne.org

R. Ayoubi
e-mail: Rafic.Ayoubi@balamand.edu.lb

G. Akkad · A. Mansour
Lab-STICC, UMR 6285, ENSTA Bretagne, Brest, France
e-mail: Ali.Mansour@ensta-bretagne.fr

B. ElHassan
Faculty of Engineering, Lebanese University, Tripoli, Lebanon
e-mail: Bachar.ElHassan@gmail.com

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_17

17.1 Introduction

Sorting is comparing and swapping array elements until a desired order is reached. The complexity of the design depends on the algorithm itself and the data stored. With the increased storage capacity of memory units and the emergence of high-level computing and data analysis applications, sorting algorithms seeped forward to become one of the most frequently executed tasks in software, thus optimizing the applications overall performance [1]. For instance, search algorithms prefers sorted data lists for maximum efficiency. Additionally, sorting is also useful in data exchange operations employed to solve problems in graph theory, computational geometry, deep learning, computer graphics, computer based simulations, and image processing in near real-time [1–5]. With this critical dependency on sorting, and the diversity of applications that embodies it, developers turned their attention to improving the efficiency of such algorithms by targeting lower-level implementations on dedicated processors and field programmable gate array (FPGA) for parallelism and accelerated convergence while combining speed and flexibility [1, 6–8].

One of the most popular architectures focused on implementing sorting algorithms sequentially, until no further significant improvement was made. Research henceforth concentrated on parallelizing these algorithms by massive pipeline and maximum resource consumption for maximum performance, hence a trade off between convergence speed and resource utilization. However, with the increase of deploy-able embedded systems, wearable devices and internet connected units, additional power consumption and resource utilization constraints have emerged [1, 9–11]. One of the simplest and frequently used sorting algorithm is the odd-even transposition sort whose performance is of the order $O(N)$ [1, 3, 12]. The odd-even transposition sort algorithm provides both parallelism and flexibility, supporting larger size arrays for a reference area size, while preserving an acceptable and efficient space-time factor [1, 3, 12, 13]. In addition, developments and improvements achieved on electronic components, resulted in minimizing transistor size and gate switching delay, allowed the use of higher clock frequencies and low complexity sequential operations. These improvements motivated re-exploring sequential, bit-serial odd-even transposition sorting network architectures to achieve flexibility, computational simplicity, minimal resource utilization and higher operating frequency while preserving parallelism, pipelining and performance [1, 3, 14–16].

Previous work and suggested architectures presented numerous ways for increasing the performance of sorting algorithms by implementing them in hardware [1, 9, 12, 13, 16, 17], and on multi-core processing units i.e. graphical processing units (GPU) [3, 4, 18, 19]. However, such improvements focuses on massive parallelism and multi-core processing for big data analysis and are not suitable for deployable systems i.e. (FPGA). Moreover, recent work in [1] proposed an optimized, shift-based, hardware implementation of the parallel-data Odd-Even Transposition sorting algorithm, with high flexibility for general purpose applications, capable of sorting arrays of length larger than two times the number of available processors. However, the suggested design in [1] increases the sorter capacity by adding additional

storage registers to temporary hold shifted data back and forth the sorting process thus increasing latency and time required for convergence making it unsuitable for limited resource devices and time critical application. In contrast to the modification presented in [1] which operates on N data bits in parallel, the motivation behind this work is to propose a sequential, bit-serial based Odd-Even transposition sorting network architecture with on the fly compare and swap. The suggested sorter is capable of sorting larger arrays for the same area size without the need of additional storage components. Additionally, this work focuses on providing higher operating frequency, minimized resource consumption and minimal computational complexity while preserving parallel operations and pipeline by employing bit level operations.

17.2 Sorting Algorithms Review

Parallel sorting algorithms were proven to provide an effective scheme to achieve accelerated performance over their sequential counterpart, however at the cost of computational complexity and increased resource utilization. In order to eliminate the trade off in computational complexity and resource utilization a modified version of the classical Odd-Even Sorting network is introduced. This section presents a brief overview of the working of the classical Odd-Even sorting algorithm, and the proposed shift-based approach [1].

17.2.1 Odd-Even Transposition Sort

The Odd-Even Transposition sort algorithm is a parallel, linear complexity $O(N)$ version of the well known sequential Bubble sort [3, 12, 13]. This modified algorithm is divided into two stages as shown in Fig. 17.1. In this process, we can see the different comparisons of each cycle. Cycle one starts by comparing the even indexed elements with their right neighbor followed by cycle two for comparing the odd indexed elements alike. Cycle one and two repeats until all data is sorted, thus the maximum input array length is directly proportional to the number of processors i.e. sorting units available [1].

17.2.2 Shift-Based Odd-Even Transposition Sort

The idea behind the following modification, is to expand the network capabilities to handling array sizes larger by a maximum of two times then the available processing cells while reducing routing complexity and interconnections. Such modification minimizes the sorting cell structure by limiting its access to two elements instead of

7	10	5	14	1	9	4	3
7	10	5	14	1	9	3	4
7	5	10	1	14	3	9	4
5	7	1	10	3	14	4	9
5	1	7	3	10	4	14	9
1	5	3	7	4	10	9	14
1	3	5	4	7	9	10	14
1	3	4	5	7	9	10	14

Fig. 17.1 Odd-Even transposition sort

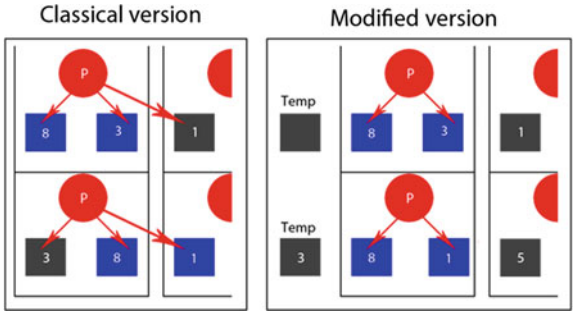


Fig. 17.2 Shift-based Odd-Even transposition

three. The additional third element is shifted back and forth within the sorter network, as shown in Figs. 17.2 and 17.3. As shown in Fig. 17.2, the blue boxes represents the working registers of processor P for a given sorting cycle. In the classical version, processor P has access to three registers, two local registers in the holding cell and one neighboring register in the right cell, hence an increase in routing complexity. In contrast, in a shift-based network, the processor P has access to only two registers i.e. cell local registers, where the additional element is shifted back in forth the sorting cell. Such modification resulted in a major reduce in routing complexity and fewer resource utilization at the cost of a temporary storage register and increased latency [1].

While the previously suggested modification, minimizes routing complexity and allows the network to sort larger array sizes, it suffers from an increased latency, slower conversion and requires additional storage registers proportional to the number of elements to be sorted. Thus it is of great interest to depict a sorting network capable of handling larger arrays for a fixed reference area with minimal routing, comparison and swap complexity.

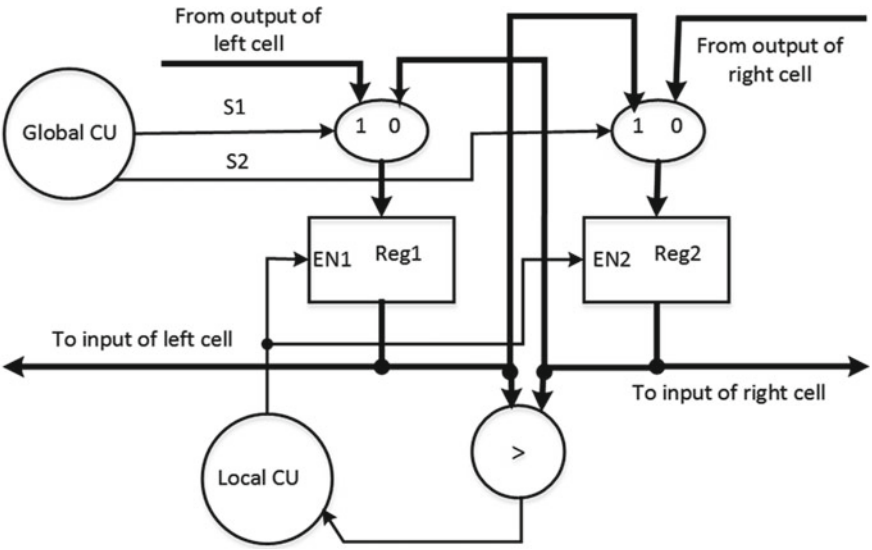


Fig. 17.3 Hardware architecture of a shift-based sorter cell

17.3 Bit-Serial Hardware Architecture

This architecture present a minimal size bit-serial odd even transposition sorting network with on the fly compare and swap capable of sorting larger array sizes in a fixed reference area for a higher clock rate and minimal routing requirements. The proposed approach preserves the parallel nature of the algorithm with $O(N)$ performance complexity.

17.3.1 Bit-Serial Odd-Even Architecture

The proposed architecture is serial, with bit-level operations, thus greatly minimizing resources utilization. Moreover, data is processed sequentially while loaded to the storage registers, most significant bit (MSB) first allowing the processing element to perform an on the fly swap the following cycle, without the need of an intermediate stage or additional storage elements. The sorting cell structure is shown in Fig. 17.4. As shown in Fig. 17.4 the cell structure is formed of three stages: Data input and routing, Storage and Processing. Moreover, each sorter cell is controlled by a local state machine to synchronize operations and re-route the input when needed. The cell operation is detailed as follows:

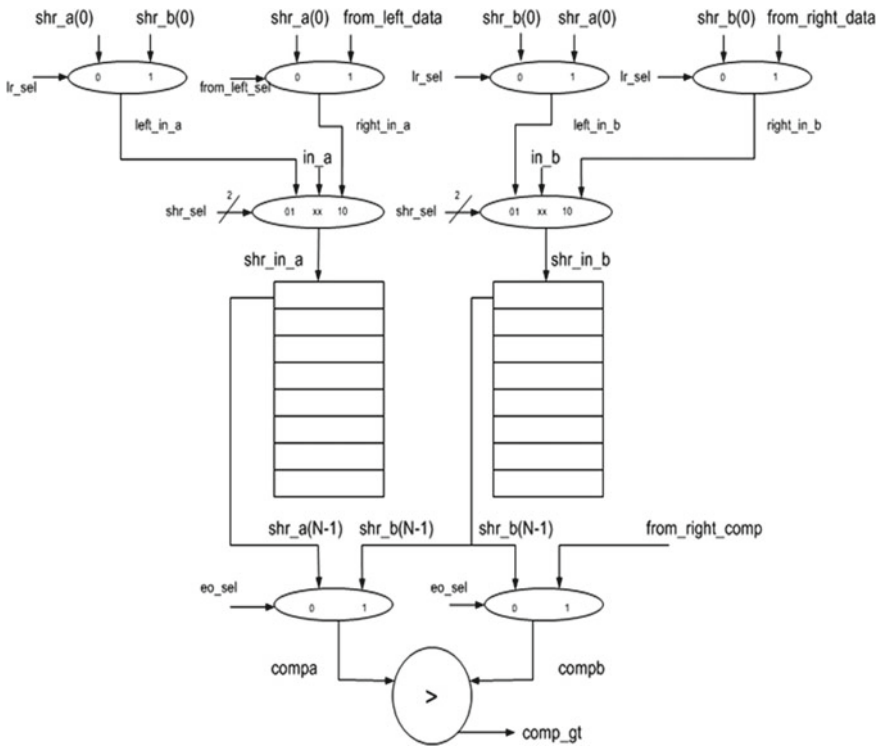


Fig. 17.4 Bit-serial sorting cell structure

1. The input stage is formed of two multiplexer levels of four and two 1-bit multiplexers respectively. The input stage routes the appropriate input data to the storage registers and perform swapping operation when required. The data is routed based on the decisions made by the local control unit i.e. from local registers, from right cell or from left cell.
2. The second stage handles the storage process and is formed of two N-bits shift registers. The input is shifted in most significant bit (MSB) first.
3. The third and final stage handles the comparison process and is formed of two input multiplexers and a reduced size 1-bit comparator. The comparison process is done MSB first and starts as soon as one input bit is shifted in lasting for N cycles. Moreover, by considering the N-th local registers as the main comparators input the swap decision can be decided at the N-th cycle i.e. when all data bits are processed hence swapping can be done on-the-fly in the next cycle by re-routing the inputs. Additionally, the comparison operation can begin comparing the swapped data directly. Such technique greatly reduces the latency of the design and eliminates the need for additional storage elements and operation cycles.

Thus the presented bit-serial structure preserves parallelism, operates on larger array sizes for a reference area given the major reduction in resource utilization i.e. using bit-level operators. Additionally, the design can be easily expanded to handle M-bits data where $M > N$ by adding additional, $M - N$ storage registers. While the increase in data bits requires additional processing cycles per iteration, the following problem is negligible by the dramatic increase in clock frequency achieved where a bit-level sequential structure can be considered as a fully pipelined architecture.

17.4 Simulation Results and Discussion

To assess the performance and resource utilization of the proposed architecture compared to the parallel classical and shift-based Odd-Even network, The design is implemented on the "Xilinx Spartan3E-XC3S1600E" FPGA. Numerical simulation, resources utilization and timing reports have been generated for each conducted experiment.

17.4.1 Classical Odd-Even Sorting Network

The classical parallel Odd-Even sorting simulation is conducted on an array p of $D = 16$ elements where $p = [8, 12, 4, 15, 2, 11, 6, 3, 5, 14, 16, 10, 1, 9, 13, 7]$. The sorting process required 13 cycles i.e.,100.152 ns to finish where a worst case scenario, requires 15 i.e 115.56 ns for an operating frequency of 129.803 MHz. Moreover, the implementation synthesis results shows the use of 1% logic slices i.e. 225 out of 14,752 and 1% 4-input look up tables (LUTs) i.e. 320 out of 29,504 for the mentioned FPGA.

17.4.2 Shift Based Odd-Even Sorting Network

Similarly the shift-based Odd-Even transposition sort simulation is conducted for the array $m = [8, 12, 4, 15, 2, 11, 6, 3, 5, 14, 16, 10, 1, 9, 13, 7, 12, 8, 10, 9, 13, 11, 15, 14]$. As shown in Fig. 17.5, the sorting process required 26 cycles for completion i.e., 133.848 ns divided in 13 sort cycles and 13 shift cycles for an operating frequency of 194.250 MHz. The synthesis results shows the use of 1% logic slices i.e. 169 out of 14,752 and less than 1% 4-input look up tables (LUTs) i.e. 200 out of 29,504 for the mentioned FPGA [1]. As 26 clock cycles were needed to sort the 16 elements array, and the clock cycle is 5.148 ns. Sorting the 16 elements requires 133.848 ns this number is 100.152 ns in the classical version. Taking the worst case scenario of 30 clocks the time needed is 154.44 ns which was 115.56 ns in the classical version. This

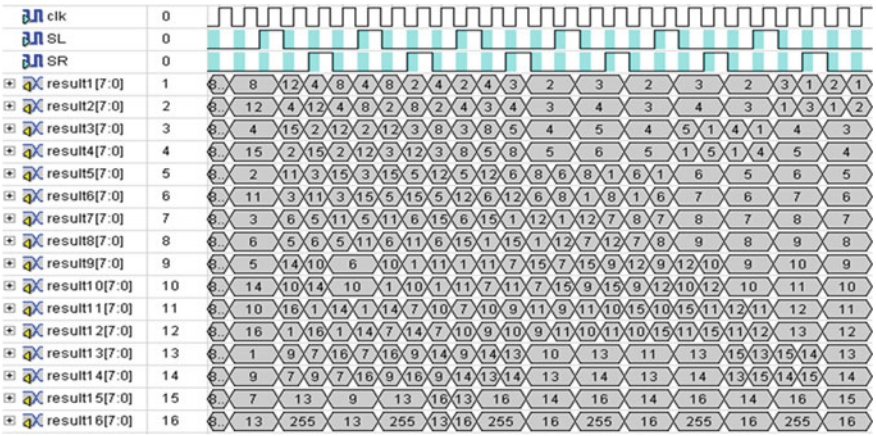


Fig. 17.5 Shift-based odd even sorter simulation

slowdown is caused by the added shift operations allowing the design to sort larger array sizes for a fixed number of processors. Such penalty increases proportionally to the added elements [1].

17.4.3 Bit-Serial Odd Even Sorting Network

The bit-serial Odd Even transposition sorting network simulation is conducted for the input array q of $D = 16$ elements with $N = 8$-bits unsigned data. $q = [150, 71, 82, 129, 24, 37, 116, 105, 18, 135, 86, 73, 148, 101, 120, 33]$ as shown in Fig. 17.6 for a simulation step size of 1 us, additional simulation cycles are caused by the data input process for initialization. Furthermore, the sorting network operated

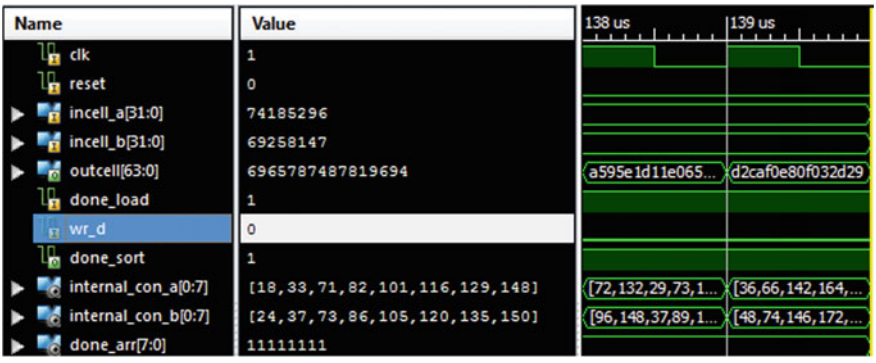


Fig. 17.6 Bit-serial odd even sorter simulation

Table 17.1 Synthesis results comparison

	Frequency (MHz)	FFs	LUT
Parallel classical sorter	129.803	225	320
Parallel shift based sorter	194.250	169	200
Bit-serial sequential sorter-8bit	620.34	80	51
Bit-serial sequential sorter-16bit	620.34	144	51
Bit-serial sequential sorter-32bit	620.34	270	76

on a maximum clock rate of 1.612 ns per cycle equivalent to 620.34 MHz. Synthesis results shows the use of 80 slice Flip-Flops (FFs) and 51 4-input look up tables (LUTs). Additionally the design was synthesized for $N = 16$, 32-bits unsigned data.

17.4.4 Results Comparison

In order to better assess the performance and resource utilization of the mentioned designs and to highlight the superior advantage of the proposed bit-serial architecture synthesis results are presented and compared in Table 17.1. Thus, as presented in Table 17.1 the proposed architecture is superior to the classical and shift based parallel network, can operate at a maximum frequency of 620.34 MHz and provides a major reduction in resource utilization. Additionally, unlike parallel computation based structures, the proposed design is flexible where a change in the number of operating bits results in a proportional increase of storage elements i.e. registers.

17.5 Conclusion and Future Work

In this paper, an optimized Bit-Serial Odd-Even Transposition sort with on the fly compare and swap hardware architecture was proposed. This implementation outperform previous parallel structures, is minimal in size, easily expandable to sort different data length i.e. bits, while preserving algorithm parallelism, complexity and pipelined structure. Additionally the presented structure can run at a much higher frequency given the simplicity of the employed bit level operations. Moreover, the sorting process begins while the data is being loaded into its memory, which means that the sorter doesn't require additional swap cycles. Further work could be made in this subject by adopting an optimized data loading technique. Improve the design to operate on signed data and fixed point representations.

References

1. Ayoubi R, Istambouli S, Abbas AW, Akkad G (2019) Hardware architecture for a shift-based parallel odd-even transposition sorting network. In: The 4th international conference on advances in computational tools for engineering applications, IEEE. Zouk Mosbeh, Lebanon
2. Batcher KE (1968) Sorting networks and their applications. In: Proceedings of the April 30–May 2, Spring joint computer conference, pp 307–314. AFIPS '68 (Spring), ACM, New York, NY, USA. https://doi.org/10.1145/1468075.1468121
3. Francis RS, Mathieson ID (1988) A benchmark parallel sort for shared memory multiprocessors. IEEE Trans Comput 37(12):1619–1626
4. Singh DP, Joshi I, Choudhary J (2018) Survey of GPU based sorting algorithms. Int J Parallel Prog 46(6):1017–1034. https://doi.org/10.1007/s10766-017-0502-5
5. Vasicek Z, Sekanina L (2008) Novel hardware implementation of adaptive median filters. In: 2008 11th IEEE workshop on design and diagnostics of electronic circuits and systems, pp 1–6
6. Akkad G, Ayoubi R, Abche A (2018) Constant time hardware architecture for a Gaussian smoothing filter. In: 2018 International conference on signal processing and information security (ICSPIS), pp 1–4
7. Akkad G, Mansour A, ElHassan B, Roy FL, Najem M (2018) Fft radix-2 and radix-4 FGPA acceleration techniques using hls and hdl for digital communication systems. In: 2018 IEEE international multidisciplinary conference on engineering technology (IMCET), pp 1–5
8. Akkad G, Mansour A, ElHassan B, Roy FL, Najem M (2018) Twiddle factor generation using Chebyshev polynomials and hdl for frequency domain beamforming. In: Applications in electronics pervading industry, environment and society, Springer lecture notes in electrical engineering. Springer
9. Chen R, Prasanna VK (2017) Computer generation of high throughput and memory efficient sorting designs on FPGA. IEEE Trans Parallel Distrib Syst 28(11):3100–3113
10. Farmahini-Farahani A, Duwe HJ III, Schulte MJ, Compton K (2013) Modular design of high-throughput, low-latency sorting units. IEEE Trans Comput 62(7):1389–1402
11. Rjabov A (2016) Hardware-based systems for partial sorting of streaming data. In: 2016 15th Biennial baltic electronics conference (BEC), pp 59–62
12. Hematian A, Chuprat S, Manaf AA, Parsazadeh N (2013) Zero-delay FGPA-based odd-even sorting network. In: 2013 IEEE symposium on computers informatics (ISCI), pp 128–131
13. Korat UA, Yadav P, Shah H (2017) An efficient hardware implementation of vector-based odd-even merge sorting. In: 2017 IEEE 8th Annual ubiquitous computing, electronics and mobile communication conference (UEMCON), pp 654–657
14. Huang C-Y, Yu G-J, Liu B-D (2001) A hardware design approach for merge-sorting network. In: ISCAS 2001. The 2001 IEEE international symposium on circuits and systems (Cat. No.01CH37196), vol 4, pp 534–537
15. Durad MH, Akhtar MN (2014) Performance analysis of parallel sorting algorithms using MPI. In: 2014 12th International conference on frontiers of information technology, pp 202–207
16. Olarlu S, Pinotti MC, Zheng SQ (2000) An optimal hardware-algorithm for sorting using a fixed-size parallel sorting device. IEEE Trans Comput 49(12):1310–1324
17. Lipu AR, Amin R, Mondal MNI, Mamun MA (2016) Exploiting parallelism for faster implementation of bubble sort algorithm using FPGA. In: 2016 2nd International conference on electrical, computer telecommunication engineering (ICECTE), pp 1–4
18. Faujdar N, Ghrera SP (2017) A practical approach of GPU bubble sort with CUDA hardware. In: 2017 7th International conference on cloud computing, data science engineering—Confluence, pp 7–12
19. Yildiz Z, Aydin M, Yilmaz G (2013) Parallelization of bitonic sort and radix sort algorithms on many core GPUS. In: 2013 International conference on electronics, computer and computation (ICECCO), pp 326–329

Chapter 18
Variable-Rounded LMS Filter for Low-Power Applications

Gennaro Di Meo, Davide De Caro, Ettore Napoli, Nicola Petra and Antonio G. M. Strollo

Abstract Precision-scalable techniques constitute an efficient solution to power consumption issues thanks to the possibility to adapt arithmetic components precision to required system-level accuracy with the aim to dynamically optimize power consumption. In this paper we propose a precision-scalable approach for the implementation of a Least Mean Square (LMS) filter. Novel solution exploits variable rounding multiplications in the learning section of the LMS filter allowing to dynamically reduce the switching activity of multipliers partial products with a minimal impact on error regime performance. Results, obtained after a Place & Route in TSMC 28 nm CMOS technology, reveal a regime precision comparable to a standard LMS implementation and a power consumption improvement up to 27%.

18.1 Introduction

Nowadays the reduction of power consumption is a key point in the design of digital circuits and important efforts are dedicated to develop new methods and techniques. Battery life, low self-heating and reliability are important design aspects in all modern electronic systems, and the problem is surely exacerbated if high operative frequencies are considered. In this scenario, precision-scalable approaches [1–3] are proposed with the assumption to tolerate some approximations for performances improvement. Audio and image processing, for instance, can leverage on limits of human senses to improve efficiency. In the area of data mining and neural network, data features are exploited to develop error-resilient applications [4, 5]. Also in the field of adaptive filters some precision-scalable techniques have been proposed for the Leas Mean Square (LMS) algorithm. Very used for applications as system identification, channel equalization or noise cancellation, it is composed by a FIR section and a learning part as shown in Fig. 18.1a. Unlike for canonical filters, LMS does not

G. Di Meo (✉) · D. De Caro · E. Napoli · N. Petra · A. G. M. Strollo
Department of Electrical Engineering and Information Technology, University of Napoli "Federico II", Naples, Italy
e-mail: gennaro.dimeo@unina.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_18

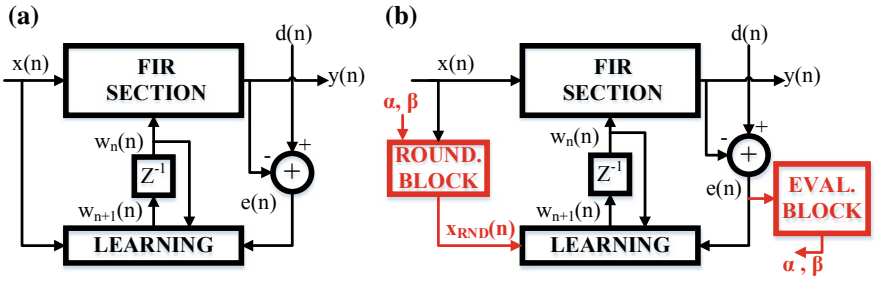


Fig. 18.1 **a** Standard LMS filter and **b** Variable-rounding LMS block diagram

have an a priori defined impulse response, but it changes its internal coefficients minimizing, in an approximate way, the mean square error (MSE) between its output and a desired signal. For this purpose, at each iteration, sum of products is executed, and multiplications between input samples and an error signal are performed to compute MSE gradient estimate (responsible for coefficients updating). As consequence LMS provides the usage of a large number of multipliers and registers, offering serious concerns from a power consumption point of view. In [6] approximate multipliers [7, 8] are used in the FIR section to reduce dissipation, but regime performances are not scalable. In [9] a run-time procedure observes coefficients magnitude and, following an external threshold, decides which terms are negligible for the output computation. Consequently, relative registers and multipliers are frozen. On the other hand, a not negligible increase in the area is due to the presence of additional blocks for regime detection and coefficients analysis, in addition to a relevant degradation of regime performances when high power saving is demanded. In this paper a variable rounding multiplication is explored in the LMS updating section for the gradient computation (as underlined in red in Fig. 18.1b). The idea is that if error signal is very small, it is possible to use a rounded version of input samples for gradient computation with negligible worsening of regime performances. In this way part of the multipliers partial products matrix is turned off, allowing power consumption saving. An advantage of this approach with respect to the technique of [9] is that it allows a power reduction in all multipliers of the learning section of the LMS filter. In addition, according to error behavior, circuit can decide between two different kinds of rounding, and the use of only one observation logic for the error signal is a very attractive solution. For a major comprehension of our proposal, in Sect. 18.2 a brief summary of LMS algorithm is offered and in Sect. 18.3 the low-power implementation is addressed. Finally, in Sect. 18.4 results and circuit implementation in TSMC 28 nm CMOS technology are discussed.

18.2 LMS Adaptive Filter

LMS computes its impulse response in an iterative way in order to minimize differences between the output signal $y(n)$ and the desired signal $d(n)$ in the mean square sense [10]. Considering input samples $x(n)$ and LMS coefficients $w_n(n)$, and defining the filter dimension DIM, $y(n)$ is given by the following expression:

$$y(n) = \sum_{i=0}^{DIM-1} w_n(i) \cdot x(n-i). \tag{18.1}$$

A comparison between $y(n)$ and $d(n)$ allows the computation of the error signal $e(n)$ used to underline the deviation respect to the desired behavior:

$$e(n) = d(n) - y(n)\,. \tag{18.2}$$

At this point, learning section updates coefficients according to the expression

$$w_{n+1}(i) = w_n(i) - \mu \cdot grad_n(i)\,. \tag{18.3}$$

where $grad_n(i)$ is the gradient estimate, given by:

$$grad_n(i) = e(n) \cdot x(n-i)\,. \tag{18.4}$$

It is worth noting that a proper choice of the step size parameter μ guarantees algorithm convergence and good regime performances [10].

18.3 Variable-Rounded LMS Filter

The key idea of this paper is approximating the gradient computation by using, in (18.4), an approximated version of $x(n-i)$, $x_{RND}(n-i)$, where some LSBs are rounded:

$$grad_{RND,n}(i) = e(n) \cdot x_{RND}(n-i)\,. \tag{18.5}$$

If we call ε_{grad} the absolute value of the gradient error, we can write:

$$\varepsilon_{grad} = |e(n)| \cdot \varepsilon_{xRND}\,. \tag{18.6}$$

Therefore, the lower is the absolute value of the error $e(n)$ the larger can be the error of $x_{RND}(n-i)$ (ε_{xRND}) for a prescribed ε_{grad} value. As shown in Fig. 18.1b, the proposed implementation provides an Evaluation Block for the error signal analysis and Rounding Blocks to obtain the approximate input $x_{RND}(n-i)$. The Evaluation

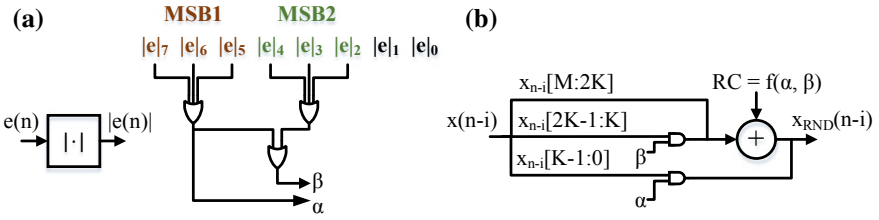


Fig. 18.2 **a** Evaluation block and **b** Rounding block schemes for the (n-i)-th acquired input sample

Block, represented in Fig. 18.2a, computes error signal module (through XOR operation between $e(n)$ and its sign bit), and divides its first most significant bits in two groups (we call them MSB1 and MSB2 group).

Starting from the two groups MSB1 and MSB2, the proposed approach uses a two-level approximation. If all the bits of MSB1 group are zero,α flag is set to zero. If also the bits of MSB2 group are all zero, the other flagβ is also set to zero. The flag α and β control the Rounding Block (represented in Fig. 18.2b). In the case $\alpha = 0$, K least significant bits of x are nullified through an AND operation. In the case in which also $\beta = 0$, additional K least significant bits of x are also nullified. In order to perform a rounding operation, a variable rounding constant RC is computed according to the following conditions:

$$\begin{aligned} RC &= x_{n-i}[K-1] \cdot 2^{-LSB+K} \; if \; \alpha = 0. \\ RC &= x_{n-i}[2K-1] \cdot 2^{-LSB+2K} \; if \; \alpha = 0 \; and \; \beta = 0. \end{aligned} \tag{7}$$

In this way, $x_{RND}(n)$ is multiplied with K (or $2K$) nullified LSBs, stacking at zero K (or $2K$) rows of the partial products matrix (see Fig. 18.3a). In addition, since gradient LSBs are zero, all coefficients LSBs are not updated and it is possible to

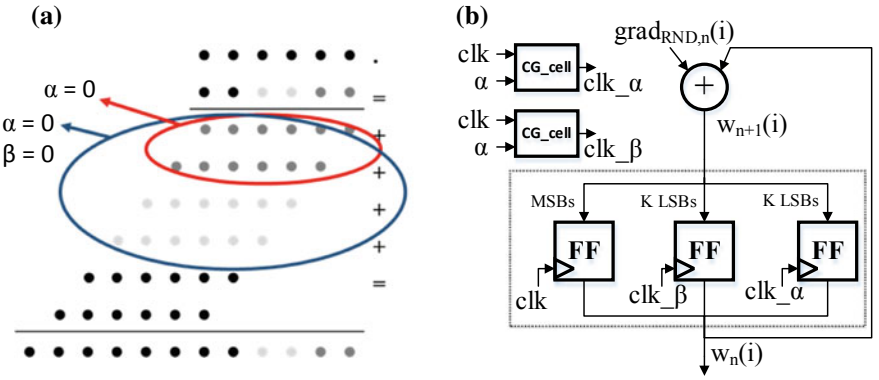


Fig. 18.3 **a** Multiplier for gradient computation and **b** clock gating for i-th feedback register. Nullified LSBs and rows are represented in gray in the figure on the left

freeze relative flip-flops. Then, as shown in Fig. 18.3b, two clock gated cells, enabled by α and β respectively, are introduced to manage all registers in the learning section.

18.4 Implementation and Results

To verify low-power properties, standard and proposed LMS are used to identify three different unknown systems. In particular a Low-pass FIR filter, a Low-pass IIR and an High-pass IIR filter are considered with order 40, 10 and 13 respectively. Convergence capabilities is investigated observing regime MSE, obtained by 25 independent simulations and averaging respective error signals. Considered length of LMS filter (DIM) is equal to 40. For low-power assessments, circuits are synthesized and routed in TSMC 28 nm CMOS technology and Post-Route results are analyzed. Inputs and coefficients are expressed in fixed-point 12-bit arithmetic, while error signal is on 18 bits. Soft rounding is demanded if 14 error MSBs are zero and hard approximation acts if 16 MSBs are nullified. We propose $K = 2$, then $x_{RND}(n)$ exhibits two or four nullified LSBs. All multipliers are synthesized with tree carry-save topology and fast vector merging adder.

Table 18.1 reports error performance. The regime MSE of proposed approach is very close to the standard LMS implementation, highlighting that the additional approximation results almost negligible with respect to other error sources. In addition, Fig. 18.4 shows the regime frequency response of the filters in the three considered cases in comparison to the frequency response of the target system. Again, we note very similar performances between standard and proposed LMS with very good in-band matching and very similar behavior in the stop-band.

Table 18.1 Regime error summary

MSE	Standard LMS	Proposed LMS
Low-pass FIR	3.52e−7	3.54e−7
Low-pass IIR	3.07e−6	3.08e−6
High-pass IIR	2.89e−5	2.90e−5

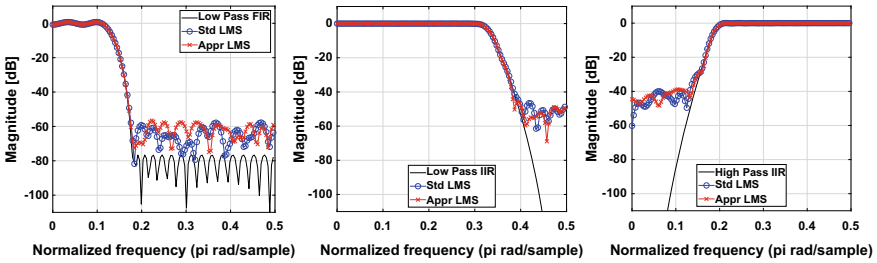


Fig. 18.4 Harmonic responses of unknown systems and LMS circuits. From the left to the right: Low-pass FIR, Low-pass IIR and High-pass IIR identification case

Table 18.2 Electrical characteristics summary

Regime P_{dyn}	Standard LMS (0.081 mm^2)	Proposed LMS (0.082 mm^2, +1.2%)
Low-pass FIR	245 μW/MHz	181 μW/MHz (−26%)
Low-pass IIR	251 μW/MHz	184 μW/MHz (−27%)
High-pass IIR	259 μW/MHz	204 μW/MHz (−22%)

Electrical post Place & Route performances are compared in Table 18.2. We can observe that proposed solution results only in a 1.2% area occupation increase (needed for additional control logic). In regime conditions, proposed LMS exhibits a sensible power dissipation reduction with respect to standard LMS. Percentage reduction is 26–27% for Low-pass FIR and IIR target systems. A lower percentage reduction (22%) is highlighted for High-pass IIR case where the regime MSE is higher.

18.5 Conclusions

A novel low-power implementation has been proposed for the LMS algorithm. A variable rounding on acquired input samples limits multipliers switching activity in the feedback section and approximation is demanded if error signal is very small. Results reveal a negligible worsening of regime MSE and area increase along with the possibility to reduce power consumption up to 27% respect to standard LMS filter.

References

1. Xu Q, Mytkowicz T, Kim NS (2016) Approximate computing: a survey. IEEE Des Test 33(1):8–22
2. Han J, Orshansky M (2013) Approximate computing: an emerging paradigm for energy-efficient design. In: 2013 18th IEEE European test symposium (ETS), Avignon, pp 1–6
3. Chippa VK, Chakradhar ST, Roy K, Raghunathan A (2013) Analysis and characterization of inherent application resilience for approximate computing. In: 2013 50th ACM/EDAC/IEEE design automation conference (DAC), Austin, TX, 2013, pp. 1–9
4. Raha A, Jayakumar H, Raghunathan V (2016 Mar) Input-based dynamic reconfiguration of approximate arithmetic units for video encoding. In: IEEE Trans Very Large Scale Integr (VLSI) Syst 24(3):846–857
5. Moons B, Verhelst M (2016) A 0.3–2.6 TOPS/W precision-scalable processor for real-time large-scale ConvNets. In: 2016 IEEE symposium on VLSI circuits (VLSI-Circuits), Honolulu, HI, pp 1–2

6. Esposito D, Di Meo G, De Caro D, Petra N, Napoli E, Strollo AGM (2018) On the use of approximate multipliers in LMS adaptive filters. In: 2018 IEEE international symposium on circuits and systems (ISCAS), Florence, pp 1–5
7. De Caro D, Petra N, Strollo AGM, Tessitore F, Napoli E (2013) Fixed-width multipliers and multipliers-accumulators with min-max approximation error. IEEE Trans Circuits Syst I Regul Pap 60(9):2375–2388
8. Petra N, De Caro D, Garofalo V, Napoli E, Strollo AGM (2011) Design of fixed-width multipliers with linear compensation function. IEEE Trans Circuits Syst I Regul Pap 58(5):947–960
9. Esposito D, Di Meo G, De Caro D, Strollo AGM, Napoli E (2019) Design of low-power approximate LMS filters with precision-scalability. In: Saponara S, De Gloria A (eds) Applications in electronics pervading industry, environment and society. ApplePies 2018. Lecture Notes in Electrical Engineering, vol 550. Springer, Cham
10. Haykin S (2002) Adaptive filter theory. Prentice-Hall

Chapter 19
A Simulink Model-Based Design of a Floating-Point Pipelined Accumulator with HDL Coder Compatibility for FPGA Implementation

Marco Bassoli, Valentina Bianchi and Ilaria De Munari

Abstract The design of an FPGA hardware architecture requires, traditionally, its description in a dedicated language (Hardware Description Language, HDL), which is often not well suited to manage wide and complex models. The design process can be simplified if the entire architecture can be described in a high abstraction level framework such as Simulink. In this paper a Simulink model-based design of a pipelined accumulator suitable for applications such as Support Vector Machine algorithms is presented. The compatibility with the HDL Coder workflow enables the direct FPGA model implementation. Moreover, the workflow output has been compared with a native VHDL equivalent floating-point accumulator intellectual property.

19.1 Introduction

Recent researches have focused on Human Activity Recognition (HAR) as a new service in the context of Smart Homes for behavioral monitoring [1]. The development of the wearable devices [2, 3] leds to implement new solutions that can be used in the field of HAR. The most advanced HAR algorithms are based on Machine Learning techniques, which are usually very computationally demanding.

To address this issue, several solutions have been proposed. An example is the decomposition of the algorithm to host the most computational-expensive parts into a cloud service [4]. On the other hand, alternative devices have been proposed [5], which equip dedicated hardware architectures (i.e. FPGAs) for the algorithm instead

M. Bassoli (✉) · V. Bianchi · I. De Munari
Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, 43124 Parma, Italy
e-mail: marco.bassoli@unipr.it

V. Bianchi
e-mail: valentina.bianchi@unipr.it

I. De Munari
e-mail: ilaria.demunari@unipr.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_19

of using general-purpose processors. This allows to have exactly the resources needed for the task and to optimize the system for performance or physical size, depending on the use case.

The design of dedicated hardware architectures is traditionally done by using a Hardware Description Language (HDL) and, after subsequent verification methods, the system is implemented on the destination platform. However, as proved by different works [6–8], dealing with high abstraction level frameworks enables the designer to eliminate the verbosity of highly typed programming languages (such as VHDL or Verilog) and to focus the attention on system functionalities only. This is possible, for example, by using MATLAB/Simulink software. A high-level, block-based design can be developed and the behavior of the system can be simulated in the same environment. Moreover, with the dedicated HDL Coder tool, an HDL code can be automatically generated from the system block diagram and hence used to program the selected platform.

This methodology is the basis of our work on the development of a Support Vector Machine (SVM) algorithm for HAR to be embedded in an FPGA-based wearable device.

Among the SVM blocks employed in our dedicated Simulink design, the accumulator is one of the most frequently used. Hence, the aim of this paper is to present a Simulink model of an accumulation circuit full compatible with the HDL Coder workflow and which exploits the advantages of a model-based design approach [9, 10].

The paper is organized as follows. In Sect. 19.2 related works are discussed while in Sect. 19.3 the designed architecture is introduced. In Sect. 19.4, results are shown and in Sect. 19.5 conclusion are drawn.

19.2 Related Works

General FPGA-based SVM architecture deals with data with high dynamic data range: thus, it is based on floating-point arithmetic, as this is the best solution with data with this requirement [11]. For this reason, we focused on floating-point accumulators' architecture. The accumulation operation becomes critical when a floating-point adder with latency is used: in this case, to produce a correct result, the input data frequency must match this latency value [12]. Many solutions have been presented in literature to face this issue. In [13], Ni and Hwang presented a version of the system in which, thanks to an articulated control logic, only one adder and a buffer are employed. In [14] a version with a better throughput has been proposed.

On a parallel-side branch, several works presented dedicated architectures for the adder part. In [15], Luo and Martonosi broke down the floating-point adder structure to embed delayed additions at the cost of a more complex control logic. A similar approach has been used in [12], and, in [16], Wang et al. presented several reduction circuits able to work with variable floating-point precision.

Table 19.1 State-of-the-art hardware architectures in reduction circuits

Method	# adders	Buffer size	Accumulator latency for a set length n
FCBT	2	$3\lceil \log n \rceil$	$\leq 3n + (p-1)\lceil \log n \rceil$
SSA	1	$2p^2$	$\leq n + 2p^2$
DB	1	$\lceil 2p/3 \rceil + \lceil p/2 \rceil$	$n + p - 1 + T_m^{DB}$
AeMFPA	2	$3\lceil \log n \rceil$	$\leq n + p\lceil \log_2 p + 2 \rceil$

In [17], Zhuo et al. presented two main architectures: the Fully Compacted Binary Tree (FCBT) and the Single Strided Adder (SSA). The FCBT is an accumulator based on two classical floating-point adders and a number of buffers k, found to be:

$$k = \log\lceil n \rceil - 1. \tag{1}$$

The purpose is to overcome limitations of solutions as [18], in which the system correctly works only for power of 2 input vector sizes. The SSA presented in [17] exploits one adder but a larger buffer size. It also introduces the ability to process multiple interleaved input sets.

In [19], Tai et al. focused their work on an area- and speed-efficient system. Starting from SSA and thanks to a complex control logic, they managed to maximize the area-time product, as shown in a reported comparison. The architecture is named Delayed Buffering (DB).

The aim of Huang and Andrews in [20] was to realize an accumulator whose output is always the running sum of the input, aspect not present in the previous works. Their architecture, called Area-Efficient Modular Fully Pipelined Architecture (AeMFPA), is characterized by a smaller buffer size and a simple control logic.

In Table 19.1 is summarized the state-of-the-art in this field, where, n is the number of input elements to be reduced, p is the accumulator latency of the reduction operation (e.g. adder, multiplier, etc.), and T_m^{DB} is the characteristic time compensation function for the DB architecture, defined in [19].

In the present work, we focus on the system presented in [19], since it offers the lowest latency for a single set. The reason is that, in an SVM context, multiple sets have to be reduced to provide the result. This means the lower the accumulator latency for a single dataset, the faster SVM result production, hence a higher system throughput.

19.3 Architecture

The proposed Simulink model is shown in Fig. 19.1. To design the proposed model, basic Simulink blocks have been used. However, since in Simulink a specific block modeling an adder with latency is missing, an *Adder With Latency* block has been created as a cascade of an adder and a delay block. This configuration also allows

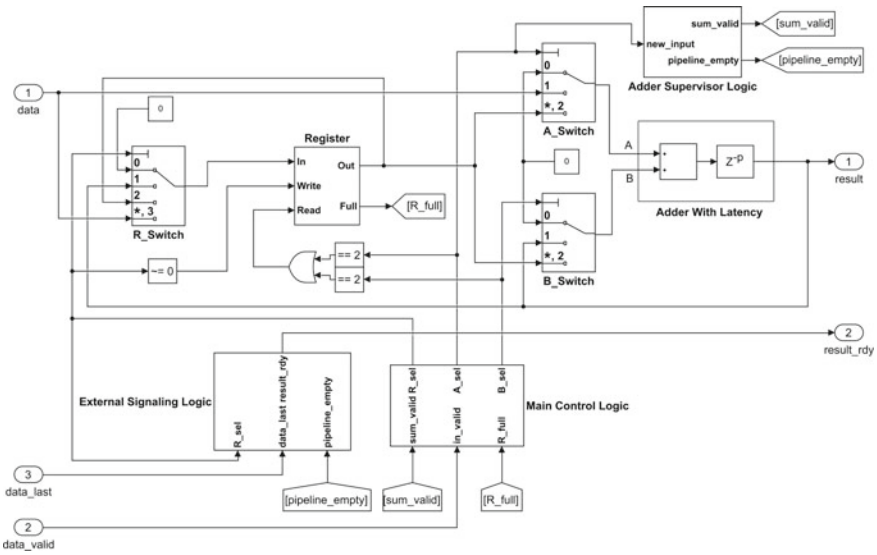


Fig. 19.1 The simulink accumulator model based on the work of [19]. In this example, the system has been configured to model a pipelined accumulator with an adder latency of p clock cycles

to configure the latency of the adder with a customizable value of p. The rest of the architecture features three *Switch* blocks (*R_Switch*, *A_Switch* and *B_Switch*) and three main *Logic* blocks (*External Signaling Logic*, *Main Control Logic* and *Adder Supervisor Logic*).

The *Switch* blocks are used as routing elements and their behavior is equivalent to the Register Transfer Level (RTL) multiplexer element. With this configuration, the *Register* can be shared by both operand A and B. Moreover, as a control logic rule, the input data can only be used as operand A while the operand B comes from the feedback path each time the adder output is valid.

The *Logic* blocks are subsystems which produce the control signals for the entire architecture. In detail:

- *Main Control Logic*: it is the core control unit of the system. As explained in [19], it controls the data path of the input data stream, the *Register* and the adder to avoid data collisions and data loss. The detailed operation of the logic is reported in Table 19.2 and an execution example is shown in Table 19.3;
- *External Signaling Logic*: it is the logic dedicated to the management of the *data_last* input flag and to produce the *result_ready* output flag. The output can be considered ready when all the input conditions are verified: *data_last* raised by the user, internal adder pipeline empty (meaning no other operands are to be processed) and last adder result placed in the *Register*. The first condition is evaluated by capturing the user *data_valid* assertion through a Set-Reset (S-R) Flip-Flop (FF), the second is directly given by the *pipeline_empty* signal from the Adder Supervisor Logic and the third is evaluated by verifying whether the *R_sel*

Table 19.2 Main control logic working behavior

	Condition	Behavior
1	Input valid	Input in register R
2	Adder output valid, data in register R	Adder output fed back to adder input, register R value to adder input
3	Input valid, data in register R	Input directly to adder, register R value to adder input
4	Input valid, adder output valid	Adder output fed back to adder input, Input directly to the adder
5	Input valid, adder output valid, data in register R	Adder output fed back to adder input, Input directly to adder, register R holding data

Table 19.3 Example of 4 input elements and a pipelined adder with a latency of 2 clock cycles

Cyc.	Data	A	B	R	Result
0	X_1			X_1	
1	X_2	X_2	X_1		
2	X_3			X_3	
3	X_4	X_4	$X_1 + X_2$	X_3	$X_1 + X_2$
4				X_3	
5		X_3	$X_1 + X_2 + X_4$		$X_1 + X_2 + X_4$
6					
7					$\sum_{i=1}^{4} X_i$

bus is equal to one. The FF S-R is reset by the *result_ready* signal delayed by one clock cycle (*reset_ready'*), so to set the system ready for the next streaming accumulation. The circuit dedicated to this task is shown in Fig. 19.2a;

- *Adder Supervisor Logic*: by checking if a new couple of inputs are presented to the adder, it notifies if any data is inside the pipeline. In addition, it signals when a sum operation has been completed and the adder output is valid. The internal logic is shown in Fig. 19.2b. The *new_input* bit signal goes high each time a new couple of operands is presented to the adder and it is used as the input of the shift

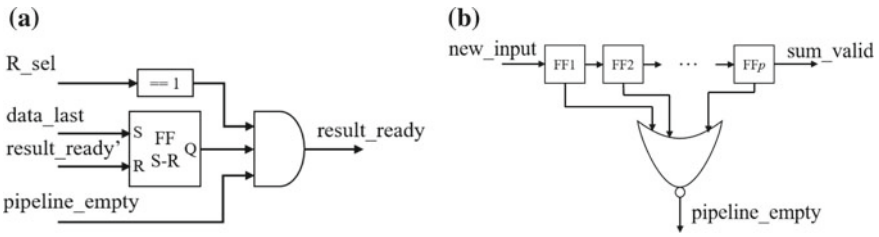


Fig. 19.2 **a** External signaling logic function; **b** Adder supervisor logic function

register represented by the FF1, FF2, …, FFp, with p the length of the internal adder pipeline. When the *sum_valid* bit goes high, p clock cycles are elapsed, meaning the addition result is ready. Moreover, if the *pipeline_empty* bit is low, means no new operands have been presented in the last p clock cycles, i.e. the internal adder pipeline is empty.

In Table 19.3, an example of the running algorithm is shown with the internal adder latency configured to be 2 clock cycles.

For simplicity, in this use case, four data elements are read, one every clock cycle, while the adder is a two-stage pipeline operator. At the cycle 0, the first element is presented. Since the adder produces a valid output only with a pair of input operands, the element is stored in the *Register*. At the next cycle, a new input data is ready and now the two operands can be pushed in the adder pipeline. The working mode repeats these steps until the first sum is generated by the adder, here at cycle 3. In this situation, the *Register* is already storing a value (X_3), so the control logic pushes into the adder the new incoming input together with the sum just generated. The *Register* is set in a hold state. At cycle 5, when a new couple of data is available, the adder is fed with the value stored in *Register* and the last generated sum. After two clock cycles (i.e. the adder pipeline latency), the final accumulation value becomes valid.

19.4 Results

The presented model has been compared with Xilinx Floating-point accumulator Intellectual Property (IP) core for FPGA implementation. To have comparable results, both architectures has been configured to have a total accumulator latency of 30 clock cycles. For the Simulink model, this means using an adder pipeline latency p of 11 clock cycles and an input streaming length n of 5 values, as found by using the DB architecture equation of Table 19.1.

In Fig. 19.3, a Simulink example of an input stream of 5 random floating-point values in the range −100 to 100 is reported.

As shown, the input flags *data_valid* and *data_last* are attached to the input stream to notify whether the value is valid and the last. After the *data_last* flag has been asserted and the whole system finishes its internal processing, the *output_ready* flag is raised for one clock cycle. This notifies the user about the result readiness.

To test the HDL Coder compatibility, a non-target-specific VHDL code generation has been carried out for an architecture based on the floating-point 32-bit format. The generate code has then been imported in Vivado software and, after synthesis and implementation elaborations for a Xilinx Artix-7 XC7A100T-CSG324 FPGA target device, results have been reported in Table 19.4. Both systems perform the same data processing: accumulation of a 32-bit floating-point input stream, with a total latency of 30 clock cycles and an input of 5 streaming values.

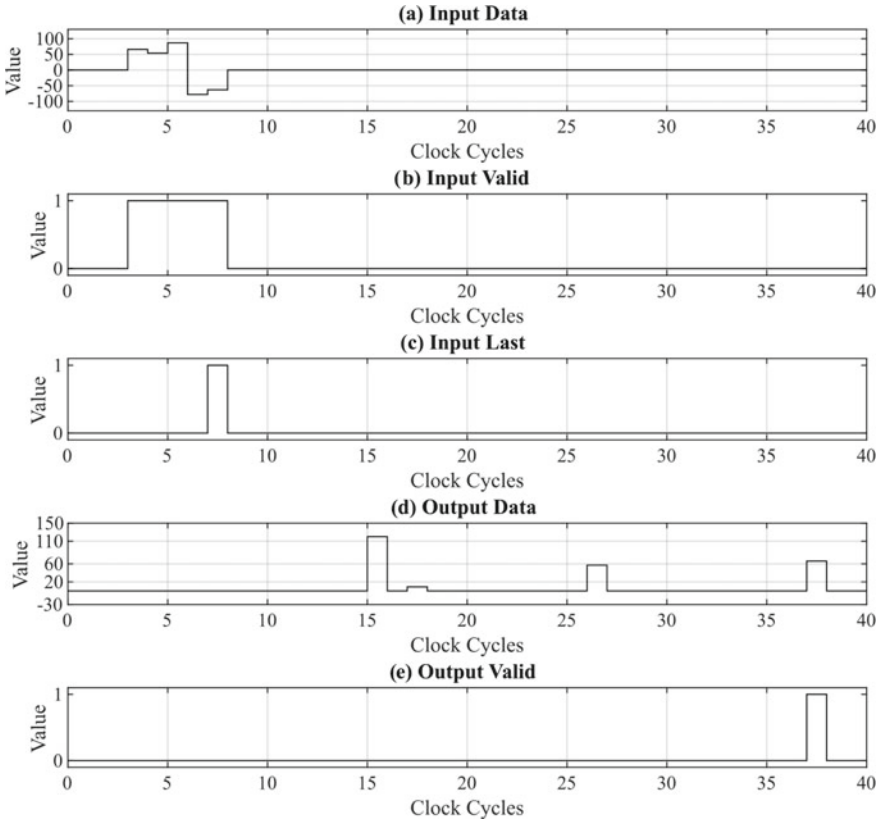


Fig. 19.3 Example of an execution of the accumulator model: **a** input data values; **b** external data valid input signal (*data_valid*); **c** external input signal to notify the last value of the set (*data_last*); **d** output value (*result*); **e** internally generated output signal to notify the output is valid (*result_ready*)

Table 19.4 Post-implementation resources usage report generated by Xilinx Vivado

	Presented accumulator	IP accumulator
Slice LUTs	635	3275
Slice Registers	723	3067

As shown, the presented model features lower resources usage then the Xilinx IP implementation. This result was expected because the internal fixed-point accumulator of the IP had to be configured to match the full data range and precision of the 32-bit floating-point format.

19.5 Conclusion

In this paper, a Simulink model-based, pipelined, and HDL Coder-compatible accumulator has been presented. The designed architecture is based on the state-of-the-art offering the lowest accumulation latency and being able to sum a data set at the clock-rate frequency. It is suitable for the integration in any design requiring this kind of arithmetic circuit, including for example SVM Machine Learning algorithms for HAR. Moreover, it can be converted to the desired HDL code for direct hardware implementation.

The behavior of the model has been verified and the full compatibility with the HDL Coder tool has been confirmed. The generated code has been imported in Xilinx Vivado software and a comparison with an IP floating-point accumulator has been performed. Results show a lower resource usage by the VDHL code generated thorough the Simulink and HDL Coder workflows.

References

1. Bassoli M, Bianchi V, De Munari I (2018) A plug and play IoT wi-fi smart home system for human monitoring. Electronics 7(9):200
2. Montalto F, Guerra C, Bianchi V, De Munari I, Ciampolini P (2015) MuSA: wearable multi sensor assistant for human activity recognition and indoor localization. Biosyst Biorobotics 11:81–92
3. Guerra C, Bianchi V, De Munari I, Ciampolini P (2015) CARDEAGate: low-cost, ZigBee-based localization and identification for AAL purposes. In: 2015 IEEE Instrumentation and Measurement Technology Conference (I2MTC)
4. Bianchi V, Bassoli M, Lombardo G, Fornacciari P, Mordonini M, De Munari I (2019) IoT wearable sensor and deep learning: an integrated approach for personalized human activity recognition in a smart home environment. IEEE Internet Things J 6(5):8553–8562
5. Gaikwad NB, Tiwari V, Keskar A, Shivaprakash NC (2019) Efficient FPGA implementation of multilayer perceptron for real-time human activity classification. IEEE Access 7(8651457):26696–26706
6. Giardino D, Matta M, Re M, Silvestri F, Spanò S (2018) IP generator tool for efficient hardware acceleration of self-organizing maps. In: International Conference on Applications in Electronics Pervading Industry, Environment and Society (APPLEPIES)
7. Hai JCT, Pun OC, Haw TW (2015) Accelerating video and image processing design for FPGA using HDL Coder and Simulink. In: 2015 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (CSUDET)
8. Michael T, Reynolds S, Woolford T (2018) Designing a generic, software-defined multimode radar simulator for FPGAs using Simulink® HDL Coder and Speedgoat real-time hardware. In: 2018 International Conference on Radar (RADAR)
9. Choe J et al (2019) Model-based design and DSP code generation using Simulink® for power electronics applications. In: 2019 10th International Conference on Power Electronics and ECCE Asia (ICPE 2019–ECCE Asia), pp 923–926
10. Perry S (2009) Model based design needs high level synthesis—a collection of high level synthesis techniques to improve productivity and quality of results for model based electronic design. In: 2009 Design, Automation and Test in Europe Conference and Exhibition (DATE '09), pp 1202–1207
11. Flynn MJ, Oberman SF (2001) Advanced computer arithmetic design

12. Nagar KK, Bakos JD (2009) A high-performance double precision accumulator. In: 2009 International Conference on Field-Programmable Technology (FPT'09)
13. Ni LM, Hwang K (1985) Vector-reduction techniques for arithmetic pipelines. IEEE Trans Comput C–34(5):404–411
14. Sips HJ, Lin H (1991) An improved vector-reduction method. IEEE Trans Comput 40(2):214–217
15. Luo Z, Martonosi M (2000) Accelerating pipelined integer and floating-point accumulations in configurable hardware with delayed addition techniques. IEEE Trans Comput 49(3):208–218
16. Wang X, Braganza S, Leeser M (2006) Advanced components in the variable precision floating-point library. In: 2006 14th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM)
17. Zhuo L, Morris GR, Prasanna VK (2007) High-performance reduction circuits using deeply pipelined operators on FPGAs. IEEE Trans Parallel Distrib Syst 18(10):1377–1392
18. Zhuo L, Morris GR, Prasanna VK (2005) Designing scalable FPGA-based reduction circuits using pipelined floating-point cores. In: 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2005)
19. Tai Y-G, Lo C-TD, Psarris K (2012) Accelerating matrix operations with improved deeply pipelined vector reduction. IEEE Trans Parallel Distrib Syst 23(2):202–210
20. Huang M, Andrews D (2013) Modular design of fully pipelined reduction circuits on FPGAs. IEEE Trans Parallel Distrib Syst 24(9):1818–1826

Chapter 20
Bitmap Index: A Processing-in-Memory Reconfigurable Implementation

M. Andrighetti, G. Turvani, G. Santoro, M. Vacca, M. Ruo Roch, M. Graziano and M. Zamboni

Abstract During the years, microprocessors went through impressive performance improvement thanks to technology development. CPUs became able to process great quantities of data. Memories also faced growth especially in density, but as far as speed is concerned the improvement did not proceed as the same rate. Processing-in-Memory (PIM) consists in enhancing the storage unit of a system, adding computing capabilities to memory cells, partially eliminating the need to transfer data from memory to execution unit. In this paper, a PIM architecture is presented for bulk bitwise operation mapped on the Bitmap Index application. The architecture is a memory array with logical computing abilities inside the cells. The array is a configurable modular architecture distributed in different banks, each bank is able to perform a different operation at the same time. This architecture has remarkable performance being faster than other solutions available in literature.

Keywords Processing-in-memory · Bitmap Index · Reconfigurable architecture

20.1 Introduction

Nowadays, data-intensive applications, such as image processing and databases ones, must process big amounts of data. This is a consequence of the speed improvement obtained throughout the years thanks to technology scaling. However, memory development did not follow the same path, resulting in a much slower performance increase. This disparity reduces the overall computing capability of the system, as memory is not able to provide data as fast as CPU demands them. This issue is known as *memory wall* or *Von Neumann bottleneck*. A possible solution to this problem is to nullify the distance between processor and memory, removing the cost of data transfer and creating a unit which is capable of storing information and performing

M. Andrighetti · G. Turvani (✉) · G. Santoro · M. Vacca · M. Ruo Roch · M. Graziano · M. Zamboni
Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
e-mail: giovanna.turvani@polito.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_20

operations at the same time. This concept is called Processing-in-Memory (PIM). There are many different approaches in literature to the Processing-in-Memory idea.

People have exploited new emerging technologies, such as NML (Nano Magnetic Logic) [3] and *Magnetic Random Access Memory* (MRAM), a non-volatile memory that uses Magnetic Tunnel Junctions (MTJs) as its basic element. Thanks to their storage and logic properties, MTJs can be used to implement hybrid logic circuits with CMOS technology ideal for a PIM architecture [7]. Another widely explored technology is *Resistive RAM*, a non-volatile memory that exploits a resistive component (metal-insulator-metal structure) to store information. ReRAM arrays are usually found in crossbar structures that enable the implementation of matrix-vector multiplication, commonly used in neural networks applications. One example of such implementation is PRIME [4], an architecture aimed at accelerating Artificial Neural Networks, which are based on operations that perfectly fit the crossbar structure. While the previous proposals shaped their approach on a particular technology, others worked on an architectural perspective, independently from the technology itself. Some tried to narrow the physical distance between memory and computation unit by stacking them on a 3-Dimensional structure, enhancing the available bandwidth by connecting the layers through True Silicon Vias [5]. Anyhow, it should be noticed that in this case even if the two units (memory and logic) are moved very close to each other, they are still distinct components. Another possible approach is to creates a system composed of an host processor surrounded by several HMC-based (Hybrid Memory Cube) units, composed of multiple memory layers stacked on a logic layer [10]. A different solution is to slightly modify the circuits controlling the memory to implement simple logic operations inside the memory array, such as Ambit [9], an in-memory accelerator which exploits DRAM technology. The DRAM array is slightly modified to perform AND, OR and NOT operations. Other examples are presented in [1, 2, 6]. Among the many proposals provided by literature, one of the best fitting representative of the PIM concept is presented in [8]. In this work the proposed architecture is a memory array where the cell itself is capable of performing several logical operations on the stored value.

In this paper, we propose a different solution of Processing-in-Memory, presenting an architecture shaped around the application of Bitmap Indexing, thus suitable for bulk bitwise operations. The proposed architecture is a memory array in which each cell is able to both store information and to be configured to execute simple logical operations such as AND, OR and XOR. The array is also distributed into banks and each bank is able to work both independently and with other banks, solving different queries, achieving flexibility and an high degree of parallelism. Since the structure is modular it can be built with as many banks as needed. The architecture synthesized is an array of 8512 kB, distributed on 16 banks. The technologies used are CMOS 45 and 28 nm. The results obtained highlight great potential as the synthesized structure can reach a maximum throughput of 2.45 Gop/s and 9.2 Gop/s for 45 nm and 28 nm respectively and it is noticeably faster than other solutions presented in literature.

20.2 The Architecture

The Processing-in-Memory paradigm requires that logic and storage elements are merged together. This paradigm is particularly suited for all those algorithms that need to perform huge amount of simple operations on data stored. To demonstrate the advantages that the PIM approach can provide, we choose to implement an architecture able to solve the Bitmap indexing problem. The Bitmap indexing is an important algorithm often used in database management systems.

The Bitmax Indexing is an algorithm used to identify, inside a database, entries that have specific characteristics. For example, inside the database of Fig. 20.1.A the query consist in the identification of how many man own a motorbike or a sport car. To reach this goal each feature is indexed using a binary representation. The gender column, for example, is divided in two sub-columns, one representing the male gender and one representing the female gender. Then each sub-column is represented using single bits. For example the first entry of the database is a female, so the M column contains '0', while the F column contains '1' (see Fig. 20.1a). Searching for a specific query inside such database means performing simple logic operations between each sub-column, as depicted in Fig. 20.1b.

In our architecture, instead of memorizing the database inside the memory following the same structure proposed in Fig. 20.1, we memorize the transpose of the matrix of bit representing the database. With this solution every row of the memory contains a column representing a specific feature. As a consequence to search for

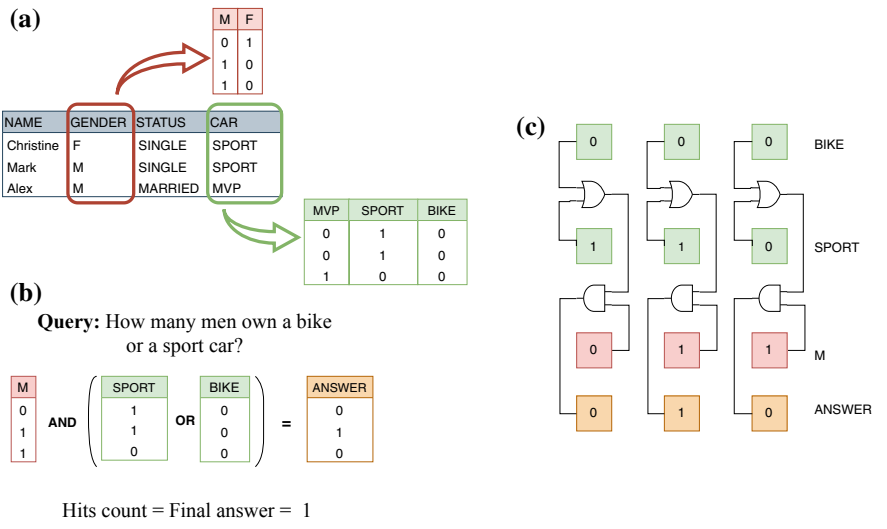


Fig. 20.1 **a** Given a table, bitmap indexing transforms each column in as many bitmap as the number of possible key-values for that column. **b** In order to answer a query logic, bitwise operations are to be performed. **c** Practical scheme of the execution of the query

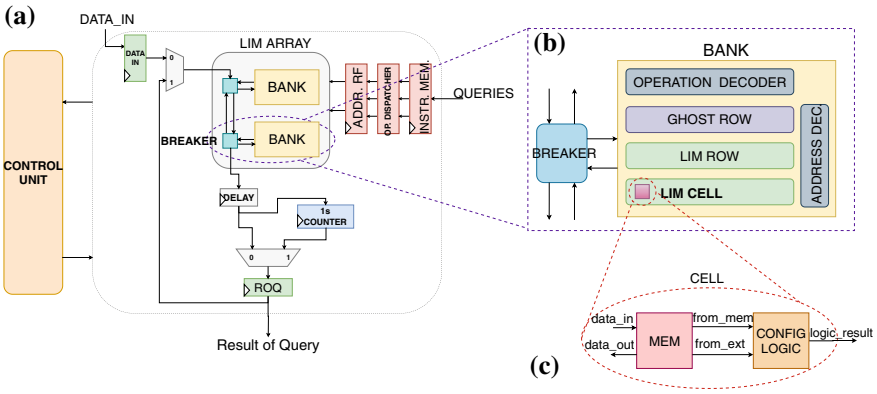


Fig. 20.2 **a** Overview of the complete architecture. **b** Structure of the duo Bank-Breaker. **c** Insight of the PIM cell

a specific query in the database it is necessary to execute logic operations between subsequent rows of the memory (Fig. 20.1c). To reach this goal we have designed a memory cell that consists of a memory element and a configurable logic block (Fig. 20.2c, more details on the implementation will be given in Sect. 20.3).

Figure 20.2 provides an overview of the complete architecture. The core part is represented by a memory storing the database. To give more flexibility to the structure the memory array is divided in banks. The circuit can be used as a standard memory if configured in that way. Otherwise it is possible to perform logic operations on stored data and to implement the Bitmap Indexing algorithm. When a query is executed all the banks in the array can eventually be activated in parallel, performing different logic operations on different rows in the bank. This is the biggest advantage of the proposed architecture because it is possible to perform a logic operation on all the data stored inside a memory bank in parallel, leading to a huge speedup in the execution of the algorithm. As depicted in Fig. 20.2a the memory array is surrounded by additional logic circuits and a control unit. For space reason we cannot describe the details of each block. The control is used to guarantee the correct execution of the algorithm according to the input queries. The instruction memory block is used to collect the queries to execute. It consists in a register file having as many registers as the number of the banks in the array. The operation dispatcher is in charge of blocking any old query. Also, since a query can take place between any couple of addresses in the array, it necessary to send the addresses to their respective bank. Thus the operation dispatcher reorders the addresses and then the address register file sends them to their own bank. As in Fig. 20.2b each memory bank contains also ghost memory rows used to store temporary results. To handle all the configuration signals needed to manage the correct execution, two decoders are needed inside each bank. The first one configures the logic operation to execute, sending it to the right row.

The second was inserted to control addresses, data flow inside the bank and select between PIM and standard memory mode. The breaker block is used to enable the communication among different banks. This structure is flexible and can be easily reconfigured to implement other algorithms.

20.3 Results and Conclusions

To evaluate the performance of the structure, a circuit composed by a 8512 kB PIM array, distributed on 16 banks with 16 bit data size, was implemented. Then, the architecture, implemented in VHDL (VHSIC Hardware Description Language), was tested with Modelsim and later synthesized with Synopsys Design Compiler using 45 nm BULK and 28 nm FDSOI CMOS technologies (Table 20.1). In this first implementation the storage elements were synthesized as latches, instead of designing a custom memory cell. As a consequence the results here presented can be greatly improved by designing a custom memory-logic cell.

Table 20.1 highlights the synthesis results. As it can be noticed the architecture is very efficient, it is capable of high clock speed but at the same time has a low power consumption.

One of the main goal this paper aimed to fulfill is the high level of concurrency. This was accomplished thanks to the internal organization of the array, that is distributed on banks which are capable of working both independently and with each other, providing flexibility in the position of the operands that are called to act in the query. To execute a simple query only one cycle is required (Table 20.2).

The maximum throughput achievable is $throughput_{max} = f_{CLK} \cdot N_{ops}$. Assuming to execute a different query in each of the 16 available banks, a maximum throughput of 2.45 and 9.2 Gop/s for 45 and 28 nm can be reached. Table 20.2 highlights the

Table 20.1 Synthesis results for 45 nm and 28 nm CMOS technologies

Parameter	45 nm	28 nm
Total area (mm^2)	2.33	1.058
f_{CLK} (MHz)	153.4	574.7
Total power (mW)	49.7	14.07

Table 20.2 Clock cycles comparison for a single query execution

	$f = A \cdot B$	$f = A \cdot (\overline{B} \cdot C)$
Pinatubo[6]	5	9
RIMPA[2]	3	5
PIMA[1]	1	3
PIM	1	2

comparison of the proposed architecture with the state of the art in terms of clock cycles required for an operation. Our architecture is always faster than the other solution proposed in literature.

It should be taken into account that even with multiple parallel operations the clock cycles required would remain constant, achieving the throughput mentioned above, meaning also that the maximum degree of parallelism reachable is equal to the number of the available banks. Moreover, thanks to its modular structure, the architecture is meant to be easily scaled to bigger dimensions and with as many banks as needed. It could also be possible to develop a 3D structure in order to increase performance. The architecture could be easily modified to implement other types of operations. In conclusion, this architecture demonstrates that a Processing-in-Memory approach leads to a great improvement of performance. The architecture here proposed achieve very good performance and has enough flexibility to be adapted to several different algorithms.

References

1. Angizi S, He Z, Fan D (2018 June) Pima-logic: a novel processing-in-memory architecture for highly flexible and energy-efficient logic computation. In: 2018 55th ACM/ESDA/IEEE design automation conference (DAC), pp 1–6
2. Angizi S, He Z, Parveen F, Fan D (2017 July) Rimpa: a new reconfigurable dual-mode in-memory processing architecture with spin hall effect-driven domain wall motion device. In: 2017 IEEE Computer Society annual symposium on VLSI (ISVLSI), pp 45–50
3. Causapruno G, Riente F, Turvani G, Vacca M, Roch MR, Zamboni M, Graziano M (2016) Reconfigurable systolic array: from architecture to physical design for NML. IEEE Trans Very Large Scale Integr (VLSI) Syst (99):1–10
4. Chi P, Li S, Xu C, Zhang T, Zhao J, Liu Y, Wang Y, Xie Y (2016 June) Prime: a novel processing-in-memory architecture for neural network computation in ReRAM-based main memory. In: 2016 ACM/IEEE 43rd annual international symposium on computer architecture (ISCA), pp 27–39
5. Kim DH, Athikulwongse K, Healy MB, Hossain MM, Jung M, Khorosh I, Kumar G, Lee YJ, Lewis DL, Lin TW, Liu C, Panth S, Pathak M, Ren M, Shen G, Song T, Woo DH, Zhao X, Kim J, Choi H, Loh GH, Lee HHS, Lim SK (2015) Design and analysis of 3d-maps (3d massively parallel processor with stacked memory). IEEE Trans Comput 64(1):112–125
6. Li S, Xu C, Zou Q, Zhao J, Lu Y, Xie Y (2016 June) Pinatubo: a processing-in-memory architecture for bulk bitwise operations in emerging non-volatile memories. In: 2016 53nd ACM/EDAC/IEEE design automation conference (DAC), pp 1–6
7. Matsunaga S, Hayakawa J, Ikeda S, Miura K, Endoh T, Ohno H, Hanyu T (2009 Apr) MTJ-based nonvolatile logic-in-memory circuit, future prospects and issues. In: 2009 design, automation test in Europe conference exhibition, pp 433–435
8. Santoro G, Turvani G, Graziano M (2019) New logic-in-memory paradigms: an architectural and technological perspective. Micromachines 10(6). https://www.mdpi.com/2072-666X/10/6/368
9. Seshadri V, Lee D, Mullins T, Hassan H, Boroumand A, Kim J, Kozuch MA, Mutlu O, Gibbons PB, Mowry TC (2017) Ambit: In-memory accelerator for bulk bitwise operations using commodity dram technology. In: Proceedings of the 50th annual IEEE/ACM international symposium on microarchitecture, pp 273–287. MICRO-50 '17, ACM, New York, NY, USA, https://doi.org/10.1145/3123939.3124544

10. Zhang D, Jayasena N, Lyashevsky A, Greathouse JL, Xu L, Ignatowski M (2014) TOP-PIM: Throughput-oriented programmable processing in memory. In: Proceedings of the 23rd international symposium on high-performance parallel and distributed computing, pp 85–98. HPDC '14, ACM, New York, NY, USA. https://doi.org/10.1145/2600212.2600213

Part V
Digital Circuits and AI Data Processing

Chapter 21
Digital Circuit for the Arbitrary Selection of Sample Rate in Digital Storage Oscilloscopes

M. D'Arco, E. Napoli and E. Zacharelos

Abstract Fine resolution selection of the sample rate is not available in digital storage oscilloscopes. They rely on offline processing to cope with such need. The paper presents an algorithm that, exploiting online processing with a digital filter characterized by dynamically generated coefficients and a memory management strategy, allows almost arbitrary selection of the sample rate from an incoming stream of samples. The paper also proposes a digital circuit implemented on FPGA to devise the possible performance of the method.

21.1 Introduction

Analogue oscilloscopes offer a discrete set of time base signals to select the time window that is analyzed. The use of a continuously variable control is possible but is in trade off with the calibration of the signal [1, 2].

In digital storage oscilloscopes (DSOs) the time base is determined by controlling the sampling rate. Again, only a discrete set of values is available [3, 4] since DSOs provide the highest sampling rate and obtain lower rates through decimation [5–7].

Flexible sample rate selection would allow more efficient usage of memory resources allowing the exact sampling rate needed for the given application. Sample rate changes can be accomplished through digital resampling approaches but the required processing power and the need of dedicated circuitry for each sampling

M. D'Arco · E. Napoli
Department Electrical and Information Technology Engineering, University of Naples Federico II, 80125 Naples, Italy
e-mail: mauro.darco@unina.it

E. Napoli
e-mail: ettore.napoli@unina.it

E. Zacharelos (✉)
Department Physics, Electronics and Electronic Computers, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
e-mail: evzachar@physics.auth.gr

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_21

rate makes this choice unfeasible [8–11]. Modern DSOs host powerful CPUs able to implement in real time: averaging, FFT spectral analysis, parameters measurements, and selection between different acquisition modes [12, 13]. Unfortunately, these CPUs cannot resample the input stream in real time.

This paper proposes a time base system that, thanks to the simplicity of its operation principle, allows fine selection of the sample rate of the digital storage scope with very fine frequency resolution up to the maximum sample rate. The proposed solution relies on a suitable memory management strategy and a dynamical digital filter.

21.2 Resampling Through Polyphase Filters

Digital resampling is common in multipurpose receivers where several different sampling rates are supported to process signals characterized by different bandwidths [14–16]. The receivers initially sample at a high sampling rate, then perform resampling by a factor, L/M, (interpolation by L, low-pass filtering, and decimation by M). Low-pass filtering removes the image frequencies; it is implemented using polyphase decomposition of both the input signal and filter coefficients.

For the sake of clarity, an example of a $\frac{3}{4}$-resampler that uses a short low pass filter with 9 coefficients, $h(n) = \{h(0), h(1), \ldots, h(8)\}$, is shown in Fig. 21.1. The input signal $y(n)$ is de-multiplexed in order to retrieve 4 consecutive samples and route them to 4 individual channels with a single operation. The output of the resampler, $z(m)$, is obtained by multiplexing the outputs produced by 3 filters, each filter defined in terms of 3 coefficients of $h(n)$ according to polyphase decomposition rules.

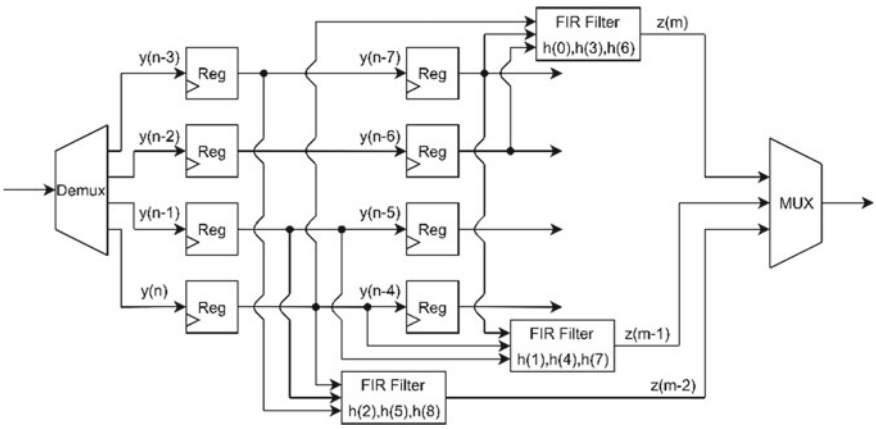


Fig. 21.1 Schematic of a digital resampler implementation based on polyphase decomposition. Resampling factor equal to 3/4 low-pass filter with 9 taps

Polyphase filters are characterized by low requirements in terms of clock frequency and can be set to both up–sample and down–sample the input stream but are not suitable for programmable resampling factors [14–17].

21.3 Proposed Resampling Algorithm

The proposed method involves the use of a digital circuit, deployed between the ADC and the acquisition memory that, depending on the chosen design parameters, allows a very fine regulation of the sample rate from half the system frequency, $1/2 \cdot f_{ck}$ up to the highest frequency, f_{ck}. The method does not lack in generality since choosing a sample rate lower than $1/2 \cdot f_{ck}$ is easily obtained by cascading the proposed circuit with a standard one that performs decimation by an integer value. It is important to highlight that the whole acquisition chain made up of ADC, digital circuit, and memory operates synchronously at the system clock rate f_{ck}.

After processing, the samples stored in the acquisition memory represent a version of the input signal resampled at a sample rate $f_s = C\, f_{ck}$, where C is an arbitrary (within limits) fractional value in the interval [1/2, 1).

21.3.1 Digital Circuit Operation

The digital circuit processes in real-time the signal $x(n)$ deriving from the ADC, and produces the output, $y(n)$. Both are produced at the highest clock rate, f_{ck}. The output is an estimation of the samples of the input signal, resampled at $f_s = C\, f_{ck}$. The value $y(n)$ is determined by combining the samples $x(n)$ and $x(n-1)$ returned by the ADC:

$$y(n) = (1 - a(n))\, x(n) + a(n)\, x(n-1) \tag{1}$$

where $a(n)$ is a time-varying coefficient, updated at every clock cycle subtracting to its current value the quantity $C^{-1} - 1$, which depends on the selection made by the user. Subtraction is skipped if the current value of the coefficient is negative, and in its place an addition by one is performed. The output of the digital circuit $y(n)$ contains, with some redundancy, the resampled version of $x(n)$. The circuit also produces a signal PTR_X, that indicates the memory location where $y(n)$ is stored. The generated sequence $y(n)$ is stored in memory at system frequency, f_{ck} but, in order to cope with the lower sampling rate, PTR_X is not incremented when the $a(n)$ coefficient is incremented by one. In this way, two consecutive outputs share the same value of PTR_X, which means that the second one overwrites the first.

An example will better clarify the meaning of $a(n)$. In Fig. 21.2 a sinusoidal signal at 54 MHz is shown. It is sampled with the 1 GHz ($T_{ck} = 1.0$ ns) system clock (sampling shown with circles). The result obtained resampling at 761 MHz

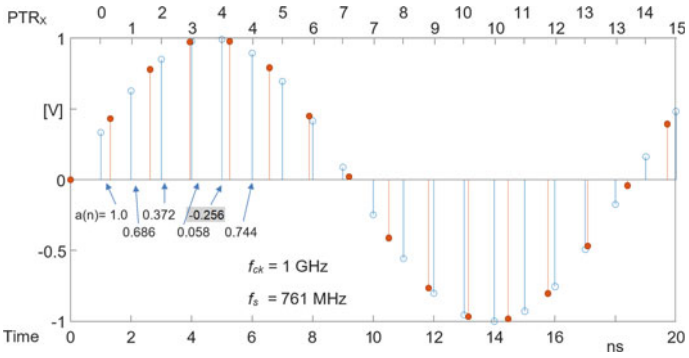


Fig. 21.2 Example sequences for *a(n)* and PTR_X

($T_s = 1.314$ ns) is shown with red bullets. The resampling factor is $C = 0.761$, and the coefficient $a(n)$ is updated subtracting $C^{-1} - 1 = 0.3141$ to the current value. $b(n) = 1 - a(n)$ represents the point inside the sampling period where resampling must be performed. The bottom axis is the time while the top axis shows the increment of the memory pointer. When *a(n)* is incremented (time: 6, 10, 14, 18 in Fig. 21.2) the memory pointer is not updated.

21.4 Performance Assessment

The proposed method suffers of a performance degradation when compared with the standard technique (zero padding, low pass, decimation). However, simulated tests with sinusoidal signals demonstrate that when the sampling clock is at least ten times higher than the signal bandwidth, the results are satisfactory.

The performances are reported in terms of standard parameters defined for a pure sine wave: signal-to-noise-and-distortion (SINAD) ratio and total-harmonic-distortion (THD). SINAD and THD are calculated: for the input signal corrupted by white Gaussian noise (rms value equal to 15% the LSB of the ADC) and quantized by and 8bit ADC; for the resampled signal.

Figure 21.3 reports the result obtained resampling at 743 MHz a 47.1 MHz signal converted with a 1GSs ADC. The original signal has SINAD = 48.49 dBc and THD = −51.50 dB. The resampled signal has SINAD = 46.97 dBc and THD = −50.04 dB showing quite limited degradation. Similar results are obtained applying 50 kHz random deviation of the input frequency.

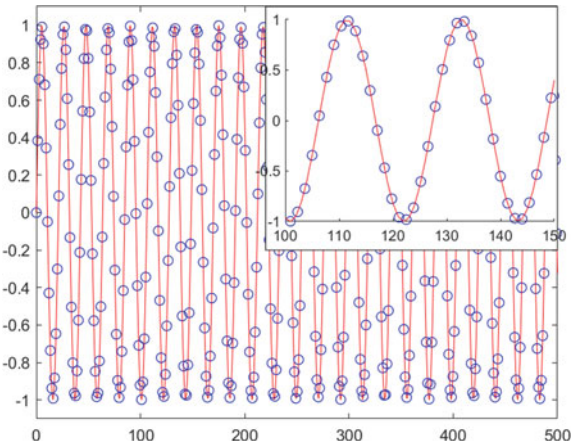


Fig. 21.3 Red line: 47.1 MHz signal digitized with an 8-bit ADC at 1 GHz sample rate. Blue circles: the same signals resampled at 743 MHz

21.5 Implementation of the Proposed Circuit

A digital circuit for the implementation of the above proposed resampling algorithm has been designed. The schematic (without pipelining) is in Fig. 21.4.

Circuit input data are the signal to be resampled x, and the resampling factor defined through the input $d = 1 - C^{-1}$. The output data are the resampled stream y, and the memory pointer, Ptr_X. The number of bits for x, d, and y is 8, while the memory pointer Ptr_X, is represented with 32 bits.

The two complementary coefficients, a and $b = 1 - a$, are multiplied by the previous value and the current value of the input signal respectively. Afterwards, the two products are summed, in order to produce the output signal, y.

The updating of the coefficient a, relies on adding either the quantity d, or in the case of exception, a unitary value to the current value of a. In the case of exception, a is negative, and the coefficient's MSB, is high, $a[9] = 1$. Otherwise, $a[9] = 0$, and d is added to the current value of a. This distinction is realized with the use of a

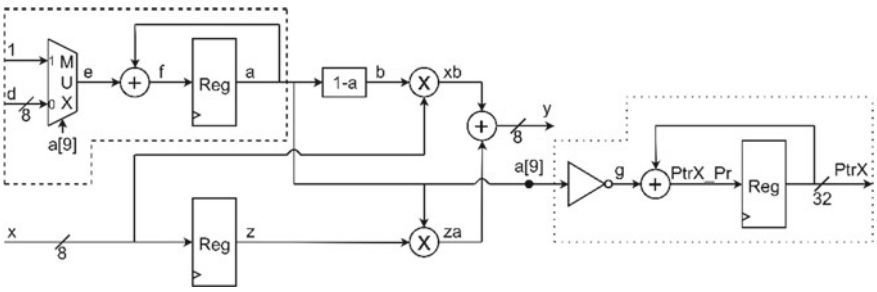


Fig. 21.4 Circuital implementation of the proposed algorithm

Table 21.1 Basic features of the resampler and FPGA resources

Name	Value
Maximum clock frequency (after pipelining)	400 MHz
Best C-step	9.76×10^{-4} (C = 0.500976)
Worst C-step	39×10^{-4} (C = 0.996094)
Combinational ALUTs	532 (<1%)
Dedicated logic registers	1432 (<1%)
DSP block 18-bit elements	3 (<1%)

multiplexer, controlled by the a's MSB. After the correct choice between "1" and "d", an accumulator is implemented for the updating of a.

A second accumulator is implemented, for the memory management. When a is positive, $a[9] = 0$, $g = 1$ and Ptr_X is incremented by a unitary value. In the case of exception, a is negative, $a[9] = 1$, $g = 0$ and Ptr_X remains unchanged.

In Table 21.1, some basic features of the circuit are presented. C-step refers to the difference between two consecutive values of the resampling factor. The limitation stems from the fact that d is represented by an 8-bit number. The resolution obtained on the resampling factor C is about 0.19%. The HDL design is implemented on a Stratix IV GX FPGA device. Table 21.1 reports the resources needed for the resampler.

21.6 Conclusion

The paper presented an algorithm and its circuital implementation, for the creation of a time base that allows fine selection of the sample rate of a digital storage scope.

The proposed algorithm shows good performances when the sampling rate, as usual, is about ten times higher than the bandwidth of the signal. The circuital implementation of the algorithm allows, as a proof of concept, to demonstrate the feasibility of the circuit and its performances when implemented on Stratix IV GX FPGA.

References

1. Oya JRG, Munoz F, Torralba A, Jurado A, Marquez FJ, Lopez-Morillo E (2012) Data acquisition system based on subsampling using multiple clocking techniques. IEEE Trans Instrum Meas 61(8):2333–2335
2. D'Apuzzo M, D'Arco M (2017) Sampling and time—interleaving strategies to extend high speed digitizers bandwidth. Measurement 111:389–396
3. Monsurrò P, Trifiletti A, Angrisani L, D'Arco M (2018) Streamline calibration modelling for a comprehensive design of ATI-based digitizers. Measurement 125:386–393

4. D'Apuzzo M, D'Arco M (2016) A wideband DSO channel based on three time-interleaved channels. JINST 11:08003. https://doi.org/10.1088/1748-0221/11/08/p08003
5. Yuan W, Jiangmiao Z, Jingyuan M (2013) Correction of time base error for high speed sampling oscilloscope. In: 2013 IEEE 11th international conference on electronic measurement & instruments, Harbin, pp 88–91
6. Angrisani L, D'Arco M, Ianniello G, Vadursi M (2012) An efficient pre-processing scheme to enhance resolution of band-pass signals acquisition. IEEE Trans Instrum Meas 61(11):2932–2940
7. D'Arco M, Genovese M, Napoli E, Vadursi M (2014) Design and implementation of a preprocessing circuit for bandpass signals acquisition. IEEE Trans Instrum Meas 63(2):287–294
8. Choi H, Gomes A, Chatterjee A (2011) Signal acquisition of high-speed periodic signals using incoherent sub-sampling and back-end signal reconstruction algorithms. IEEE Trans Very Large Scale Integr (VLSI) Syst 19(7):1125–1135
9. Kirchner M, Bohme R (2008) Hiding traces of resampling in digital images. IEEE Trans Inf Forensics Secur 3(4):582–592
10. Popescu AC, Farid H (2005) Exposing digital forgeries by detecting traces of resampling. IEEE Trans Signal Process 53(2):758–767
11. Porwal S, Katiyar SK (2014) Performance evaluation of various resampling techniques on IRS imagery. In: 2014 7th international conference on contemporary computing (IC3), Noida, pp 489–494
12. Xu T, Fumagalli A, Hui R (2018) Real-time DSP-enabled digital subcarrier cross-connect based on resampling filters. IEEE/OSA J Opt Comm Netw 10(12):37–946
13. Oya JRG, Muñoz F, Torralba A, Jurado A, Garrido A, Banos J (2011) Data acquisition system based on subsampling for testing wideband multistandard receivers. IEEE Trans Instrum Meas 60(9):3234–3237
14. Fiala P, Linhart R (2014) High performance polyphase FIR filter structures in VHDL language for software defined radio based on FPGA. In: 2014 international conference on applied electronics, Pilsen, pp 83–86
15. Johansson H, Harris F (2015) Polyphase decomposition of digital fractional-delay filters. IEEE Signal Process Lett 22(8):1021–1025
16. Laddomada M (2008) On the polyphase decomposition for design of generalized comb decimation filters. IEEE Trans Circ Syst I 55(8):2287–2299
17. Porteous M (2011) Introduction to digital resampling. RF Engines white paper lit. number 4216984, 15 June 2011. https://www.techonline.com/electrical-engineers/education-training/tech-papers/4216984

Chapter 22
An Intelligent Informative Totem Application Based on Deep CNN in Edge Regime

Paolo Giammatteo, Giacomo Valente and Alessandro D'Ortenzio

Abstract In this paper we present an application targeting an informative totem, with a discussion about its possible usage and the requirements it needs to satisfy. In this regard, we propose a Machine Learning algorithm, a Convolutional Neural Network, performing computation on images taken from a camera on an edge-computing platform. Performance tests on two different edge processors are reported, respectively for a CPU and a GPU, and a comparison with the principal competitors is provided. Our final goal is to lay the foundation for the application of an informative totem in an edge computing regime, which is able to recognize the age and the gender of the person approaching it in order to give a better presentation of its contents.

Keywords Age and gender estimation · Convolutional neural networks · Edge computing · Embedded systems

22.1 Introduction

Informative totems are tools that can provide useful information, such as finding your way around a building or buying a train ticket at the train station in few passages. With the advent of Artificial Intelligence (AI), and in particular of Machine Learning (ML) techniques, these devices can improve their performance by providing more

P. Giammatteo (✉) · G. Valente
Centro di Eccellenza DEWS, Universitá degli Studi dell'Aquila, L'Aquila, Italy
e-mail: paolo.giammatteo@univaq.it

G. Valente
e-mail: giacomo.valente@univaq.it
URL: http://dews.univaq.it/index.php?id=dewshome

A. D'Ortenzio
Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica, Universitá degli Studi dell'Aquila, L'Aquila, Italy
e-mail: alessandro.dortenzio@graduate.univaq.it
URL: https://www.disim.univaq.it/main/index.php

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_22

effective support to the user and facilitating their purpose [1]. In addition, informative totems are an example of what are nowadays called *smart-edge* devices, bringing the attention to the topic of *edge-computing* [2].

Let us consider a significant application of *smart-totems*: supposing that there is a teenage kid lost inside a shopping mall and no longer able to find his mother. He needs to find a way in order to rejoin with her, and he knows his mother could be inside a specific shop of the mall. While looking around, his gaze is caught by an informative totem that is asking him if he needs help. The kid approach to the totem and, as he touches the screen, the totem offers him some actions tailored for the situation, among which "search for a shop". The kid choose this option and the totem shows him a map with the correct path to reach the shop from his position. The kid memorizes the path and proceeds to the shop.

From this example scenario, it is possible to identify some functional requirements (FR) for applications targeting *smart-totems*, such as: (*FR1*) recognizing the age and the gender of a person that is approaching to the totem and (*FR2*) producing information basing on them. Together with functional requirements, even non-functional (NFR) ones can be identified: (*NFR1*) a system response within a certain time, possibly under real-time requirements, and (*NFR2*) the adaptation to sudden and continuous changes of physical entities with which the system interacts.

Given these requirements, the current trend to address them is represented by AI for age and gender recognition [3] and automatic adaptation and by edge-computing for the real-time response [2, 4]. However, nowadays most of AI applications are implemented on cloud: for example, considering ML [5] (one of the AI techniques to perform image computation), the ML algorithm for age and gender recognition are developed for cloud applications, not considering the limited resources of an edge-computing system.

We place our contribution within a growing research trend: the porting of ML algorithms on edge-devices (NNs) [6]. Our proposal is an informative totem able to recognize the age and the gender of a person that is coming toward it and provide a response basing on this. Our goal is to satisfy FR1, FR2, NFR1 and NFR2 above described: we developed a ML algorithm (specifically, a *neural network*, NN) that works on some images taken with a camera (representing the edge of our system), and we implemented the NN on an edge-computing platform located close to the camera. We tested the proposed application on two different edge-computing platforms, one with a CPU and one with a GPU, and we compared results with the principal competitors.

The paper is organized in the following way: Sect. 22.2 gives an overview about related works on the topic, Sect. 22.3 describes our system and experimental results, together with a discussion with other competitors. In Sect. 22.4, some conclusions and future works are reported.

22.2 Related Works

Age and gender classification are very important for advertising and marketing, but other potential uses include also automatic ticket office or informative totems. Classifying age and gender of people basing on their face image is a well known problem in academic literature: to this end, several algorithms have been proposed. An exhaustive survey on methods and approaches in age and gender estimation is given in the paper of Atallah [5], providing an overview on the issue from 2010 to 2017. From this work, it emerges that *Deep Learning* (DL), and in particular *Convolutional Neural Networks* (CNN), nowadays provide the best performance on age and gender recognition, and we witness to a gradual shifting of the use of classic ML methods to those of DL. The work in [7] presents one of the first methods adopting DL, the CNN, and it showed improved performance compared with traditional feature-based methods [8], such as *Support Vector Machines* (SVM).

In the context of cloud-computing, there are several private companies that provide this service by calling an API in cloud, such as Google [9], Amazon [10] and Sighthound [11]. The latter also contributed to academic literature with the paper [12], with a 61% of accuracy in age estimation.

On the other hand, in the context of edge-computing devices, there have been implementations of age and gender recognition algorithms with ML, especially DL. In the commercial field it is possible to find some applications that address this issue: Axis enterprise proposes the system called Demographic Identifier [13]. Pyramics does the same with Pysense [14]. The latter, in particular, exploits the age and gender recognition software developed by Fraunhofer IIS [15] for embedded platforms, which was also used for other embedded platforms respect to the one considered by Pysense.

At the best of our knowledge, there are few implementations with CNN in academic literature. Azarmehr [16] proposes one of the most significant approaches using an SVM algorithm, implemented on a quad-core Snapdragon 600. On the other hand, Chen [17] addressed the problem only for the gender recognition using a CNN, executed on a custom architecture implemented on FPGA. Irick [18] also reported an Artificial Neural Network (ANN) based system executed on an architecture implemented on an FPGA, that achieves an accuracy of 83.3%, roughly processing 30 images per second.

A schematic summary is reported in Table 22.1.

Table 22.1 Paper comparison

Author	Title	Scope
Axis	Demographic identifier [13]	Commercial
Pyramics	Pysense [14]	Commercial
Fraunhofer IIS	Shore [15]	Commercial
Azarmehr	Real-time embedded... [17]	Academic
Chen	Hardware/software... [18]	Academic

22.3 Application and Results

In this section, we present our system. Firstly we present our goal, and then we move to description of the system components. Then, the performed tests are shown and a final discussion on results is reported.

22.3.1 The Goal and System Description

The main idea we want to present in this paper is the transfer of a CNN, previously developed [19], able to recognize the age and gender from an image of an individual, on edge devices, in order to test the relative performance of the final system. We conceived a comparison between a CPU and a GPU edge device(s), performing the classification made by the CNN, by observing how much execution time is needed to accomplish it at edge conditions for both devices.

In this way, we want to lay the foundations for a more detailed study of the problem of moving ML algorithms, in particular as NNs, for the recognition of age and gender of individuals on edge devices, which notoriously possess greater constraints of computational resources if compared with cloud devices. The CNN and the edge devices used for comparison are described in the following paragraphs.

The proposed neural network The considered NN is a CNN, in particular a VGG16-like from an architectural point of view [19]. Our algorithm classifies an individual image in ten different classes, binding together the information of age and gender, in order to have one NN able to perform the age/gender prediction, limiting as much as possible the memory occupation. Indeed, by owning two networks that separately perform age and gender prediction, would increase the occupied memory space, which is a non-trivial aspect for an edge-computing device. In particular, our solution occupies approximately 600 MB, and it is able to recognize people according to the classification method defined in [19], with an accuracy of 40% and an off-by-1 accuracy of 70%. Nevertheless, our attention, is currently focused on the inference process of the CNN, rather than on the question of training phase, already addressed in paper [19]. Therefore, our interest is the time performed by the CNN in doing a prediction.

The CNN is written in Python 3.6, exploiting the ML libraries Tensorflow and Keras. Further detail on the CNN are reported in [19].

The edge device The Edge-Computing revolution makes it necessary to seek alternatives to the use of low-profile microcontrollers, as it has been traditionally done in wireless sensor networks. When algorithms become more computing intensive, architectures over classical CPUs, such as GPUs and circuits implemented on FPGAs, can prove beneficial when used as processing platforms. Moreover, System-on-Programmable Chips (SoPCs), integrating FPGAs with microcontrollers on the same device, allow combining the flexibility of software with the performance of hardware.

In this paper, we consider the comparison between a CPU and a GPU edge platform. In particular, the accounted device is the Nvidia Jetson Nano board [20]. This platform represents a valid solution considering the prospective of transferring ML applications to edge-computing devices. It consists of a Quad-core ARM® Cortex®-A57 MPCore CPU, together with a Nvidia Maxwell™ GPU architecture, with 128 Nvidia CUDA® cores and a RAM memory size of 4 GB. The comparison is made on the same board, performing the classification of our CNN firstly on the ARM® processor and then on the GPU, with the aim of obtaining the execution times, of the same ML algorithm, on both architectures.

22.3.2 Results

We executed our CNN algorithm on both processing element of the same Nvidia Jetson Nano board, firstly on the ARM, then on the GPU. The results obtained are shown in Table 22.2.
As expected, the ARM provides a worse performance than the GPU, emphasizing the importance of using hardware accelerators also at the edge.

22.3.3 Discussion

Results shown in Sect. 22.3.2 are preliminary and further refinements are needed in order to get better timing performance. As mentioned in Sect. 22.2, our direct competitors are Demographic Identifier [13] and Pysense [14], which propose a commercial solution with an application oriented to retail. However, Pysense exploits the recognition software developed by the Fraunhofer IIS [15], which provides further results of its software on other edge-computing platforms. Finally, from the academic literature, we consider the paper [16]. We summarized all these information, publicly available, in Table 22.3, where we compare our solution with the features of the competitors.

Looking at the table, it can be seen that not all information are available. Despite this observation, our solution compared to others still maintains attractiveness. At the moment, it is the only one that contemplates a single algorithm for the estimation

Table 22.2 Processor comparison

Processor	Execution time (s)	Frame per second (fps)
ARM Cortex A57	10.08	0.10
Jetson Nano 128 CUDA Core	0.13	7.76

Table 22.3 Competitor comparison

Hardware platform	Timing performance (fps)	Age accuracy (%)	Gender accuracy (%)	Image size (pixel)	Power dissipation (W)	Software used for estimation
Artpec-6 2 core ARMv7 [13]	–	–	–	1920 × 1080	–	–
Snapdragon 805 4 core ARMv7 [14]	8.92	–	94.30	1280 × 720	–	–
Jetson TX2 256 core CUDA [15]	29.40	–	94.30	1280 × 720	15.0	–
Snapdragon 600 4 core ARMv7 [16]	20.00	83.87[a]	95.79	1280 × 720	–	Two for age and gender (SVM)
Cyclone V FPGA [17]	20.73	-	97.20	32 × 32	–	Only for gender (CNN)
Jetson Nano 128 core CUDA (Us)	7.76	39.40[b]	–	1280 × 720	10.0	One both for age and gender (CNN)

[a]This percentage refers to the average between the female and male results
[b]This percentage refers to age and gender together because the CNN classification is bound, see [19] for the classes details

of both age and gender, so this means less memory space occupied on the edge device; furthermore, it exploits a Deep Learning (DL) approach, unlike the approach proposed by Azarmehr [16], which uses an SVM algorithm. CNN methods generally outperform SVM methods, this because it is known that deep learning performs well when large training sets are being used [21]. At this regards, Chen [17] provides an application on FPGA with discrete results, but its algorithm is only able to recognize the gender of an individual.

About Axis solution [13], we do not have information, so a consistent comparison is not possible. For the Pyramics solution [14], and so the Fraunhofer IIS software [15], we have no information about the age and gender prediction software used. The timing performance are better in their case, but we have no percentage about age estimation.

The terms of comparison with competitors are still unclear, due the limited availability of information on the two requirements considered (Age and Gender recognition on edge devices). Indeed, in some cases, the other solutions do not report details on the ML algorithm used for age and gender recognition. Others, on the other hand,

do only recognition of age or gender separately. Apparently, we are at the beginning of a study that gives a more general look at this particular issue. From our part, there is still work to do in CNN optimization and percentage prediction. Our aim is to increase the timing performance as well as the prediction accuracy, keeping in consideration the hardware constraints of the edge platform we are going to use, a sort of hardware and NN architecture co-design. Our attention is also focused on testing our solution on FPGA-based platforms, exploring the flexibility opportunities given by this kind of devices [22].

22.4 Conclusions

In this paper we presented our idea targeting an informative totem, its possible usage and the requirements it needs to satisfy. We presented our solution implementing a CNN on an edge device. Tests on timing performance have been done and a comparison with principal competitors is reported. The latter has not been easy due to the fact that not all the information are available for each competitor. Our solution still lacks of accuracy but we foresee to improve it according to the edge platform we are going to use. A hardware/software co-design of the entire system is required, taking into account the CNN architecture, the hardware and compression techniques for the NN. Surely, in future, we will widen the comparison with FPGA-based platforms.

References

1. Di Mascio T, Gennari R, Melonio A, Tarantino L (2014) Engaging New users into design activities: the TERENCE experience with children. In: Smart organizations and smart artifacts, pp 241–250
2. Satyanarayanan M (2017) The emergence of edge-computing. Computer 50(1):30–39
3. Shi W, Cao J, Zhang Q, Li Y, Xu L (2008) Artificial intelligence techniques: an introduction to their use for modelling environmental systems. Math Comput Simul 78:379–400
4. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) Edge-computing: vision and challenges. IEEE Intern Things J 3:637–646
5. Atallah RR, Kamsin A, Ismail MA, Abdelrahman SA, Zerdoumi S (2018) Face recognition and age estimation implications of changes in facial feature: a critical review study 6:28290–28304
6. Li H, Ota K, Dong M (2018) Learning IoT in edge: deep learning for the internet of things with edge-computing. IEEE Netw 32–1:96–101
7. Levi G, Hassner T (2015) Age and gender classification using convolutional neural networks. In: 28th IEEE conference on computer vision and pattern recognition (CVPR), pp 34–42, IEEE Press, Boston
8. Eidinger E, Enbar R, Hassner T (2014) Age and gender estimation of unfiltered faces. IEEE Trans Inf Forensics Secur 9:2170–2179
9. Google Vision API https://cloud.google.com/vision/?source=post_page
10. Amazon Rekognition https://aws.amazon.com/it/rekognition/?source=post_page
11. Sighthound Recognition API https://www.sighthound.com/products/cloud
12. Dehghan A, Ortiz EG, Shu G, Masood SZ (2017) DAGER: deep age, gender and emotion recognition using convolutional neural networks. arXiv:1702.04280

13. AXIS, Demographic Identifier, https://www.axis.com/it-it/products/axis-demographic-identifier
14. Pyramics Pysense https://pyramics.com/en/products/
15. Fraunhofer IIS Shore. https://www.iis.fraunhofer.de/en/ff/sse/ils/tech/shore-facedetection.html
16. Azarmehr R, Laganire R, Lee WS, Xu C, Laroche D (2015) Real-time embedded age and gender classification in unconstrained video. In: 28th IEEE conference on computer vision and pattern recognition (CVPR), pp 57–65, IEEE Press, Boston
17. Chen ATY, Biglari-Abhari M, Wang KIK, Bouzerdoum A, Tivive FHC (2016) Hardware/software co-design for a gender recognition embedded system. In: International conference on industrial, engineering and other applications of applied intelligent systems (IEA/AIE), pp 541–552, Morioka
18. Irick K, DeBole M, Narayanan V, Sharma R, Moon H, Mummareddy S (2007) A unified-streaming architecture for real-time face detection and gender classification. In: International conference on field programmable logic and applications, pp 267272. IEEE Press, New York
19. Giammatteo P, Fiordigigli FV, Pomante L, Di Mascio T, Caruso F (2019) Age & gender classifier for edge computing. In: 2019 8th mediterranean conference on embedded computing (MECO), IEEE Press, Budva
20. Nvidia Jetson Nano Developer Kit. https://developer.nvidia.com/embedded/jetson-nano-developer-kit
21. Lemley J, Abdul-Wahid S, Banik D, Andonie R (2016) Comparison of recent machine learning techniques for gender recognitionfrom facial images. In: 27th modern artificial intelligence and cognitive science conference (MAICS), pp 97–102, Dayton
22. Meloni P, Capotondi A, Deriu G, Brian M, Conti F, Rossi D, Raffo L, Benini L (2018) NEURAghe: exploiting CPU-FPGA synergies for efficient and flexible CNN inference acceleration on Zynq SoCs. ACM Trans Reconfigurable Technol Syst 11:18:1–18:22

Chapter 23
FPGA-Based Clock Phase Alignment Circuit for Frame Jitter Reduction

Dario Russo and Stefano Ricci

Abstract Frame jitter occurs when the delay between a trigger and the start of a signal acquisition or signal generation is different among subsequence data frames. Test bench waveform signal generators features low frame jitter (e.g. 400 ps rms), but this performance is still insufficient for the instrument to be used in sensitive applications like Doppler velocimetry. In this work a circuit is presented that synchronizes on-the-fly an internal clock to every occurrence of an external trigger. It is implemented in a Field Programmable Gate Array (FPGA) and features a frame jitter lower than 100 ps rms.

23.1 Introduction

Frame jitter occurs in instruments or systems that acquire or produce frames of data synchronized by a trigger [1]. This is the case, for example, a waveform function generator that produces sinusoidal bursts triggered by an external pulse sequence. Small temporal differences between the trigger active edge and the actual start of the burst generation represent the frame jitter.

Applications like interferometric radar [2] or Doppler velocimetry are quite sensitive to this problem. For example, in ultrasound Doppler for biomedical [3] or industrial velocimetry [4], bursts of ultrasounds are transmitted every Pulse Repetition Interval (PRI). The target produces an echo whose phase changes depending on its position among subsequent PRIs. Target velocity is detected by reading the phase changes that occur in subsequent data frames (PRIs). Unfortunately, frame jitter alters directly the signal phase, affecting the accuracy of the velocity measurement.

Test bench function generators like, e.g. 33612A from Keysight Technologies Inc. (Santa Rosa, CA, USA) features a frame jitter as low as 320 ps rms. However, a jitter

D. Russo (✉) · S. Ricci
Information Engineering Department, University of Florence, 50123 Florence, Italy
e-mail: dario.russo@unifi.it

S. Ricci
e-mail: stefano.ricci@unifi.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_23

lower than 100 ps rms is desirable in most of the aforementioned applications, so the use of test-bench instrumentation can be not feasible.

In this paper a resynchronization circuit is presented that produces a clock whose phase is synchronized on-the-fly to an input trigger. The synchronization occurs for every trigger pulse. The generated clock can be used to generate/acquire data frames with low frame jitter. The circuit is implemented in a Field Programmable Gate Array (FPGA) of the Cyclone III family (Altera-Intel, Santa Clara, CA USA) [5]. Next section describes the architecture of the proposed circuit, and in Sect. 23.3 experiments show how the proposed circuit limits the frame jitter below 100 ps rms.

23.2 Architecture of the Synchronizer Circuit

System A and System B work with the independent and asynchronous clocks clk_A and clk_B, respectively (see Fig. 23.1). System A generates periodic events to System B signaled by the active edge of the Sync signal. Every time an edge on Sync is received, the Phase Alignment Circuit (PAC), embedded in the FPGA of the System B, tunes the phase of the clk_S to the phase of the trigger.

System B exploits clk_S for generating and/or acquiring data with low frame jitter with respect to the trigger.

The proposed architecture of the PAC is sketched in Fig. 23.2. A Tapped–Delay-Line (TDL), typically employed in Time-to-Digital converters [6], performs a fine measurement of the temporal delay between the Sync signal and the clk_{TDL} signal. The delay is represented by the number N of delay elements crossed in the TDL (see following Sect. 23.2.1) by Sync before the clk_{TDL} edge occurs. N is represented by a thermometric code, converted in a binary value by the following encoder. The Calibration RAM (C-RAM) stores at address N the number of phase steps necessary for the correction. The clk_S phase is tuned by accessing the Phased Locked Loop (PLL) through the Phase Shift Control interface [5]. This operation affects the phase of clk_S only, thus the PLL never loses the lock condition. The Calibration Unit (CU) populates the C-RAM once at system switch-on. Details of the mentioned blocks are given in the following sections.

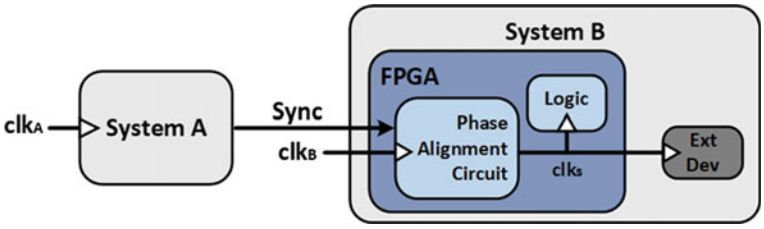


Fig. 23.1 Configuration setup of the proposed system

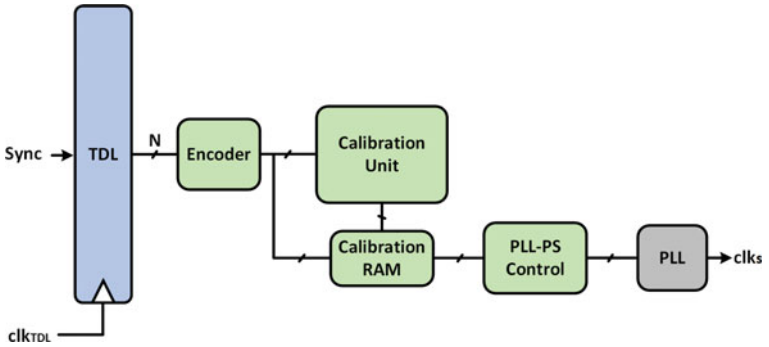


Fig. 23.2 Architecture of the proposed phase alignment circuit

23.2.1 Tapped Delay Line (TDL)

A TDL consists of a set of delay elements followed by registers, as shown in Fig. 23.3. Each pair delay element-register represents a TDL Cell and returns the status ("Cell Status" register) of that cell at clk_{TDL} rate [7]. The TDL is fed by the Sync signal that propagates in the delay line as sketched on the left of the Fig. 23.3. At time t_0, a Sync edge enters the TDL and crosses the delay elements as it propagates (t_1, …, t_n), leading to a variation of the Cell Status register. At the first rising edge of clk_{TDL} after the Sync edge fed the delay line, the Cell Status register represents the number of elements crossed by Sync edge. In particular, the phase information is stored in the position of the transition 0-to-1 in the bits of the register, that is detected by the following encoder. In the example shown in Fig. 23.3 the clk_{TDL} edge stops

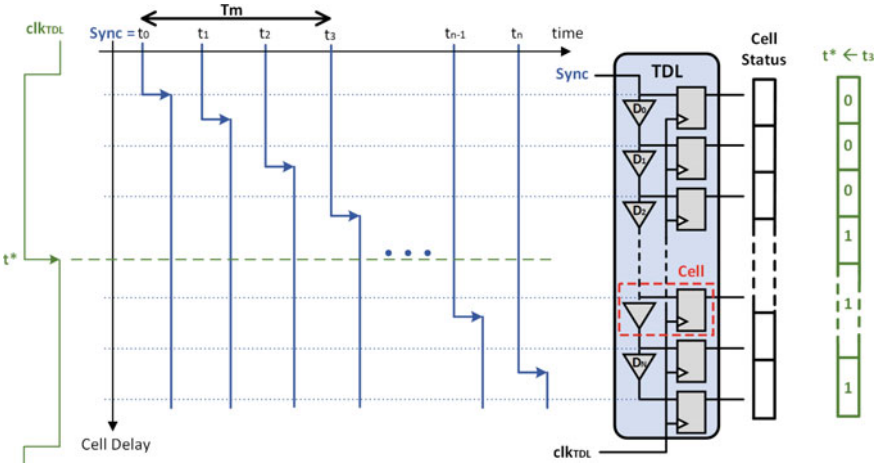


Fig. 23.3 Example of signal propagation in the TDL structure

the register sampling at t_3, after Sync crossed 3 delay elements. The delay T_m is quantified given the delay of each Cell, which is obtained through the calibration process detailed in next section.

The FPGA implementation of a TDL requires a deep knowledge of the target device architecture and of the tools for constraining the physical placement. The realization of "small" and harmonized delay elements (order of tens of ps) is the main issue. A typical solution consists in exploiting the "carry" logic normally used to realize adders, counters, etc. [7]. Indeed, the carry routing paths are more matched with each other than the general routing paths of the FPGA, and grant delays below 70–80 ps, depending on the target device. The carry logic can be used by realizing an adder and by forcing its inputs to "0" and "1" so the output is dependent on the adder carry-in value only. Then, a N-bit adder realizes a N-Cell TDL. Moreover, each bit of the adder can be implemented in a Logic-Element (LE), which is the basic unit of the Cyclone III FPGA, that includes a register (FF) as well. The latter must be used as Cell register to reducing the path between adder and register, minimizing the skew between the outputs of the Cells. In the Cyclone III device, the LEs are grouped into groups of 16 called Logic Array Block (LAB) [5]: to realize a N-Cell TDL with N > 16, more LABs are necessary. Specific constraints should be set in "Design Partition Planner" and "Chip Planner" tools of Quartus II software to direct the fitter to use consecutive LABs that have a carry delay similar to that among LEs. Constraints are also given so that the fitter will use LUT and FF of the same LE to implement each single TDL Cell [5].

All these considerations let to implement a reliable and reproducible structure that can't be obtained with the typical FPGA design flow, where the fitter is free to place the logic according to general optimization strategies.

23.2.2 Calibration Unit (CU)

In order to know the delay associated to each Cell of the TDL a calibration is required. Moreover, the calibration compensates for the TDL delay deviations due to temperature and power supply variations. There are two different calibration processes: "double registration" and "statistical" calibration [8]. The first approach is the simplest, but only the mean value of Cell delays is estimated. Although this is not sufficient for an accurate phase measurement, it is useful for an initial estimation of how many delay cells are required in the TDL:

$$N_{TDL} = \frac{t_{TDL}}{t_{cell}^{mean}} \tag{1}$$

where t_{TDL} is the period length of clk_{TDL}.

The second calibration process, i.e. the statistical calibration, lets to estimate the delay of each single Cell. It is based on a Code-Density-Test [8], where several thousands of Sync input hits, evenly spread in a time equal to t_{TDL}, feed the delay

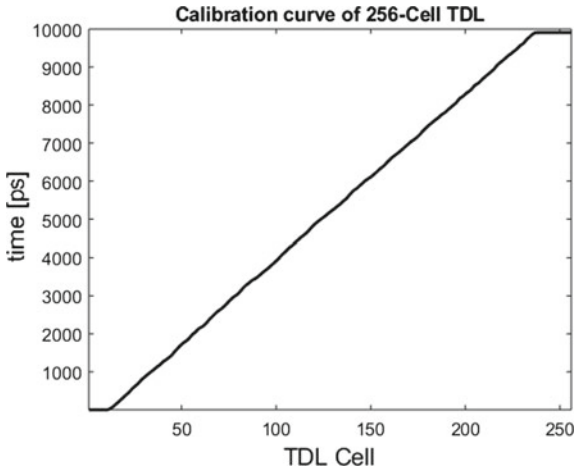


Fig. 23.4 Calibration curve of a 256-Cell TDL implemented in a Cyclone III FPGA

line. This approach results in a "calibration curve", like the one shown in Fig. 23.4. This procedure is implemented in the Calibration Unit of Fig. 23.2, which stores the calibration curve in the C-RAM.

23.2.3 PLL-PS Control

The fine phase measurements performed by the TDL and converted by the encoder are used to tune an Altera-Intel PLL. The PLL allows the dynamic shift of the phase of its outputs clocks relative to the reference. This is achieved through the "dynamic phase shifting" interface [5]. The phase shifting is performed by steps whose resolution depends on the voltage controlled oscillator frequency f_{VCO}:

$$res_{shift} = \frac{1}{8 \cdot f_{VCO}} \tag{2}$$

The "PLL PS Control" block of Fig. 23.2 connects to the phase shift interface. The commands for the shift steps (step up/down, output selection, step strobe) are serialized with the interface clock, clk_{PSI}. Each step requires 5 clock cycles of clk_{PSI}, which corresponds to 50 ns for $clk_{PSI} = 100$ MHz. For example, a phase rotation of 20 steps takes 1 μs. This time can be reduced by rising the clk_{PSI} frequency.

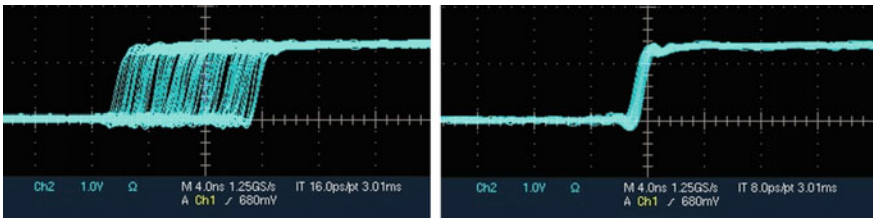


Fig. 23.5 High persistence display of scope during jitter measurements. Tests were done with the re-phasing circuit was not active (left) and active (right). Time scale is 4 ns/division

23.3 Experiments and Results

The proposed circuit was implemented in the Cyclone III FPGA of the house-made board [9]. The circuit included a 256-Cell TDL working with clk_{TDL} of 100 MHz. The VCO frequency was set to 600 MHz, corresponding to a phase step $res_{shift} =$ 208 ps. The clk_{PSI} was 100 MHz, thus in the worst case the phase was aligned in 2.4 μs after the Sync pulse. The rephrased clock, i.e. clk_{S}, was 100 MHz as well.

The "double registration" was performed to assess the mean delay of the Cells, which resulted in $t_{cell}^{mean} = 45$ ps. Being the TDL constituted by 256 Cells, the total delay was $256 \cdot 45$ ps $= 11.52$ ns, suitable to cover the 10 ns of the clk_{TDL} period. For the experiment the circuit was connected to the function generator 33612A (Keysight Technologies Inc. Santa Rosa, CA, USA) and the oscilloscope TDS5104 (Tektronix, Inc. Beaverton. OR, USA). In particular, the function generator produced a pulse every 1 ms connected to the Sync input of the proposed circuit and the trigger input of the scope. A pulse generated by the proposed circuit from the re-phased clk_{S} was visualized and acquired by the scope, triggered by the original pulse. Figure 23.5 shows on the left the output pulse when the resynchronization circuit was not enabled, i.e. $\text{clk}_{\text{S}} = \text{clk}_{\text{TDL}}$. As expected the positive edge position varies with respect to the trigger in a 10 ns range, i.e. the period of the sampling clock. Once the resynchronization is enabled, the range of variation of the pulse edge reduces significantly, like shown on the right of Fig. 23.5. In this last case, the jitter measured was less than 90 ps rms.

23.4 Conclusion

The proposed circuit is able to dynamically adjust the phase of an internally generated clock to the rising edge of an input trigger. The tuning occurs at every trigger pulse. The rephrased clock features a jitter lower than 100 ps, making the circuit suitable for sensitive applications like, for instance, Doppler velocimetry [10] or Time of Flight (ToF) measurements [11].

References

1. Kalashnikov AN, Challis RE, Unwin ME, Holmes AK (2005) Effects of frame jitter in data acquisition systems. IEEE Trans Instrum Meas 54(6):2177–2183. https://doi.org/10.1109/TIM.2005.858570
2. Pieraccini M, Miccinesi L (2019) Ground-based radar interferometry: a bibliographic review. Remote Sens 11(9):1029. https://doi.org/10.3390/rs11091029
3. Ricci S, Ramalli A, Bassi L, Boni E, Tortoli P (2018) Real-time blood velocity vector measurement over a 2D region. IEEE Trans Ultrason Ferroelect Freq Control 65(2):201–209. https://doi.org/10.1109/TUFFC.2017.2781715
4. Birkhofer B, Debacker A, Russo S, Ricci S, Lootens D (2012) In-line rheometry based on ultrasonic velocity profiles: comparison of data processing methods. Appl Rheol 22(4):44701. https://doi.org/10.3933/ApplRheol-22-44701
5. Cyclone III Device Handbook, CIII 5V1-4.2, Altera Corp (2012)
6. Roberts GW, Ali-Bakhshian M (2010) A brief introduction to time-to-digital and digital-to-time converters. IEEE Trans Circ Syst II-Express Briefs 57(3):153–157. https://doi.org/10.1109/TCSII.2010.2043382
7. Dadouche F, Turko T, Uhring W, Malass I, Dumas N, Le Normand JP (2015) New design-methodology of high-performance TDC on a low cost FPGA targets. Sens Transducers J 193(10):123–134
8. Wu J (2010) Several key issues on implementing delay line based TDCs using FPGAs. IEEE Trans Nucl Sci 57(3):1543–1548. https://doi.org/10.1109/TNS.2010.2045901
9. Ricci S, Meacci V, Birkhofer B, Wiklund J (2017) FPGA-based system for in-line measurement of velocity profiles of fluids in industrial pipe flow. IEEE Trans Ind Electron 64(5):3997–4005. https://doi.org/10.1109/TIE.2016.2645503
10. Ricci S, Vilkomerson D, Matera R, Tortoli P (2015) Accurate blood peak velocity estimation using spectral models and vector doppler. IEEE Trans Ultrason Ferroelect Freq Control 62(4):686–696. https://doi.org/10.1109/TUFFC.2015.006982
11. Marino-Merlo E, Bulletti A, Giannelli P, Calzolai M, Capineri L (2018) Analysis of errors in the estimation of impact positions in plate-like structure through the triangulation formula by piezoelectric sensors monitoring. Sensors 18(10):E3426. https://doi.org/10.3390/s18103426

Chapter 24
Real-Time Embedded System for Event-Driven sEMG Acquisition and Functional Electrical Stimulation Control

Fabio Rossi, Ricardo Maximiliano Rosales, Paolo Motto Ros and Danilo Demarchi

Abstract The analysis of the surface ElectroMyoGraphic (sEMG) signal for controlling the Functional Electrical Stimulation (FES) therapy is being widely accepted in the active rehabilitation field due to the high benefits in the restoration of functional movements for subjects affected by neuro-muscular disorders. Portability and real-time functionalities are major concerns, and, among the others, two correlated challenges are the development of an embedded system and the implementation of lightweight signal processing approaches. In this respect, the event-driven nature of the Average Threshold Crossing (ATC) approach, considering its high correlation with the muscle force and the sparsity of its representation, could be an optimal solution. In this paper we present an embedded ATC-FES control system equipped with a multi-platform software featuring an easy-to-use Graphical User Interface (GUI). The system has been tested on 5 healthy subjects in order to test its effectiveness: we obtained a correlation coefficient value of 0.86 ± 0.07, as similarity index between the healthy movement and the stimulated one during the elbow flexion exercise.

Keywords Surface Electromyography · Event-driven · Functional Electrical Stimulation · Embedded system

24.1 Introduction

Nowadays, an increasing number of active rehabilitation techniques are moving to the *bio-mimetic* approach, which relies on the analysis of the surface ElectroMyoGraphy (sEMG) signal for, e.g., the application of Functional Electrical Stimulation

F. Rossi (✉) · R. M. Rosales · D. Demarchi
Dipartimento di Elettronica, Politecnico di Torino, Turin, Italy
e-mail: fabio.rossi@polito.it

D. Demarchi
e-mail: danilo.demarchi@polito.it

P. Motto Ros
Electronic Design Laboratory (EDL), Istituto Italiano di Tecnologia (IIT), Genoa, Italy
e-mail: paolo.mottoros@iit.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_24

(FES) [1], with the aim of physiologically control the muscle functional restoration as much as possible [2]. In particular, FES employs low energy current pulses to modulate the muscle contraction [3] following this approach: a complex stimulation pattern, useful to activate the group of muscles involved in a movement, is regulated by sEMG envelope evaluation or by muscle force indicators (e.g., RMS, ARV) [4].

In a practical application, the sEMG processing and FES control is a fundamental task to be carried out in real-time [5]. Since the run-time performance bottleneck could be easily related to the use of a general purpose computer for the FES control (often concurrently running, or loaded with, many other unrelated applications or functionalities, leading to unpredictable performances), here the idea is to replace it with a dedicated embedded system. In this regard, major concerns will be the effectiveness and safety of the stimulation and the resulting performances, i.e., a latency short enough to fulfill the real-time constraints and the quality of the stimulated movement.

We propose an embedded bio-mimetic FES system based on the Average Threshold Crossing (ATC) event-driven technique applied to the sEMG signal. The ATC, which essentially compares the sEMG signal with a threshold, enables the implementation of a low-complexity on-board feature extraction process directly in hardware [6, 7], able to support, e.g., the recognition of different gestures [8]. The minimal data size of the ATC information [7] and its sparsity (due to its event-driven nature) perfectly matches the low computational capabilities of an embedded system. Evolving from the architecture presented in the previous work [7], with the aim of making the system portable and improving the run-time performance, we replaced the personal laptop, and the software based on the Matlab® and Simulink® environment, with a Raspberry Pi 3 B+ as the processing and control core of the system, running a multi-platform software. Its main tasks are the management of the sEMG multi-channel wireless acquisition, the computation and update of the FES parameters from the ATC data, and the safe control of the stimulator. The software features a Graphical User Interface (GUI) as well, to monitor and control every aspect of the system, eventually guiding the user into setup different stimulation sessions.

24.2 System Architecture

24.2.1 Hardware

The developed system represented in Fig. 24.1 can be conceptually divided into three main parts: the sEMG acquisition modules and the articular electro-goniometers as inputs, the Raspberry Pi acting as central control and processing unit, and the FES stimulator. The sEMG acquisition can be performed using two different types of device, depending on the application-case: we provide a complete four-channels board (*a*), suitable for multiple-muscle monitoring on the same limb, or four single sEMG modules (*b*), that can be employed individually or in group on different body regions.

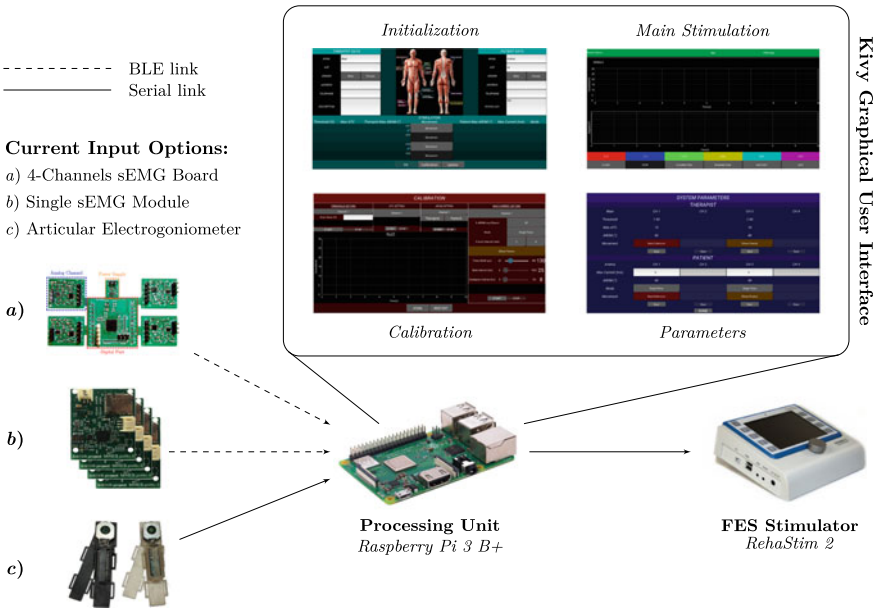


Fig. 24.1 System hardware and graphical user interface architecture

In both cases the ATC is implemented in hardware, using the standard window of 130 ms [6], and the data are wirelessly transmitted via Bluetooth Low Energy (BLE) to the Raspberry Pi. Moreover, we developed digital articular electro-goniometers (*c*) that can be employed as optional input in the case the user needs a visual feedback on the angular limb motion helpful to evaluate the running stimulation.

On the other side, we employ the commercial medical-certified RehaStim2 stimulator device provided by the HASOMED GmbH company, which is able to generate biphasic rectangular current pulses on up to eight channels simultaneously [9]. The stimulator is interfaced with an external device by means of the ScienceMode2 bidirectional communication protocol [10], which supports the control of complex stimulation patterns and training scenarios since intensity, pulse-width, and frequency are user-selectable pulse-by-pulse. The Raspberry Pi runs the main software, including the GUI, controls and acquires data from the input devices, processes the data and generates the stimulation patterns in real-time, and controls the FES stimulator.

24.2.2 Software

The software has been based on a object-oriented design in order to promote flexibility and modularity [11] (e.g., leveraging encapsulation, inheritance, and composition features), both to enable a seamless integration of different devices (e.g., input de-

vices, see Sect. 24.2.1) and to enable the future development of new processing algorithms. A multi-threaded architecture has been developed in order to map the functional tasks onto different running threads [12], so to optimize the use of computational resources and to avoid complex (run-time) code interdependencies. From the development standpoint, we based the software on the Python language, because of its cross-platform nature, its widespread adoption, and the large availability of third-party multi-platform libraries (in particular, we used the standard library for implementing the multi-threading features, and the Kivy library [13] for the GUI).

Referring to Fig. 24.1, the GUI is organized in four full-screen views, through which the user is able to properly configure and perform the system actions. After the login, the *Initialization* view allows the user to set the acquisition and stimulation parameters directly or by retrieving them from a database or through a calibration procedure. In the last case a dedicated *Calibration* view guides the user through the specific steps. Once the parameters have been set, the user can modify or save them using the *Parameters* view. The *Main Stimulation* view is the core of the GUI, allowing the user to start/stop the stimulation session, and providing visual feedback by showing both the FES intensity and the angular information.

A calibration procedure, divided into four sub-steps, is essential to define the ATC-FES control parameters on a per-user basis. First, the ATC threshold is set just above the sEMG baseline in order to maximize the threshold crossing events with the minimal muscle effort. Then, the maximum ATC value and the maximal current intensity are evaluated in way to create the proper relationship between acquisition and stimulation data. In the end, the Angular Range Of Motion (AROM) is evaluated. In this way, we are able to obtain a calibrated set of parameters enabling the implementation of a simple, yet effective, ATC-FES control algorithm based on lookup tables.

The multi-threading structure of the system and the running state of the involved threads during a typical stimulation session is reported in Fig. 24.2. The *Main Thread* runs all along the session waiting for the user input and creating child threads: the *FES Control* manages the communication with the RehaStim2, also providing the watchdog timer function; ATC_{th}, ATC_{max}, $AROM_{max}$ and I_{max} represent the four calibration steps which trigger the *acq* (data acquisition) threads.

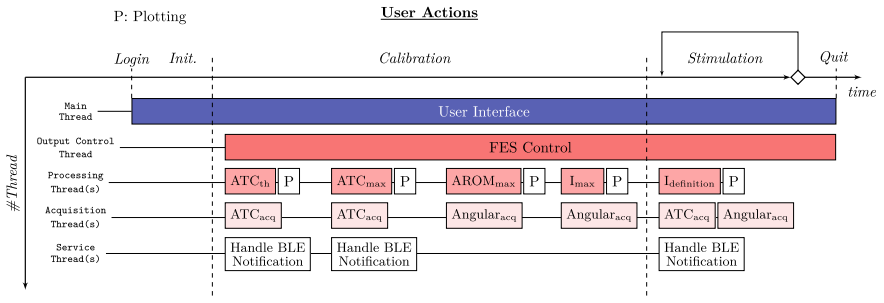


Fig. 24.2 Multi-threading structure during a typical stimulation session

24.3 Results and Discussion

The characterization of the computational resources consumption, including CPU usage percentage and RAM utilization, has been carried out using the *htop* GNU/Linux common tool available on the Raspberry Pi system. During the idle state, the CPU use is less than 5.3%, and it increases up to a 50% to 70% range during a normal operation (i.e, stimulation session), considering one or four ATC active channels as minimum/maximum configuration, respectively. The RAM utilization memory is in the 84 MB to 92 MB range. Real-time performance has been estimated by considering the duration of the processing time, which is defined as the elapsed time between the 4-channels ATC data reception and the consequent update on the FES device. The analysis, performed both on a GNU/Linux platform (Raspberry Pi) and a Microsoft® Windows® one (laptop equipped with an Intel® Core® i3-3227U clocked at 1.9 GHz and with 4 GB of RAM), is graphically reported in the boxplots on Fig. 24.3a. As shown in the left boxplot, the 95% of the data is lower than 50 ms with a mean value of 11.8 ms. On the other side, the computational power of a personal computer allowed us to obtain a mean value of 3.6 ms. In both cases the real-time requirement is satisfied (even if considering the ATC window as part of the overall system latency), and this confirms the benefits of using an sEMG event-driven approach.

The system has been then tested on 5 healthy subjects (3 males, 2 females, 24–27 years old) performing the elbow flexion, as a functional rehabilitative movement, in a therapist-patient scenario. In order to quantify the effective movement reproducibility, the limb motion has been acquired by means of the described electro-goniometers and the correlation coefficient between the signals is used as similarity measurement. The majority of the correlation values were above 0.8 (median value, mean value of 0.86 ± 0.07), which proves the high-fidelity reproduction of the movement, as shown in the example in Fig. 24.3b.

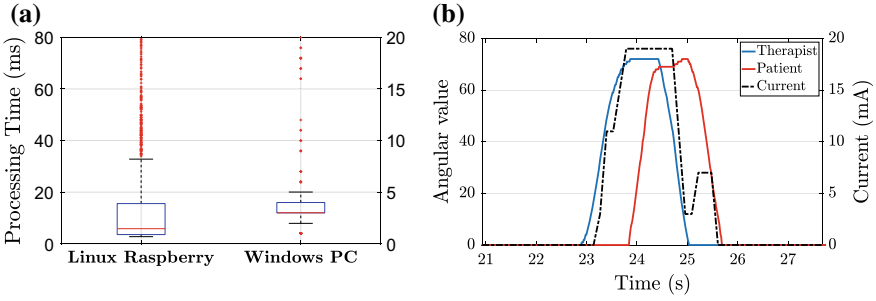


Fig. 24.3 In **a** the distribution of the elapsed time between ATC data reception and FES control update is shown. In **b** the recorded angular signals (blue: therapist, red: patient) associated to a single movement repetition, along with the applied stimulation current (black dashed line), are represented

24.4 Conclusion

The paper proposes an implementation of a multi-channel real-time embedded system (running a multi-platform software) for event-driven (ATC) sEMG-FES control. The promising results in terms of real-time processing, computation resources usage, and high-fidelity movement reproduction show the advantage of an event-driven approach w.r.t. literature sEMG-driven-FES system. Future investigations about optimal FES parameters computation and multi-channel cross-processing information will further improve the quality of the FES control.

References

1. Ferrante S, Chia Bejarano N, Ambrosini E, Nardone A, Turcato AM, Monticone M, Ferrigno G, Pedrocchi A (2016) A personalized multi-channel FES controller based on muscle synergies to support gait rehabilitation after stroke. Frontiers Neurosci 10:425
2. Perruchoud D, Pisotta I, Carda S, Murray MM, Ionta S (2016) Biomimetic rehabilitation engineering: the importance of somatosensory feedback for brain-machine interfaces. J Neural Eng 13(4):041001
3. Sweeney JD (1992) Skeletal muscle response to electrical stimulation. In: Reilly JP (ed) Electrical stimulation and electropathology, 1st edn. Press Syndicate of the University of Cambridge
4. Toledo C, Martinez J, Mercado J, Martín-Vignon-Whaley AI, Vera-Hernández A, Leija-Salas L (2018) sEMG signal acquisition strategy towards hand FES control. J Healthc Eng 2018:1–11
5. Li Z, Guiraud D, Andreu D, Benoussaad M, Fattal C, Hayashibe M (2016) Real-time estimation of FES-induced joint torque with evoked EMG. J NeuroEng Rehabil 13(1):60
6. Fernandez Guzman DA, Sapienza S, Sereni B, Motto Ros P (2017) Very low power event-based surface EMG acquisition system with off-the-shelf components. In: IEEE biomedical circuits and systems conference (BioCAS). https://doi.org/10.1109/BIOCAS.2017.8325152.
7. Rossi F, Motto Ros P, Sapienza S, Bonato P, Bizzi E, Demarchi D (2019) Wireless low energy system architecture for event-driven surface electromyography. In: Saponara S, De Gloria A (eds) Applications in electronics pervading industry, environment and society. Springer International Publishing, Cham, pp 179–185
8. Sapienza S, Motto Ros P, Fernandez Guzman DA, Rossi F, Terracciano R, Cordedda E, Demarchi D (2018)On-line event-driven hand gesture recognition based on surface electromyographic signals. In: 2018 IEEE international symposium on circuits and systems (ISCAS), pp 1–5
9. HASOMED GmbH, Operation Manual RehaStim 2, RehaMove 2, Sept 2011
10. Kuberski B (2012) ScienceMode2—description and protocol
11. Philips D (2015) Python 3 Object-oriented Programming, 2nd edn. Packt Publishing
12. Lott S (2015) Functional python programming. Packt Publishing.
13. Kivy (2019) Kivy home. [Online]. Available: https://kivy.org/#home

Chapter 25
A Fast Approximation of the Hyperbolic Tangent When Using Posit Numbers and Its Application to Deep Neural Networks

Marco Cococcioni, Federico Rossi, Emanuele Ruffaldi and Sergio Saponara

Abstract Deep Neural Networks (DNNs) are being used in more and more fields. Among the others, automotive is a field where deep neural networks are being exploited the most. An important aspect to be considered is the real-time constraint that this kind of applications put on neural network architectures. This poses the need for fast and hardware-friendly information representation. The recently proposed Posit format has been proved to be extremely efficient as a low-bit replacement of traditional floats. Its format has already allowed to construct a fast approximation of the sigmoid function, an activation function frequently used in DNNs. In this paper we present a fast approximation of another activation function widely used in DNNs: the hyperbolic tangent. In the experiment, we show how the approximated hyperbolic function outperforms the approximated sigmoid counterpart. The implication is clear: the posit format shows itself to be again DNN friendly, with important outcomes.

Keywords Deep neural networks (DNNs) · Posit · Activation functions

25.1 Introduction

The use of deep neural networks (DNN) as a general tool for signal and data processing is increasing both in industry and academia. One of the key challenge is the cost-effective computation of DNNs in order to ensure that these techniques can be implemented at low-cost, low-power and in real-time for embedded applications in IoT devices, robots, autonomous cars and so on. To this aim, an open research field is devoted to the cost-effective implementation of the main operators used in DNN, among them the activation function. The basic node of a DNN implements the sum of products of inputs (X) and their corresponding Weights (W) and then applies an

M. Cococcioni (✉) · F. Rossi · S. Saponara
Department of Information Engineering, University of Pisa, 56122 Pisa, Italy
e-mail: marco.cococcioni@unipi.it

E. Ruffaldi
MMI Spa, 56011 Calci, Pisa, Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_25

activation function $f(\cdot)$ to it to get the output of that layer and feed it as an input to the next layer. If we do not apply an activation function then the output signal would simply be a simple linear function, which has a low complexity but is not power enough to learn complex mappings (typically non-linear) from data. This is why the most used activation functions like Sigmoid, Tanh (Hyperbolic tangent) and ReLu (Rectified linear units) introduce non-linear properties to DNN [1, 2]. Choosing the activation function for a DNN model must take into account various aspects of both the considered data distribution and the underlying information representation. Moreover, for decision critical applications like machine perception for robotic and autonomous cars, also the implementation accuracy is important.

Indeed, one of the main trend in industry to keep low the complexity of DNN computation is avoiding complex arithmetic like double-precision floating point (64-bit), but relying on much more compact formats like BFLOAT or Flexpoint [3, 4] (i.e. a revised version of the 16-bit IEEE-754 floating point format adopted by Google Tensor Processing Units and Intel AI processors) or transprecision computing [5, 6] (e.g. the last Turing GPU from NVIDIA sustains INT32, INT8, INT4 and fp32 and fp16 computation [5]). To this aim, this paper presents a fast approximation of the hyperbolic tangent activation function combined with a new hardware-friendly information representation based on Posit numerical format.

Hereafter, Sect. 25.2 introduces the Posit format and the CppPosit library implemented at University of Pisa for the computation of the new numerical format. Section 25.3 introduces the hyperbolic tangent and its approximation. Implementation results when the proposed technique is applied to DNN with known benchmark dataset are reported in Sect. 25.4, where also a comparison with other known activation functions, like sigmoid, is discussed. Conclusions are drawn in Sect. 25.5.

25.2 Posit Arithmetic and the CppPosit Library

The Posit format as proposed in [7–9] is a fixed-length representation composed by at most 4 fields as shown in Fig. 25.1.: *1-bit* sign field, variable-length regime field, variable-length (up to *es-bits)* exponent field and a variable-length fraction field. The overall length and the maximum exponent lengths are decided a-priori. Regime length and bit-content is determined as by the number of consecutive zeroes or ones terminated, respectively, by a single one (negative regime) or zero (positive regime) (see Fig. 25.2).

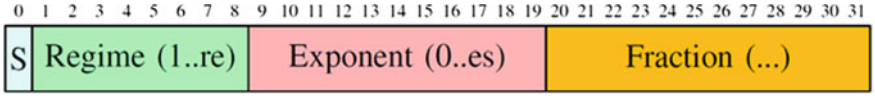


Fig. 25.1 An example of Posit data type

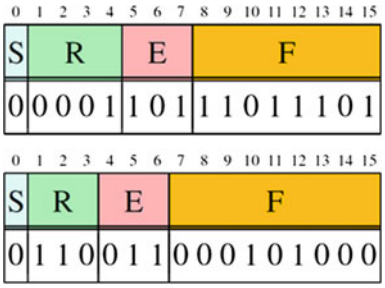


Fig. 25.2 Two examples of 16-bit Posit with 3 bits for exponent (es = 3). In the upper the numerical value is: $+256^{-3} \cdot 2^{5} \cdot (1 + 221/256)$(221/256 is the value of the fraction, 1 + 221/256 is the mantissa). The final value is therefore $1.907348 \times 10^{-6} \cdot (1 + 221/256) = 3.55393 \times 10^{-6}$. In the lower the numerical value is: $+256^{+1} \cdot 2^{3} \cdot (1 + 40/512)$(40/512 is the value of the fraction, 1 + 40/512 is the value of mantissa). The final value is therefore $2048 \cdot (1 + 40/512) = 2208$

In this work we are going to use the cppPosit library, a modern C++ 14 implementation of the original Posit number system. The library identifies four different operational levels (L1–L4):

- L1 operations are the ones involving bit-manipulation of the posit, without decoding it, considering it as an integer. L1 operations are thus performed on ALU and are fast.
- L2 operations involve unpacking the Posit into its four different fields, with no exponent computation.
- L3 operations instead involve full exponent unpacking, but without the need to perform arithmetic operations on the unpacked fields (examples are converting to/from float, posit or fixed point).
- L4 operations require the unpacked version to perform software/hardware floating point computation using unpacked fields.

L1 operations are the most interesting, since they are the most efficient ones. L1 operations include inversion, negation, comparisons and absolute value. Moreover, when $esbits = 0$, L1 operations also include doubling/halving, 1's complement when the specific Posit representation falls within the range [0, 1] and an approximation of the sigmoid function, called here fast Sigmoid, and described in [9]. Table 25.1 reports some implemented L1 operations stating whether the formula is exact or an approximation and the operation requirements in terms of Posit configuration and value. It is important to underline that every effort put in finding an L1 expression for some functions or operations has two advantages: a faster execution when using a software emulated PPU (Posit Processing Units), and a lower area required (i.e. less transistors) when the PPU is implemented in hardware.

Table 25.1 L1 operations summary

Operation	Approximation	Requirements
$2 \cdot x$	No	Esbits = 0
$x/2$	No	Esbits = 0
$1/x$	No	None
$1 - x$	No	Esbits = 0, x in $[-1, 1]$
FastSigmoid [9]	Yes	Esbits = 0
FastTanh (see below)	Yes	Esbits = 0

25.3 The Hyperbolic Tangent and Its Approximation FastTanh

The hyperbolic tangent is a non-linear activation function typically adopted as a replacement to the sigmoid activation function. The advantage of the hyperbolic tangent over the sigmoid is the higher enhancement given to the negative values. In fact, the output of the hyperbolic tangent spans in $[-1, 1]$ while the sigmoid outputs are only half of the previous, lying in $[0, 1]$. Furthermore, this difference in output range heavily impacts performances when using small-sized number representation, such as Posits with 10 or 8 bits. If we consider the sigmoid function applied to a Posit with x bits, we are actually using, as output, a Posit with $x - 1$ bits, since we are discarding the range $[-1, 0]$, which is significantly dense when using the Posit format (see Fig. 25.3).

However, as already mentioned before, the sigmoid function

$$\text{sigmoid}(x) = 1/\left(e^x - 1\right)$$

has a fast and efficient L1 approximation when using Posits with 0 exponent bits [9] (FastSigmoid). In order to exploit a similar trick for the hyperbolic tangent, we first introduced the *scaled sigmoid function*:

$$\text{sSigmoid}_k(x) = k \cdot \text{sigmoid}(k \cdot x) - k/2 \tag{25.1}$$

Particularly interesting is the case $k = 2$, when the scaled sigmoid coincides with the hyperbolic tangent:

$$\text{sSigmoid}_2(x) = \left(e^{2 \cdot x} - 1\right)/\left(e^{2 \cdot x} + 1\right) = \tanh(x) \tag{25.2}$$

Now that we can express the hyperbolic tangent as a linear function of the sigmoid one, we must rework the expression in order to provide a fast and efficient approximation to be used with Posits.

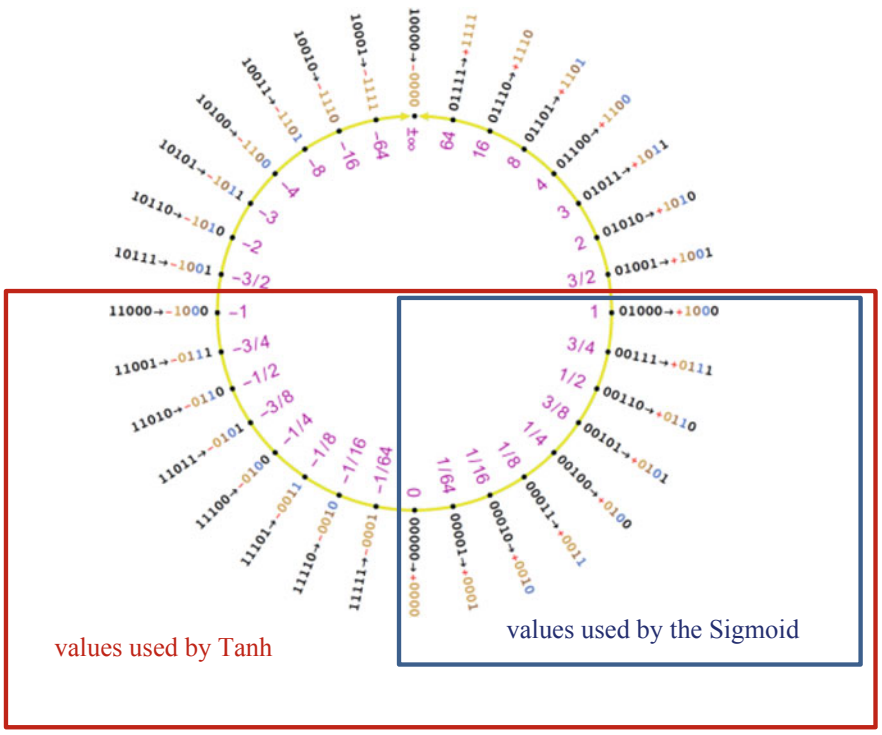


Fig. 25.3 The posit circle when the total number of bits is 5. The hyperbolic tangent uses all the numbers in $[-1, 1]$, while the sigmoid function only the ones in $[0, 1]$

We know that Posit properties guarantee that, when using 0 exponent bits format, doubling the Posit value and computing its sigmoid approximation is just a matter of bit manipulations, so they can be efficiently obtained. The subtraction in Eq. (25.1) does not come with an efficient bit manipulation implementation as-is. In order to transform it into an L1 operation we have to rewrite it as:

$$\text{FastTanh}(x) = 2 \cdot \text{FastSigmoid}(2 \cdot x) - 1 \tag{25.3}$$

Then let us focus on negative values for x only. For these values, the expression $2 \cdot \text{FastSigmoid}(2 \cdot x)$ is inside the unitary region $[0, 1]$. Therefore, the L1 *1's complement* can be applied. Finally, the negation is always an L1 operation, thus for all negative values of x the hyperbolic tangent approximation can be computed as an L1 operation. Moreover, thanks to the anti-symmetry of the hyperbolic tangent, this approach can also be extended to positive values. The following is a possible pseudo-code implementation:

```
FastTanh(x) →y
   x_n = x > 0? -x:x
```

```
s = x > 0
y_n = neg(compl1(twice(FastSigmoid(twice(x_n)))))
y = s > 0? -y_n:y_n
```

where `twice` is an L1 operation which computes $2 \cdot x$ and `compl1` is the L1 function that computes the 1's complement, again as an L1 operation.

Since we are also interested in training neural networks, we also need an efficient implementation of the hyperbolic tangent derivative:

$$d(\tanh(x))/d(x) = 1 - \tanh(x)^2$$

Let $y = \tanh(x)^2$, we know that $1 - y$ is always a L1 operation when $esbits = 0$, since $\tanh(x)^2$ is always in $[0, 1]$.

25.4 Experimental Results

We compared the approximated hyperbolic tangent to the original version in terms of execution time and precision. Figure 25.4 shows the precision comparison, reporting also for Posit8 and Posit16 the mean squared error between the approximated and the original form (for both types, we used 0 bits of exponent). Figure 25.5 shows execution time comparison for several repetitions. Each repetition consists in computing about 60,000 hyperbolic tangents with the approximated formula and the exact one. As reported, the precision degradation is in the order of 10^{-3} while the gain in speed is around a factor 6 (six time faster). In Figs. 25.4 and 25.5 *fast appr tanh* is the Posit-based implementation, using L1 operations, of the Tanh function, by using the FastTanh formula in Eq. 25.3. This corresponds to the column labeled has FastTanh in Table 25.2.

Then we tested the approximated hyperbolic tangent as activation function for the LeNet-5 convolutional neural network, replacing the exact hyperbolic tangent used in the original implementation proposed in [10, 11] and comparing results against the original activation. The network model has been trained on MNIST [11] and Fashion-MNIST datasets [12].

Table 25.2 shows performance comparison between the two activation functions (FastTanh and Tanh) on the two datasets. Moreover, also the results obtained with Sigmoid and ReLu are reported, since they are widely adopted in literature as activation functions for DNN. The results in Table 25.2 in terms of accuracy show that the FastTanh outperforms both the ReLu and the FastSigmoid (a well-known approximation of the sigmoid function) which are widely used in state-of-art to implement activation functions in DNN.

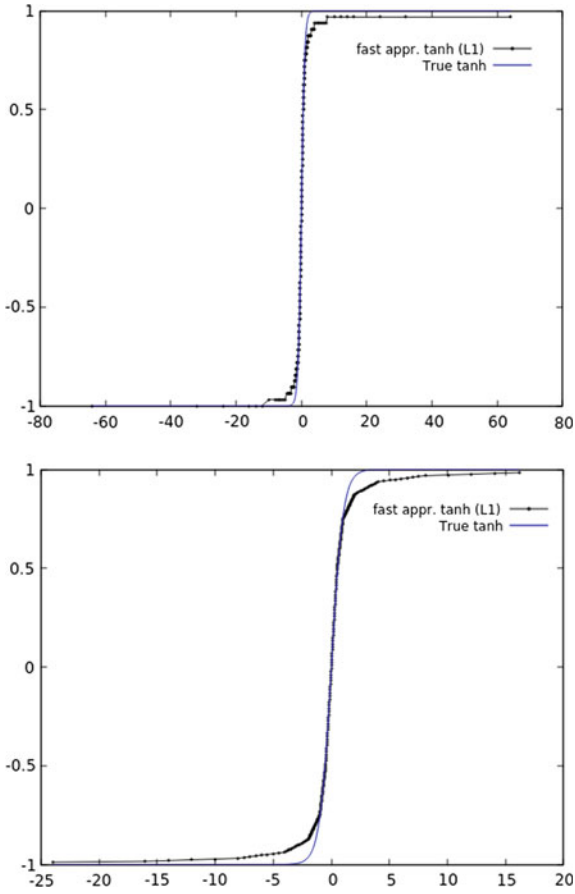


Fig. 25.4 Comparison between exact hyperbolic tangent (*True tanh*, in blue) and FastTanh (*fast appr. tanh*, in black), for Posit<8,0> (top) and Posit<16,0> (bottom). For Posit<8,0> the mean squared error is 2.816×10^{-3}, while for Posit<16,0> it is 2.947×10^{-3}

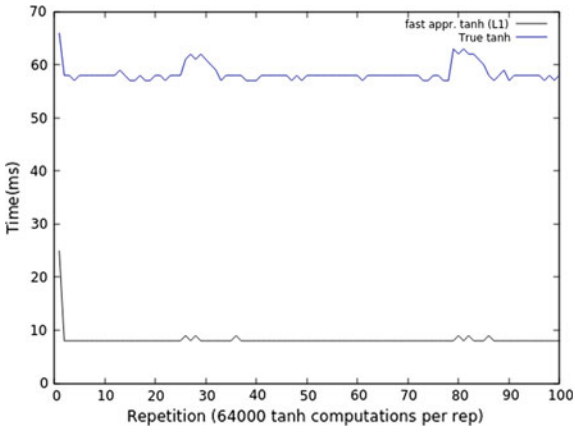


Fig. 25.5 Comparison of execution time of multiple consecutive executions between exact hyperbolic tangent (*True tanh*, in blue) and FastTanh (*fast appr tanh*, in black)

Table 25.2 Accuracy (%) and inference time (ms) comparison between different activation functions and different Posit configurations (MNIST and Fashion-MNIST data set)

Activation	FastTanh (this paper)		True Tanh		FastSigmoid [9]		ReLu	
	%	ms	%	ms	%	ms	%	ms
	MNIST							
Posit16,0	98.5	3.2	98.8	5.28	97.1	3.31	89.0	2
Posit14,0	98.5	2.9	98.8	4.64	97.1	3.09	89.0	1.9
Posit12,0	98.5	2.9	98.8	4.66	97.1	3.04	89.0	1.9
Posit10,0	98.6	2.9	98.7	4.62	96.9	3.08	89.0	1.9
Posit8,0	98.6	3.01	98.4	4.84	94.2	3.01	88.0	1.9
	FASHION-MNIST							
Posit16,0	89.6	3.4	90.0	5.5	85.2	3.4	85.0	2.1
Posit14,0	89.6	2.9	90.0	5.0	85.2	3.2	85.0	1.9
Posit12,0	89.7	2.9	90.0	5.1	85.2	3.1	85.0	1.9
Posit10,0	89.7	2.9	89.7	5.1	85.1	3.2	85.0	1.9
Posit8,0	89.6	3.1	89.3	5.2	84.3	3.0	84.0	1.9

25.5 Conclusions

In this work we have introduced FastTanh, a fast approximation of the hyperbolic tangent for numbers represented in Posit format which uses only L1 operations. We have used this approximation to speed up the training phase of deep neural networks. The proposed approximation has been tested on common deep neural network benchmarks. The use of this approximation resulted in a slightly less accurate neural network, with respect to the use of the slower true hyperbolic tangent, but with better performance in terms of inference time of the network. In our experiment, the FastTanh also outperforms both the ReLu and the FastSigmoid, which is a well-known approximation of the sigmoid function, a de facto standard activation function in neural networks.

Acknowledgements Work partially supported by H2020 European Project EPI (European Processor Initiative) and by the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project (Departments of Excellence program), granted to the Department of Information Engineering of the University of Pisa.

References

1. Pedamonti D (2018) Comparison of non-linear activation functions for deep neural networks on MNIST classification task. arXiv:1804.02763

2. Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: 27th International conference on international conference on machine learning (ICML) 2010, pp 807–814
3. Köster U et al (2017) Flexpoint: an adaptive numerical format for efficient training of deep neural networks. In: NIPS 2017, pp 1740–1750
4. Popescu V et al (2018) Flexpoint: predictive numerics for deep learning. In: IEEE symposium on computer arithmetics, 2018
5. "NVIDIA TURING GPU Architecture, graphics reinvented", White paper n. WP-09183-001_v01, pp 1–80, 2018
6. Malossi A et al (2018) The transprecision computing paradigm: concept, design, and applications. In: IEEE DATE 2018, pp 1105–1110
7. Cococcioni M, Rossi F, Ruffaldi E, Saponara S (2019) Novel arithmetics to accelerate machine learning classifiers in autonomous driving applications. In: IEEE ICECS 2019, Genoa, Italy, 27–29 Nov 2019
8. Cococcioni M, Ruffaldi E, Saponara S (2018) Exploiting posit arithmetic for deep neural networks in autonomous driving applications. IEEE automotive 2018, pp 1–6
9. Gustafson JL, Yonemoto IT (2017) Beating floating point at its own game: posit arithmetic. Supercomput Front Innov 4(2):71–86
10. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
11. LeCun Y, Jackel L, Bottou L, Brunot A, Cortes C, Denker J, Drucker H, Guyon I, Muller U, Sackinger E, Simard P, Vapnik V (1995) Comparison of learning algorithms for handwritten digit recognition. In: Fogelman F, Gallinari P (eds) International conference on artificial neural networks, Paris. EC2 and Cie, pp 53–60.
12. Xiao H, Rasul K, Vollgraf R (2017) Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv:1708.07747

Part VI
Sensors and Sensing Electronic Systems

Chapter 26
2-D Acoustic Particle Velocity Sensors Based on a Commercial Post-CMOS MEMS Technology

Andrea Ria, Massimo Piotto, Mattia Cicalini, Andrea Nannini and Paolo Bruschi

Abstract A 2-dimensional acoustic sensor with programmable directivity is proposed. The sensor is formed by the combination of two orthogonal detectors of acoustic particle velocity. Differently from previous versions, the proposed device is fabricated using a commercially available post-CMOS technology, opening the way to low cost applications. Characterization of the frequency response and directivity is presented. The possibility of producing a sensor with electronically programmable directivity is demonstrated.

26.1 Introduction

The development of silicon micromachining technologies has allowed the fabrication of miniaturized thermal sensors capable of a direct measurement of the acoustic particle velocity (APV). These sensors, combined with a traditional microphone, enable full knowledge of the local acoustic field. Furthermore, their intrinsic directionality makes them useful for applications requiring sound source localization [1–3] or noise suppression [4, 5].

The first MEMS APV sensor, named Microflown™, was proposed by De Bree et al. [6] and it was based on the heat transfer between two micro-hotwires placed

A. Ria · M. Piotto · M. Cicalini · A. Nannini · P. Bruschi
Department Ingegneria Dell'Informazione, Università di Pisa, via Caruso 16, 56122 Pisa, Italy
e-mail: andrea.ria@ing.unipi.it

M. Cicalini
e-mail: mattia.cicalini@phd.unipi.it

A. Nannini
e-mail: andrea.nannini@unipi.it

P. Bruschi
e-mail: paolo.bruschi@unipi.it

M. Piotto (✉) · P. Bruschi
CNR-IEIIT, via Caruso 16, 56122 Pisa, Italy
e-mail: massimo.piotto@unipi.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_26

on suspended bridges. Since then, the original device has been developed leading to commercial compact probes allowing detection of multiple APV components. Other devices based on multiple micro-wires integrated on the same chip have been proposed to perform 2-D [7, 8] and 3-D [9] measurements. These sensors have been fabricated with a dedicated micromachining technology that was not compatible with the standard IC fabrication processes. The possibility of fabricating APV sensors with a CMOS process followed by a simple post-processing applied in a research laboratory has been demonstrated [10, 11]. This technique enabled the integration of two APV sensors with orthogonal sensitivity axis on the same chip with the possibility of obtaining a 2-D APV sensor with a programmable directivity [12]. Recently, an APV sensor fabricated by means of a commercial post-CMOS technology [13] available to small-medium enterprises has been proposed [14].

This work expands the experiments presented in [14] by combining the signals of two orthogonal APV sensing structures to obtain a single-chip sensor with programmable directivity. Comparison of the response of two independent structures also provides preliminary indication of the matching properties of this novel fabrication flow.

26.2 Sensor Description

The basic structure that forms the APV sensors is shown in Fig. 26.1. The device is formed by two parallel polysilicon wires placed at a micrometric distance (L_{gap} in the figure). Each wire is split into three identical segments, supported by U-shaped silicon dioxide cantilevers, which are suspended into a single deep cavity etched into the silicon substrate. The wires are self-heated by an electrical current (bias current) and the heat exchange between them takes place through both conduction and forced convection. The latter mechanism depends on the local APV, which induces oscillating temperature variations between the wires. Due to the temperature coefficient

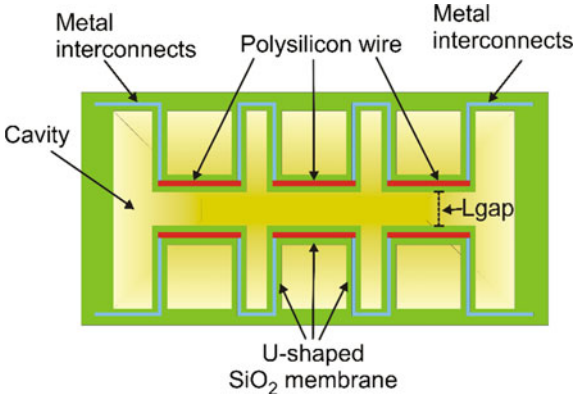


Fig. 26.1 Elemental sensing structure used to compose the APV sensors. Polysilicon wires are depicted in red, while metal interconnects (metal 1) are represented in blue. All the six U-shaped cantilevers are suspended into a single deep cavity in order to introduce effective thermal insulation from the substrate

of the wire resistance (TCR $\cong 1 \times 10^{-3}$ K^{-1}), the bias current converts the temperature variations into an electrical signal (voltage), which is proportional to the APV. Separation of the wires into smaller segments allows keeping an optimal wire length and resistance, while increasing stiffness and reducing etching times.

The sensors have been designed using a standard 0.35 μm CMOS process provided by Austria Micro System, followed by a post-CMOS front-side micromachining step, aimed at etching the cavities. The whole fabrication flow was provided by the CMP consortium [13]. The sensor dimensions were optimized using an original simulation approach [11, 14] based on the COMSOL™ environment. In practice, parametric simulations performed by varying the main dimensions (e.g. the Lgap distance) has been used to find the configuration that provides the maximum sensitivity.

The designed test chip, 2.88 mm × 2.88 mm wide, included several different APV sensors. In this work, we have used the two identical APV sensors shown in the micrograph of Fig. 26.2, indicated by SX and SY. Each one of these sensors is formed by two elemental structures as that in Fig. 26.1, connected to form a Wheatstone bridge as schematically shown for SY on the right of Fig. 26.2. Connection is made in such a way that the resistance variations of all wires induced by the APV give in-phase contributions to the output voltage. SX and SY differ only for their orientation, which is such that SX and SY are sensitive to the APV component along the X-and Y-axis, respectively.

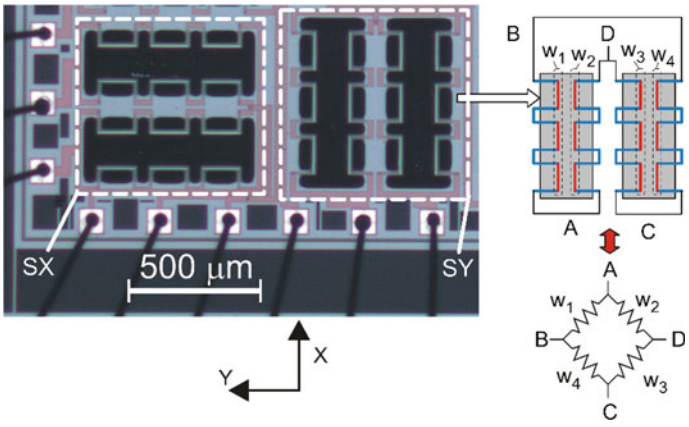


Fig. 26.2 Micrograph of the test chip portion where the two APV sensors used in this work (SX and SY) are located. Sensors SX and SY are sensitive to APV components located along the X and Y-axis, respectively, indicated below the micrograph. The way the four wires W_{1-4} that form SY are connected and the equivalence with a Wheatstone bridge are shown on the right

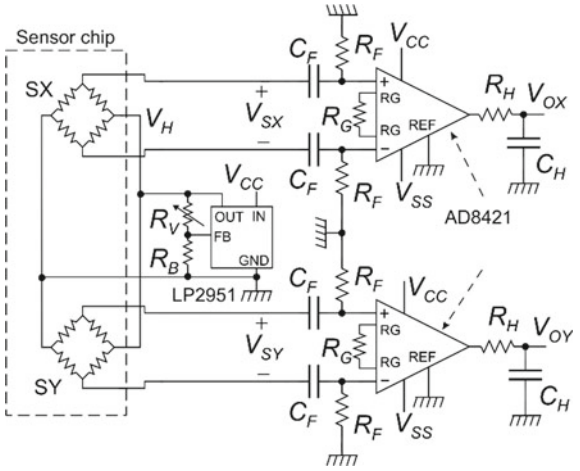


Fig. 26.3 Readout configuration used in the experiments. The dashed box includes components and connections that are integrated into the test chip. All other components are placed in a purposely-built printed circuit board (PCB)

26.3 Measurement Setup

The chips are packaged into 44 pin cases (JLCC44) resulting in overall dimensions of 24 mm × 24 mm × 8 mm for each sample under test. The readout configuration is shown in Fig. 26.3. A dc supply voltage (V_H) is applied to one diagonal of both SX and SY bridges, while the corresponding output signals (V_{SX} and V_{SY}) of the sensor are taken on the other diagonal. Voltage V_H, provided by a LP2951 regulator, can be varied in the 1.25–12 V range by changing the variable resistor R_V. For conventional acoustic intensity, the output signals are very small (tens of microvolts), thus amplification with ultra-low noise instrumentation amplifiers (Analog Devices AD8421, set to gain = 200) is required. The dc component of the bridge output voltages is removed by high pass filters C_F–R_F, with a cut-off frequency of 10 Hz. The amplified output signals V_{OX} and V_{OY} are low pass filtered (roll-off frequency 15 kHz) by R_H, C_H. A 16-bit digitizer (Picotech PicoScope 4262) is used to acquire the output voltages.

Signal processing is performed using programs running on a personal computer. Frequency response measurements are obtained using the stationary wave tube approach. A rotating sample-holder allows varying the orientation of the sensors with respect to the direction of the APV, which is parallel to the tube axis.

26.4 Experimental Results and Discussion

Preliminary characterization of the sensors involved determination of the dependence of the average wire overheating (ΔT) as a function of the voltage V_H applied to the bridges. The temperature was estimated from the resistance through the TCR. The result is shown in Fig. 26.4 (left) for SX and SY. Note that the higher the static

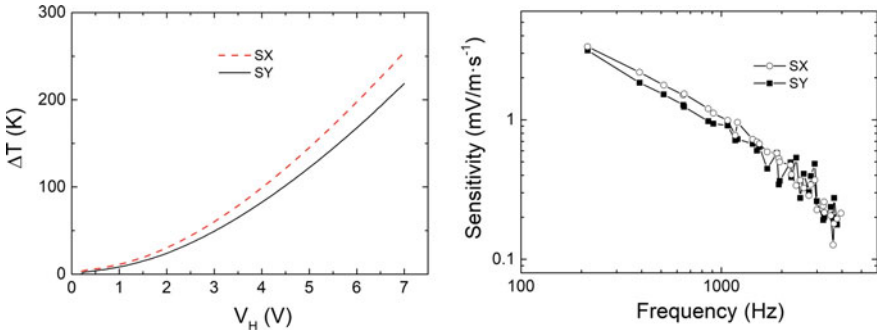


Fig. 26.4 Overheating temperature versus supply voltage (left) and frequency response, measured for $V_H = 7$ V (right)

overheating, the higher the temperature oscillations induced by the APV, and then the higher the sensitivity. A mismatch between the two sensors is clearly visible, probably due to difference in the thermal insulation. The frequency response of the sensor sensitivity is shown in Fig. 26.4 (right) for $V_H = 7$ V. The strongly low-pass characteristic of the response is practically the same for SX and SY, while the former presents a higher sensitivity, which is consistent with the above-mentioned higher temperature reached by the wires.

The directivity of the SX and SY sensors is shown in Fig. 26.5 (left). The typical figure of eight, deriving from a cosine-like response, is visible. We have synthetized a response with maximum sensitivity along an arbitrary axis by simply calculating a linear combination of the output signals V_{OX} and V_{OY}, according to:

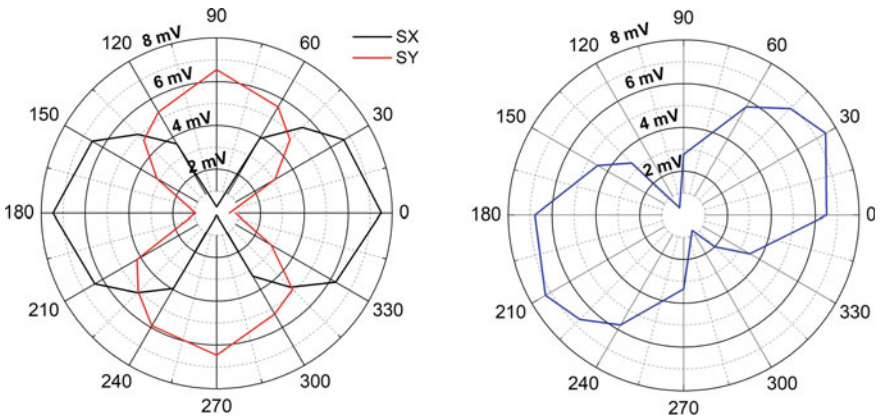


Fig. 26.5 Left: polar plots of the output voltages (V_{OX}, V_{OY}) as a function of the sensor orientation, measured for $V_H = 7$ V. Right: polar plot of the composite output voltage $V_{O\theta}$ defined by Eq. (26.1). Coefficients a and b that appear in Eq. (26.1) are chosen to obtain an angle of maximum sensitivity of 30°

$$V_{O\theta} = aV_{OX} + bV_{OY} \quad with \quad a = \cos(\theta), \quad b = \sin(\theta) \tag{26.1}$$

where θ is the desired direction of maximum sensitivity. The result, for a preferential sensitivity direction of 30°, is shown in Fig. 26.5 (right).

In conclusion, the experimental results confirm that it is possible to program the axis of maximum sensitivity by simply changing the coefficients of the linear combination. This goal has been obtained with sensors that did not require completing the chip fabrication in a research lab, making the development of this kind of devices attractive even for small enterprises. A current limit of the proposed sensors is the frequency response, which is considerably worse than previous devices developed with a research-grade post-CMOS procedure [10, 11]. This drawback is tied to the larger thermal mass of the wires, due to the lower resolution of the commercial micromachining technology with respect to the research process. Nevertheless, the sensitivity at frequencies up to 1 kHz is sufficient to encourage the experimentation in sound source selection and localization scenarios. Furthermore, application to acoustic impedance measurements for low frequency material characterization can also be envisioned.

Acknowledgements This research has been supported by DELTATECH Italy and the Italian municipality of Sogliano al Rubicone (Italy) in the framework of the SIHT (Sogliano Industrial High Technology) project.

References

1. Ramamohan KN, Comesaña DF, Leus G (2018) Uniaxial acoustic vector sensors for direction-of-arrival estimation. J Sound Vib 437:276–291
2. Song Y, Li YL, Wong KT (2015) Acoustic direction finding using a pressure sensor and a uniaxial particle velocity sensor. IEEE Trans Aerosp Electron Syst 51(4):2560–2569
3. Song Y, Wong KT, Li Y (2015) Direction finding using a biaxial particle-velocity sensor. J Sound Vib 340:354–367
4. Garcia-Bonito J, Elliott SJ (1999) Active cancellation of acoustic pressure and particle velocity in the near field of a source. J Sound Vib 221(1):85–116
5. Felisberto P, Santos P, Jesus SM (2018) Acoustic pressure and particle velocity for spatial filtering of bottom arrivals. IEEE J Oceanic Eng 99:1–14
6. De Bree HE, Leussink P, Korthorst T, Jansen H, Lammerink TS, Elwenspoek M (1996) The μ-flown: a novel device for measuring acoustic flows. Sens Actuators A 54(1–3):552–557
7. Pjetri O, Wiegerink RJ, Lammerink TS, Krijnen GJ (2013) A crossed-wire 2-dimensional acoustic particle velocity sensor. In: Sensors. IEEE, pp 1–4
8. Pjetri O, Wiegerink RJ, Krijnen GJ (2016) A 2D particle velocity sensor with minimal flow disturbance. IEEE Sens J 16(24):8706–8714
9. Yntema DR, Van Honschoten JW, Wiegerink RJ, Elwenspoek M (2008) A complete three-dimensional sound intensity sensor integrated on a single chip. J Micromech Microeng 18:115004
10. Bruschi P, Butti F, Piotto M (2011) CMOS compatible acoustic particle velocity sensors. In: IEEE Sensors 2011 conference proceedings, Limerick, 28–31 Oct 2011, pp 1405–1408

11. Piotto M, Butti F, Zanetti E, Di Pancrazio A, Iannaccone G, Bruschi P (2015) Characterization and modeling of CMOS-compatible acoustical particle velocity sensors for applications requiring low supply voltages. Sens Actuators A 229:192–202
12. Piotto M, Butti F, Bruschi P (2014) Acoustic velocity sensors with programmable directivity. In: Sensors, Springer, New York, pp 271–275
13. CMP website. URL: https://mycmp.fr/datasheet/mems-bulk-micromachining-frontside-bulk-micromachining
14. Piotto M, Ria A, Stanzial D, Bruschi P (2019) Design and characterization of acoustic particle velocity sensors fabricated with a commercial post-CMOS MEMS process. In: The 20th International conference on solid-state sensors, actuators and microsystems (Transducer 2019), Berlin, 23–27 June 2019, pp 1839–1842

Chapter 27
A High-SNR Distributed Acoustic Sensor Based on ϕ-OTDR Using a Scalable Phase Demodulation Scheme Without Phase Unwrapping

Yonas Muanenda, Stefano Faralli, Claudio J. Oton and Fabrizio Di Pasquale

Abstract We propose and experimentally demonstrate a high-SNR Distributed Acoustic Sensing (DAS) scheme using an array of Ultra-Weak Fiber Bragg Gratings (UWFBGs) with dynamic phase extraction in direct detection based on delayed interferometry and a scalable phase-generated carrier demodulation algorithm requiring no phase unwrapping.

27.1 Introduction

Distributed optical fiber sensors have interesting applications in many environmental safety and integrity monitoring systems involving the measurements of parameters over long distances [1]. They have been used in among others the transportation, power generation & distribution sectors and industrial process control systems. Recently, distributed acoustic sensing, which is the use of an optical fiber for measuring vibrations over an extended region, has attracted significant attention in the fiber sensing community [2]. It is based on the sensitivity of coherent Rayleigh scattering in an optical fiber to external perturbations such as vibrations which alter the phase of the backscattered light [3, 4]. The specific environmental parameter can be monitored by observing the local change in intensity [5], to obtain information on the position and frequency of the vibration or retrieving additional information by either quantifying the phase change [6] or using wavelength shift methods [4], for more precise quantitative measurement of the perturbation. So far, various techniques have been proposed for the demodulation of the phase but there are still some limitations which need to be addressed. The intensity of coherent Rayleigh backscattering from a standard single mode fiber is very low which means that the SNR is in general low and requires additional components or advanced signal manipulation methods. Recently, there have been some investigations to address this issue by using

Y. Muanenda (✉) · S. Faralli · C. J. Oton · F. Di Pasquale
Institute of Communication, Information and Perception Technologies (TeCIP), Scuola Superiore Sant'Anna, Via Moruzzi 1, 56124 Pisa, Italy
e-mail: y.muanenda@santannapisa.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_27

an array of Ultra-weak Fiber Bragg Gratings (UWFBGs) which are gratings with a very small reflectivity in the order of 40 dB or below. They allow multiplexing of hundreds or even thousands of them in a single mode fiber without cumulative losses detrimental to measurement SNR reduction [6, 7]. Schemes using time-division multiplexing [8] or using laser sweeping and phase-unwrapping (a mechanism used to retrieve the values of the phase beyond the range of the arctangent function) has also been proposed and demonstrated [9].

Other techniques akin to the ones used for interrogation of standard single mode fibers include ones based on distributed phase demodulation employing a table lookup operation [10] and a 3 × 3 coupler [11], which necessitates duplicate receivers. Others employing coherent detection or I-Q demodulation can be used but most of them involve phase unwrapping operations which are typically computationally heavy [12]. Using such algorithms in a distributed sensor would significantly lower the dynamic performance. In the differentiate-square-multiply (PGC-DMS) algorithm [13], demodulation computations involve division operations which are susceptible to division-by-zero errors. In this work, we propose and experimentally demonstrate a DAS based on UWFBGs utilizing a homodyne demodulation scheme with delayed interferometry and direct detection using the phase-generated carrier (PGC) demodulation with the differentiate-cross-multiply (PGC-DCM) [7] algorithm. The method offers a high SNR, does not require phase-unwrapping, introduces less errors coming from division-by-zero operations in PGC-DMS, involves computations readily implementable with analogue or digital processing and uses a simple direct detection receiver.

27.2 Demodulation Scheme and Experimental Setup

Considering the case of a disturbance introducing a phase change of ϕ (t) between two consecutive gratings, it can be shown that the delay and mixing of the response of two UWFBGs in an interferometer having a relative phase shift of $\Delta\delta = C\omega_0(t)$, with c being the modulation depth, yields an output composed of two terms which appear at the odd and even multiples of the controlled modulating frequency ω_0 [14–16]. After mixing the output of the delayed interferometer with the first and second harmonic of the phase modulating frequency, ω_0 and $2\omega_0$ and subsequent low-pass filtering of the remaining terms, two intermediate components can be obtained:

$$\begin{aligned} s_1(t) &\propto G J_1(C) \sin\phi(t), \\ s_2(t) &\propto H J_2(C) \cos\phi(t). \end{aligned} \tag{27.1}$$

In (27.1), G and H are the amplitudes of the RF mixing signals, and J_1 (C) and J_2 (C) are Bessel function terms of the modulation depth. While the arctan of the ratio of the two intermediate components can be used, letting G = H, it also requires J_1 (C) = J_2 (C), which is true only for a specific value

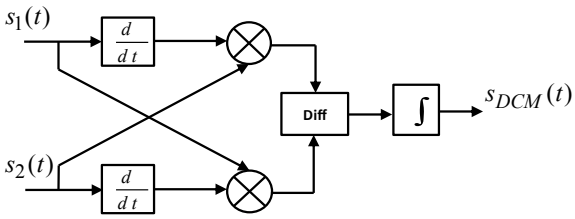


Fig. 27.1 Computations involved in phase demodulation using the PGC-DCM

of the modulation depth C = 2.63. Other modulation points introduce errors and, most importantly, the arctan function requires computationally costly phase unwrapping. The proposed method in this contribution is the PGC-DCM whose schematic is shown in Fig. 27.1. The computations in the diagram yield:

$$\frac{d}{dt}s_{DCM}(t) \propto [GHJ_2(C)J_1(C)] \times [\sin^2\phi(t) + \cos^2\phi(t)]\frac{d}{dt}\phi(t). \tag{27.2}$$

After simplification using trigonometric identity, the demodulated phase can be extracted by integration of both sides of (27.2), yielding:

$$s_{DCM}(t) = GHJ_2(C)J_1(C)\phi(t). \tag{27.3}$$

Hence, after normalization of (27.3) to handle the scaling factors, the proposed algorithm can be used to obtain the demodulated phase without the use of costly two-dimensional phase unwrapping in a distributed system and reduced errors due to division-by-zero operations. Note also that PGC-DCM is known to exhibit lower harmonic distortions than the PGC-arctan algorithm [13].

The experimental setup used to validate the proposed technique is shown in Fig. 27.2. As shown, light from a narrowband laser of 200-kHz linewidth is amplified using an Erbium-Doped Fiber Amplifier (EDFA) and filtered using an Optical Band-pass Filter (OBPF) before being gated with an Acousto-optic Modulator (AOM) to generate the interrogating pulses. After another round of amplification and filtering, the pulses are sent through a three-port optical circulator into the fiber under test

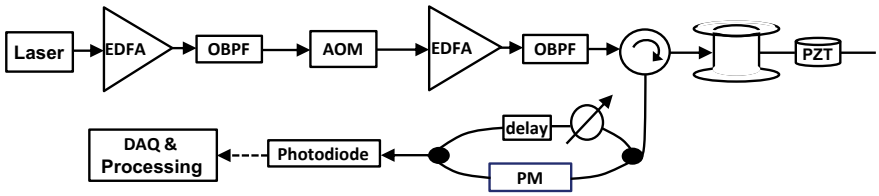


Fig. 27.2 Experimental setup of proposed ϕ-OTDR: Erbium-doped fiber amplifier (EDFA); optical band-pass filter (OBPF); acousto-optic modulator (AOM), digital acquisition (DAQ); phase modulator (PM); piezoelectric actuator (PZT)

(FUT), which is composed of 200 UWFBGs each with reflectivity of ~−43 dB, FWHM of ~3.4 nm and spaced 5 m apart along a 1-km standard singlemode fiber.

The fiber between the two gratings at the end of the FUT is wound around a pizeoelectric (PZT) actuator on which controlled vibrations have been applied using a voltage amplifier driven with a waveform generator. The backscattering from the FUT is collected at the return port of the circulator, which feeds the unbalanced interferometer where a delay is applied in one arm and the phase of the light on the other one is modulated using the Phase Modulator (PM). The beating is then detected using a direct detection receiver with a simple pin photodiode of 125 MHz bandwidth and acquired using a real-time digital acquisition system for processing using the PGC-DCM technique.

27.3 Experimental Results and Discussions

The first set of measurements involved the observation of raw back-reflection traces from the UWFBG array when pulses of 20 ns were sent into the sensing fiber. Since the aim is to observe the visibility of back-reflections, this has been done by disconnecting the arm of the interferometer having the PM. A sample set of 500 raw traces is depicted in Fig. 27.3a and b which show that the raw traces from the gratings have consistent and enhanced visibility suitable for measurement with a high SNR.

This is also confirmed with a comparison of the traces with a "singlemode" operation of the FUT, which was done by shifting the emission wavelength of the laser away from the passband of the UWFBGs. The traces given in Fig. 27.4 show the backscattering from 20-ns pules for both cases and 120-ns ones only for standard "singlemode" operation. (Note that using a 120-ns pulse with the UWFBG array would have resulted in the interference of the backscattering from adjacent gratings.) As shown, both at the near- and far-end, the intensity of the response from the singlemode operation is close to the noise floor and won't enable distributed measurements at the end of the fiber while the UWFBGs exhibit higher visibility even for narrow pulses.

Subsequently, the full setup was used by connecting the PM and 10,000 traces were acquired for a segment of the fiber at the far-end, for a total duration of 200 ms, to observe the evolution of the traces in the presence of the overlap and beating from the delayed interferometer. In this case, upon photodetection, a clear beating of the interfering fields is observed on the oscilloscope, which is also verified in the acquired traces.

A sample measurement done when a phase modulation of 10 kHz and vibration of 2.5 kHz is applied to the PZT is depicted in Fig. 27.5, both in time and frequency domain. As shown, the two intermediate components centered at the modulating angular frequency ω_0 and its double $2\omega_0$ are seen with a spacing of 2.5 kHz, consistent with the spectrum of a typical phase modulated signal.

Subsequently, the PGC-DCM algorithm was used to obtain the demodulated phase from the intermediate components as shown in Fig. 27.6. It can be observed in

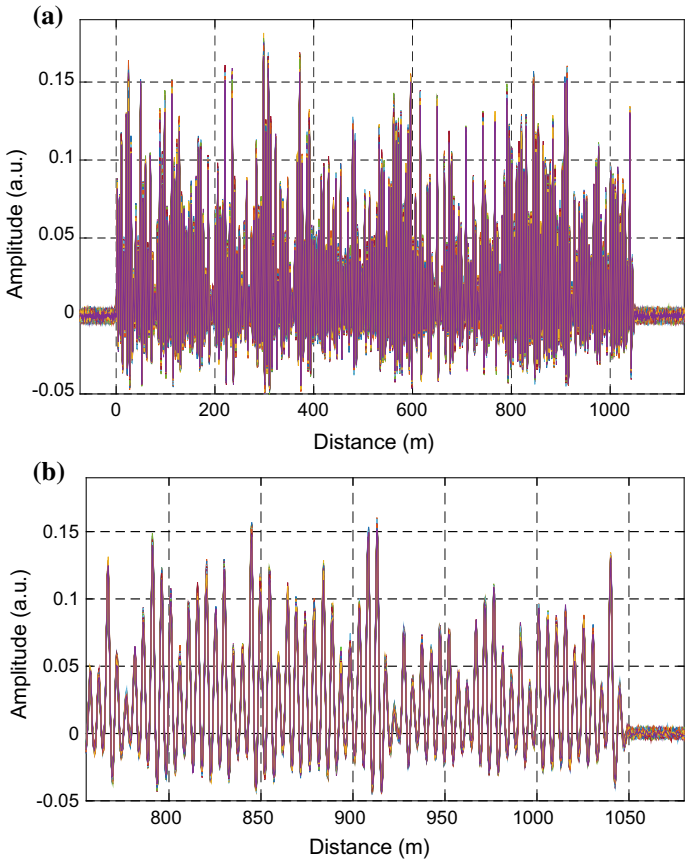


Fig. 27.3 **a** Overlapped raw backscattering traces showing high visibility and consistent reflection from ultra-weak gratings **b** traces at the far end showing capacity to measure with high SNR

Fig. 27.6a that there is consistent demodulation with the proposed technique even at the zero crossing points of the intermediate signals along the whole 200-ms duration of the 2.5-kHz signal (500 cycles). In addition, before high-pass filtering, the demodulated phase exhibits slow variations due to environmental drifts, which are perturbations of interest in many structural health monitoring applications.

The demodulated and high-pass filtered phase response depicted in Fig. 27.6b has an SNR of ~34 dB, thanks to the use of UWFBGs. Note that, the trace for singlemode operation using 20 ns probing pulses does not enable vibration measurement as the intensity at the far-end is equal to the noise floor as shown in the plot in Fig. 27.4b.

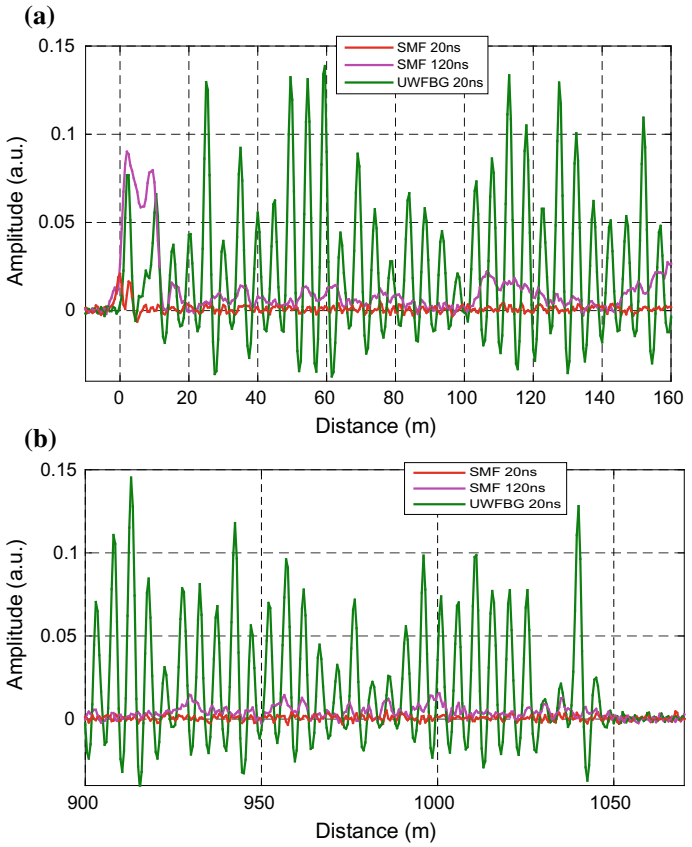


Fig. 27.4 Comparison of the backscattering traces from the UWFBG array with a singlemode operation at the **a** near-end, and **b** far-end of the responses, confirming significantly higher visibility due to gratings

27.4 Conclusions

In summary, we have proposed and experimentally demonstrated a DAS based on ϕ-OTDR based on identical UWFBGs for high-SNR phase demodulation. A UWFBG array with 200 identical gratings of reflectivity ~−43 dB each spaced at 5 m has been interrogated using a direct detection scheme involving the PGC-DCM demodulation technique. The response of the grating array has been shown to exhibit significantly higher visibility with the single pulse equivalent. In addition, using a single pulse probe and a direct detection receiver after a delayed interferometer involving relative phase modulation in one arm, the dynamic phase change induced by a 2.5 kHz vibration at the end of a 1-km fiber is retrieved with an SNR of ~34 dB.

The phase demodulation technique does not require computationally costly phase unwrapping which would otherwise degrade the dynamic performance of DAS when

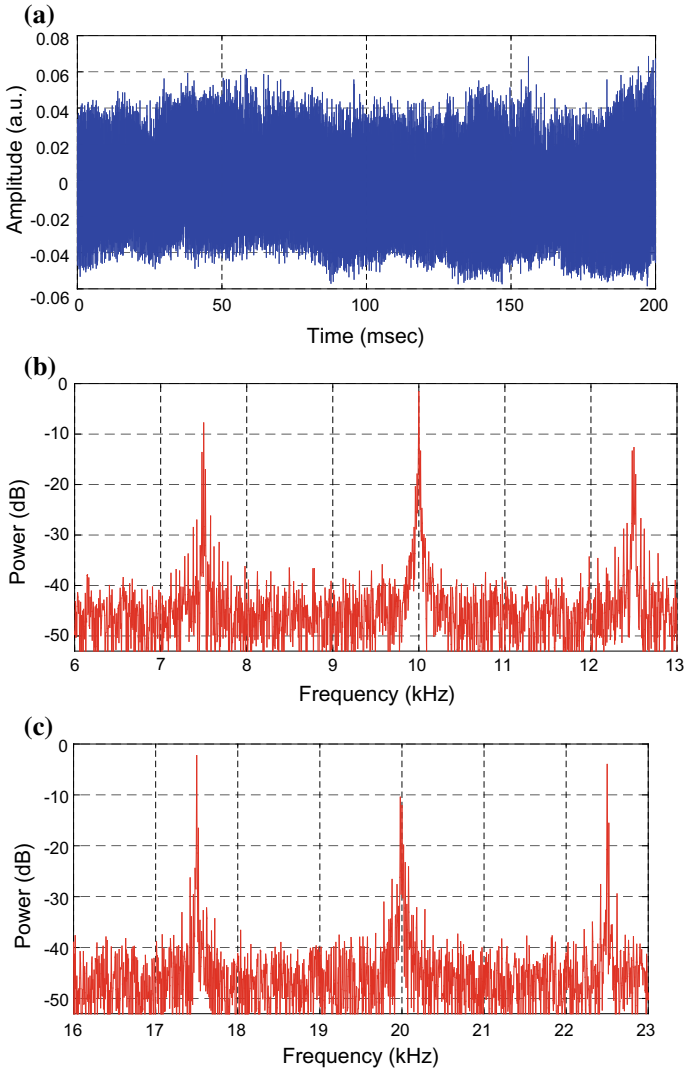


Fig. 27.5 Observed beating from the delayed interferometer at the point of the PZT: **a** time domain evolution, and intermediate components **b** centered at ω_0 and **c** centered at $2\omega_0$

phase retrieval for quantitative measurements are involved. It is also robust against errors due to division-by-zero operations. In addition, the method is scalable as it involves integration and differentiation operations which, thanks to the ubiquity of systems for performing fractional order calculus, can be realized with digital signal processing schemes based on FPGAs or analog ones using operational amplifiers, both of which are also candidates for small-scale integration. Hence, the proposed

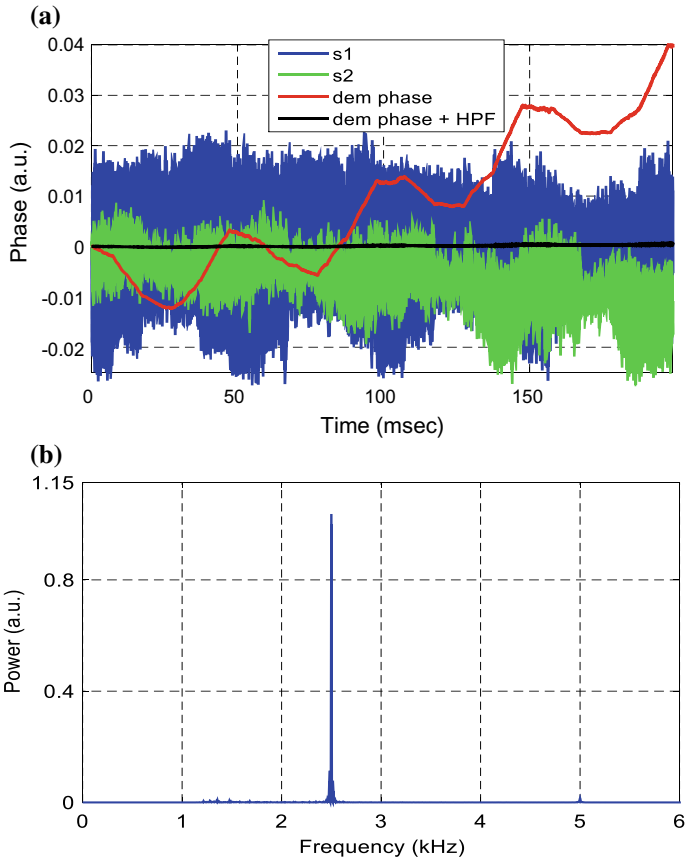


Fig. 27.6 **a** Intermediate components and demodulated phase before and after high pass filtering showing consistent phase retrieval at zero-crossing points. **b** Spectrum of the demodulated phase change induced by 2.5 kHz vibration of a PZT actuator

scheme offers a high-SNR DAS based on a scalable and consistent homodyne phase demodulation technique suitable for distributed dynamic measurements.

References

1. Muanenda Y, Oton CJ, Di Pasquale F (2019) Application of Raman and Brillouin scattering phenomena in distributed optical fiber sensing. Front Phys 7:155
2. Muanenda Y (2018) Recent advances in distributed acoustic sensing based on phase-sensitive optical time domain reflectometry. J Sens 2018:3897873
3. Muanenda Y, Oton CJ, Faralli S, Di Pasquale F (2015) High performance distributed acoustic sensor using cyclic pulse coding in a direct detection coherent-OTDR. In: Proceedings of 5th Asia-Pacific Optical Sensors Conference Conference, pp 965547 (SPIE)

4. Liehr S, Muanenda Y, Münzenberger S, Krebber K, Wavelength-modulated C-OTDR techniques for distributed dynamic measurement. In: 26th International conference on optical fiber sensors, OSA technical digest, (Optical Society of America, 2018), paper TuE15
5. Muanenda Y, Oton CJ, Faralli S, Di Pasquale F (2017) A ϕ-OTDR sensor for high-frequency distributed vibration measurements with minimal post-processing. In: Proceedings of 19th Italian national conference on photonic technologies, pp. 1–4. (IEEE)
6. Tang J, Li L, Guo H, Yu H, Wen H, Yang M (2017) Distributed acoustic sensing system based on continuous wide-band ultra-weak fiber Bragg grating array. In: Proceedings of 25th optical fiber sensors conference, pp. 1–4. (IEEE)
7. Muanenda Y, Faralli S, Oton CJ, Cheng C, Yang M, Di Pasquale F (2019) Dynamic phase extraction in high-SNR DAS based on UWFBGs without phase unwrapping using scalable homodyne demodulation in direct detection. Opt Express 27(8):10644–10658
8. Han P, Li Z, Chen L, Bao X (2017) A High-speed distributed ultra-weak FBG sensing system with high resolution. IEEE Photonics Technol Lett 29(15):1249–1252
9. Zhu F, Zhang Y, Xia L, Wu X, Zhang X (2015) Improved Φ-OTDR sensing system for high-precision dynamic strain measurement based on ultra-weak Fiber bragg grating array. J Lightwave Technol 33(23):4775–4780
10. Zhang X, Sun Z, Shan Y, Li Y, Wang F, Zeng J, Zhang Y (2017) A high performance distributed optical fiber sensor based on Φ-OTDR for dynamic strain measurement. IEEE Photonics J 9(3):1–12
11. Wang C, Shang Y, Liu X, Wang C, Yu H, Jiang D, Peng G (2015) Distributed OTDR-interferometric sensing network with identical ultra-weak fiber Bragg gratings. Opt Express 23(22):29038–29046
12. Cheng Z, Liu D, Yang Y, Ling T, Chen X, Zhang L, Bai J, Shen Y, Miao L, Huang W (2015) Practical phase unwrapping of interferometric fringes based on unscented Kalman filter technique. Opt Express 23(25):32337–32349
13. Zhang A, Zhang S (2016) High stability fiber-optics sensors with an improved PGC demodulation algorithm. IEEE Sens J 16(21):7681–7684
14. Muanenda Y, Faralli S, Oton CJ, Di Pasquale F (2018) Dynamic phase extraction in a modulated double-pulse ϕ-OTDR sensor using a stable homodyne demodulation in direct detection. Opt Express 26(2):687–701
15. Muanenda Y, Faralli S, Oton CJ, Di Pasquale F (2018) Stable dynamic phase demodulation in a DAS based on double-pulse ϕ-OTDR using homodyne demodulation and direct detection. Proc SPIE 10654:106540B
16. Dandridge A, Tveten AB, Giallorenzi TG (1982) Homodyne demodulation scheme for fiber optic sensors using phase generated carrier. IEEE Trans Microw Theory Tech 30(10):1635–1641

Chapter 28
Silicon Nanowires as Contact Between the Cell Membrane and CMOS Circuits

P. Piedimonte, D. A. M. Feyen, M. Mercola, E. Messina, M. Renzi and F. Palma

Abstract We describe an innovative approach to sensing bioelectric signals at high space-time resolution with low invasiveness based on growing small Silicon Nano Wires (SiNW) at low-temperature (200 °C). The resulting SiNWs are compatible with ICs, allowing on-site amplification of bioelectric signals. We report our preliminary results showing biocompatibility and neutrality of SiNWs used as seeding substrate for cells in culture. With this technology, we aim to produce a compact device allowing on-site, synched and high signal/noise recordings of a large amounts of biological signals from networks of excitable cells (e.g. neurons) or distinct subdomains of the cell membrane, thus providing super-resolved descriptions of the propagation of electric waveforms within living cells and networks.

28.1 Introduction

Ionic currents across membranes are crucial in both excitable and non-excitable cells; their accurate measurement requires efficient coupling between cell membrane and measuring electrodes. An elective (though somehow 'classic') approach

P. Piedimonte
Basic and Applied Sciences for Engineering Department, Sapienza University of Rome, Rome, Italy

D. A. M. Feyen · M. Mercola
Medicine and Cardiovascular Institute Department, Stanford University School of Medicine, Palo Alto, California, USA

E. Messina
Policlinico Umberto I, Sapienza, Rome, Italy

M. Renzi
Physiology and Pharmacology Department, Sapienza University of Rome, 00184 Rome, Italy

F. Palma (✉)
Electronics and Telecommunications Engineering Department, Sapienza University of Rome, Rome, Italy
e-mail: fabrizio.palma@uniroma1.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_28

to investigate membrane currents and (action) potentials in details, from network to single-channel activity, is the patch-clamp technique [1]. However, in most configurations patch-clamp relies on accessing (thus, perturbing) the interior of single cells, which limits the recording output both in duration and n value. Extracellular recording methods, such as multielectrode arrays [2] and multitransistor arrays [3], are noninvasive and allow long-term and multiplexed measurements. However, extracellular recording not only sacrifices the one-to-one correspondence between cells and electrodes, but also suffers significantly in signal strength and quality.

So, high-resolution investigations of the molecular mechanisms underlying cell excitability and pharmacological screening of ion-channel drugs is still usually performed by low-throughput, intracellular recording methods [4]. Optical methods based on ion indicators [5], styryl potentiometric dyes [6] and voltage-sensitive proteins [7] represent a valid alternative for their ability to deliver massively parallel recordings, and yet are still suffering for perfectible resolution and applicability. Realizing an all-electrical device for electrophysiology, a closely packed microelectrode array (MEA) capable of high-precision intracellular recording from a large network of cells has long been a major pursuit in bioengineering, neuro- and cardio-technology [8]. In this scenario, the use of nanowire transistors [9] or nanotube-coupled transistors [10] could significantly improve the strength of the signal recorded from living cells.

A most ambitious goal pursued to date was to realize all-electrical electrophysiological imaging by CMOS-MEA, which should allow massively parallel recording of cellular networks. However, attempts till now limited to adopt extracellular approaches affected by relatively low sensitivity. Recently, it has been shown that vertical nanopillar electrodes can record both the extracellular and the intracellular action potentials of cardiomyocytes in culture over a long period of time with excellent signal strength and quality [11]. Also, it was possible to repeatedly switch between extracellular and intracellular recording by nanoscale electroporation/resealing processes and to detect subtle changes of action potentials induced by ion channel drugs.

However, the combination of this technique with large-scale integration typical of microelectronics has not been attempted yet, due to the difficulty to combine the usual CMOS technology with the nanotechnology needed to grow small-sized nanowires. MEMS technology capable to create pillars and wires, cannot reach the minimal dimension required, and cannot be extended to the large area of wafer used in the IC technology. The characteristics of the device we intend to set up (high-density, small-sized sensing units with unprecedented sensitivity - 6 electron/s - and SiNWs deposited after the creation of the IC) would thus allow to disclose unparalleled in-deep description of the integration and propagation of bio-electrical signals, from sub-neuronal domains to a superior mapping of the spike activity in complex networks, and eventually enabling the adoption of novel algorithms for spike waveforms identification.

28.2 Description of the Integrated Circuit Structure

The project we propose aims to realize a CMOS-nanoelectrode array (CMOS MEA) designed to work as a high fidelity, all-electrical imager amenable for thousands of parallel recordings with highest temporal and spatial resolution. By doing so, our CMOS MEA will bridge the gap between currently available MEAs (ensemble recordings at relatively low resolution) and patch-clamp (highly-resolved recordings limited to single cells). In particular, we intend to produce a compact device allowing on-site, synched, high signal/noise recordings of very large amounts of biological signals from e.g. neurons or, as a further step, hundreds of spots on the surface of any (excitable) cell, thus providing super-resolved descriptions of the electric waveforms along cellular microdomains. A main caveat of this approach is the signal dampening due to the screening effect of the extracellular medium. We intend to overcome this limitation by using nano-structured electrodes. In particular, we will grow silicon nanowires (SiNWs) directly on the surface of already existing image-sensing devices (Fig. 28.1). Conductive SiNWs will mirror the charge at membrane surface onto the surface of the underlying CMOS integrated circuit (IC). This will perform the on-site amplification of the bioelectric signals, which we predict to be unaltered thanks to the intimate contact nanowire/cell membrane. Moreover, such tight interaction would predictably allow to switch repeatedly between extracellular and intracellular recording by nanoscale electroporation and resealing processes.

In practice, three main preliminary requirements needed to be verified prior to considering going ahead with the implementation of our device: (i) adapting the CMOS-IC to accommodate the growth of SiNWs must not interfere with its signal sensitivity; (ii) SiNWs must be broadly bio-compatible; and, (iii) SiNWs must indeed interact very tightly with the cell membrane, so to predictably by-pass the problem of signal dampening.

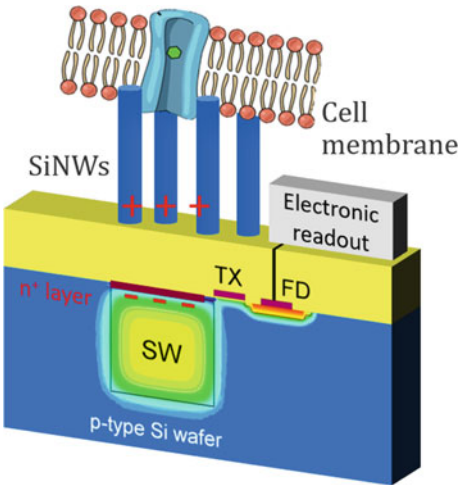


Fig. 28.1 Sketch of the nanowires grown on the back-side of the LFoundry pixel structure, with the addressing transistors schematically shown

Here, we show our preliminary results addressing these points.

28.3 Experimental Results

28.3.1 SiNWs on Custom-Designed ICs

We plan to use custom-designed ICs (e.g. LFoundry technology LF11iS-BSI, 0.11-μm CMOS with 4 Mpixels, 2 μm pitch). The technology does involve the thinning of the back surface of the chip, with the active structure moved in close proximity of the chip surface. Of note, in this first phase, there will be no need of intervention on the pixel design aimed to optimize its sensitivity to the external charge. The sensitivity tests we run made us confident that the pixels bearing the back thinning technology do already have a degree of sensitivity sufficient to demonstrate the feasibility of the idea (data not shown).

Growth of SiNWs with 15 nm-diameter (Fig. 28.2) is obtained on a regular basis at the plant installed at the Sapienza Laboratory for Nanotechnology and Nanoscience using the non-invasive, low temperature (200 °C) process known as MWCVD (MicroWave Chemical Vapor Deposition) developed at CNIS, Sapienza [12].

We plan to optimize the MWCVD process to minimize the possible alteration of the detection chip and to standardize the quality of SiNWs produced.

The chip will then be packaged to allow the connection to an interface board used for charge image-data transfer. Notably, the use of an open package will allow for cell culturing directly on the surface of the chip.

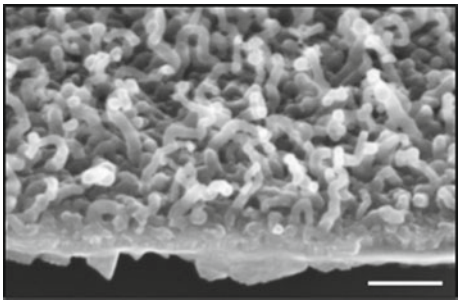


Fig. 28.2 Typical Silicon Nano Wires grown at low temperature (200 °C) and used as seeding substrate for cells in culture. Scale bar: 200 nm

28.3.2 Broad Biocompatibility of Silicon Nanowires

Our preliminary results [13–15] show that different cell types have unaltered morphology and functional properties when seeded on SiNWs compared to control condition.

Current-clamp recordings from NG18CC15 neural like cells on SiNWs showed passive properties and excitability profile typical of cells on control substrates. Likewise, voltage-clamp experiments from BV-2 cells revealed same membrane current profile and density across different substrates.

We also tested BV-2 cells grown on SiNWs (both silicon and silicon oxide substrates) using Ca^{2+} epifluorescence imaging and found that both basal intracellular $[Ca^{2+}]$ and ATP-elicited $[Ca^{2+}]_i$ rise were typical of BV-2 cells in physiological conditions.

Primary hippocampal neurons and microglial cells from mice also showed biocompatibility with SiNWs as shown by immunofluorescence.

To further broaden our characterization of SiNW biocompatibility we tested iPS-derived human cardiomyocytes (at Stanford University School of Medicine, CA).

iPS-derived cardiomyocytes developed normally on SiNW substrates and showed good adhesion to the seeding surface. Notably, we could record typical calcium fluxes (not shown) and contractile activity from the iPSC-CMs grown on the nanowires (Fig. 28.3), indicating that SiNWs are compatible with regular growth and physiological behavior of these human-derived cells.

Beside the functional studies on biocompatibility, the quality of the cell membrane/nanostructures interface is crucial to design bio-devices. Recent findings in the fabrication of artificial bio-interfaces added much to our understanding of such interface; however, it is still open the question on how the cell membrane accommodate the presence of sharp objects at the nano-scale.

To address this, we investigated the morphology of the cell/SiNWs interface at high-resolution using SEM on fixed cultures grown for 48 h on SiNWs in standard

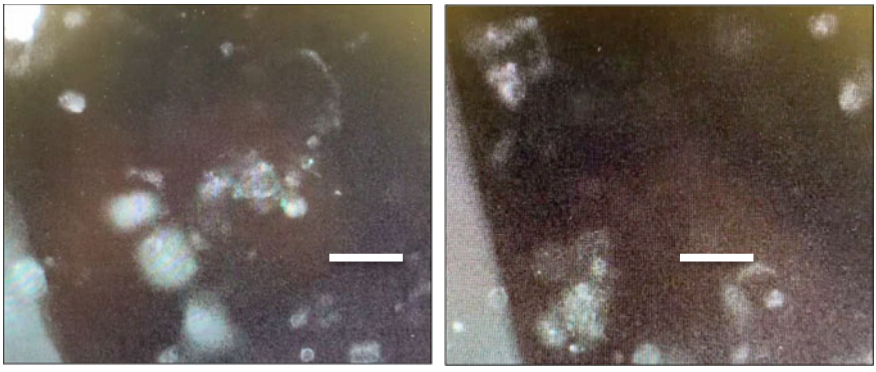


Fig. 28.3 Beating cardyomiocites before (left) and after (right) contraction (Scale bars: 20 μm)

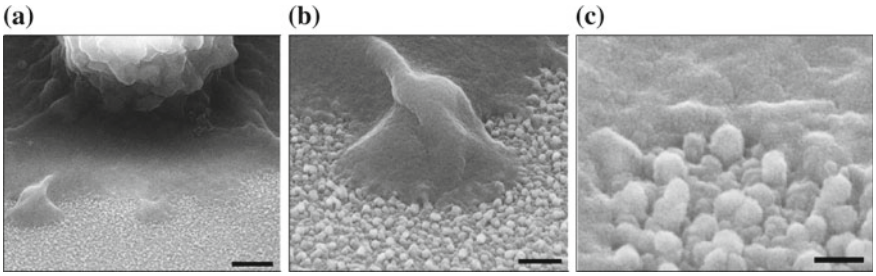


Fig. 28.4 SEM images of GH4C1 neuron-like cell cultures on silicon nanostructures. (Scale bars: **a** 1 μm, **b** 200 nm, **c** 50 nm)

conditions. Figure 28.3 depicts SEM of GH4C1 neuron-like cells and Fig. 28.4 BV-2 microglial cells on silicon nanowires.

First, we noticed that both culturing (and fixative) standard procedures did not interfere with the presence of SiNWs, indicating that intact nanostructures were present and preserved also during our functional recordings. In fact both individual cells and SiNWs could be readily and clearly identified and appeared unaffected using SEM. Furthermore, the overall cell morphology appeared unaltered and the cellular membrane is shown to interact very closely and tightly to the engineered substrates. Our SEM images show that independent of the nanostructure size and orientation the cellular membrane tends to grip to and engulf the substrate along its full profile, including the sharp edge at the nanowires base (Fig. 28.5).

Altogether, Silicon NanoWires do interact tightly with cell membranes and do not appear to alter normal survival, morphology and functional properties of several cell types in vitro, thus resulting amenable for non-interfered biological measures and conditioning.

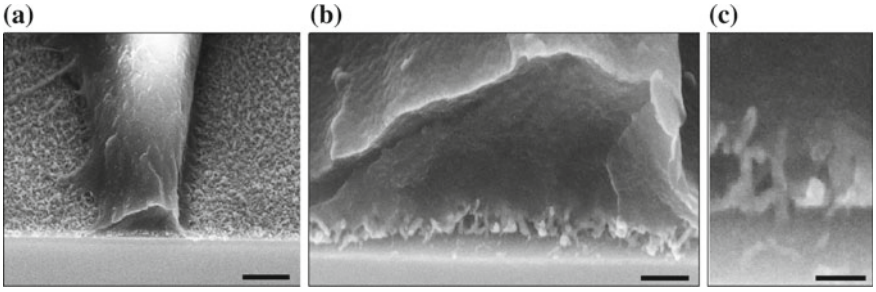


Fig. 28.5 SEM images of BV-2 microglial cell cultures on silicon nanowires. (Scale bars: **a** 1 μm, **b** 200 nm, **c** 50 nm)

28.4 Conclusions

We demonstrated that SiNWs grown on the back-surface of an image-sensing chip are compatible with normal growth and physiological response of several cell types, both excitable (neurons; cardiomyocites) and non-excitable (microglia; GH4C1), obtained from immortalized cell-lines, mouse primary cultures, or human iPSCs. Also, we demonstrated that indeed the interaction between SiNWs and cell membranes is extremely tight, thus predictably allowing little or no signal filtering.

We believe that, though still preliminary, our piece of evidence represent an encouraging, crucial step on the road of setting-up a compact device allowing on-site, synched and high signal/noise recordings from living cells or tissues (e.g. brain slices). Provided that we will success, possible future applications of our novel technology might be (i) to investigate and condition, at highest resolution, propagation and integration of electrical signals in living cells; and (ii) the design of prosthetic implants working as engineered interfaces.

Acknowledgements Authors wish to thank LFoundry for the information on the technology LF11iS-BSI.

References

1. Sakmann B, Neher E (2009) Single-Channel Recording, 2nd edn. Springer, New York
2. Pine J (1980) Recording action potentials from cultured neurons with extracellular microcircuit electrodes. J Neurosci Methods 2:19–31 [PubMed: 7329089]
3. Lambacher A et al (2004) Electrical imaging of neuronal activity by multi-transistor-array. Appl Phys Mater Sci Process 79:1607–1611
4. Zheng W, Spencer RH, Kiss L (2004) High throughput assay technologies for ion channel drug discovery. Assay Drug Dev Technol 2:543–552 [PubMed: 15671652]
5. Herron TJ, Lee P, Jalife J (2012) Optical imaging of voltage and calcium in cardiac cells & tissues. J Circ Res 110:609–623
6. Cheng H, Lederer WJ, Cannell MB (1993) Calcium sparks: elementary events underlying excitation-contraction coupling in heart muscle. Science 262:740–744
7. Matiukas A et al (2007) Near-infrared voltage-sensitive fluorescent dyes optimized for optical mapping in blood-perfused myocardium. Heart Rhythm 4:1441–1451
8. Fromherz P (2002) Electrical interfacing of nerve cells and semiconductor chips. ChemPhysChem 3:276–284
9. Timko BP et al (2009) Electrical recording from hearts with flexible nanowire device arrays. Nano Lett 9:914–918 [PubMed: 19170614]
10. Duan X et al (2011) Intracellular recordings of action potentials by an extracellular nanoscale field-effect transistor. Nat Nanotechnol
11. Xie C, Lin Z, Hanson L, Cui Y, Cui B (2012) Intracellular recording of action potentials by nanopillar electroporation. Nat Nanotechnol 7(3):185
12. Patent filed on the 22/09/2017, Ref code: IT0549-17- UNIVERSITÀ DEGLI STUDI DI ROMA-CB)
13. Piedimonte P et al (2019) Silicon nanowires to detect electric signals from living cells. Mater Res Express 6(8)

14. Piedimonte P et al (2019) Silicon Nanowires as Biocompatibile Electronics-Biology Interface. In: 2019 20th International conference on solid-state sensors, actuators and microsystems & eurosensors XXXIII
15. Piedimonte P et al (2019) Biocompatibility of silicon nanowires: A step towards IC detectors. In: Nanoinnovation 2018- AIP conference proceedings

Chapter 29
Ultra-Low Power Displacement Sensor

Alessandro Bertacchini, Marco Lasagni and Gabriele Sereni

Abstract In this paper a proof of concept of ultra-low power Eddy-Current Displacement Sensor (EDCS) is presented. Measurements carried out on the realized prototype, show that the sensor has a resolution up to 6 μm, with a power consumption of 28 μW. The ultra-low power consumption of the sensor could pave the road towards the realization of a new generation of smart wireless displacement and proximity sensors able to gather the needed energy directly from the environment where they operate thanks to the introduction of energy harvesting-based devices.

29.1 Introduction

The demand of displacement sensors is rapidly increasing with the proliferation of industrial automation. At the same time, there is a growing request for smart, lower power, wireless and low-cost devices. Displacement sensors have an important role in many applications because of their intrinsic capabilities to estimate indirectly also other quantities such as pressure, acceleration, etc. For example, they can be used to easily obtain redundant sensors in several safety critical applications.

These sensors exploit mainly optical, capacitive, and Eddy-Current principles. Optical sensors have usually high resolution with large measurement range, but they are often expensive, sensitive to optical contaminations, and above all, power hungry. Capacitive sensors can achieve extremely high resolution, but target grounding

A. Bertacchini (✉)
DISMI—Department of Sciences and Methods for Engineering,
University of Modena and Reggio Emilia, 42122 Reggio Emilia, Italy
e-mail: alessandro.bertacchini@unimore.it

M. Lasagni · G. Sereni
IndioTECH srl, via Roma 4, 42014 Castellarano, Reggio Emilia, Italy
e-mail: marco.lasagni@indiotech.com

G. Sereni
e-mail: gabriele.sereni@indiotech.com

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_29

is necessary and they are sensitive to the properties of dielectrics placed in the measuring gap. Conversely, Eddy-Current Sensors are contactless devices operating on the principle of magnetic induction, and they can precisely measure the position (displacement or proximity), x, of a metallic target in contaminated environments (e.g. dust, oil particles, etc.) and also through non-metallic materials such as plastics, dirt, etc. The main drawback of these sensors is the thermal drift that under uncontrolled conditions can affect the measurement.

Moreover, as mentioned previously, Eddy-Current Sensors are widely used in industrial applications not only for direct displacement/proximity measurements but also to estimate material property and detect crack fatigue inspection [1–4], bearing wear [5], thicknesses [6, 7], etc.

29.1.1 Main Contributions of This Work

Different methods have been developed to enhance the resolution and reduce the thermal drift, such as measuring the working frequency, precise amplitude demodulation, etc. However, none of these methods can ensure ultra-low power performance. In fact, many works [8–11] show that the continuous power consumption cannot be lower than 5 mW. This amount of power can be usually considered low in many applications but in many others where the sensing device is battery-powered (e.g. wireless sensor nodes) leads to an unacceptable rate of the battery substitution. The purpose of this work is to demonstrate that it is possible to obtain a low cost 10 μm-resolution ECDS with power consumption in the order of tens of microwatts. In order to eliminate any thermal drift issue, in this first implementation, all the measurements have been carried out at constant temperature. In future works, of course, once a complete smart sensor device will be realized by adding an ultra-low power wireless microcontroller, a temperature compensation algorithm can be included.

29.2 Operating Principle and System Description

The operating principle of a typical ECS is based on magnetic induction. The main components of the sensor are a conductive target, a sensing coil and an electronic circuit interface, as sketched in Fig. 29.1 *left*. The sensor coil, driven by an ad hoc AC current, generates an alternating magnetic field, which concatenates with the nearby conductive target inducing Eddy Currents. In turn, Eddy Currents generate a magnetic field, which is opposite to the one generated by the coil. This causes a magnetic flux reduction and energy dissipation in the sensor coil.

With reference to Fig. 29.1 *center*, the coil-target air coupling can be considered as an equivalent transformer. The primary of the transformer is the sensing coil and is comprised of the inductor L_x and the series resistor R_x. The secondary of the

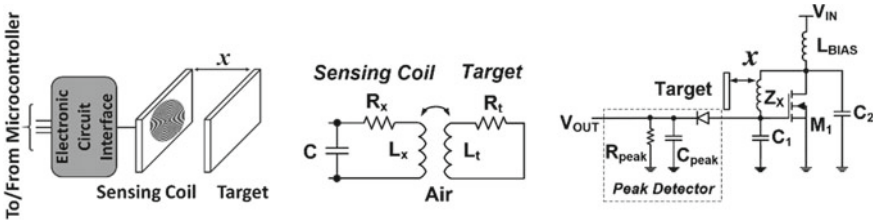


Fig. 29.1 Eddy current sensing system: simplified operating principle of the sensor (*left*), transformer model and equivalent circuit (*center*) and implemented electronic circuit interface (*right*)

transformer, in which the Eddy Current flows, is comprised of L_t and R_t, representing the target.

The proximity of the conductive target influences both the inductance L_x and the series resistance R_x of the sensing coil. In particular, by moving the target closer to the sensing coil, i.e. the distance x decreases, L_x increases, whereas R_x decreases. Vice versa if the target moves away from the sensing coil, i.e. the distance x increases, L_x decreases and R_x increases. Simply by adding a capacitance C in parallel to L_x, it is possible to form an LC-tank oscillator needed to generate the alternating magnetic field. A change of x results, in a variation of the equivalent sensor impedance, $Z_{eq} = Z_x//\mathrm{C}$, where $Z_x = R_x + j\omega L_x$. The relationship between Z_x and the target position x depends on the characteristics of the sensor coil, on the working frequency, and on the properties of the target.

In order to guarantee the proper operation of the LC-tank oscillator by keeping the oscillation stable, an active circuit acting as a negative resistance is needed to restore the power loss in the tank as a consequence of the change in R_x due to the x variations.

Several circuit interfaces have been presented in the literature (e.g. [8–12]), but all the proposed solutions show a continuous power consumption larger than 5 mW, which is quite high for battery-operated systems. In order to obtain a sensor with good resolution (few micrometers) with power consumption in the order of tens of microwatts, the simple circuit shown in (Fig. 29.1 *right*) has been realized. It combines an LC-tank oscillator with a peak detector providing an output voltage proportional to the distance between the sensing coil and the target.

By exploiting the results of LTSpice simulations, the optimized circuit has been implemented by using a zero threshold MOSFET ALD212900 as M_1 and high-quality passive devices. The chosen M_1 has both high forward transconductance and output conductance at very low supply voltages, allowing lowering the voltage supply level of the circuit, and reducing consequently the circuit power consumption. The inductor chosen as L_x (22R106C Murata) has a high Q (>140) at the circuit oscillation frequency (≈160 kHz). To ensure the validity of the Barkausen's criteria needed for a proper oscillation of the LC-Tank, $C_1 = 10$ nF and $C_2 = 100$ pF has been chosen, while the peak detector components used were $R_{peak} = 882$ kΩ and $\mathrm{C}_{peak} = 100$ nF with a classic 1N4148 diode. Finally, $L_{BIAS} = 65$ mH has been used.

29.3 Measurement Setup and Experimental Results

All the test have been carried out by exploiting the setup sketched in Fig. 29.2 *left*. The implemented Electronic Interface Circuit (EIC) has been fixed by means of a mounting bracket to the fixed end of a commercial manual outside micrometer, while the target has been positioned on its own mounting bracket fixed to the moving end of the micrometer. In this way, the distance between target and sensing coil can be varied of micrometers in a repeatable way. The target has been realized using common FR4 for PCB production with a 35 μm copper layer and an area of 15 × 15 mm, larger than the diameter of the inductor used as sensing coil. The EIC's output voltage, V_{OUT}, has been measured by means of an Agilent DSO9254A oscilloscope, while the EIC's supply voltage has been provided by an Agilent N6715B DC Power Analyzer to measure precisely also the power consumption of the sensor.

In order to limit any thermal drift issues, all the tests have been carried out at the same temperature ($T_{room} = 23$ °C).

Figure 29.3 shows an example of V_{OUT} over time for a given dynamic micrometric-displacement profile of the target with respect to the coil in case of EIC's supply voltage of 100 mV. The same behavior over time under the same dynamic displacement profile has been obtained for different EIC's supply voltage in the range of 75–200 mV. In particular, Fig. 29.3 shows how V_{OUT} changes in response to different displacements in the range 0–5000 μm. As discussed previously, when the target is close to the sensing coil, the oscillation amplitude decreases due to the larger power losses, with a consequent decrease of V_{OUT}. Vice versa, by moving the target away from the coil V_{OUT} increases. Results show good linearity in the range 0–3000 μm (Fig. 29.3 *right*).

The sensor's sensitivity S, expressed in [mV/μm], can be defined as $S = \Delta V_n/\Delta x_n$ by discretizing the whole measurement range Δx in n sub-ranges. In the same way, the resolution R, expressed in [μm], can be defined as $R = N_n/S$. Where N_n is the

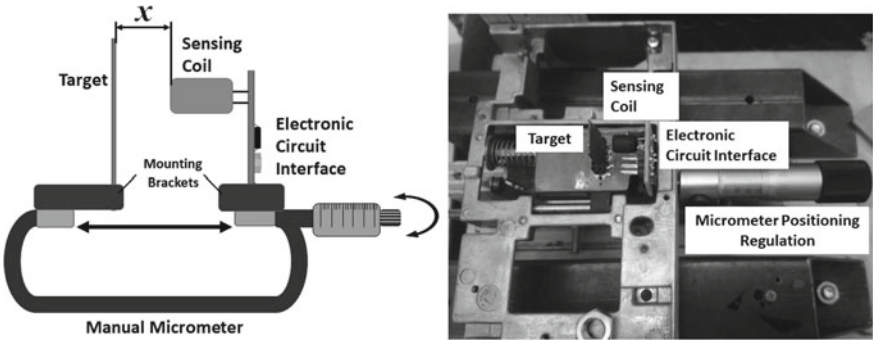


Fig. 29.2 Measurement setup. Simplified sketch with mounting brackets omitted (*left*) and real setup (*right*). By rotating the micrometer's knob it is possible to change the distance between sensing coil and target in a controlled and repeatable way

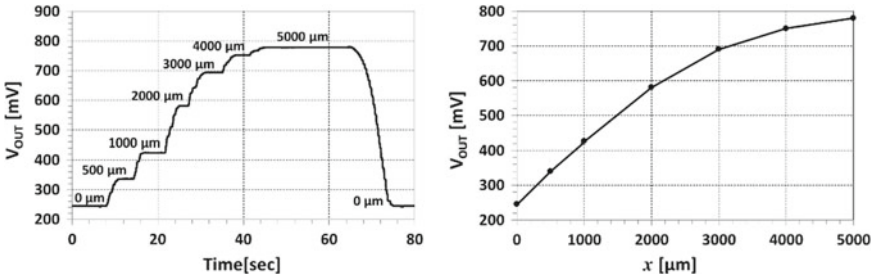


Fig. 29.3 Example of Sensor Output Voltage, V_{OUT}, over time during micrometric target's displacements x with respect to the coil (*left*) and V_{OUT} versus. x (*right*). In the example the EIC's is supplied with $V_{IN} = 100$ mV, corresponding to an overall power consumption of $P_{IN} = 28$ µW

RMS voltage noise of the signal in the *n-th* sub-range, ΔV_n is the output voltage variation in the *n-th* sub-range and Δx_n is the considered sub-range.

As shown in Fig. 29.4, the larger is S, the better is the R of the sensor. Over the considered measurement range Δx [0–5000 µm], R is better than 10 µm over a distance range from 0 to 3 mm. In particular, in the displacement range $\Delta x_n =$ 0–500 µm, the resolution can slightly improve to 6 µm.

Figure 29.5 shows the relationship between R and the power consumption of the sensor, P_{IN}, for different supply voltages, V_{IN}, in the sub-range $\Delta x_n = 0$–500 µm, that is the most interesting one for many industrial applications. As it is possible to note, an increase of V_{IN}, hence of P_{IN}, improves the achievable resolution. In particular, with $V_{IN} = 200$ mV, P_{IN} rises to 140 µW, but the resolution improves down to 3 µm. For the considered subrange an acceptable linearity R-P_{IN} can be achieved for a V_{IN} in the range 100–200 mV. Similar results can be obtained for the other considered Δx_n sub-ranges.

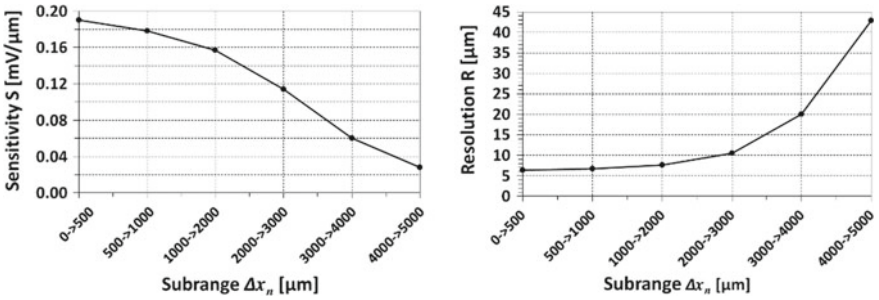


Fig. 29.4 Voltage-referred Sensitivity (*left*) and Resolution (*right*) with respect to the Δx_n in case of EIC's fed by a 100 mV supply voltage corresponding to a power consumption of only 28 µW

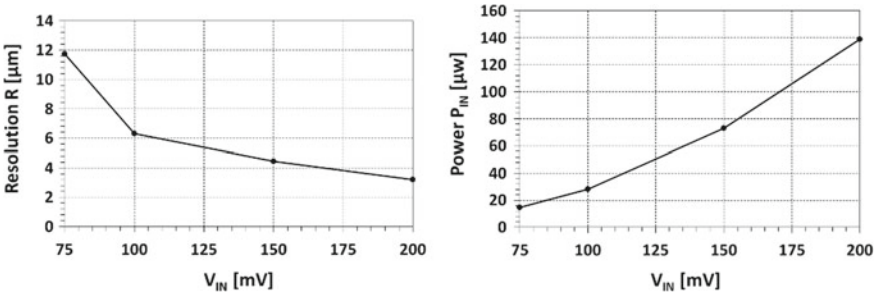


Fig. 29.5 Power consumption P_{IN} (left) and resolution R (right) of the realized electronic interface circuit for different supply voltages V_{IN} in the $\Delta x_n = 0$–500 μm sub-range

29.4 Conclusions

An ultra-low power Eddy-Current displacement sensor has been presented in this paper. The measured prototype shows a resolution of 6 μm with an input supply voltage of 100 mV and a total power consumption of only 28 μW.

The extremely low power consumption could be the enabling factor towards the realization of a new generation of displacement sensors for several industrial applications. The new sensors indeed could be self-powered by taking advantages from the state-of-the-art energy harvesting techniques and gather their needed energy directly from the environment where they operate. At the same time, the ultra-low power consumption allows designers adding wireless capabilities to the new sensors widening their applications range.

From an application point of view, a complete characterization over temperature needs to be carried out in order to limit the effect of thermal drift that is the main drawback of Eddy Current Sensors. In this way, useful information needed to implement a self-compensation algorithm of the measurements can be collected.

References

1. Johnston DP, Buck JA, Underhill PR, Morelli JE, Krause TW (2018) Pulsed eddy-current detection of loose parts in steam generators. IEEE Sens J 18(6):2506–2512
2. Alatawneh N, Underhill PR, Krause TW (2018) Low-frequency eddy current testing for detection of subsurface cracks in CF-188 stub flange. IEEE Sens J 18(4):1568–1575
3. Stott CA, Underhill PR, Babbar VK, Krause TW (2018) Pulsed eddy current detection of cracks in multilayer aluminum lap joints. IEEE Sens J 15(2):956–962
4. Pereira D, Clarke TGR (2015) Modeling and design optimization of an eddy current sensor for superficial and subsuperficial crack detection in inconel claddings. IEEE Sens J 15(2):1287–1292
5. Yamaguchi T, Ueda M (2007) An active sensor for monitoring bearing wear by means of an eddy current displacement sensor. Meas Sci Technol 18(1):311–317
6. Cheng W (2017) Thickness measurement of metal plates using swept frequency eddy current testing and impedance normalization. IEEE Sens J 17(14):4558–4569

7. Li W, Ye Y, Zhang K, Feng ZH (2017) A thickness measurement system for metal films based on eddy-current method with phase detection. IEEE Trans Ind Electron 64(5):3940–3949
8. Nabavi MR, Nihtianov S (2011) Eddy-current sensor interface for advanced industrial applications. Ind Ele IEEE Trans 58(9):4414–4423
9. Nabavi MR, Pertijs MAP, Nihtianov S (2013) An interface for eddy-current displacement sensors with 15-bit resolution and 20 mHz excitation. IEEE Solid-State Circ J (48)11
10. Wang H, Liu H, Li W, Feng Z (2014) Design of ultrastable and high resolution eddy-current displacement sensor system. In: IECON 2014 annual conference of the IEEE, pp 2333–2339
11. Welsby SD, Hitz T (1997) True position measurement with eddy current technology. Sens J Appl Sens Tech 14(11):30–41
12. Nabavi MR, Nihtianov SN (2012) Design strategies for eddy-current displacement sensor systems: review and recommendations. IEEE Sens J 12(12):3346–3355

Chapter 30
Simulation of an Optical-to-Digital Converter for High Frequency FBG Interrogator

Vincenzo Romano Marrazzo, Francesco Fienga, Michele Riccio, Luca Maresca, Andrea Irace and Giovanni Breglio

Abstract In this paper, design and simulations of an optoelectronic circuit for the conversion of the optical signal, coming from an interrogation system for FBG sensors, into a digital signal, is presented. The approach is divided into an optical introduction of the interrogation system, an analog section and, finally, digital considerations. The analog processing part is mainly based on the realization of a double stage transimpedance amplifier to obtain, in the working conditions, the best performances required in terms of high gain and wide bandwidth. The output voltage from the analog section is then converted to digital via a 12-bit ADC and sent to an FPGA that processes the defined algorithm in order to obtain the needed optical-electrical linear conversion. The circuit simulations, digital stability and other consideration, including the stability to optical power variability obtained by the numerically simulated interrogation system, are performed, highlighting the peculiarities of this new type of high frequency FBG interrogator.

V. R. Marrazzo (✉) · M. Riccio · L. Maresca · A. Irace · G. Breglio
Department of Electrical Engineering and Information Technologies, University of Naples "Federico II", Naples, Italy
e-mail: vincenzoromano.marrazzo@unina.it

M. Riccio
e-mail: michele.riccio@unina.it

L. Maresca
e-mail: luca.maresca@unina.it

A. Irace
e-mail: andrea.irace@unina.it

G. Breglio
e-mail: breglio@unina.it

F. Fienga
National Institute for Nuclear Physics (INFN), Napoli Section, Naples, Italy
e-mail: francesco.fienga@na.infn.it

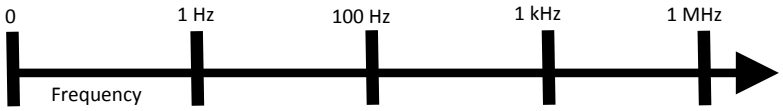
S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_30

30.1 Introduction

In the last decades, Fiber Bragg Grating (FBG) sensors have been studied and employed in several environments due to many advantages that characterize this kind of optical sensors such as: low cost, small size, immunity to electromagnetic interference, bio compatibility and no toxic material among the most important [1]. Combined with an interrogator, a FGB measurement system becomes a reliable and accurate sensing method for temperature and mechanical strain in a huge variety of fields: from minimally invasive microsurgery in which FBG sensors are employed as force sensors for the surgeon [2], to automotive field where FBG sensors are glued in a tyre monitoring circumferential and longitudinal strain [3, 4]. Despite all these advantages, the high cost of the interrogation systems limits the usage of FBG sensors: the higher is the accuracy that is needed the higher is the expensiveness of the interrogator. Furthermore, it depends also on the type of variation (Fig. 30.1): very fast variations, as impact damage detections or hydrophones, can be appreciated with an interrogation system that is optically and electrically more complex and definitely more expensive than a normal interrogator employed to monitor temperature or vibration of high buildings or bridges.

Many schemes for wavelength interrogation are reported in literature with different detection algorithm based on Fabry-Pérot or Mach-Zehnder interferometer [5, 6], spectroscopic charge coupled devices (CCD), or using the power ratio between optical filters [7, 8]. Nevertheless, no one is enough satisfactory to have appeal in the market.

In this paper a high frequency wavelength interrogation concept is presented and analyzed, focusing on the electronic circuit and its characteristics. This system is briefly descripted from the analytical point of view to understand the working principle and its optical characteristics; then circuital simulations follow, focusing on the device choice with stability considerations.



Application fields	- **Strain, temperature;** - **Surveillance monitoring for infrastructure.**	- **Vibration monitoring in SHM.**	- **Seismograph;** - **Vibration monitoring;** - **Acceleration monitoring.**	- **Hydrophones;** - **Acustic emission spectroscopy;** - **Impact damage detection.**
Detection system	- OSA; - Tunable laser & power meter.	- FP tunable filter with piezoelectric actuator.	Interrogation system without mechanical moving parts	

Fig. 30.1 Detection system in function of application field in function of frequency

30.2 Sensing System

The FBG works as an optical filter with a very narrow band whose central wavelength is called Bragg wavelength and is dependent with the modulation period of the refractive index, created inside the fiber core, and the refractive index of the mode propagating inside the core. The Bragg wavelength is very sensitive to temperature or strain variation, experiencing a wavelength shifting which has to be detected from the interrogator. The proposed interrogation concept is depicted in Fig. 30.2: a broadband spectrum generates the light irradiated towards a FBG sensor through an optical circulator. The FBG reflects part of the signal which becomes, again through the optical circulator, the input signal of the Arrayed Waveguide Grating (AWG) whose working principle is to separate a polychromatic spectrum in many output channels depending on the wavelength, as an integrated prism. Due to the AWG, when the Bragg wavelength shifts, it will space among the channels, becoming very easy to detect. Every AWG channel, four in this case study, is connected to a photodiode to transduce the signal from optical to electrical. The signal is then converted from current to voltage through a transimpedance amplifier (TIA) and digitalized with an Analog to Digital Converter (ADC), ready to be read by a FPGA performing the interrogation algorithm and detecting the wavelength deviation.

Assuming the FBG spectrum as apodized, the side lobes are eliminated and the reflectance spectrum can be approximated with a Gaussian shape [9]. The Transmittance of every AWG channel can be approximated as Gaussian as well. The output signal from the generic m_{th} AWG channel can be calculated integrating in the whole spectrum of wavelength containing FBG reflectance $B(\lambda)$, AWG transmittance $A(\lambda)$ and a parameter that take care of the light source spectrum $S(\lambda)$:

$$I_m(\lambda_{FBG}) = \int A_m(\lambda) B(\lambda) S(\lambda) d\lambda \tag{1}$$

The λ_{FBG} is in a form as $\lambda_{FBG} = F\alpha + \beta$, in linear dependence with the interrogation function F that is defined in dependence of two adjacent AWG output channels (C_m contains information about the AWG mth channel):

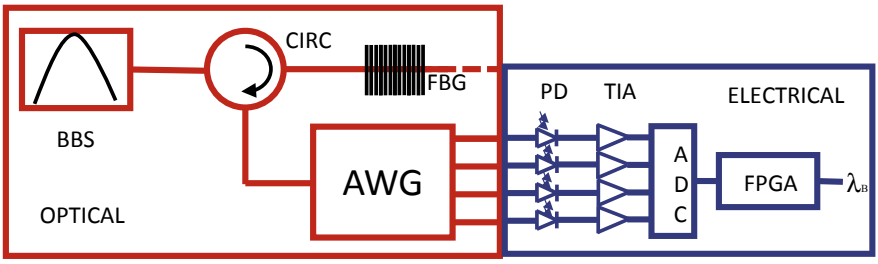


Fig. 30.2 Interrogation system block diagram

$$F_{m,m+1}(\lambda_{FBG}) = \arctan h\left(\frac{I_m C_{m+1} - I_{m+1} C_m}{I_m C_{m+1} + I_{m+1} C_m}\right) \quad (2)$$

30.3 Circuital Simulations

As shown in the previous analysis, the Arrayed Waveguide Grating is definitely the most important device in the interrogation system: from its characteristics, the resolution and reliability of the interrogator are defined. This is right from the optical point of view, but the proposed system is also composed by an electrical side which has the task to convert the information stored into an optical signal, to a digital one and cannot be ignored for the correct working of the whole interrogator. For this reason, electrical consideration must be done as well.

30.3.1 Photodiode

The first component to choose to design the Optical-to-Digital Converter (ODC) is the photodiode and the way to use it: for this application, in which the aim is to design a circuit working at about 10 MHz, the speed and so the photoconductive mode is necessary. In Photoconductive mode the photodiode is biased (typically with −5 V), this increases the dark current but reducing the parasitic capacitance and, thus, increasing speed and responsivity (SCR zone wider, hence more photons absorbed). The photodiode chosen is made of InGaAs material in order to absorb photons in C band (1550 nm), with these characteristics:

– 2 GHz response;
– 20 pA dark current at −5 V;
– 0.95 A/W responsivity at −5 V;
– 1 pF parasitic capacitance at −5 V;
– 3×10^{-15} W/Hz$^{1/2}$ Noise Equivalent Power.

These values are needed to simulate the photodiode in the circuital analysis that follows.

30.3.2 Transimpedance Amplifier

Due to the parasitic capacitance, photodiodes provide in output current at high impedance (high at DC). If this current flows into a resistor to generate a voltage, two problems will come: if high gain is needed, a large resistor is needed as well reducing the response, hence increasing time-constant; with a small resistor, the gain

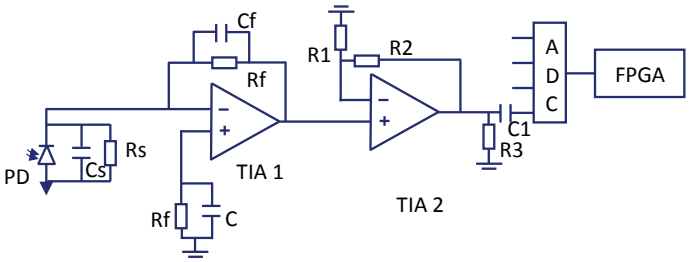


Fig. 30.3 Proposed circuit ODC whose aim is to convert optical signal in digital

is lower, increasing the speed, but the signal to noise ratio (SNR) might be unacceptable. The solution is to feed the photodiode's output current into the summing point of a transimpedance amplifier. Now the response time is not dependent on the photodiode parasitic capacitance, allowing to use large resistor for high gain and improving SNR too.

The proposed ODC is shown in Fig. 30.3: a two-stage employing low noise, low input current and low input capacitance operational amplifier is used since high gain and wide bandwidth is required. The gain can be calculated as:

$$\text{Gain (V/W)} = R_f\left(1 + \frac{R_2}{R_1}\right)R_\lambda = 121k \tag{30.3}$$

Where R_λ represents the responsivity. This value comes from the consideration that the AWG output optical power has to be converted in a voltage that matches the ADC input characteristics. The two stages include a first transimpedance preamplifier in inverting configuration for current to voltage conversion, then a non-inverting amplifier for the remaining voltage amplification. First stage gain is directly determined by Rf; second stage gain is determined by R1 and R2. An issue comes from the instability, that can affect also a simple configuration of an operational amplifier if the delay created by amplifier's input capacitance reacts with the feedback resistance. This can be avoided moving the pole created at higher frequency or deleting it with a zero. The best solution is to connect a feedback capacitor Cf in parallel with Rf limiting the frequency response and avoiding gain peaking that can lead to overshooting.

The Rf and C parallel on the non-inverting input is needed to compensate the thermal DC drift due to the temperature coefficient of the amplifier input current.

As depicted in Fig. 30.4a, b, in which impulsive response and bode diagram are shown, the optimal feedback capacitance value is 0.5 pF, allowing to have a 10 MHz bandwidth and a fast response without overshooting.

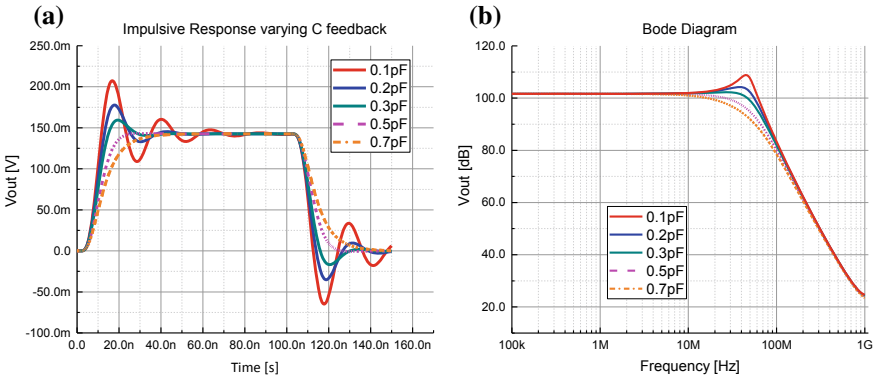


Fig. 30.4 **a** Impulsive response with many values of the feedback capacitance; **b** bode diagram with the same values of feedback capacitance

30.3.3 Digital Considerations

The second stage of transimpedance amplifier is followed by an RC filter whose aim is to neglect the intrinsic DC offset before to get the signal ready to be processed by ADC. In Fig. 30.5a, b is shown the optical power variation from the interrogator, got with a numerical simulation, and the TIA output voltage experienced employing the optical output power as current source in the ODC circuital simulation. The shape is the same as the current one (hence the optical one), with a maximum voltage of about 1 V which is the maximum value allowed by the ADC (2 Vpp, 12-bit, 65 MSPS). The ADC was simulated on MATLAB environment and implemented on Quartus Altera synthetizing the interrogation algorithm.

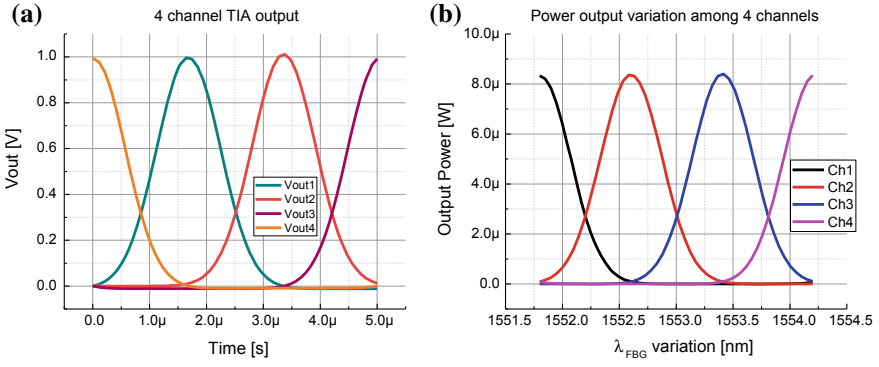


Fig. 30.5 **a** Channel output voltage from four TIA; **b** output optical power from four AWG channels during FBG interrogation

30.4 Conclusions

An Optical-to-Digital Converter has been simulated. From the results shown here, it is evident that the proposed approach is able to transduce the optical signal, which comes from some output channels of the AWG system employed in this kind of interrogation system, in voltage with a linear dependence. The amplifier gain has been chosen in function of the ADC: with a 12-bit and a maximum of 1V allowed, a TIA gain of 121 k is needed. With these characteristics, the ADC quantum is about 0.244 mV that means 2 nA in terms of current and about 2 nW in terms of optical power, which determines a wavelength resolution below one picometer. The circuit works up to 10 MHz, allowing to the proposed interrogator to become a valid competitor for this kind of measurement systems.

References

1. Culshaw B, Kersey A (2008) Fiber-optic sensing: a historical perspective. J Lightw Technol 26(9):1064–1078
2. Selvaggio M, Fontanelli GA, Marrazzo VR, Bracale U, Irace A, Breglio G, Villani L, Siciliano B, Ficuciello F (2019) The musha underactuated hand for robot-aided minimally invasive surgery. Int J Med Rob Comput Assist Surg. e1981
3. Breglio G, Fienga F, Irace A, Russo M, Strano S, Terzo M (2017) Fiber bragg gratings for strain and temperature measurements in a smart tire, lecture notes in engineering and computer science. In: Proceedings of the world congress on engineering 5–7 2017, London, U.K. pp 759–763
4. Coppo F, Pepe G, Roveri N, Carcaterra A (2017) A multisensing setup for the intelligent tire monitoring. Sensors
5. Mallinson SR (1987) Wavelength-selective filters for single-mode fiber WDM system using Fabry-Perot interferometers. Appl Opt 26:430–436
6. Usbeck K, Ecke W, Hagemann V, Mueller R, Willsch R (1999) Temperature referenced fiber bragg grating refractometer sensor for on-line quality control of petrol products. In: Proceedings 13th OFS, Kyongju, Japan, pp 163–166
7. Song M, Yin S, Ruffin PB (2000) Fiber bragg grating strain sensor demodulation with quadrature sampling of Mach-Zehnder interferometer. Appl. Opt. 39:1106–1111
8. Davis MA, Kersey AD (1994) All-fiber bragg grating strain-sensor de-modulation technique using a wavelength division coupler. Electron Lett 30:75–77
9. Tosi D, Olivero M, Perrone G, Vallan A (2009) Improved fibre bragg grating interrogation for dynamic strain measurement. Joint Meeting DGaO/SIOF, Brescia

Chapter 31
Wireless Sensors for Intraoral Force Monitoring

M. Merenda, D. Laurendi, D. Iero, D. M. D'Addona and F. G. Della Corte

Abstract A device for wireless intraoral forces monitoring is presented. Miniaturized strain gauge sensors are used for the measurements of forces applied by tongue and lips. A sensor interface IC is able to multiplex among four sensors and a low energy transmission module, equipped with an ARM Cortex–M0 core, is used for signal elaboration and remote wireless data transmission using Bluetooth® Low Energy standard protocol. The main novelty rely in the dynamic correction of the output corrupted by the prestrain issue. Moreover, the device shows a reduced dimension and the ability to transmit data wirelessly, without the use of external cables normally used in state-of-the-art intraoral monitoring devices.

31.1 Introduction

In the fields of odontology and maxillofacial surgery, information about intraoral forces could be used for monitoring of dental and occlusal pathologies, for judging the functional state of the masticatory system and for the comparison of alternative treatments in post-surgical evolution [1–3].

M. Merenda (✉) · D. Laurendi · D. Iero · F. G. Della Corte
Dipartimento di Ingegneria dell'Informazione, delle Infrastrutture e dell'Energia Sostenibile (DIIES) Università Mediterranea, Reggio Calabria, Italy
e-mail: massimo.merenda@unirc.it

D. Iero
e-mail: demetrio.iero@unirc.it

F. G. Della Corte
e-mail: francesco.dellacorte@unirc.it

HWA srl, Reggio Calabria, Italy

D. M. D'Addona
DICMaPI, Università degli Studi di Napoli Federico II, Piazzale V. Tecchio 80, 80125 Naples, Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_31

A peculiar characteristic of a sensor for monitoring intraoral forces is, clearly, the dimensions [4]. In fact, it must be either positioned inside the mouth or in contact with a very limited surface, such as that of the tooth.

Another characteristic is the resolution of the sensor, which must detect forces of few grams [5]. Sensors should be compatible with standard CMOS technology or industrialized processes [4, 6–8].

Last requirement for this kind of device is to overcome the prestrain problem that affects every strain gauge sensor, that is caused by their mechanical placement and led to an altered rest condition.

31.2 System Description

With the aim of creating a completely wireless and size constrained system, a custom circuit was conceived, designed and prototyped with a form factor of 2 × 1 cm, as shown in Fig. 31.1. The system will be embedded using an EPO-TEK® MED-301 biocompatible epoxy from Epoxy Technology Inc. to be used inside human mouth. The circuit consists of four main blocks (Fig. 31.3):

- **Sensors**: an analysis of the state-of-the-art literature and the search for the most performing electronic components available on the market [4], led us to the selection of the model 015LW by VPG Inc. (Fig. 31.2), a 120 Ω strain gauge with sizes of 1.90 × 1.37 mm. The sensor block also includes a MEMS accelerometer for future use (Fig. 31.3).
- **Power Conditioning**: it contains the supply source regulation block and a DC/DC converter able to boost a 1.5 V coin battery source.

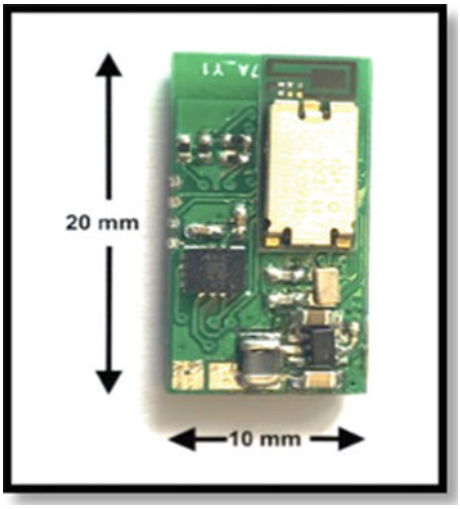


Fig. 31.1 PCB overview

Fig. 31.2 015LW by VPG Inc

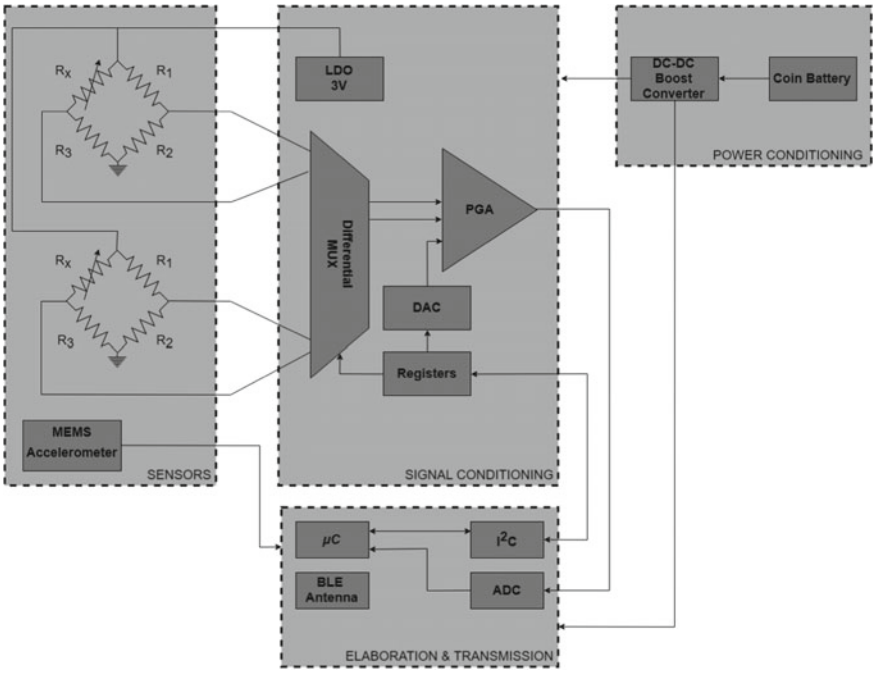


Fig. 31.3 Block diagram of the system

- **Sensor conditioning**: it contains a sensor interface IC with 16:1 differential multiplexer for interfacing multiple bridge sensors. It connects the output of one of the 4 bridges to a programmable gain amplifier (PGA). The prestrain problem is compensated and overcome by using a 10-bit DAC which dynamically generates an offset voltage added to the sensor signal, equal and opposite to that generated by the effect of the prestrain.
- **Elaboration and transmission**: The PCB host the SoC EYAGJNZXX by Taiyo Yuden, an ANT + Bluetooth® low energy transmission module with an ARM Cortex-M0 core. This block allows the connection with a hub or external smartphone application (Fig. 31.4). The Bluetooth was selected for the characteristics of extreme low power consumption required by the protocol and the reduced size of the module.

It is possible to calculate the variation of the resistance ΔR as shown below:

$$\Delta R = R[4(V_{OUT} - 1.5) \mp 4GV_{DAC}]/[GV_{BRDG} - 2(V_{OUT} - 1.5) \pm 2GV_{DAC}] \tag{31.1}$$

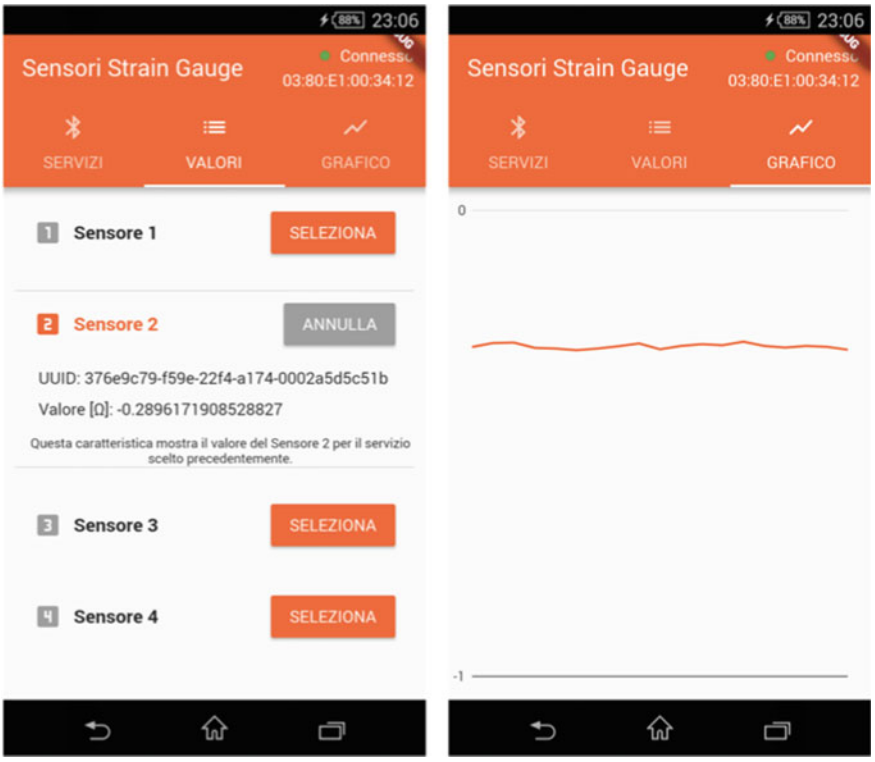


Fig. 31.4 Application screenshoots

Fig. 31.5 Application of a strain gauge sensor in an intraoral appliance

where R is the nominal value of the sensors resistance ($R = R_1 = R_2 = R_3$), V_{BRDG} is the supply voltage of the bridges (3 V), G is the gain of the PGA, V_{DAC} is the output of the 10-bit DAC, V_{OUT} is the value read by the ADC of the microcontroller and $\Delta R = R_X – R$.

31.3 Experimental Analysis

Due to the high amplification value of the PGA (from 2 to 760 V/V), prestrain [9] could saturate the output leading to a corruption of the measurements of the forces coming from the strain gauges, embedded in resin as shown in Fig. 31.5. With the use of a proper software routine, the DAC output is adjusted and can dynamically compensate for prestrain issues, as shown in Fig. 31.6.

The measurements are sent to a Smartphone using a Bluetooth connection, and showed in a custom application where the user can select, for each single sensor, to see raw or elaborated data, showed in Fig. 31.4.

31.4 Conclusions

In this work, a wireless intraoral sensor device has been proposed. It is well suited to extrapolate information about intraoral forces because of its reduced size, the use of BLE protocol instead of wire communications and the ability to dynamically compensate for prestrain issues.

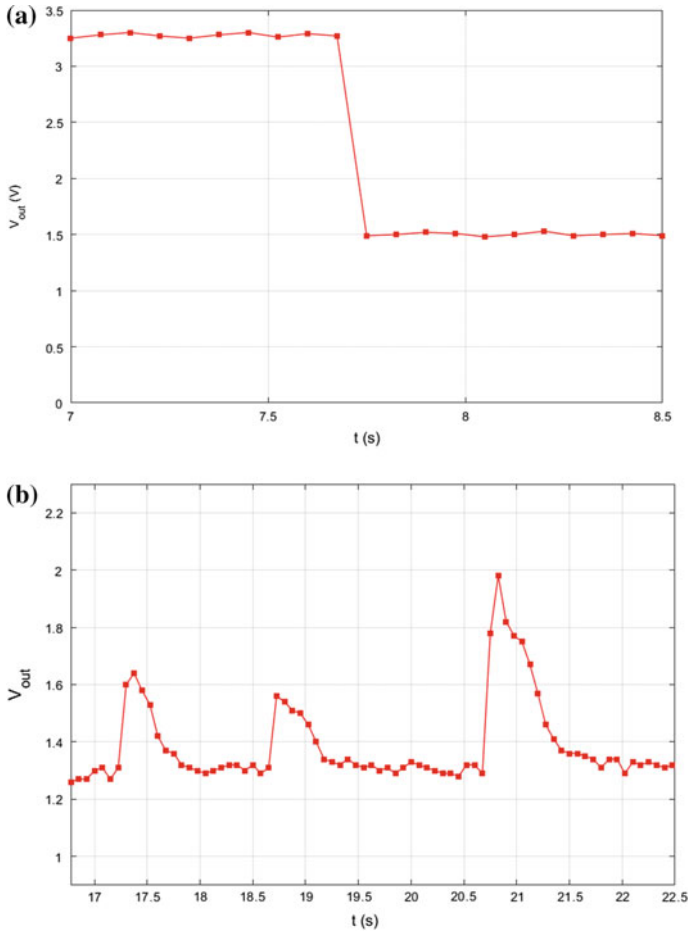


Fig. 31.6 Saturation of the output (**a**) and prestrain overcome (**b**). Voltage output of the signal conditioning block after DAC offset addition

Acknowledgements The research results presented in this paper are based on the activities carried out in the framework of the project "MoSSY—Cyber Physical System Technology for the Monitoring of Stomatognathic System" (00008– ALTRI_DR_408_2017_Ric.di_Aten-DADDONA) funded by the University of Naples Federico II within the "Programma per il finanziamento della ricerca di Ateneo" (2016–2019).

References

1. Koc D, Dogan A, Bek B (2010) Bite force and influential factors on bite measurement: a literature review. Eur J Dent 4:223–232

2. Pereira LJ, Pastore MG, Bonjardim LR, Castelo PM, Gavião MBD (2007) Molar bite force and its correlation with signs of temporomandibular dysfunction in mixed and permanent dentition. J Oral Rehabil 34:759–766
3. Lantada AD, Bris CG, Morgado PL, Maudes JS (2012) Novel system for bite-force sensing and monitoring based on magnetic near field communication. Sensors 12(9):11544–11558
4. D'Addona DM, Merenda M, Della Corte FG (2019) Electronic sensors for intraoral force monitoring: state-of-the-art and comparison. Procedia CIRP 79:730–733
5. D'Addona DM, Rongo R, Teti R, Martina R (2018) Bio-compatible cyber-physical system for cloud-based customizable sensor monitoring of pressure conditions. Procedia CIRP 67:150–155
6. Aquilino F, Della Corte FG, Fragomeni L, Merenda M, Zito F (2009) CMOS fully-integrated wireless temperature sensors with on-chip antenna. In: European microwave week 2009, 39th European microwave conference, art. no. 5296138, pp 1117–1120
7. Merenda M, Felini C, Della Corte FG (2018) A monolithic multisensor microchip with complete on-chip RF front-end. Sensors (Switzerland), 18(1). Article no 110
8. Aquilino F, Della Corte FG, Merenda M, Zito F (2008) Fully-integrated wireless temperature sensor with on-chip antenna. In: Proceedings of IEEE sensors, Article no 4716552, pp 760–763
9. Rees DWA (1986) The sensitivity of strain gauges when used in the plastic range. Int J Plast 2(3):295–309

Part VII
Power and High Voltage Electronics

Chapter 32
Reinforced Galvanic Isolation: Integrated Approaches to Go Beyond 20-kV Surge Voltage (invited)

Egidio Ragonese, Nunzio Spina, Alessandro Parisi and Giuseppe Palmisano

Abstract This paper provides a survey about alternative approaches to implement silicon–integrated galvanic isolators with very high isolation rating (i.e., compliant with the reinforced isolation requirements). Traditional integrated galvanic isolators are based on chip–scale isolation capacitors or transformers, whose performance is limited by the adopted isolation technology (i.e., the dielectric material and its thickness). In this paper, two approaches for data and power transfer are discussed, which exploit the RF coupling between two isolated interfaces, while packaging/assembling techniques are used to guarantee high galvanic isolation.

32.1 Introduction

Reliability and safety issues require galvanic isolation in several application fields, such as the automotive (i.e., electric and hybrid vehicles) the industrial (i.e., motor control, automation, etc.), the medical (i.e., implanted devices, defibrillators, patient monitoring, etc.), the consumer (i.e., home appliance, inductive cooking, etc.) and the communication one (i.e., sensors, wire line networks, etc.). A general block-diagram of a galvanically isolated system is depicted in Fig. 32.1a. Data signals are transferred across the galvanic isolation barrier to enable bidirectional communication between the two interfaces A and B, while an isolated power supply for interface B is provided from interface A by a power transfer technique. Recent standardization

E. Ragonese (✉) · G. Palmisano
DIEEI, Università di Catania, Viale A. Doria 6, 95125 Catania, Italy
e-mail: egidio.ragonese@unict.it

G. Palmisano
e-mail: giuseppe.palmisano@unict.it

N. Spina · A. Parisi
STMicroelectronics, Stradale Primosole 50, 95121 Catania, Italy
e-mail: nunzio.spina@st.com

A. Parisi
e-mail: alessandro.parisi@st.com

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_32

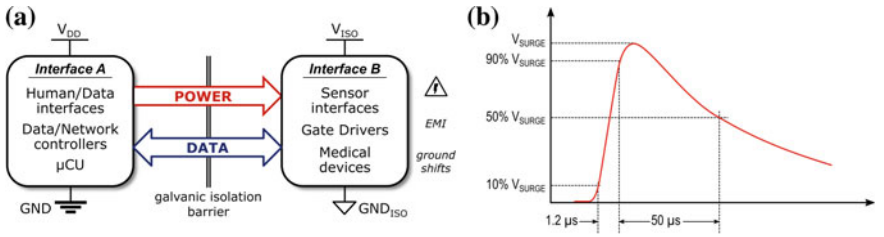


Fig. 32.1 **a** Simplified block-diagram of a galvanically isolated system. **b** Simplified surge test profile according to [1]

for semiconductor isolators defines accurate testing for the maximum transient isolation voltage, V_{IOTM}, and the maximum repetitive voltage, V_{IORM}, which measure the capacity to withstand high voltages for very short periods of time and throughout the device lifetime, respectively [1]. Another important specification is the maximum surge isolation voltage, V_{SURGE}, that quantifies the capability of the isolator to withstand very high voltage impulses of a certain transient profile, which can arise from indirect lightning strikes or faults, as shown in Fig. 32.1b. The highest level of isolation, namely reinforced isolation, is achieved at the component level, only if it passes the surge test with a V_{SURGE} greater than 10 kV. At the present time, both industrial and automotive applications are moving towards 10 kV, some applications (e.g., patient monitoring systems) already require V_{SURGE} higher than 15 kV, while in the near future, galvanic isolation up to 20 kV will be required.

State–of–the–art galvanic isolators are based on electromagnetic (EM) coupling (i.e., capacitive or inductive) across a dielectric layer (i.e., the galvanic barrier). An integrated galvanic barrier can be implemented by using silicon dioxide (SiO_2), which exhibits a breakdown voltage (BV) of about 1000 V/µm [2], sometimes in combination with silicon nitride (Si_3N_4) and oxynitride (SiON) to further improve its isolation rating [3]. In the last years, oxide isolation has been successfully exploited for highly integrated isolated data [4–6] and isolated power transfer [7–11] by means of on–chip capacitors or stacked transformers. However, oxide insulation can reliably provide a limited surge capability (typically 5–6 kV), since increasing the oxide thickness produces wafer mechanical stress and second–order BV effects. The use of two series–connected galvanic isolation barriers, namely double isolation, is exploited to improve the overall isolation rating. It can be a viable solution for digital isolators (i.e., data transfer) [12], with a maximum V_{surge} of 12.8 kV by using a couple of isolation capacitors [13]. However, in isolated dc–dc conversion this approach is affected by a power efficiency degradation, which can be slightly mitigated by adopting integrated LC resonant barriers [14].

The galvanic barrier can be also implemented with other dielectric layers, such as the polyimide, traditionally used in semiconductor industry for stress relief. In this case, the isolation device (typically a stacked transformer) is built as a stand–alone chip by using post–processing fabrication steps at the cost of reducing the integration

level (i.e., from two to three chips per each isolated channel). This approach guarantees high data rates with high isolation rating (up to 20 kV) and common–mode transient immunity (CMTI) performance (better than 200 kV/μm) [15], while being also suited to power transfer up to several hundreds of mW with maximum power efficiencies higher than 30% [16]. Since polyimide BV is about 250 V/μm, typically the isolation layer is about 3x thicker to sustain the same isolation voltage of an oxide barrier. On the other hand, very thick polyimide layers can be manufactured with a record of 32.5-μm thickness able to withstand 20-kV surge voltage [15], which is not practical using silicon dioxide layers. In any case, isolation approaches based either on integrated SiO_2 barriers or post–processed polyimide transformers have inherent limitations in terms of isolation that can be improved only by means of expensive and time–consuming technological advances.

Sections 32.2 and 32.3 describe two alternative isolation techniques based on radio frequency (RF) coupling between two isolated interfaces, which are suited for isolated data and isolated power transfer, respectively [17, 18]. In these approaches, the galvanic isolation is provided by packaging/assembling techniques, which guarantee design flexibility. Indeed, the distance through insulation (DTI), which is responsible for the isolation rating, can be properly increased to guarantee to the required V_{SURGE}.

32.2 RF Galvanic Isolators Based on Planar Coupling

Breakdown limitations of traditional isolation approaches can be overcome without using expensive or exotic technologies by exploiting planar near–field coupling between two micro-antennas (i.e., spiral antennas) [17]. The latter are integrated on two side-by-side co–packaged chips (i.e., Chip 1 and Chip 2), as shown in Fig. 32.2a, while a standard molding compound is exploited as isolation material between them. With a DTI of about 500 μm, an isolation rating higher than 20 kV can be achieved. EM simulations are required to evaluate the weak coupling between the antennas, as shown in Fig. 32.2b. Using a standard CMOS substrate (i.e., with a conductivity of about 10 S/m) the simulated antenna coupling at 650 μm (corresponding to a DTI of 500 μm) is about −30 dB at 1 GHz. This is the starting point for the design of the galvanic isolator according to the block–diagram of Fig. 32.2c. Data transmission adopts a pulse width modulation (PWM) technique with an RF carrier (typically a few gigahertz according to the adopted CMOS node), which is more robust than traditional ASK modulation, since the bit information is the time–length of the RF burst rather than its amplitude. In a single isolated data channel, Chip 1 includes an integrated spiral antenna and an RF transmitter front-end (TX) driven by a base–band interface (i.e., PWM modulator). Chip 2 includes a second spiral antenna and an RF receiver front-end with a base-band interface for data demodulation (PWM demodulator). Both PMW modulator and demodulators are simple digital circuits with very low power consumption. On the other hand, TX and RX blocks are very critical, especially if the application sets stringent power consumption specifications.

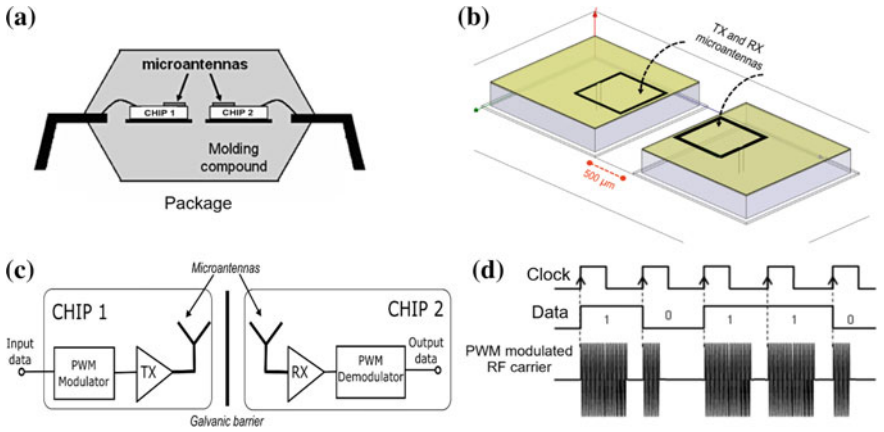


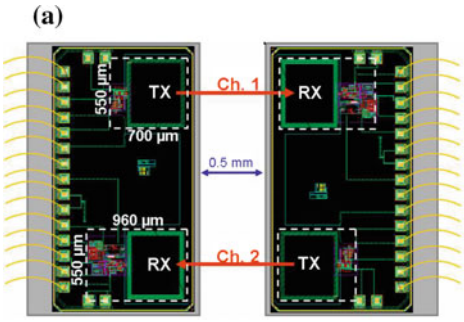
Fig. 32.2 **a** Package structure. **b** HFSS simulation of RF microantennas. **c** Block-diagram of the single–channel galvanic isolator. **d** Clock, data and PWM modulated RF carrier

The TX block can be implemented by means of an RF oscillator exploiting the spiral antenna inductance for the resonant tank. The oscillator can be properly turned on and off using switches controlled by the PWM signal. Main TX design issues are related to the co–design between the RF oscillator and the TX antenna, as well as the reduction of the current consumption at a fixed level of the oscillation voltage. The wireless channel between the two antennas considerably attenuates the TX signal (of about 30 dB) and therefore the RX front-end must raise it by using a low–noise amplifier (LNA). Then, a rectifier stage with an output filter (typically a simple Gilbert cell with an RC load) can be used to draw the envelop of RX signal. Finally, the rectified signal is compared with a threshold level to rebuild the initial TX PWM signal. The choice of the threshold is a critical issue since it determines the RX immunity to noise.

A bidirectional RF galvanic isolator based on the above described approach has been designed in a standard 0.35-μm CMOS technology provided by STMicroelectronics. The RF galvanic isolator layout and its main performance are reported in Fig. 32.3a, b, respectively. Despite the very high isolation rating capability, this solution has intrinsic limitations in terms of both silicon area and power consumption, which can limit its applicability. On the other hand, the adoption of scaled CMOS node and therefore the consequent increase of the RF carrier frequency can reduce the antenna size, while improving the data rate and the power consumption.

32.3 RF Galvanic Isolators Based on Face-to-Face Coupling

The key parameter for galvanically isolated power transfer systems is power efficiency. Indeed, it must be traded off with the isolation performance being both related



(b)

Parameter	Value
Carrier frequency	990 MHz
Current consumption	8.5 mA @ 20 Mb/s
Maximum data rate	40 Mb/s
TX antenna	14 nH
RX antenna	34 nH
Isolation rating	> 20 kV (estimated)
Transient immunity	> 200 kV/µs (simulated)
Technology	0.35-µm CMOS

Fig. 32.3 RF galvanic isolator based on planar coupling: **a** layout **b** performance

to the dielectric thickness of the isolation transformer, either using an integrated barrier or a post-processed device. When a very high isolation rating is required, a different isolation approach can be applied to the traditional power transfer architecture [18]. It takes advantage of a well-established technology, known as "wafer-to-wafer bonding", already used in different environments (e.g., integration of MEMs). The idea is to exploit face-to-face coupling between metal spirals by opposing two silicon wafers, interposing a dielectric layer between them and using proper vias called "trough silicon vias" (TSV) to implement external connections, as depicted in Fig. 32.4a. For better understanding, the 3D views and the cross–section of the isolator are reported in Fig. 32.4b, c, respectively. A face-to-face isolation transformer is implemented by using the wafer–to-wafer bonding technique. The top chip contains the primary coil of the transformer (to be connected to the oscillator), while the bottom chip includes the secondary winding that drives the rectifier, according

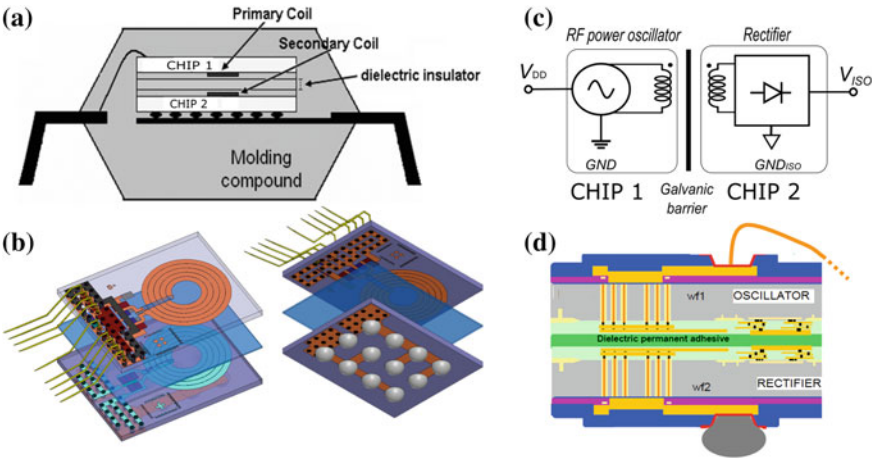


Fig. 32.4 **a** Package structure. **b** Simplified block diagram of the RF galvanic isolator. Face-to-face 3D assembly: 3D top and bottom views **c**, **d** cross-section

to the simplified architecture in Fig. 32.4d. Both transformer windings are therefore fabricated using the top metal layer of the adopted technology. The overall vertical structure can be mounted into a proper package by using bonding wires and solder bumps for top and bottom wafer connections, respectively. This approach allows the interposed dielectric layers to be properly chosen in order to guarantee the required isolation rating without degrading the transformer efficiency. To this aim, high dielectric strength materials can be used. Moreover, different isolation ratings may be adjusted by controlling the distance and/or the type of materials used when attaching the first and second silicon chips on wafer level. Compared to traditional isolated power transfer system based on integrated or post–processed power transformers, this approach is more flexible and can provide both higher isolation rating and smaller isolator size.

References

1. DIN VDE Semiconductor Devices-Magnetic and Capacitive Coupler for Basic and Reinforced Isolation, VDE Verlag VDE V 0884–11, Jan 2017
2. Palumbo V, Ghidini G, Carollo E, Toia F (2015) Integrated transformer. US Patent App 14733009, filed June 8 2015
3. Mahalingam P, Guiling D, Lee S (2007) Manufacturing challenges and method of fabrication of on-chip capacitive digital isolators. In: Proceedings of international symposium on semiconductor manufacturing, Oct 2007, pp 1–4
4. Krone A et al. (2001) A CMOS direct access arrangement using digital capacitive isolation. In: Proceedings IEEE international solid-state circuits conference digital technology papers, Feb 2001, pp 300–301
5. Moghe Y, Terry A, Luzon D (2012) Monolithic 2.5 kV RMS, 1.8 V–3.3 V dual-channel 640 Mbps digital isolator in 0.5 μm SOS. In: Proceedings of IEEE international SOI conference, Oct 2012, pp 1–2
6. Kaeriyama S et al (2012) A 2.5 kV isolation 35 kV/us CMR 250 Mbps digital isolator in standard CMOS with a small transformer driving technique. IEEE J Solid-State Circ 47:435–443
7. Spina N, Fiore V, Lombardo P, Ragonese E, Palmisano G (2015) Current-reuse transformer coupled oscillators with output power combining for galvanically isolated power transfer systems. IEEE Trans Circ Syst I: Reg Papers 62:2940–2948
8. Lombardo P, Fiore V, Ragonese, E, Palmisano G (2016) A fully-integrated half-duplex data/power transfer system with up to 40 Mbps data rate, 23 mW output power and on-chip 5 kV galvanic isolation. In: IEEE international solid-state circuits conference digital technology papers, Feb 2016, pp 300–301
9. Greco N, Spina N, Fiore V, Ragonese E, Palmisano G (2017) A galvanically isolated dc–dc converter based on current-reuse hybrid coupled oscillators. IEEE Trans Circuits Syst II: Exp Brief 64:56–60
10. Fiore V, Ragonese E, Palmisano G (2017) A fully-integrated watt-level power transfer system with on-chip galvanic isolation in silicon technology. IEEE Trans Power Electron 32:1984–1995
11. Ragonese E et al (2018) A fully integrated galvanically isolated DC-DC converter with data communication. IEEE Trans Circ Syst I: Reg Pap 65:1432–1441
12. Javid M, Ptacek K, Burton R, Kitchen J (2018) CMOS bi-directional ultra-wideband galvanically isolated die-to-die communication utilizing a double-isolated transformer. In: Proceedings of IEEE international symposium on power semiconductor devices and ICs, May 2018, pp 88–91

13. Texas Instruments. ISO7841x High-Performance, 8000-V_{PK} Reinforced Quad-Channel Digital Isolator. Accessed: 2018. (Online). Available: http://www.ti.com
14. Greco N, Parisi A, Lombardo P, Spina N, Ragonese E, Palmisano G (2018) A double–isolated DC-DC converter based on integrated LC resonant barriers. IEEE Trans Circ Syst I: Reg Pap 65:4423–4433
15. Yun R, Sun J, Gaalaas E, Chen B (2016) A transformer-based digital isolator with 20 kV_{PK} surge capability and >200 kV/μS common mode transient immunity. In: Proceedings of IEEE symposium on VLSI circuits, June 2016, pp 1–2
16. Qin W et al (2019) An 800 mW fully integrated galvanic isolated power transfer system meeting CISPR 22 Class-B emission levels with 6 dB margin. In: IEEE international solid-state circuits conference digital technology papers, Feb 2019, pp 246–248
17. Spina N, Girlando G, Smerzi SA, Palmisano G (2013) Integrated galvanic isolator using wireless transmission. US Patent 8364195 B2, Jan 2013
18. Renna CMA, Scuderi A, Magro C, Spina N, Ragonese E, Marano B, Palmisano G (2015) Microstructure device comprising a face to face electromagnetic near field coupling between stacked device portions and method of forming the device. US Patent 9018730 B2, granted 28 Apr 2015

Chapter 33
Experimental Characterization of a Commercial Sodium-Nickel Chloride Battery for Telecom Applications

Federico Baronti, Roberto Di Rienzo, Roberto Roncella, Gianluca Simonte and Roberto Saletti

Abstract Preliminary results coming from the experimental characterization of a commercial Sodium-Nickel chloride battery for telecom applications are shown in this paper. The battery was instrumented to collect all the possible current, voltage and temperature data from the paralled strings that constitute it. The aim is to have a better insight of a battery technology that seems an appealing candidate as alternative to Li-ion technology in some stationary applications. Charge/discharge cycles carried out with different loads show that the entire battery energy cannot fully be exploited at low load currents, when the internal battery heater dissipation is not negligible, and at high loads, when the internal dissipation leads to a dangerous increase of the internal temperature and to the battery disconnection. Pulse current tests useful for the validation of improved battery models are finally shown.

Keywords Sodium-nickel chloride battery · Battery Management Systems · Battery modeling

33.1 Introduction

Sodium-Nickel chloride batteries, usually called ZEBRA, are a very interesting alternative to Lithium-Ion (Li-Ion) ones [1]. ZEBRA technology yields energy density comparable with the Li-Ion one, but shows higher coulombic efficiency and it is

F. Baronti · R. Di Rienzo · R. Roncella · G. Simonte · R. Saletti (✉)
Dipartimento di Ingegneria dell'Informazione, University of Pisa, via Girolamo Caruso 16, 56122 Pisa, Italy
e-mail: roberto.saletti@unipi.it

F. Baronti
e-mail: federico.baronti@unipi.it

R. Di Rienzo
e-mail: roberto.dirienzo@unipi.it

R. Roncella
e-mail: roberto.roncella@unipi.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_33

inherently safer [2]. Moreover, this technology may lead to rather inexpensive batteries because it employs chemical substances abundant in nature, unlike Lithium [3]. The drawback is that a ZEBRA battery must work with an internal temperature in the 250–350 °C range. Thus, it must be equipped with a heater that increases the energy losses and makes these batteries not suitable to every application. However, the above mentioned advantages make ZEBRA very appealing as alternative to Lithium in many applications, such as in some automotive cases [4, 5], as energy storage system for renewable energy [6, 7] and as energy source in telecommunication applications [8, 9].

Unfortunately, the production process of the ZEBRA battery and the full exploitation of its capabilities present some open challenges which limit its penetration in the battery market. Many studies have been presented in the last years to address the production problems [10–12]. Instead, the literature is poor about the full exploitation of this technology that improved and accurate battery modeling and battery state estimation can provide [13]. A large amount of experimental data, still not available, would be needed to address the problem.

The aim of this work is to carry out an experimental characterization of a commercial ZEBRA battery, the FZSonick® 48TL200, in order to collect and make available data useful for a better modeling of a sodium-nickel chloride battery. As the 48TL200 battery does not provide access to all its internal data, additional sensors have been applied to the battery and an experimental setup has been developed to carry out the test campaign.

33.2 The ZEBRA Battery Under Test

The basic cell of a ZEBRA battery is composed by a liquid sodium negative electrode, and a positive electrode which consists of nickel surrounded by a mixture of nickel chloride ($NiCl_2$), nickel (Ni), iron sulphide (FeS) and liquid electrolyte. The electrodes are separated by a solid electrolyte made of beta alumina that allows the conduction of the sodium ions [5]. The main chemical reactions occurring in the cell are:

$$2\mathrm{Na} + \mathrm{NiCl_2} \underset{\text{Charge}}{\overset{\text{Discharge}}{\rightleftharpoons}} 2\mathrm{NaCl} + \mathrm{Ni} \qquad E = 2.58\ \mathrm{V} \tag{33.1}$$

$$\mathrm{FeCl_2} + 2\mathrm{Na} \underset{\text{Charge}}{\overset{\text{Discharge}}{\rightleftharpoons}} \mathrm{Fe} + 2\mathrm{NaCl} \qquad E = 2.35\ \mathrm{V} \tag{33.2}$$

The cell is operational at high temperature, because the electrolyte liquefies at 157 °C. Usually, the best performance is obtained in a temperature range of about 270–350 °C, in which the beta alumina resistance becomes comparable with the other contributions [14].

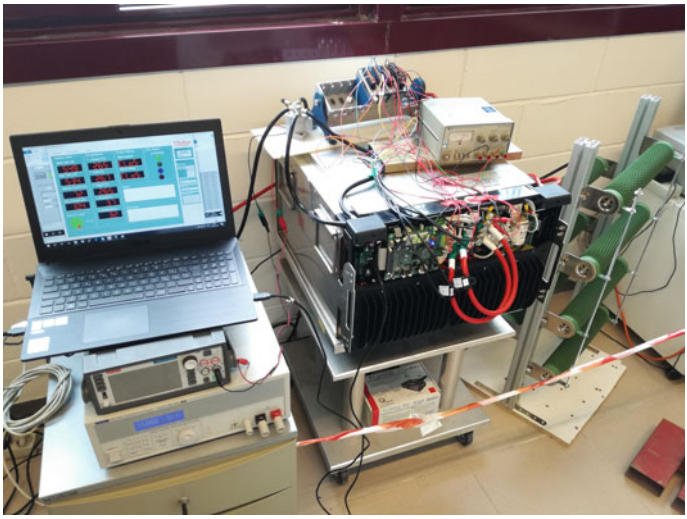


Fig. 33.1 Photograph of the experimental setup

The 48TL200 is a commercial ZEBRA battery for telecommunication applications with nominal voltage of 48 V, capacity of 200 Ah, and maximum current of 200 A and 150 A as peak and continuous values, respectively. The battery weight is 105 kg and its volume is 496 mm × 558 mm × 320 mm. The battery consists of 100 cells grouped in 5 strings of 20 series-connected cells, which are connected in parallel by the Battery Management System (BMS). The BMS is an electronic system that controls the battery behavior avoiding unsafe situations. In fact, it is able to stop the charge and discharge processes, to control the battery power switch, and to manage three heaters to maintain the battery internal temperature at least at 265 °C. The battery case is thermally insulated to limit the energy lost to keep the minimum internal temperature required. On the other hand, the thermal insulation is an obstacle to the dissipation of the heat internally generated during heavy discharge phases.

33.3 Experimental Setup and Results

Only few inner battery data are available on the 48TL200 user interface. Therefore, we first developed an experimental setup, shown in Fig. 33.1, to monitor and save as many as possible internal quantities of the battery, without accessing the battery inside and so with no risk of damage or alteration of its behavior. In particular, we added further sensors to measure the battery current and voltage, and the voltage and current of each of the 5 strings.

The setup employs a National Instrument cDAQ-9178 chassis equipped with one NI9219 and three NI9215 modules. The NI9219 provides 4 general purpose channels

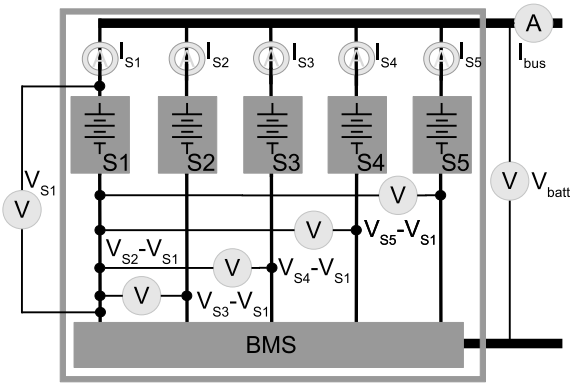


Fig. 33.2 Block diagram of the instrumented battery

that we used to measure the battery voltage, the battery current with a shunt resistor of 0.5 mΩ, and the voltage of the first string. The three NI9215 provide 4 analog input voltage channels each, in the range of ±10 V. We used them to acquire the voltage difference between the first string and the other ones, according to the block diagram shown in Fig. 33.2. The individual current of each string is measured by means of 5 open loop Hall effect sensors HCT-0010-050.

The measurement system is controlled by a custom interface developed in LabView, which runs on a laptop. The interface allows us to control the power supply (QPX1200SP) used to charge the battery and the load composed by a power resistor bank and a relay. In particular, 7 resistors of 2.2 Ω and 1.5 kW can be dynamically connected in parallel, to obtain a load resistance from about 315 mΩ up to 2.2 Ω and thus a discharge current from about 135 A down to 22 A, respectively. The PC is also connected via USB to the battery BMS and stores the few quantities measured by the BMS such us the internal temperature.

Several tests were carried out to extract the main features of the ZEBRA battery and to understand more deeply the BMS behavior. These experiments provide data by which very important information can be extracted for a better battery modeling and the development of a BMS with more accurate algorithms to estimate the internal state of the battery, such us the State of Charge (SoC), as it happens for Li-ion batteries [15]. The availability of separate data for each string makes it possible a deeper view of the inner behavior of the battery.

Let us show in Fig. 33.3 an example of experimental result. It reports the voltage and current of the first string and the entire battery during a test in which the battery was completely discharged with a constant load of 2.2 Ω and then recharged with a power supply setting of 55 V. We note that the battery and the first string voltages are overlapped in the discharge phase. The same happens for the other strings that are parallel connected possibly with ideal diodes. Instead, the two voltages are different in the charge phase. The battery voltage is 55 V as set by the external supply, whereas the first string voltage and current resemble a classic Constant Current/Constant Voltage (CC/CV) profile [16]. A similar behavior is found on the other strings. This

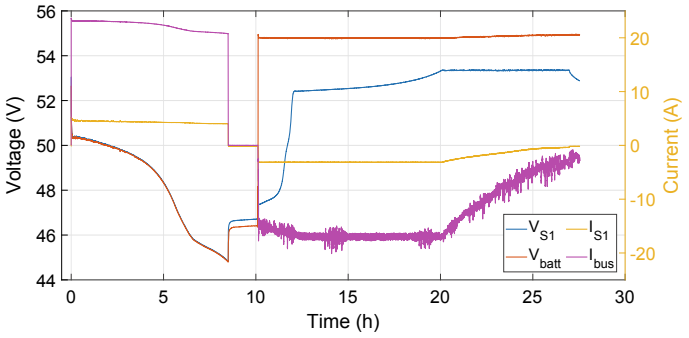


Fig. 33.3 Current and voltage of the battery and the first string during a test consisting of a full discharge with a constant load of 2.2 Ω followed by a recharge phase

Table 33.1 Discharge test results with different constant loads

Load (mΩ)	Avg. current (A)	Charge (Ah)	Energy (kWh)	Time (h:min)	Temp. (final) (°C)
2200	22	189	9.2	8:30	281
1100	43	200	9.4	4:39	316
733	63	200	9.2	3:06	337
550	80	197	8.8	2:26	350
440	99	176	7.8	1:46	350
367	118	161	7.0	1:20	350
314	135	148	6.3	1:05	350

observation suggests to us that the parallel connection is removed and the strings are individually charged with dedicated DC/DC converters by the BMS.

This test was repeated for all the load values and the available results are summarized in Table 33.1. These results suggest a couple of interesting conclusions. First, the extracted charge and energy are maximum for an average current of about 50 A. This happens because the energy loss in the battery heaters to maintain the minimum value of the battery temperature decreases when the battery current increases. The task of maintaining the minimum temperature level is also sustained by the heat generated in the parasitic series resistance of the battery, and the heaters may eventually be turned off. Second, the heat generated in the battery series resistance at high discharge currents is larger than that dissipated through the case. The net effect is that the battery temperature increases too much and overcomes its maximum value, eventually causing the battery disconnection. We end up with the very disappointing result that the discharge phase is interrupted by the BMS before the battery is fully discharged and the energy contained is not fully exploitable by the application.

Finally, let us show the results of a pulsed current test (PCT), i.e. a test in which the battery load is switched on and off at controlled time instants, useful to investi-

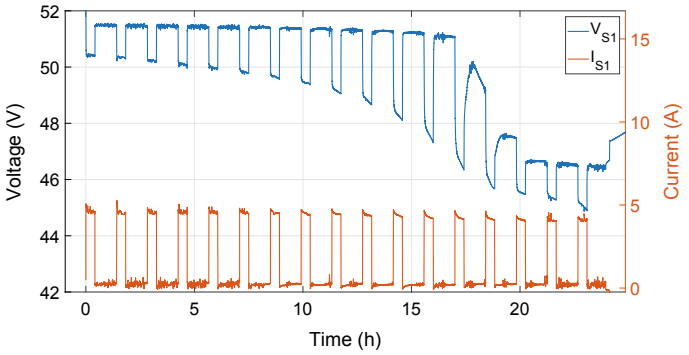


Fig. 33.4 Current and voltage of the first string in a pulsed current test

gate the Open Circuit Voltage (OCV) battery response and the behavior at different Depths of Discharge (DOD). Figure 33.4 shows the first string voltage and current during a PCT test with SoC steps of about 6%. It is worth noting that the battery behavior is straightforward down to about 70% DOD, with an almost constant OCV and an internal resistance gradually increasing (larger voltage steps for the same current pulse). The behavior changes significantly at deeper DOD. This indicates that a different chemical reaction involving the iron component of the cell is taking place [14]. The availability of these data makes it possible the investigation on more accurate electrical models of the battery [17].

33.4 Conclusions

This paper shows the preliminary results coming from the experimental characterization of a commercial nickel-chloride sodium battery for telecom applications. This kind of battery seems an appealing candidate for Li-ion technology replacement, particularly in stationary applications, due to the intrinsic safety of the involved chemistry. However, research efforts are still needed in battery modeling and BMS implementations. The battery was instrumented to collect all the possible data, as far as temperature, voltage and current of the paralled strings are concerned. The experimental setup is described and the results coming from charge/discharge cycles with different loads are discussed. It results that the entire battery energy can fully be exploited only when the heater dissipation is negligible at low load currents. At high loads, the battery does not provide the whole energy because the internal dissipation leads to a dangerous increase of the internal temperature, which forces the BMS to disconnect the battery before the end of the test. Finally, experimental results coming from pulse current tests useful for the validation of improved battery models are shown.

References

1. Bhatt DK, El Darieby M (2018) An assessment of batteries form battery electric vehicle perspectives. In: 2018 6th IEEE international conference on smart energy grid engineering, SEGE 2018, pp 255–259
2. Hueso KB, Palomares V, Armand M, Rojo T (2017) Challenges and perspectives on high and intermediate-temperature sodium batteries. Nano Res 10(12):4082–4114
3. Gruber PW, Medina PA, Keoleian GA, Kesler SE, Everson MP, Wallington TJ (2011) Global lithium availability: a constraint for electric vehicles? J Ind Ecol 15(5):760–775
4. Capasso C, Veneri O (2014) Experimental analysis of a zebra battery based propulsion system for urban bus under dynamic conditions. Energy Procedia 61:1138–1141
5. O'Sullivan TM, Bingham CM, Clark RE (2006) Zebra battery technologies for the all electric smart car. In: International symposium on power electronics, electrical drives, automation and motion, 2006. SPEEDAM 2006, Nov 2006, pp 244–248
6. Lu X, Yang Z (2014) Molten salt batteries for medium- and large-scale energy storage
7. Bignucolo F, Coppo M, Crugnola G, Savio A (2017) Application of a simplified thermal-electric model of a sodium-nickel chloride battery energy storage system to a real case residential prosumer. Energies 10(10):1497
8. Restello S, Lodi G, Miraldi AK (2012) Sodium nickel chloride batteries for telecom application: a solution to critical high energy density deployment in telecom facilities. In: INTELEC, international telecommunications energy conference (proceedings), pp 1–6
9. Restello S, Spa F, Maggiore M, Zanon N, Spa F, Maggiore M (2013) Sodium nickel batteries for telecom hybrid power systems. 2. Sodium nickel chloride technology, vol 2, pp 324–328
10. Lu X, Coffey G, Meinhardt K, Sprenkle V, Yang Z, Lemmon JP (2010) High power planar sodium-nickel chloride battery, pp 7–13
11. Lu X, Li G, Kim JY, Lemmon JP, Sprenkle VL, Yang Z (2013) A novel low-cost sodium-zinc chloride battery. Energy Environ Sci 6(6):1837–1843
12. Chang HJ, Lu X, Bonnett JF, Canfield NL, Son S, Park YC, Jung K, Sprenkle VL, Li G (2018) "Ni-less" cathodes for high energy density, intermediate temperature Na-$NiCl_2$ batteries. Adv Mater Interfaces 5(10):1–8
13. Benato R, Dambone Sessa S, Necci A, Palone F (2016) A general electric model of sodium-nickel chloride battery. In: AEIT 2016—international annual conference: sustainable development in the Mediterranean area, energy and ICT networks of the future
14. Sudworth JL (2001) The sodium/nickel chloride (ZEBRA) battery. J Power Sources 100(1–2):149–163
15. Morello R, Di Rienzo R, Roncella R, Saletti R, Baronti F (2018) Hardware-in-the-loop platform for assessing battery state estimators in electric vehicles. IEEE Access
16. Dung LR, Chen CE, Yuan HF (2016) A robust, intelligent CC-CV fast charger for aging lithium batteries. In: IEEE international symposium on industrial electronics
17. Musio M, Damiano A (2015) A non-linear dynamic electrical model of sodium-nickel chloride batteries. In: 2015 international conference on renewable energy research and applications, ICRERA 2015

Chapter 34
Design and Development of a Prototype of Flash Charge Systems for Public Transportation

Adriano Alessandrini, Riccardo Barbieri, Lorenzo Berzi, Fabio Cignini, Antonino Genovese, Fernando Ortenzi, Marco Pierini and Luca Pugi

Abstract As a part of a National research program (Ricerca di Sistema 2015 and 2017), ENEA has proposed an innovative hybrid storage system that allows a fast recharge of electric public transportation vehicles. In order to demonstrate the feasibility of this idea, researchers of University of Florence have implemented the proposed system on a existing electric bus in order to demonstrate the feasibility of the proposed system. In this work authors introduce main features of the proposed prototype.

34.1 Introduction: The Proposed Flash Carge System

Fast recharge of electro-chemical accumulators (the so called batteries) is substantially limited by the maximum applicable recharge currents [1, 2].

Technology of Super-Capacitors offers the possibility of a fast recharge thanks to a very high specific power. However specific energy of capacitor is relatively small and it's not able to assure a prolonged autonomy. The proposed hybrid system described in Fig. 34.1 offers a reasonable compromise: capacitors are used to transfer at each bus-stop a reasonable amount of energy to travel between two consecutive recharge stations. A conventional electro-chemical battery is used to assure a residual autonomy to the vehicle assuring a sufficient reliability to complete the mission even in off design conditions. This solution in different shapes have been proposed also by other research groups as the one lead by Yu and Tarsitano [3]. However, what make really the difference between systems proposed in literature and the proposed

A. Alessandrini
DICEA, University of Florence, Florence, Italy

R. Barbieri · L. Berzi · M. Pierini · L. Pugi (✉)
DIEF, University of Florence, Via di Santa Marta 3, 50139 Florence, Italy
e-mail: luca.pugi@unifi.it

F. Cignini · A. Genovese · F. Ortenzi
ENEA Centro di Ricerca di Casaccia, Rome, Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_34

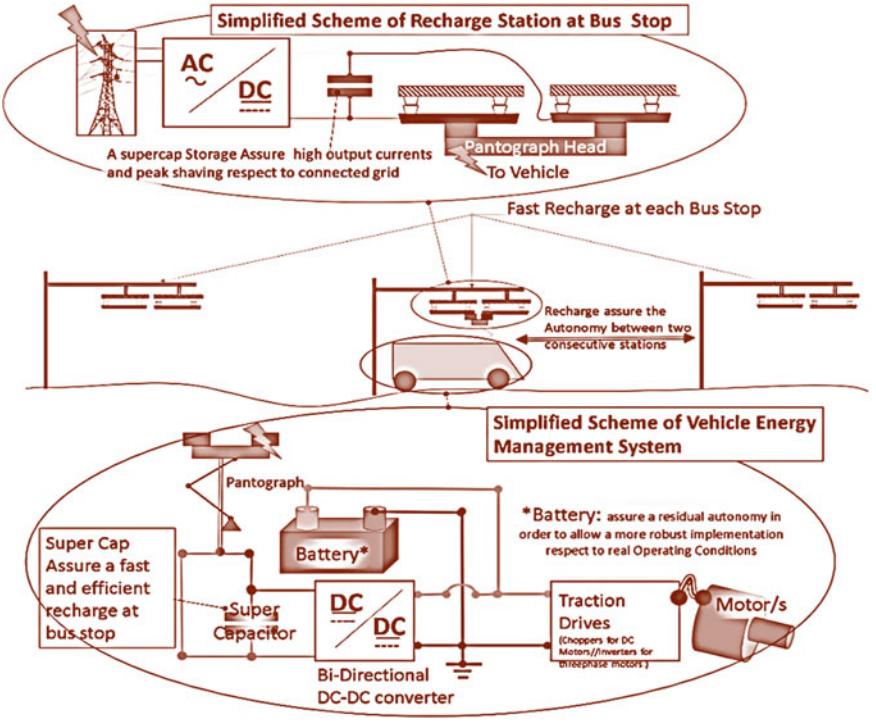


Fig. 34.1 Simplified scheme of the proposed flash charge system

implementation is the development of a working prototype able to physically demonstrate that the proposed solution is really able to work in a physical demonstrator [4] assembled using commercial components. In this activity previous experiences in design of electric vehicles have greatly influenced the design of the system [5]

For this reasons authors revamped an existing electric bus from Tecnobus™ which main features are visible in Table 34.1.

As visible in Fig. 34.2, in order to collect the current from the stationary recharge station the bus was equipped with a pantograph for fast recharge developed by Schunk, capacitors and static converters (provided by TAME™) needed to properly couple supercapacitor with batteries and traction DC bus have been installed in the same encumbrances that were designed for the original power-pack. In this way authors proved that the system can be easily installed without increasing overall vehicle weight or reducing space for passengers and more generally for vehicle payload.

Table 34.1 Main features of revamped electric vehicle (Gulliver-Tecnobus)

Vehicle parameters[a]	Value
Weight (tare/maximum)	4270/6045 [kg]
Batteries (kind/capacity/weight)	Lead Acid/ 595 [Ah] @72 [V]/1500 [kg]
Inst. traction system (nom. peak power)	21 [kW]/25 [kW]
Lights and on board instr. (power)	200 [W]
Air conditioning system[b]	2 [kW]
Estimated autonomy with only super cap	1000 [m] (one bus stop)
Recharge time	35 [s] (less than the time needed to carry people on a bus stop)

[a]Parameters are referred to the original Gulliver bus before authors retrofit
[b]The bus used for the first demonstrator was not equipped with an air conditioning unit, however authors have preliminary established a feasible size of the system

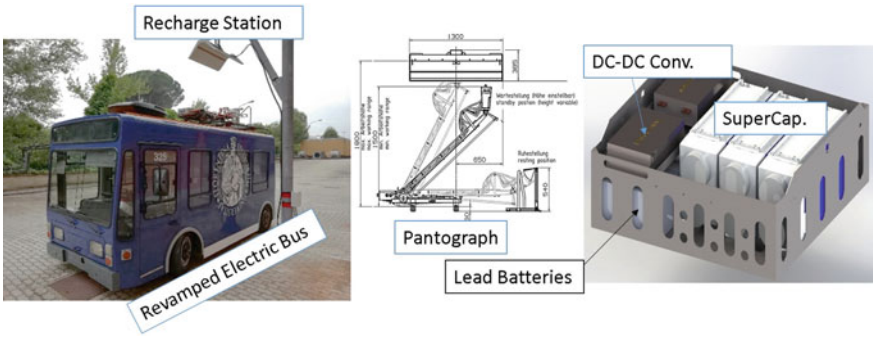


Fig. 34.2 Scheme of performed revamping activities

34.2 Proposed Control System

Respect to systems currently proposed in literature [3], authors have deliberately chosen a very simple and robust controller which is very innovative for the proposed layout since it allows a fast calibration and a robust implementation using cheap commercial micro-controllers. Proposed control scheme as visible in Fig. 34.3 can be briefly described in the following way: first, according the current state of charge of the battery, it's calculated a value of current (I_{ref}) that should be applied by the DC-DC converters, to the battery to provide a corresponding smooth efficient recharge using the energy stored in super-capacitors. The system is regulated using a current closed loop whose actuators/active elements are the DC-DC converters connected to super capacitors. DC-DC converters are configured as current servo-amplifiers; they provide the desired output current taking the power from connected super capacitors.

Since it's a closed loop scheme, there is no need to measure additional loads due to other vehicle systems since they are treated as disturbances which must be rejected

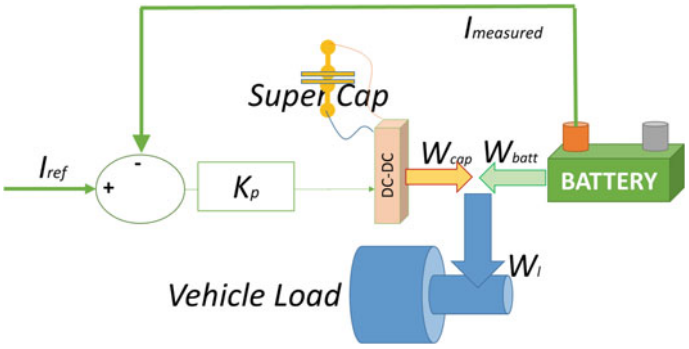


Fig. 34.3 Proposed control scheme

by the loop. The proposed regulator is a simple proportional one. A scheduling of the proportional gain K_p respect to battery or capacitors state of energy/charge can be easily used to better calibrate the system.

To better understand the behaviour of the proposed control system the following simplifications/assumptions must be supposed valid:

- System can be approximated and solved as an LTI one (Linear Time Invariant);
- Energy conversion losses and other nonlinear phenomena are neglected;
- Bandwidth of external disturbances is very low respect to the one of the control systems (supposed stable) so they can be treated as static contributions.

Assuming as valid the over cited simplifications, the power required by vehicle loads W_L can be treated as an equivalent static disturbance and all the system can be investigated using standard properties and rules used for LTI systems. In this way it's possible to calculate approximately how, in steady state conditions, the load W_L should be distributed between capacitors and batteries (34.1).

$$\begin{aligned} \overbrace{W_{cap}}^{\substack{power\\ from\\ Capacitors}} &\approx W_L \frac{K_p}{1+K_p} + I_{ref} V_{batt} \frac{K_p}{1+K_p}; \\ \overbrace{W_{batt}}^{\substack{power\\ from\\ Batteries}} &\approx W_L \frac{1}{1+K_p} - I_{ref} V_{batt} \frac{K_p}{1+K_p}; \end{aligned} \tag{34.1}$$

According Relation (34.1) by rising the gain K_p most of the power should be provided by capacitors, otherwise by batteries. By specifying directy I_{ref} als the user has the possibility of choosing a specific amount of power flowing from or to batteries forcing their recharge or discharge according system logic.

34.3 Final Prototype and Preliminary Testing Activities

The final prototype of the system, visible in Fig. 34.4 was finally assembled and tested first using a simulation test bench and then by driving in an internal road circuit available at the same research center of ENEA in Casaccia (Rome, Italy).

In Fig. 34.5, some experimental results are shown: vehicle performs a mission (start and stops acceleration braking etc.) and the total current required by various connected loads is measured and compared with the contribution of supercapacitors (DC-DC Current) and of the batteries: proposed approach involves a smooth behavior of the system which is able to largely exploit the energy stored in supercapacitors, maintaining small and smooth power demands on batteries.

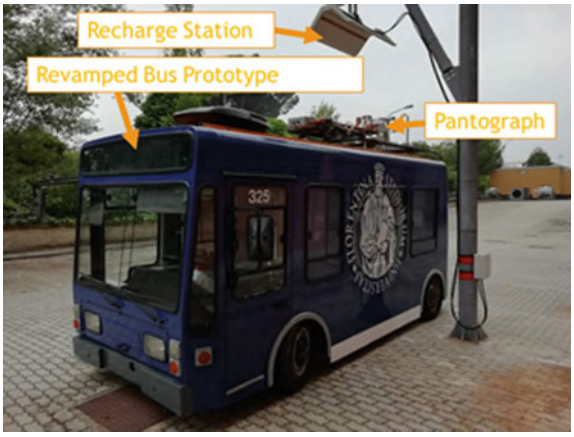


Fig. 34.4 UNIFI prototype during testing activities (June 2019)

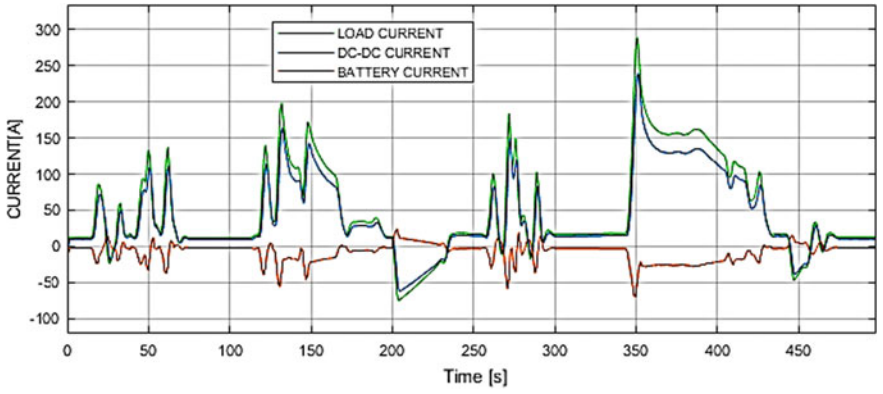


Fig. 34.5 Preliminary results, showing the capability of the proposed system to smoothly reduce the applied load on batteries exploiting energy provided by super-capacitors

It's interesting to notice that the behavior of the system can be easily tuned acting on the gain of the proposed battery current controller. More complicated or nonlinear calibration can be performed simply be introducing a tabulated gain scheduling. It's interesting to notice that the proposed controller runs on ATMEL 16 Bit micro-controllers with limited performances and most of the computational effort are exploited by communication and diagnostic tasks. This is a demonstration of the simplicity of the proposed approach which give wide margin for future implementations on more performing hardware.

34.4 Conclusion and Future Developments

Authors have successfully assembled a prototype of the proposed flash charge systems. Preliminary results are encouraging. Currently authors are completing experimental activities and these complete results will probably the object of a future publication. Authors are also planning some improvements of the current system:

First, the substitution of current lead-acid batteries with more performing lithium ones. In this way stored energy can be further increased. Also, Lithium batteries should support a further increase of transferred energy during the flash charge process since their specific power is quite higher respect to lead ones. It's interesting to notice that from previous research experiences [6, 7] durability of this kind of batteries should also be greatly improved by the adoption of the proposed system. Also the recharge static station should be further improved by introducing an active system able to further increase the transferred power. Also authors plan the possibility of testing wireless recharge systems on the bus which also in their previous experiences have proven to be a valid solution for the recharge of road vehicles [8, 9].

Acknowledgements authors wish to thank all the people that have cooperated to the success of this activity. In particular, authors wish to mention all the people of ENEA and University of Florence which have provided a fundamental support in terms for preliminary testing activities and for the final assembly of the vehicle.

References

1. Ceraolo M, Barsali S, Lutzemberger G, Marracci M (2009) Comparison of SC and high-power batteries for use in hybrid vehicles. SAE Technical Paper 2009-24-0069
2. Barrade P, Rufer A (2002) Supercapacitors as energy buffers: a solution for elevators and for electric busses supply. In: Proceedings of the power conversion conference-osaka 2002 (Cat. No. 02TH8579), vol 3. ages, IEEE, pp 1160–1165
3. Yu H, Tarsitano D, Hu X, Cheli F (2016) Real time energy management strategy for a fast charging electric urban bus powered by hybrid energy storage system. Energy 112:322–331. https://doi.org/10.1016/j.energy.2016.06.084

4. Ortenzi F, Pasquali M, Pede G, Lidozzi A, Di Benedetto M (2018) Ultra-fast charging infrastructure for vehicle on-board ultracapacitors in urban public transportation applications. EVTeC and APE Japan on Sept 30-Oct 3 2018
5. Pugi L, Grasso F, Pratesi M, Cipriani M, Bartolomei A (2017) Design and preliminary performance evaluation of a four wheeled vehicle with degraded adhesion conditions. Int J Electr Hybrid Vehicles 9(1):1–32. https://doi.org/10.1504/IJEHV.2017.082812
6. Locorotondo E, Pugi L, Berzi L, Pierini M, Pretto A (2018) Online state of health estimation of lithium-ion batteries based on improved ampere-count method. In: Proceedings—2018 IEEE international conference on environment and electrical engineering and 2018 IEEE industrial and commercial power systems Europe, EEEIC/I and CPS Europe 2018 https://doi.org/10.1109/eeeic.2018.8493825
7. Locorotondo E, Pugi L, Berzi L, Pierini M, Lutzemberger G (2018) Online identification of thevenin equivalent circuit model parameters and estimation state of charge of lithium-ion batteries. In: Proceedings—2018 IEEE international conference on environment and electrical engineering and 2018 IEEE industrial and commercial power systems Europe, EEEIC/I and CPS Europe 2018, art. no. 8493924, https://doi.org/10.1109/eeeic.2018.8493924
8. Pugi L, Reatti A, Corti F (2018) Application of wireless power transfer to railway parking functionality: preliminary design considerations with series-series and LCC topologies. J Adv Transport, art. no. 8103140. https://doi.org/10.1155/2018/8103140
9. Pugi L, Reatti A, Corti F (2019) Application of modal analysis methods to the design of wireless power transfer systems. Meccanica 54(1–2):321–331. https://doi.org/10.1007/s11012-018-00940-x

Chapter 35
Unsupervised Monitoring System for Predictive Maintenance of High Voltage Apparatus

Christian Gianoglio, Andrea Bruzzone, Edoardo Ragusa and Paolo Gastaldo

Abstract The online monitoring of a high voltage apparatus is a crucial aspect for a predictive maintenance program. The insulation system of an electrical machine is affected by partial discharges (PDs) phenomena that—in the long term—can lead to the breakdown. This in turn may bring about a significant economic loss; wind turbines provide an excellent example. Thus, it is necessary to adopt embedded solutions for monitoring the insulation status. This paper introduces an online system that exploit fully unsupervised methodologies to assess in real-time the condition of the monitored machine. Accordingly, the monitoring process does not rely on any prior knowledge about the apparatus. Nonetheless, the proposed system can identify the relevant drifts in the machine status. Notably, the system is designed to run on low-cost embedded devices.

Keywords Predictive maintenance · Embedded systems · Partial discharges

35.1 Introduction

A machine under voltage is customarily subject to electrical, mechanical and thermal stresses, which lead to an aging of the insulation system. To prevent a breakdown, a scheduled, periodical maintenance is required. As maintenance involves a momentary shutdown of an apparatus, economic losses might be consistent. Automated predictive maintenance aims at optimizing such process, thus increasing productivity and reducing maintenance costs. The goal is to take advantage of automatic online systems that can monitor the electrical insulation status and eventually support a suitable scheduling of the maintenance. The literature provides several approaches to online PDs monitoring. In general, monitoring first involves the acquisition of signals by conducted or irradiated sensors. Such signals fed a classification system

C. Gianoglio (✉) · A. Bruzzone · E. Ragusa · P. Gastaldo
Department of Electrical, Electronic, Telecommunication Engineering and Naval Architecture (DITEN), University of Genoa, Genova, Italy
e-mail: christian.gianoglio@edu.unige.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_35

that is entitled to identify the category of the defect affecting the apparatus. Usually machine learning methodologies support the classification system of the PD sources. In [2] an ultra high frequency (UHF) antenna has been used to sample the discharges; then, a set of features has been extracted from the raw data. The K-means algorithm supported a clustering according to the defects into the feature space. In [7] the defects have been classified using a neural network on features extracted from signals picked up by a high frequency current transformers (HFCT). In [4] the DBSCAN clustering is applied on features extracted from the wavelet decomposition of signals coming from a HFCT sensor. Recently, in [1] a gas insulated substation (GIS) has been monitored with both antenna and HFCT sensors; PD sources recognition has been performed by exploiting clustering techniques. The major weakness of those approaches is that the classification system requires a training procedure. This in turn means that a large training set should be available.

Other approaches are based on the statistical analysis of the PD signals. In [8] the supervised classification approach utilizes histogram similarity. Accordingly, a PD pattern is suitably represented as an histogram. The system relies on a database of histograms, where each defect is represented. As a result, online monitoring applies an hypothesis test to compare the input histogram with each histogram included in the database. This approach again requires a large dataset of labeled histograms. Nonetheless, multi-defects PD patterns represent an issue.

In this paper, online monitoring relies on a fully unsupervised approach that does not require a pre-built training set. The model is entitled to track in real time the drift affecting the insulation system. To this purpose, aging phenomena are detected by using the approaches commonly adopted in anomaly detection. Hence, in principle, no prior knowledge of the apparatus under investigation is needed. Moreover, the whole system is designed to target low-resources devices, such as low-cost embedded systems.

35.2 Detecting Aging Phenomena via Anomaly Detection

The proposed monitoring model basically assesses in real-time the aging of the insulation system. Since a fully unsupervised approach is targeted, the model is designed to detect significant changes in the status of the apparatus without knowledge base. Hence, the monitoring process is approached as an anomaly detection problem. In any instant T the goal is to check whether the apparatus shows an anomalous behavior with respect to the past. The status of the insulation system at time T is characterized by sampling—at a given frequency—the amplitude of the signal sensed by the HCFT in a time window δ. Such signal is converted in a vector $\mathbf{x}$ by using the same process that leads to a PD pattern; in this case, though, the phase information is discarded to obtain a vector. As a result, the size of $\mathbf{x}$ depends on the occurrences of PDs in the time window δ.

Anomaly detection is implemented by applying hypothesis testing. Two well-known statistical tests are utilized to this purpose: Chi-Square (Chi2) and Kolmogorov–Smirnov (KS). The null hypothesis is that the vector $\mathbf{x}$ measured at time T and any vector measured before T come from the same population. In other words, one has an anomaly when the measurement at time T is not consistent with the previous measurements. The underlying hypothesis is that aging phenomena lead progressively to significant changes in the distribution of PDs. Hence, such discontinuities can be detected even in the absence of trained classifiers or any knowledge base. In general, hypothesis testing involves two vectors: vector $\mathbf{x}_T$ reports on the measurement at time T, while vector $\tilde{\mathbf{x}}$ represents the reference, i.e., a representative sample of the population with specific distribution that characterized the apparatus before time T. The latter vector obviously should be updated every time an anomaly, i.e., a major discontinuity in the apparatus status is detected.

In the case of Chi2 test, hypothesis testing involves two discrete distributions. Thus, both the vectors should be expressed as histograms with $nbins$ number of bins. Let O_i be the value of the histogram of $\mathbf{x}_T$ for the ith bin. Analogously, let $\tilde{R}_i$ be the value of the histogram of $\tilde{\mathbf{x}}$ for the ith bin. The χ^2 statistical quantity that can be computed as:

$$\chi^2 = \sum_{i=1}^{nbins} \frac{(K_1 O_i - K_2 \tilde{R}_i)^2}{O_i + \tilde{R}_i} \tag{35.1}$$

where K_1 and K_2 are scaling constants that are used to adjust for unequal sample sizes. Given χ^2 and the degrees of freedom DF, which corresponds to the number of non-empty bins, the p-value can be computed from the $Chi2$ distribution. The null hypothesis is accepted if p-value $< \alpha$, where α is the significance level set a-priori (usually $\alpha = 0.05$).

The KS processes the empirical cumulative distribution functions (ECDF) of $\mathbf{x}_T$ and $\tilde{\mathbf{x}}$, respectively. Given N points p_n ordered from smallest to largest value, the ECDF is defined as

$$FO_n = i_n / N \tag{35.2}$$

where i_n is the number of elements smaller than p_i in $\mathbf{x}_T$. Analogously, FE is the ECDF of $\tilde{\mathbf{x}}$. As a result, the KS test statistic relies on

$$D = \sup_n |FO_n - FE_n| \tag{35.3}$$

Given D, the p-value can be calculated as in [5]. Eventually, the null hypothesis is rejected if the significance level α is lower than the p-value.

In the following, $FindAnomaly(\mathbf{x}_T, \tilde{\mathbf{x}}, \alpha)$ will denote a function that returns 1 when the null hypothesis is rejected, and 0 otherwise. Such function may exploit either the Chi2 test or the KS test.

35.3 Online Monitoring for Predictive Maintenance

The proposed monitoring systems is designed to identify in real time the abrupt discontinuities in the status of the apparatus, which are reported as alerts. Accordingly, the monitoring procedure is organized as follows. In standard mode, the monitoring systems continuously get the vector $\mathbf{x}_t$ measured at time t and verifies the occurrence of an anomaly (as per Sect. 35.2). If an anomaly is revealed, a second procedure starts; its goal is to verify that an alert should be activated. Let T^* denote the instant in which an anomaly as occurred. The procedure sets an alert only if—in the time window between T^* and $T^* + \Delta$—the anomaly ($number\ Of\ Anomalies$) occurs again at least thr times, where thr and Δ have been fixed empirically. Algorithm 1 gives the pseudo-code of this procedure. In practice, the monitoring system is designed to set an alert only when a sequence of anomalies is detected. Thus, it can detect only the significant discontinuities in the status of the apparatus. It is worth noting that the approach is fully unsupervised and does not require any previous knowledge about the apparatus. Moreover, the computational complexity of the whole monitoring procedure is negligible. This in turn means that the monitoring system can be hosted by low-cost, resource-limited embedded systems.

The proposed monitoring system can indeed become part of an IoT-based predictive maintenance, as per Fig. 35.1. First, the alerts can allow one to categorize all the data associated to an apparatus according to well defined landmarks. Hence, a remote database can collect structured information provided by all the monitored apparatuses. In practice, the landmarks generated via an unsupervised process can be exploited to eventually label the data provided by online measurements. As a result, a remote warehouse can exploit such database and machine learning methodologies to make inferences—given a measure $\mathbf{x}_t$ for an apparatus—on the exact type of defect affecting the insulation system under investigation. Under the paradigm of edge computing, the local embedded system can be designed to implement the inference function, which receives its parameters from the cloud. In this regard, the literature offers a few examples of hardware-friendly implementations of inference functions [3, 6].

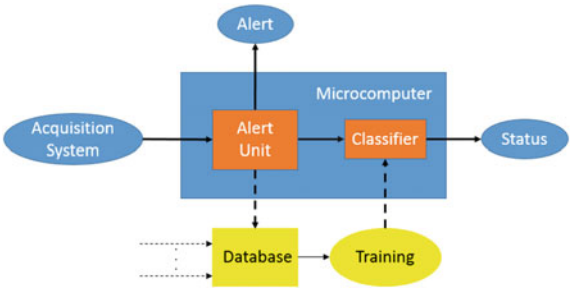


Fig. 35.1 Embedded system for alert detection and classification

Algorithm 1 Alert generator

Input: $\tilde{\mathbf{x}}$, Δ, thr
Online
numOfAnomalies=1;
for $t = T^* + 1$ to $T^* + \Delta$ **do**
get $\mathbf{x}_t$
flag=FindAnomaly($\mathbf{x}_t$,$\tilde{\mathbf{x}}$,α);
if flag **then**
numOfAnomalies++;
if numOfAnomalies > thr **then**
Alert(T^*);
$\tilde{\mathbf{x}} = \mathbf{x}_t$
return
end if
end if
end for

35.4 Experimental Results

The experimental session involved two twisted pair specimens that underwent aging tests according to standard IEC 60851-5. A HFCT with a bandpass behavior placed around the ground cable provided the sensor. Signals were sampled by a Picoscope with a bandwidth in the range 0–200 MHz and a maximum sample frequency of 1 GSamples/s. The monitoring system was deployed on a Raspberry Pi. In the first experimental session, the Picoscope was configured with a fullscale of 20 V and 12 bit resolution. The status of the specimen was monitored every minute, with a sampled time window $\delta = 0.5$ s. Figure 35.2 shows—on a time scale—the alerts generated by monitoring system before the specimen breakdown (a total of 20 h). The red marks refer to a monitoring system relying on the KS test; the blue marks refer to the Chi2 test. After about four hours, both the setup started to produce alerts. In the following four hours the effects of aging phenomena assumed almost a periodic pattern. Then, the gap between successive alerts progressively increased. The Chi2 test actually produced two more alerts than the KS test. Possibly in the case of Chi2 test sensitivity also depends on $nbins$; in this experiment it has been chosen empirically as $nbins = 25$. The second test, made on another specimen, aimed at evaluating the impact of the analog to digital converter (ADC) resolution on the monitoring system. In this test, the status of the specimen was monitored every 2 min ($\delta = 0.5$ s); the breakdown occurred after 22.5 h. Figure 35.3 gives the alerts produced with a 8 bit resolution (in blue) and with a 12 bit resolution (in red). Overall, the system generated more alerts when adopting a 8 bit resolution. In fact, it is reasonable to assume that a coarser quantization makes anomaly detection more prone to errors. Under such assumption, the alerts generated with a 8 bit resolution respectively 1 hour after the start and half an hour before the breakdown might represent false alarms.

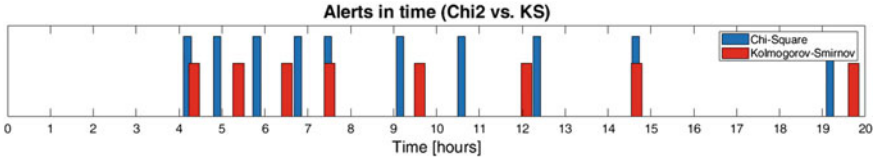


Fig. 35.2 Alerts produced by Chi2 and KS tests on the first specimen

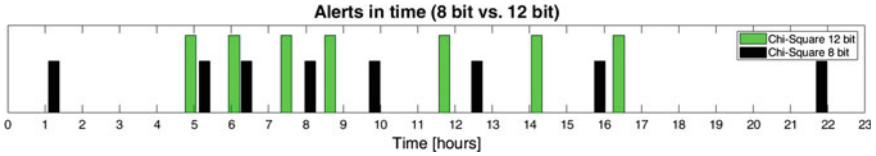


Fig. 35.3 Alerts produced by Chi2 test with resolutions of 8 and 12 bit

35.5 Conclusions

This paper shows that anomaly detection paradigms can support a fully unsupervised monitoring system for predictive maintenance of high voltage apparatus. The proposed method is computationally light and fit IoT solutions that rely on edge computing. The monitoring system can identify in real-time the significant changes in the status of the apparatus, thus revealing aging effects. At the same time, the system enables the automated labeling of acquired data, which become structured information to be stored and processed by deep analytics.

References

1. Álvarez F, Garnacho F, Ortego J, Sánchez-Urán M (2015) Application of HFCT and UHF sensors in on-line partial discharge measurements for insulation diagnosis of high voltage equipment. Sensors 15(4):7360–7387
2. Gao W, Ding D, Liu W (2011) Research on the typical partial discharge using the UHF detection method for GIS. IEEE Trans Power Deliv 26(4):2621–2629
3. Gianoglio C, Guastavino F, Ragusa E, Bruzzone A, Torello E (2018) Hardware friendly neural network for the PD classification. In: 2018 IEEE conference on electrical insulation and dielectric phenomena (CEIDP). IEEE, pp 538–541
4. Hao L, Lewin P, Hunter J, Swaffield D, Contin A, Walton C, Michel M (2011) Discrimination of multiple PD sources using wavelet decomposition and principal component analysis. IEEE Trans Dielectr Electr Insul 18(5):1702–1711
5. Massey FJ Jr (1951) The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc 46(253):68–78
6. Ragusa E, Gianoglio C, Gastaldo P, Zunino R (2018) A digital implementation of extreme learning machines for resource-constrained devices. IEEE Trans Circuits Syst II: Express Briefs 65(8):1104–1108

7. Wang MH (2005) Partial discharge pattern recognition of current transformers using an ENN. IEEE Trans Power Deliv 20(3):1984–1990
8. Yazici B (2004) Statistical pattern analysis of partial discharge measurements for quality assessment of insulation systems in high-voltage electrical machinery. IEEE Trans Ind Appl 40(6):1579–1594

Chapter 36
Control System Design for Cogging Torque Reduction Based on Sensor-Less Architecture

Dini Pierpaolo and Sergio Saponara

Abstract In this work a sensor-less architecture based on the Extended Kalman Filter observer and on feedback linearization control system is proposed. This work proves that also with a sensor-less architecture (where only the measurements of the machine's electrical dynamic quantities are available) it is possible to assume a previously control solution, proposed by the authors, to reduce the intrinsic problem of the Cogging Torque. Results in term of the trajectory tracking control problem on the direct current component and on the rotor axis position are presented. An analysis on the initial condition of the rotor axis position revels that the architecture is robust in term of variation of start condition of the global system.

36.1 Introduction

Increasingly, servo drives based on brushless motors are used. This type of electric motors has a high efficiency, good capacity to deliver relatively high torques and excellent characteristics in dynamic regime. These features make brushless motors the most suitable in applications such as the implementation of operating machines and industrial robots in assembly lines. However, the presence of permanent magnets creates some limitations in the use of this type of motor. One of the problems with brushless motors is the presence of an intrinsic phenomenon called Cogging.

This phenomenon is caused by the magnetic interaction between the two main parts of the machine that from an operational point of view can be interpreted as a torque oscillation. This phenomenon is therefore a problem in those applications where a great deal of precision is required. Moreover, it is the cause of unwanted noise for the entire drive system. The result of this interaction is an additive pair which causes an undesired oscillation on the rotation of the rotor axis even in the

D. Pierpaolo (✉) · S. Saponara (✉)
Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Pisa, Italy
e-mail: pierpaolo.dini@phd.unipi.it

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

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_36

absence of power supply to the electric machine. This problem is partially solved by requiring physical modifications in the production phase of the electric motor, which provides for different shapes of the stator slots and of magnets [1, 2] and therefore the replacement of the motor itself. However, these solutions are often expensive as they require customized procedures. Therefore, is simpler from an operational and functional point of view to design a electronic control system that rejects this torque disturbance. The interest part of the following work is the usage of a sensor-less architecture in the context of the Cogging Torque. Exploiting the non-linear control technique designed in our previously work, that is based on the mathematical model of the Cogging Torque as a function of the rotor position. The challenge is to verify if our control algorithm works also with sensor-less architecture in which the measure of the encoder/resolver is not available. There are many works in the literature that propose a control system for synchronous motors with permanent magnets based on sensor-less architecture. In [3–5] are reported the results of implementation of sensor-less architectures based on EKF (Extended Kalman Filter), in discrete time, for brushless motors with air gap induction type trapezoidal, i.e. Brushless DC motors (BLDC). In [6] a solution is presented which makes use of a modern variant of Kalman Filter called Unscented Kalman Filter (UKF). In [7, 8] the sensor-less architecture used involved a sliding mode state observer; in [8] also the controller is developed with a sliding mode technique. In [9] the project of a sensor-less architecture is described exploiting the criteria of the H-infinity control theory while in [10] the project of a state observer based on the theory of neural networks is proposed. In [11] the possibility of using a sensor-less architecture to reduce torque ripple is presented.

Compared to the other works to which reference is made, regarding sensor-less architectures based on state observer, in this article a non-linear control system based on EKF has been developed in continuous time instead of discrete time.

Furthermore, compared to the previous works, the proposed EKF model refers to the dynamics of the motor expressed in three-phase axes instead of in direct-quadrature axes. This is because from an operational point of view the current and voltage measurements available as observer inputs are actually deriving from the three-phase dynamics. Furthermore, compared to the previous works, the reduction of an intrinsic disturbance of the machine which depends directly on the variables estimated by EKF is included.

36.2 Cogging Torque

The architecture of a Brushless motor is composed of:

i. a stator identical to an Asynchronous motor which inside the slots has arranged the copper windings
ii. a rotor (typically both with cylindrical symmetry) on which permanent magnets are arranged.

There are two main configurations, the first identified with the abbreviation SPM (surface permanent magnets) in which the magnets are arranged on the outer surface of the rotor, the second is indicated with IPM (internal permanent magnets) where the permanent magnets are set in the iron of rotor. The problem of Cogging Torque is present for both SPM and IPM brushless motors. In this work, as in [12], reference is made to a brushless motor with SPM architecture. The Cogging Torque is born from the magnetic interaction between the permanent magnets placed on the rotor surface and the teeth of the stator slots. Further, since it is completely due to the field produced by the permanent magnets, it is also present when the electric motor is not powered. This magnetic interaction causes the birth of magnetic forces that can be represented as force vectors applied in the centre of the magnets themselves, which create a mechanical torque that induces the rotation of the rotor. The direction of the magnetic force vectors depends on how the flow lines of the magnetic field produced by the magnets are closed in the stator iron through the teeth of the slots. It is intuitive that the way in which the magnetic flux lines generated by the magnets pass through the stator teeth, depends on the reciprocal position between the rotor and the stator.

Each magnetic flux line can be interpreted as a line passing between the centre of a magnet and the centre of a stator tooth, closed in the rotor and stator iron, as shown schematically in Fig. 36.1.

Since the internal structure of a brushless motor is symmetrical, it is also intuitive that the same tooth-magnet configuration is repeated several times throughout a

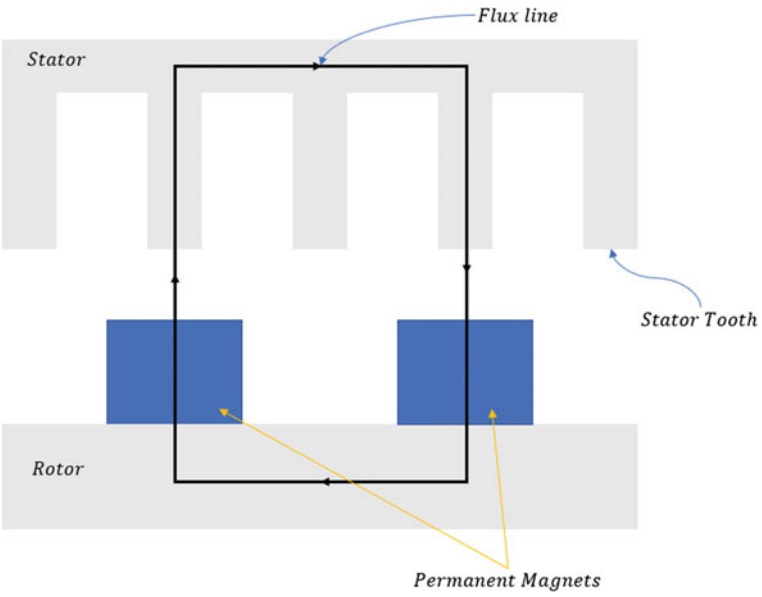


Fig. 36.1 Schematic representation of the path of the magnetic flux lines

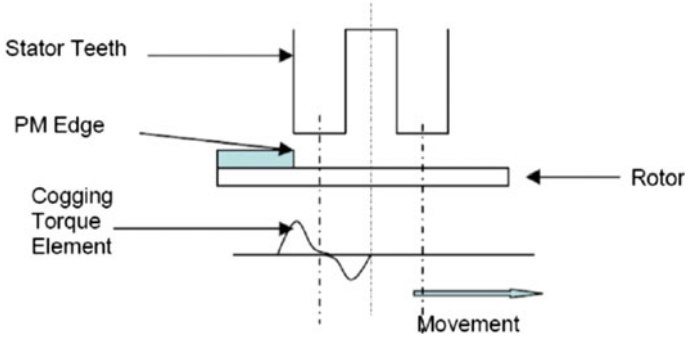


Fig. 36.2 Schematic representation of the contribution to the Cogging Torque of a single tooth-magnet pair

corner. Further it is possible to think of the total Cogging Torque as the overlapping of the effects of "cogging torque elements" associated with each tooth-magnet pair.

As schematically shown in Fig. 36.2, during the rotation, each magnet (due to the direction of the closing of the flow) sees a force applied which attracts it towards the tooth and once the centre of the magnet has passed the centre of the tooth, it sees a force that rejects it. Locally there is a symmetrical situation in terms of attraction and repulsion forces which suggests that the Cogging Torque can be represented as a periodic function of the rotation angle, whose period depends on the number of magnets and stator teeth, with average zero value. The Cogging Torque depends only on the magnetic interaction between magnets and teeth, due to the field produced by the magnets themselves, so this torque can be interpreted as an additive disturbance with respect to the electromagnetic torque which instead depends only on the currents supplied by the motor when it is powered.

As in [12], this article also refers to a closed description of the cogging pair. In particular, reference is made to [13] which describes the mathematical model reported in Eq. (36.1).

$$T_{cog} = \sum_{k=1}^{m} T_k \sin(kZ\theta + \alpha_k) \tag{36.1}$$

where T_k and α_k represent the coefficients relative to the kth harmonic of the Fourier series, Z is the number of stator teeth and m is the number of harmonics necessary to approximate the behaviour of the Cogging Torque.

36.3 Control System Design

In this article we refer to our previous work [12] on the design of a control system that solves the problem of Cogging Torque, extending it to the case in which a sensor-less architecture is used. In general, sensor-less architecture involves the design of a system for estimating the angular position and angular velocity of the rotor that exploits the measurement of the three-phase voltages that supply the motor and the three-phase currents supplied. In this work a continuous-time EKF as a state observer was considered (Fig. 36.3).

With reference to the unified theory of electrical machines [14], for the design of a control system for three-phase motors it is advisable to describe the dynamic equilibrium of the machine in terms of two-phase equivalent circuits. As shown in Fig. 36.3, the direct transformations of Blondel (block B) and Park [block $A(\hat{\theta})$] are applied to the current vector, while the inverse transformations ($A^T(\hat{\theta})$ and B^T) are applied to the voltage vector output from the FLC block. In this work the design of the EKF refers to the dynamic model expressed in the three-phase coordinate reference while the control laws are expressed in the direct and quadrature axis system (which are obtained after the application of both direct coordinate transformations $A(\theta)$ and B).

For completeness, in Eqs. (36.2) and (36.3) are reported the dynamics equation of the electro-mechanic equilibrium both in three-phase and direct-quadrature frames respectively.

$$\begin{pmatrix} u_a \\ u_b \\ u_c \end{pmatrix} = R_s \begin{pmatrix} i_a \\ i_b \\ i_c \end{pmatrix} + L_{eq} \frac{d}{dt} \begin{pmatrix} i_a \\ i_b \\ i_c \end{pmatrix} - p\omega k_\varphi \begin{pmatrix} \sin p\theta \\ \sin\left(p\theta - \frac{2\pi}{3}\right) \\ \sin\left(p\theta - \frac{4\pi}{3}\right) \end{pmatrix}$$

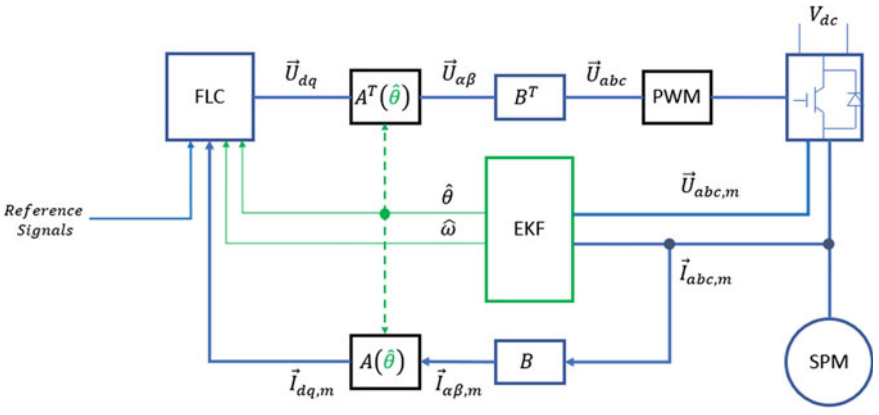


Fig. 36.3 Schematic representation of the complete system

$$J_m \frac{d\omega}{dt} + \beta\omega = -pk_\varphi \sum_{k=1}^{3} i_k \sin\left(p\theta - \frac{2(k-1)\pi}{3}\right) + T_{cog} \tag{36.2}$$

$$\begin{pmatrix} u_d \\ u_q \end{pmatrix} = R_s \begin{pmatrix} i_d \\ i_q \end{pmatrix} + L_{eq} \frac{d}{dt} \begin{pmatrix} i_d \\ i_q \end{pmatrix} + p\omega \begin{pmatrix} -L_{eq} i_q \\ L_{eq} i_d + i_d \end{pmatrix}$$

$$J_m \frac{d\omega}{dt} + \beta\omega = \frac{3}{2} pk_\varphi i_q + T_{cog} \tag{36.3}$$

where u_a, u_b, u_c and i_a, i_b, i_c are the three-phase domain voltage and current components, meanwhile u_d, u_q and i_d, i_q are the voltage and current components in the direct-quadrature reference frame. With R_s and L_{eq} are indicated the stator electrical resistance and equivalent inductance respectively. With p is indicated the pole pairs, J_m represents the inertia moment of the rotor and β is the friction coefficient; θ and ω represent the rotor angular position and speed respectively.

From an operational point of view, considering the three-phase axis model of the motor for the EKF project is advantageous compared to the design of the motor model in direct-quadrature axes as it avoids having a feedback of the estimated angle for the calculation of the transformation of Park on current measurements, which instead is present in [3–5]. As in our previous work [12], in the dynamics of the motor we refer to a mathematical model of Cogging Torque with seven harmonics in the Fourier development of Eq. (36.1), while in the model used for the development of the FLC control reference is made to the four-harmonic model.

In Eqs. (36.4) and (36.5) the control laws (the vector of the voltages that supply the motor expressed in axes dq) and the update dynamics of the state observer are summarized.

$$\begin{pmatrix} u_d \\ u_q \end{pmatrix} = \frac{\begin{pmatrix} L_{g_2} h_2 & -L_{g_2} L_f^2 h_1 \\ -L_{g_1} h_2 & L_{g_1} L_f^2 h_1 \end{pmatrix}}{L_{g_1} L_f^2 h_1 L_{g_2} h_2 - L_{g_1} h_2 L_{g_2} L_f^2 h_1} \left[\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} - \begin{pmatrix} L_f^3 h_1 \\ L_f h_2 \end{pmatrix} \right] \tag{36.4}$$

$$\begin{aligned} \frac{d\hat{x}}{dt} &= F(\hat{x})\hat{x} + G(\hat{x})u + K_{EKF}(y - C\hat{x}) \\ \frac{dP}{dt} &= F(\hat{x})P + PF(\hat{x})^T + Q - K_{EKF} R K_{EKF}^T \\ K_{EKF} &= PC^T R^{-1} \end{aligned} \tag{36.5}$$

In Eq. (36.2), $L_f^k h_j$ is the kth Lie derivative of the generic output function h_j along the direction of the vector field f, while the vector $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ represents the auxiliary control in the new base identified through the operating procedure of the technique used. Reference is made to a proportional auxiliary control, a simple static feedback of the state variables defined in the new base [12].

In Eq. (36.5) we indicate with $\hat{x}$ the vector of the estimated state, referring to the state representation derived from the dynamics expressed in three-phase axes, $F(\hat{x})$

and $G(\hat{x})$ are the Jacobian matrices calculated in the current estimated value, y is the vector of the variables of measurable state (current vector), C is the vector that maps the state vector in the output vector, K_{EKF} is the gain of the observer, Q and R are the covariance matrices of the noises that act on the vector of state x and on the vector of input u and P is the covariance matrix of the estimated state, which solves the *Riccati* matrix differential equation.

Compared to our previous work, the expression reported in Eq. (36.4) will depend on the estimated state (relative to the state representation in axes dq). It means that the model of the Cogging Torque that appears in the dynamics of the motor is in relation to the true position while the part of the controller that uses this model is a function of the estimated position.

36.4 Simulation Results

To verify the validity of the proposed architecture, the results of the trajectory tracking problem in terms of desired position and current in quadrature axis are presented below.

The results in terms of estimation of the currents expressed in three-phase reference and the result of the estimation of the angular velocity are also reported in the following.

For all the simulations that we present, reference is made to the following parameters for the motor and EKF models (Tables 36.1, 36.2).

Table 36.1 Parameters of the brushless model

R_s	3.3 Ω
L_{eq}	50 mH
k_ϕ	0.5 Wb
J_m	0.02 Kgm^2
p	3
β	0.01 Ns/m

Table 36.2 Cogging Torque model parameters

T_1	4.85 N
T_2	2.04 N
T_3	0.3 N
T_4	0.06 N
α_1	0.009 rad
α_2	0.01 rad
α_3	0.017 rad
α_4	α_3

For the model of the measurement noises we assume additive Gaussian signal for the current and voltage components. In particular we set null mean value for both current and voltage noise and 1 [A] and 5[V] standard deviation respectively.

The first type of desired trajectory for rotor positioning is a general shape in which are presents some change of rotative direction without steady state phases.

The covariance matrixes $P(t_0)$, Q and R are set to identity matrix of just dimension.

Figure 36.4 shows the estimation result for what concerns three-phase current components, that is basically the first check since that are the available measures.

One of the issues linked with the usage of the EKF as estimation system, is the initial condition settings for what concern the updating dynamic equation of the filter itself.

Clearly if the initial conditions of the estimated state vector are to much different respect the initial conditions of the reference process dynamics model, the EKF cannot realizes a good estimation.

It is mandatory to verify the EKF performance with different initial conditions further than with different desired trajectory in terms of rotor position.

In particular, we have verified that the estimated position converged to real position set as initial condition $\theta(t_0) = 0.5$, $\theta(t_0) = 1.0$, $\theta(t_0) = 1.5$ and $\theta(t_0) = 2.0$ fixing initial condition for the other state variable (current components and speed).

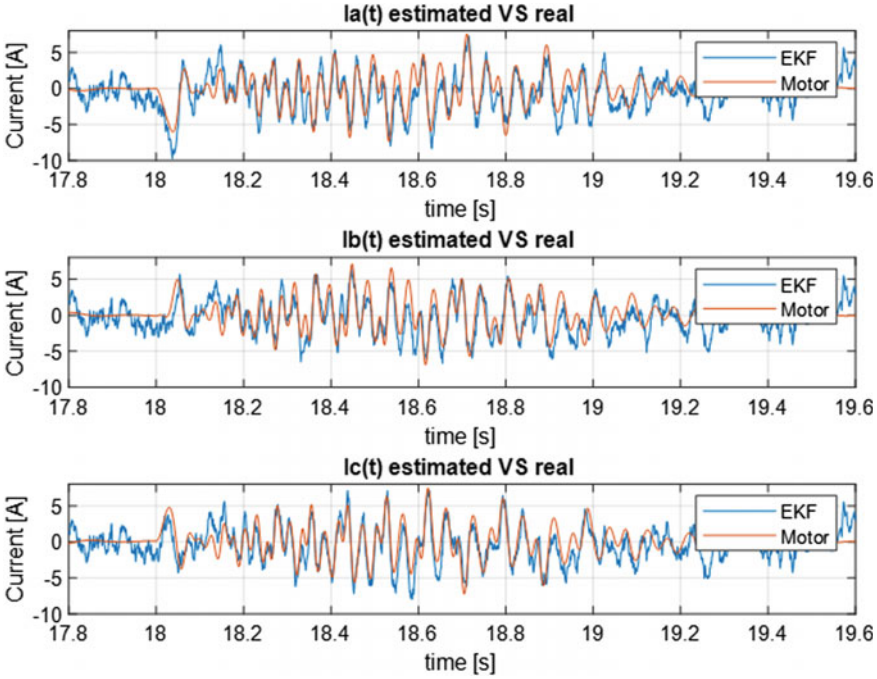


Fig. 36.4 Three-phase currents estimation

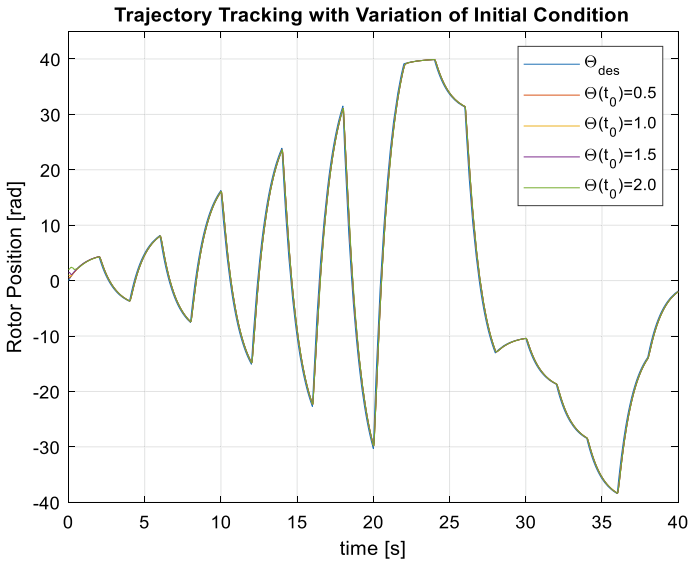


Fig. 36.5 Trajectory tracking of rotor axis position

Figure 36.5 represents the result of the trajectory tracking for the rotor axis positioning to desired behaviour, with different initial conditions in terms of rotor start position.

Figure 36.6 shows that the estimated position is convergent to the real position, in particular is reported the behaviour of the position error related to the results shown in Fig. 36.5.

Figures 36.7 and 36.8 shows the result in term of rotor angular speed estimation with different initial conditions, in order to verify the robustness of the EKF as a function of the motor speed variation. Figure 36.9 shows the behaviour of the estimation error of angular speed of the rotor, for different speed initial condition, fixing the initial condition for the other variables. Figure 36.10 shows the result of the trajectory tracking control problem for the direct current component. From the theory of the Brushless control [14], to emulate a FOC (Field Oriented Control) architecture, it is imposed a null reference signal for this current component.

36.5 Conclusion and Future Work

In this work a sensor-less architecture based on the EKF observer and on feedback linearization control system is proposed. We have verified that also with a sensor-less architecture it is possible to assume our previously control solution [12] to reduce the intrinsic problem of the Cogging Torque. Results in term of the trajectory tracking control problem on the direct current component and on the rotor axis position are

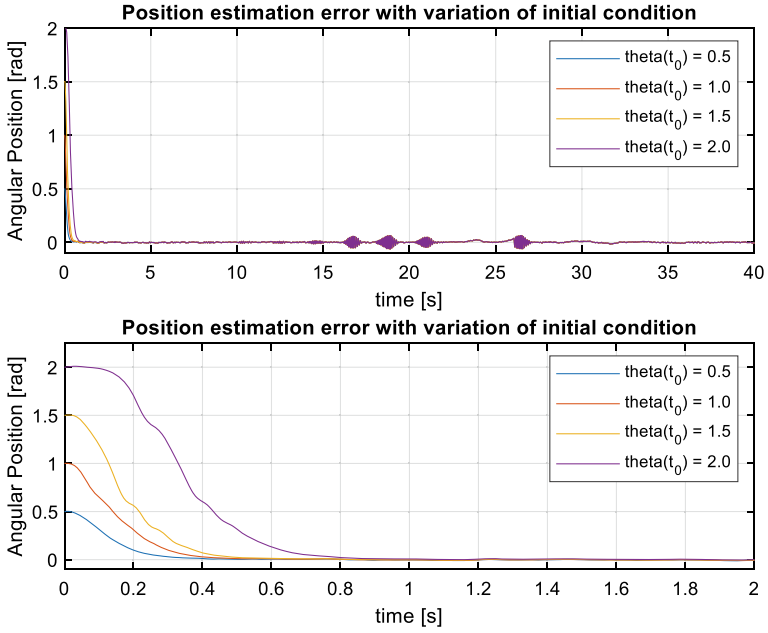


Fig. 36.6 Position estimation error

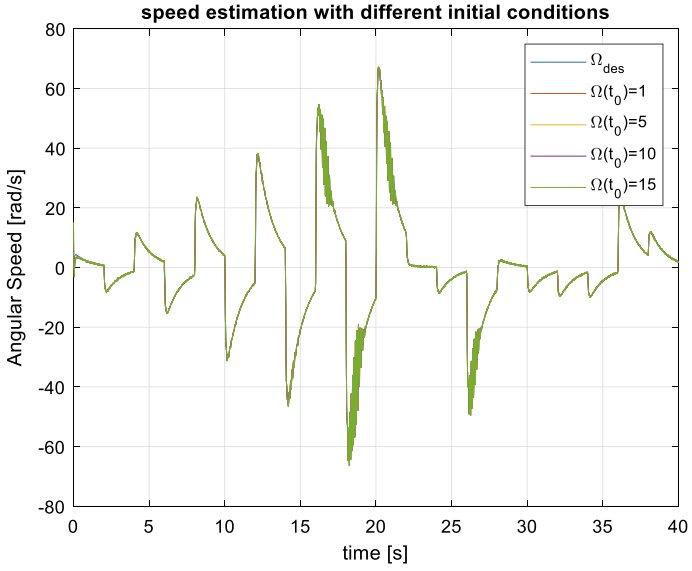


Fig. 36.7 Rotor speed estimation

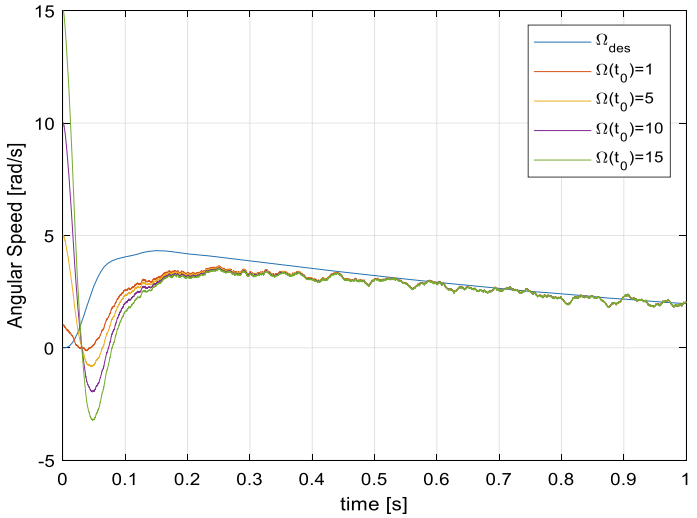


Fig. 36.8 Transitory phase in rotor speed estimation

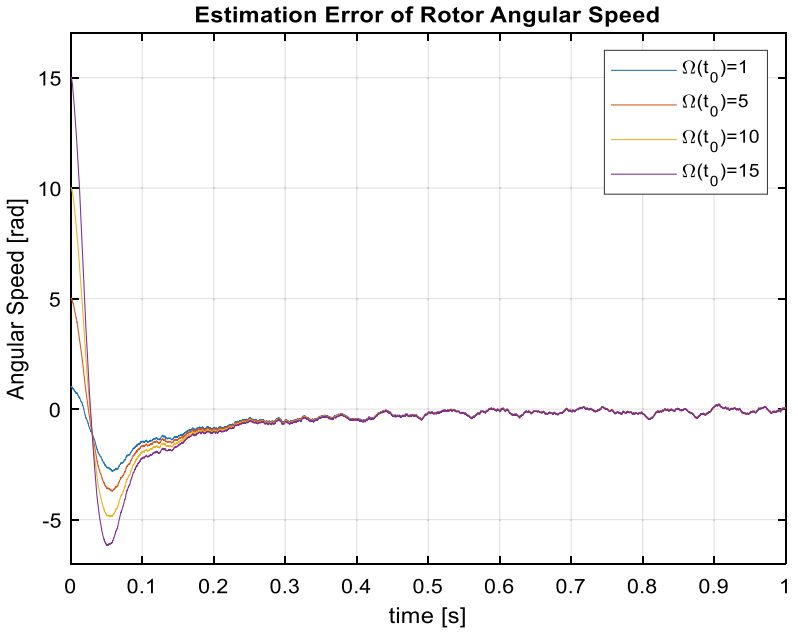


Fig. 36.9 Transitory phase of the angular speed estimation error

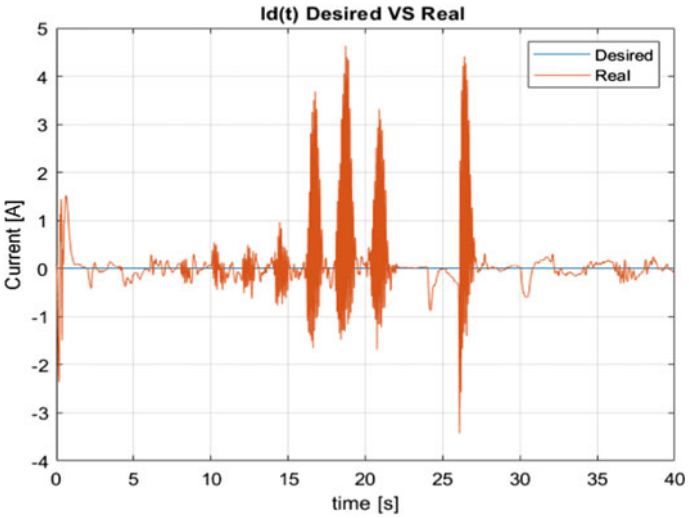


Fig. 36.10 trajectory tracking result in term of direct current $i_d(t)$

presented. An analysis on the initial condition of the rotor axis position reveals that the architecture is robust in term of variation of start condition of the global system.

In summary, the contribution of this work is to demonstrate that through a continuous-time EKF observer that refers to the dynamics of the motor in three-phase axes, it is possible to realize a sensor-less architecture that exploits the design of a FLC controller for Cogging Torque reduction, which instead is derived by referring to the dynamics expressed in direct-quadrature axes.

As future work, it could be interesting to verify if the new architecture solution can be implemented on a real embedded system as the sensor-based version proposed in [12]. Clearly the introduction of the EKF increase the complexity of the global control system. With respect to the sensor based FLC version, in which only algebraic operation was required, in this case the EKF requires to solve iteratively a differential matrix equation (*Riccati* equation) that reasonable requires higher computational capability than a low-cost embedded system like Arduino Uno, used in [12].

Another extension to this work could be the comparison between different types of state observatories, UKF or Sliding Mode or Neural Networks, to verify that the reduction of the cogging pair, based on FLC, can be implemented in any type of sensorless architecture and what is the best configuration.

References

1. Caruso M et al (2015) Analysis, characterization and minimization of IPMSMs cogging torque with different rotor structures. In: 2015 International conference on ecological vehicles and renewable energies (EVER)

2. Hwang Myeong-Hwan, Lee Hae-Sol, Cha Hyun-Rok (2018) Analysis of torque ripple and cogging torque reduction in electric vehicle traction platform applying rotor notched design. Energies 11(11):3053
3. Zhang Z, Feng J (2008) Sensorless control of salient PMSM with EKF of speed and rotor position. In: IEEE international conference on electrical machines and systems
4. Termizi MS et al (2017) Sensorless PMSM drives using Extended Kalman Filter (EKF). In: 2017 IEEE conference on energy conversion (CENCON). IEEE
5. Tian G et al (2018) Rotor position estimation of sensorless PMSM based on Extented Kalman Filter. In: 2018 IEEE international conference on mechatronics, robotics and automation
6. Lv H, Wei G, Ding Z (2014) UKF—based for sensorless brushless DC motor control. In: IEEE 2014 international conference on mechatronics and control (ICMC)
7. Tingna S, Na L, Li W (2010) Sensorless control for brushless DC motors using adaptive sliding mode observer. In: 29th Chinese control conference
8. Zheng C, Li Y (2016) Sensorless speed control for brushless DC motors system using sliding-mode controller and observers. In: 2016 8th International conference on intelligent human-machine systems and cybernetics (IHMSC), vol 1. IEEE
9. Vinida K, Chacko M (2016) A novel strategy using H infinity theory with optimum weight selection for the robust control of sensorless brushless DC motor. In: 2016 IEEE symposium on sensorless control for electrical drives (SLED). IEEE
10. Jinquan Z, Deying G (2014) Control method of sensorless brushless DC motor based on neural network. In: IEEE 26th Chinese control and decision conference
11. Sun X et al (2016) A sensorless method with torque ripple suppression for brushless DC motors. In: 19th International conference on electrical machines and systems (ICEMS). IEEE
12. Dini P, Saponara S (2019) Cogging torque reduction in brushless motors by a nonlinear control technique. Energies 12(11):2224
13. Tudorache T et al (2012) Improved mathematical model of PMSM taking into account cogging torque oscillations. Adv Electr Comput Eng 12(3):59–64
14. Krause P (2017) Introduction to electric power and drive systems. Wiley, New York

Part VIII
Signal and Data Processing

Chapter 37
Acoustic Emissions Detection and Ranging of Cracks in Metal Tanks Using Deep Learning

Gian Carlo Cardarilli, Luca Di Nunzio, Rocco Fazzolari, Daniele Giardino, Marco Matta, Marco Re and Sergio Spanò

Abstract This work proposes a new method for the estimation of the distance of cracks in pressure metal tanks. This method is obtained coupling the acoustic emissions analysis and the deep learning techniques. Using a 2D CNN we are able to estimate the distance between a crack and an acoustic emission piezoelectric sensor. The CNN is trained on images representing the spectrogram of acoustic emission located at distances of 2, 20, 40, 60, 80, 100, 120 and 140 cm. We obtained a RMSE of 2.54 cm.

37.1 Introduction

The method of acoustic emissions (AE) is a commonly used non-destructive testing (NDT) technique to detect defects in mechanically loaded structures and parts. If a system is exposed to mechanical load or pressure, the occurrence of structural

G. C. Cardarilli · L. Di Nunzio (✉) · R. Fazzolari · D. Giardino · M. Matta · M. Re · S. Spanò
Department of Electronic Engineering, University of Rome Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy
e-mail: di.nunzio@ing.uniroma2.it

G. C. Cardarilli
e-mail: cardarilli@ing.uniroma2.it

R. Fazzolari
e-mail: fazzolari@ing.uniroma2.it

D. Giardino
e-mail: giardino@ing.uniroma2.it

M. Matta
e-mail: matta@ing.uniroma2.it

M. Re
e-mail: re@ing.uniroma2.it

S. Spanò
e-mail: spano@ing.uniroma2.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_37

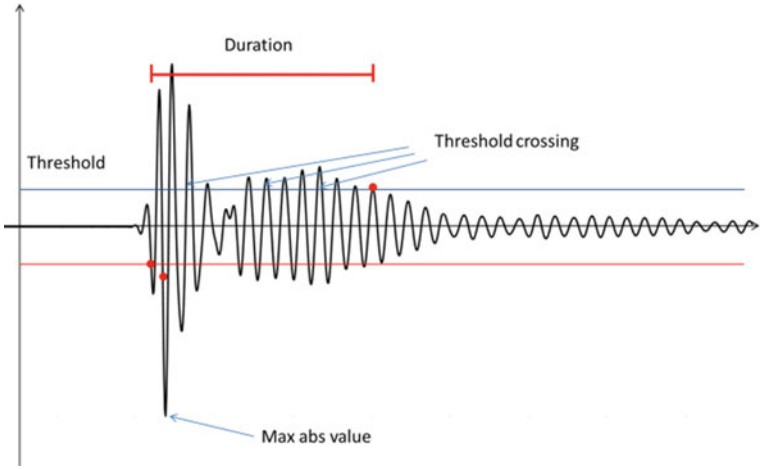


Fig. 37.1 Acoustic emission signal and its feature set

discontinuities generates power through the constituent material in the manner of sound emissions. The AE technique enables to check the integrity of a broad range of bodies by evaluating information from piezoelectric sensors and hardware devices.

Testing the mechanical properties of pressure tanks is one of the most common implementations in this field. The current legislation provides for the use of this methodology in some countries (e.g. Italy) [1]. Other standard checking protocols are based on periodic inspections that does not allow continuous monitoring and require the use of very unwieldy instrumentation. AE refers to the generation of transient elastic waves (Fig. 37.1) by a sudden redistribution of stress in a material [2, 3]. The analysis of the AE waves can be used to detect damage on structures.

During the periodic checks, a typical AE based procedure requires the installation of specific sensors on the structure under test by specialized operators. These sensors are generally wired and connected to a central processor that collects and analyzes the data for rupture events detection [4]. The introduction and development of embedded systems for Wireless Sensor Networks (WSN) [5, 6] and IoT [7, 8] allowed the Acoustic Emission procedures to be capable of real time monitoring and enhanced the data management. In addition to the detection, the automatic localization of the crack is another crucial aspect. As a matter of fact, tanks size can reach hundreds of meters in length and the non-automated localization can prove to be very difficult.

The literature presents different solutions for automatic localization of cracks. The most used consists in the use of a great number of sensors equally distributed forming a grid on the tank [9]. Localization is performed by applying triangulation on data coming from the sensors. Although this method is the most commonly used and approved by the laws of many countries, it requires a large number of sensors and the timing synchronization of all them. For this reason, in the last years, other methods based on the dispersion of the modes of the acoustic emission signal have been proposed [10]. Unfortunately, these methods do not work correctly on surfaces

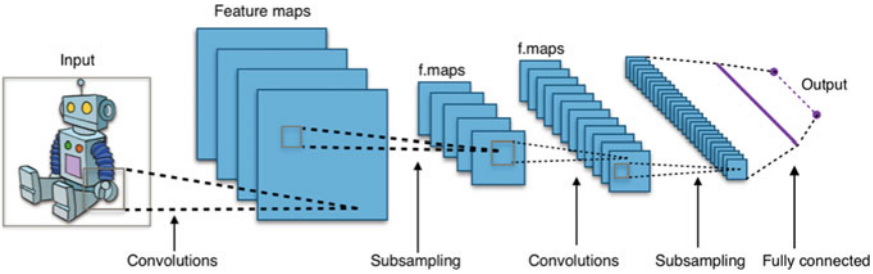


Fig. 37.2 Architecture of a convolutional neural network for image classification or regression

with obstacles and junctions. In fact, in these cases, echoes signals generated by rebounds interfere with the analysis.

In recent times, the development of Artificial Intelligence and Machine Learning (ML) algorithms enabled the use of such techniques in several applications, being the signal processing and analysis of sensor network data two of them.

The main task categories that are addressed with a ML approach can be divided in three groups: classification of data, pattern recognition and data regression. Generally, a ML design flow requires the developer to select a suitable feature set that characterizes the data, the algorithm training stage and, finally, its deployment in inference stage. The features (Fig. 37.1) that are usually employed in AE applications are the maximum of absolute value of the amplitude, the signal duration, the signal energy and the number of crossings of a given threshold [7].

In recent years, among the numerous AI model categories, Deep Learning (DL) became a trending topic, both in research and industry [11–16]. The main advantage of a DL approach with respect to a traditional ML one is the capability to learn and extract automatically the feature maps that are needed to achieve the solution of the task. Deep Learning gained popularity after the achievements of Convolutional Neural Networks (CNN), shown in Fig. 37.2, which are typically used in the computer vision field of applications for classification and regression of image data [17].

In this paper we propose a method based on Deep Learning and CNN to process the image of an Acoustic Emission in a metal pressured tank in the form of its spectrogram. By designing a regression model, the algorithm is able to detect a rupture event and to estimate its distance (ranging) from the piezoelectric sensor. Implementing this technique in multiple sensor nodes of the network, it is also possible to estimate the position of the event by triangulation (localization). Other research works apply similar CNN approaches in other fields of application, such as seismic events detection [18] and biomedical [19].

37.2 Materials and Methods

To design a regression AI system, both in ML and DL it is necessary to collect and/or generate a suitable labeled database to train the model. In our case we arranged a mixed mechanical and electronic framework to build the AE training and test dataset. The framework included a 5 mm thick plate (160 × 40 cm) made of steel (that simulates the structure under test), a VALLEN 150M piezoelectric sensor and a 2H pencil lead which diameter is 0.3 mm. The pencil was used to generate the acoustic emission by stimulating the plate using an incident angle of 45°, this is a standard setup for such AE experiments [20].

The signals were sampled at 2 Mega samples per second (Msps) generating the waveforms at different distances from the sensor, as shown in Fig. 37.3.

We considered the following distances: 2, 20, 40, 60, 80, 100, 120 and 140 cm. For every distance, we took 10 measurements obtaining a total dataset of 80 instances. In order to expand our database, we replicated 10 times each measure and added a White Gaussian Noise (AWGN) with Signal to Noise Ratio (SNR) 20 dB. We then obtained a final dataset with a total of 800 instances.

37.2.1 Deep Learning Framework

In order to apply 2D Convolutional Neural Networks to our dataset, we generated the spectrograms to be fed to the AI systems. The spectrograms used a Hamming window of size 128, an overlap of 127 samples and an FFT size of 512. Figure 37.4 shows some examples of the obtained spectrograms. Subfigure (a) refers to a crack event at a distance of 20 cm, (b) 40 cm, and (c) 100 cm. Spectrograms size is 257 × 397 points.

The architecture of the 2D CNN was developed empirically using MATLAB 2019a. Figure 37.5 shows the building layers of our proposed network.

All the maxpool functions act on 2 × 2 windows with a stride factor of 2. All the convolutional layers have 3 filters: the size of the filters of the first one is 8, 16 in

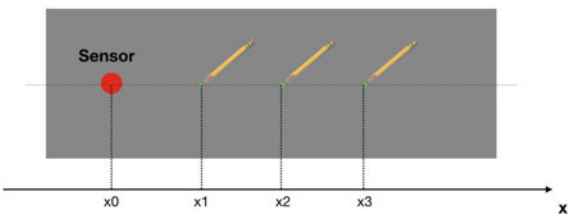


Fig. 37.3 Experimental setup framework: generation of acoustic events on the metal plate at different distances

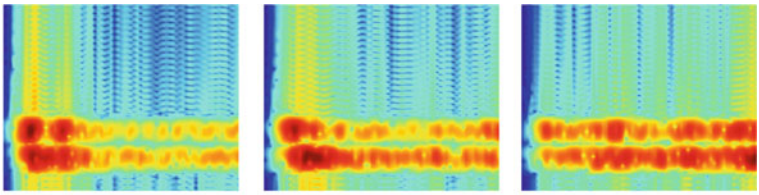


Fig. 37.4 Examples of spectrograms

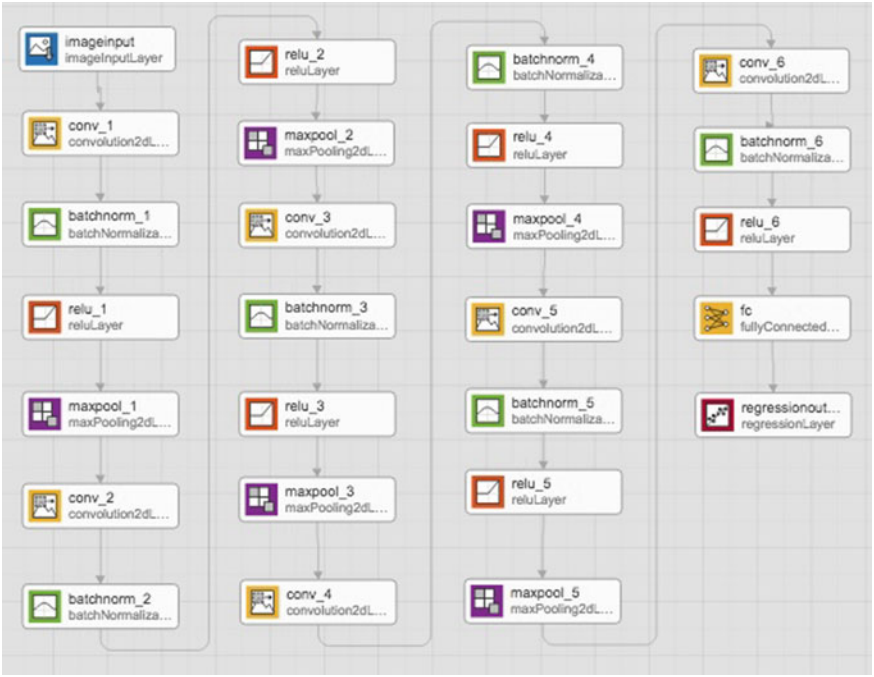


Fig. 37.5 Convolutional neural network layers stack

the second one, and 32 in the others. The final regression layer outputs the estimated distance of the AE waveform received by the sensor.

37.3 Experimental Results

We trained the convolutional network using our 800 spectrograms dataset. We applied the cross-validation technique considering 75% of the instances for the training set

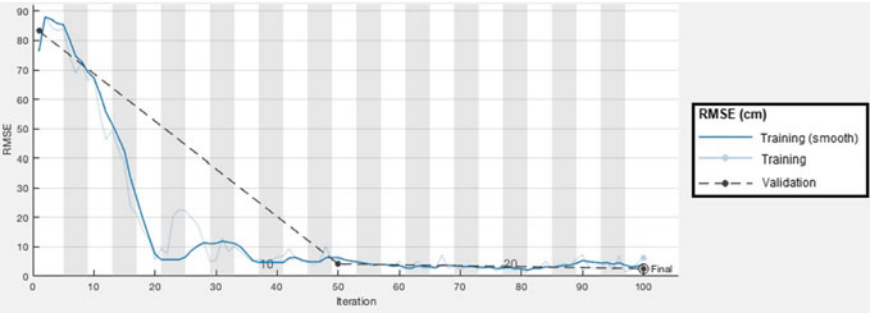


Fig. 37.6 RMSE (cm) and loss after the training process

and the other 25% for the validation set. As shown in Fig. 37.6, we obtained an optimal convergence after about 50 iterations.

The results are presented in terms of Root Mean Square Error (RMSE). We obtained a RMSE of 2.54 cm. Considering the distance steps between the dataset instances we can state that the network is able to have a good generalization and to locate the rupture very accurately.

37.4 Conclusions

We designed a Convolutional Neural Network that is able to generalize the distance between an Acoustic Emissions sensor and a rupture point in a pressure metal tank. Our proposed network presents a RMSE of 2.54 cm meaning that the system can locate the defects very accurately. Our method, if applied to multiple sensors connected to a same network, would be able perform a geometrical triangulation and to locate the rupture point precisely. A possible evolution of the work is in the direction of applying additional advanced signal processing techniques to allow accurate detection even in very noisy environments. This technique can be improved using algorithms based both on Pulse Compression and on adequate windowing techniques [21]. Moreover, the use of unsupervised processing tools to dynamically improve the signal-to-noise ratio allows a more accurate estimate of the parameters that characterize the AE. In sight of a practical implementation of our method, it is important to gather the acoustic emission sensor data from the tank during its installation. This is because the AE characteristics differ from one tank to another, due to microscopic structural imperfections of the material.

References

1. Attuazione della direttiva 97/23/CE in materia di attrezzature a pressione. dlgs 93/2000 Italian Legislation
2. Cardarilli GC, Di Nunzio L, Massimi F, Fazzolari R, De Petris C, Augugliaro G, Mennuti C (2018) A wireless sensor node for acoustic emission non-destructive testing. Lect Notes Electr Eng
3. Bechhoefer E, Qu Y, Zhu J, He D (2013) Signal processing techniques to improve an acoustic emissions sensor. In: Proceedings of the annual conference of the prognostics and health management society. pp 581–58
4. Grosse Christian U, Ohtsu M (eds) (2008) Acoustic emission testing. Springer Science & Business Media, Berlin
5. Akyildiz Ian F et al (2002) Wireless sensor networks: a survey. Comput Netw 38(4):393–422
6. Perumalla V, Ramanjaneyulu BS, Kolli A (2017) Simulation study of topological structures and node coordinations for deterministic WSN with TSCH. Int J Inform Vis 1(4)
7. Giardino D, Matta M, Spanò S (2019) A feature extractor IC for acoustic emission non-destructive testing. Int J Adv Sci Eng Inf Technol 9(2):538–543
8. Giuliano R, Mazzenga F, Neri A, Vegni AM (2017) Security access protocols in IoT capillary networks. IEEE Internet Things J 4(3):645–657
9. Riqualificazione serbatoi GPL con metodo EA, Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro (INAIL) (2019). https://www.inail.it/cs/internet/attivita/ricerca-e-tecnologia/certificazione-verifica-e-innovazione/certificazione/riqualificazione-serbatoi-gpl-con-metodo-ea.html
10. Ni Q-Q, Iwamoto M (2002) Wavelet transform of acoustic emission signals in failure of model composites. Eng Fract Mech 69(6):717–728
11. Lu Y (2017) Industry 4.0: A survey on technologies, applications and open research issues. J Ind Inf Integr 6:1–10
12. Matta M, Cardarilli GC, Di Nunzio L, Fazzolari R, Giardino D, Nannarelli A, Re M, Spanò S (2019) A reinforcement learning based QAM/PSK symbol synchronizer. IEEE Access
13. Cardarilli GC, Di Nunzio L, Fazzolari R, Nannarelli A, Re M, Spano S (2019) N-dimensional approximation of euclidean distance. IEEE Trans Circuits Syst II Express Briefs
14. Cardarilli GC, Di Nunzio L, Fazzolari R, Re M, Spano S (2019) AW-SOM, an algorithm for high-speed learning in hardware self-organizing maps. IEEE Trans Circuits Syst II: Express Briefs
15. Cardarilli GC, Di Nunzio L, Fazzolari R, Giardino D, Matta M, Re M, Silvestri F, Spanò S (2019) Efficient ensemble machine learning implementation on FPGA using partial reconfiguration. Lect Notes Electr Eng 550:253–259
16. Hordri NF, Yuhaniz SS, Shamsuddin SM (2016) Deep learning and its applications: a review. In: Postgraduate annual research on informatics seminar 2016, Universiti Teknologi Malaysia
17. Russakovsky O et al (2015) Imagenet large scale visual recognition challenge. Int J Comput Vis 115(3):211–252
18. Zhu L, Peng Z, McClellan J (2018) Deep learning for seismic event detection of earthquake aftershocks. In: 2018 52nd asilomar conference on signals, systems, and computers, IEEE
19. Zhang J et al (2019) Fine-grained ECG classification based on deep CNN and online decision fusion. Preprint at arXiv:1901.06469
20. ASTM E-976, Standard Guide for Determining the Reproducibility of Acoustic Emission Sensor Response, ASTM International
21. Burrascano P, Laureti S, Senni L, Ricci M (2018) Pulse compression in nondestructive testing applications: reduction of near sidelobes exploiting reactance transformation. IEEE Trans Circuits Syst I Regul Pap (99):1–11. https://doi.org/10.1109/tcsi.2018.2862868

Chapter 38
Recognizing Breathing Rate and Movement While Sleeping in Home Environment

Maksym Gaiduk, Ralf Seepold, Natividad Martínez Madrid, Simone Orcioni and Massimo Conti

Abstract The recovery of our body and brain from fatigue directly depends on the quality of sleep, which can be determined from the results of a sleep study. The classification of sleep stages is the first step of this study and includes the measurement of vital data and their further processing. The non-invasive sleep analysis system is based on a hardware sensor network of 24 pressure sensors providing sleep phase detection. The pressure sensors are connected to an energy-efficient microcontroller via a system-wide bus. A significant difference between this system and other approaches is the innovative way in which the sensors are placed under the mattress. This feature facilitates the continuous use of the system without any noticeable influence on the sleeping person. The system was tested by conducting experiments that recorded the sleep of various healthy young people. Results indicate the potential to capture respiratory rate and body movement.

M. Gaiduk (✉) · R. Seepold
HTWG Konstanz, Alfred-Wachtel-Str. 8, 78462 Konstanz, Germany
e-mail: maksym.gaiduk@htwg-konstanz.de

R. Seepold
e-mail: ralf.seepold@htwg-konstanz.de

M. Gaiduk
University of Seville, Av. Reina Mercedes s/n, 41012 Seville, Spain

R. Seepold · N. Martínez Madrid
Department of Information and Internet Technology, Sechenov University, Bolshaya Pirogovskaya st., 119435 Moscow, Russian Federation
e-mail: natividad.martinez@reutlingen-university.de

N. Martínez Madrid
Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany

S. Orcioni · M. Conti
Department of Information Engineering, Università Politecnica delle Marche, Via Brecce Bianche, 12, 60131 Ancona, Italy
e-mail: s.orcioni@univpm.it

M. Conti
e-mail: m.conti@univpm.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_38

38.1 Introduction

Sleep is necessary for everybody and sleeping an adequate time with good quality ensures that people feel good and have more energy for their daily tasks. The National Sleep Foundation (NSF) recommends that adults sleep 7–8 h a day [1, 2].

The sleep phases can be divided into two main categories: Non-Rapid Eye Movement (NREM) and Rapid Eye Movement (REM) phase. The REM phase occurs after an initial stage of deep sleep. It is a phase in which dreams occur while the eyes move rapidly in different directions, with the heart and respiratory rate becoming irregular. The REM phase alternates with light and deep stages of the NREM phase and becomes longer with each sleep cycle. Adults with healthy sleeping habits spend about 20% of their time in the REM phase, while this percentage decreases with age [3].

Typically, sleep is analyzed in a sleep laboratory using polysomnography (PSG). Here the electrophysiological signals are recorded and interpreted. However, sleeping in this environment differs from a "normal" sleep at home. For a person to be monitored over a longer period of time, home installation is the only possible solution. An additional aspect is that the costs of a "home" system should be significantly lower, while the most important relevant sleep parameters can still be collected [4]. For example, monitoring movement and breathing will support the detection of apnea [5].

The main objective of the method presented in this work is to track and analyze a person's movement, breathing, and heart rate during sleep. The main difference of this system compared to other approaches is the innovative way of placing the sensors under the mattress to ensure the familiar sleeping comfort.

38.2 Methodology

There are several types of pressure sensors that could be used for this project [6]. Force Sensing Resistor (FSR) is a type of material whose resistance changes under pressure. Several types of research have been carried out with this type of sensor and they have proven their reliability and accuracy when used in a sensor grid [7–9]. Therefore, the FSR sensor was selected for use in this method.

16 FSR sensors can be individually connected to each sensor node, while the sensor nodes are connected to each other via a system-wide bus (I^2C) with address arbitration. Due to this fact, a simple expansion of the system by connecting additional sensor nodes with sensors is possible. All sensors will automatically receive an address in the system one after another and therefore no manual adjustment is necessary.

A node is implemented as a small and simple PCB. It features an ATMEL SAM D21 microcontroller based on 32-bit ARM architecture. The advantage of this microcontroller is the large number of 12-bit resolution AD pins and the compatibility with

a lot of widely-used frameworks and tools. The firmware is based on ARM mbed framework and is written in C++. In regular intervals, the node measures voltage value on sensor pins and saves that data in a local dynamic buffer. When a request arrives via a system bus, the microcontroller processes the request and returns the latest measurements.

The "Endpoint" is a device that acts as an interface between the network of sensor nodes and external clients and services. In the context of the presented project, an Intel Edison was used as "Endpoint". Periodically, an endpoint queries the network for the latest data. Received sensors' values are being stored along their timestamps and node locations in a local database, which is used for possibility of easy access from multiple devices, connected wirelessly to the "Endpoint".

To achieve a flexible sensor mesh, automatic address arbitration is implemented. By every system start, all nodes reset their addresses and wait for a high signal on input pin. When this happens, the node takes the offered address and responds to the bus. After that, the endpoint instructs that node to rise its sense output pin high so that the next node can catch the address. This arbitration algorithm allows users to place boards in any order for their sensor mesh network.

Figure 38.1 shows the system structure. Its estimated cost is about 150 € based on costs of single components.

To conduct the study, 24 FSR sensors (FSR 406) were connected to sensor nodes (see Fig. 38.2). The positions of sensors can be changed depending on experiment aims. In the first step, experiments were conducted with three subjects in different age groups (18–25, 26–30 and 31–35). Two male and one female subjects participated in the test. Body Mass Index of test persons was 22 ± 2.5 kg/m^2. No significant health disorders were present on the test subjects. A total of approximately two hours were spent in bed, simulating sleep in different positions to collect the movement and breathing data.

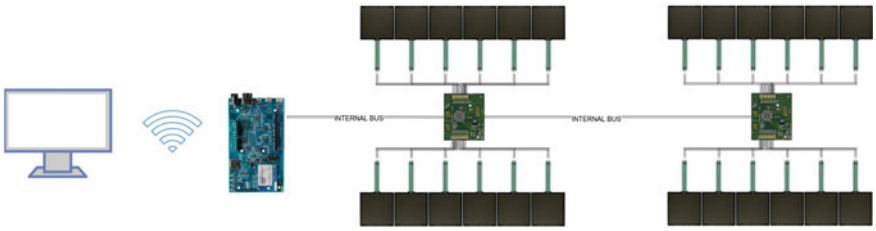


Fig. 38.1 System architecture

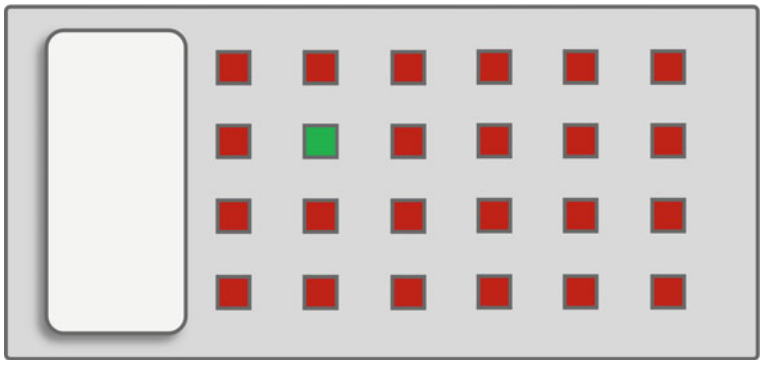


Fig. 38.2 Sensors' positions

38.3 Results

Signals from only one sensor (position is presented by green colour at Fig. 38.2) are displayed for simple and clear representation in Fig. 38.3. Some notable events and corresponding positions are also presented in this Figure.

All movements are recognizable, and also the periodic signal is easy to recognize. Figure 38.4 shows the enlarged representation of this periodic signal with blue dots as peaks from reference device Zephyr. It is necessary to mention that the frequency of the signal recording was 1 Hz due to the chosen architecture. This is the main reason why the periodic signal does not seem to be very clear and why it is not possible to detect exact peaks.

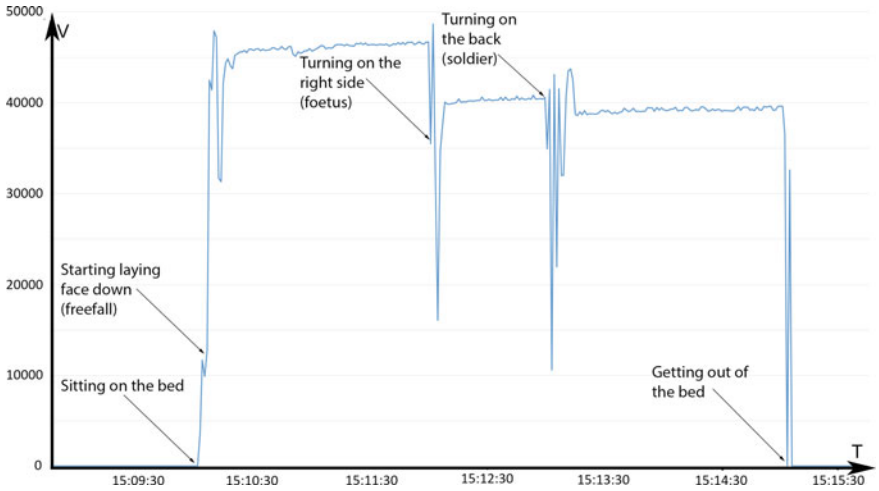


Fig. 38.3 Visualization of one sensor

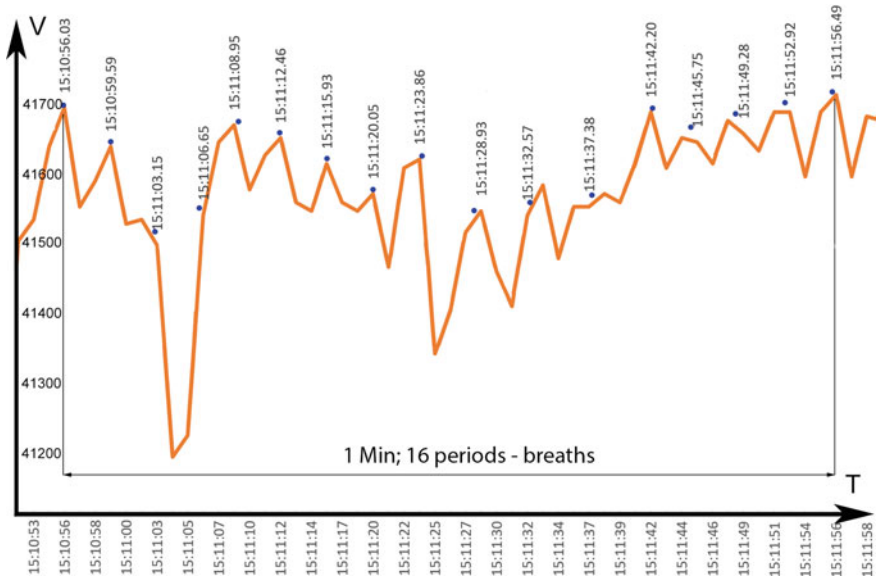


Fig. 38.4 Zoom-in of a periodical signal

When evaluating the respiratory signal, it is necessary to consider the subject's movement and accordingly reconstruct some of the peaks during the movement. A Zephyr BioHarness chest strap sensor was used as the reference device for evaluating respiratory rate detection [10]. The participants used this device during the test and the time stamps of the respiration peaks of the reference device, and the sensors had always less than 0.5 s difference. It is accurate enough for the respiratory rate detection used in a sleep study because the data processing is typically performed at 30 s intervals in sleep medicine.

38.4 Discussion

The developed system prototype provides the measurement of different vital parameters relevant for sleep stages analysis. The system architecture consists of a network of pressure sensors that are installed in a bed. Using the system does not cause inconvenience during sleep as the sensors are placed under the mattress. The sensors could detect body movements and respiratory signals. In this respect, the system seems to be suitable for sleep analysis.

One of main novel points of the proposed system is its possible application in home environment under the bed mattress for the measurements in non-obtrusive way. Another important point is the using of automatically address arbitration, which allows changing the number and positions of sensors in a very fast and easy way.

Using of FSR sensors for the measurements should also be mentioned. And the estimated price of about 150 € for the components in sale by retail, which means a much lower price in case of mass production is also an important aspect in comparison with other similar devices.

The evaluation of the system operation with the reference device has confirmed that breathing can be detected even at a frequency of 1 Hz. The next step is to perform a long-term test with night monitoring and evaluate the results. At the same time, work on increasing the system frequency has already begun. This can open the possibility to improve the results and to enable the recognition of heart rates.

The next step will be to connect the hardware system to a sleep stage classification algorithm [11], to experiment in a sleep laboratory where the results can be evaluated in collaboration with sleep medicine experts.

Acknowledgements This research was partially funded by the EU Interreg V-Program "Alpenrhein-Bodensee-Hochrhein": Project "IBH Living Lab Active and Assisted Living", grants ABH040, ABH041 and ABH066.

References

1. Kendall S (2015) National sleep foundation recommends new sleep times (Online). Available: http://www.sleephealthjournal.org
2. Pilcher J, Huffcutt A (1996) Effects of sleep deprivation on performance: a meta-analysis (Online). Available: www.watermark.silverchair.com
3. Cherry K The 4 stages of the sleep (NREM and REM sleep cycles) (Online). Available: www.verywell.com
4. Muzet A (1988) Dynamics of body movements in normal sleep. Sleep'86, pp 232–234
5. Zeidler MR et al (2015) Predictors of obstructive sleep apnea on polysomnography after a technically inadequate or normal home sleep test. J Clin Sleep 11:1313
6. Bicking RE Fundamentals of pressure sensor technology (Online). Available: www.sensorsmag.com/components/fundamentals-pressure-sensor-technology
7. Lokavee S, Suwansathit W, Tantrakul V, Kerdcharoen T (2014) Unconstrained detection of respiration rate and efficiency of sleep with pillow-based sensor array. In: 2014 11th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON), p 1–6 (Online). Available: https://doi.org/10.1109/ecticon.2014.6839779
8. Lokavee S, Watthanawisuth N, Mensing JP, Kerdcharoen T (2011) Sensor pillow system: monitoring cardio-respiratory and posture movements during sleep. In: The 4th 2011 biomedical engineering international conference, pp 71–75 (Online). Available: https://doi.org/10.1109/bmeicon.2012.6172021
9. Sundar A, Das C (2015) Low cost, high precision system for diagnosis of central sleep apnea disorder. In: 2015 international conference on industrial instrumentation and control (ICIC), pp 354–359 (Online). Available: https://doi.org/10.1109/iic.2015.7150767

10. Kim J-H, Roberge R, Powell JB, Shafer AB, Williams WJ (2013) Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the Zephyr BioHarnessTM. Int J Sport Med 34(6):497–501. http://doi.org/10.1055/s-0032–1327661
11. Gaiduk M, Penzel T, Ortega JA, Seepold R (2018) Automatic sleep stages classification using respiratory, heart rate and movement signals. Physiol Meas 39(12). https://doi.org/10.1088/1361-6579/aaf5d4

Chapter 39
A Fast Face Recognition CNN Obtained by Distillation

Luca De Bortoli, Francesco Guzzi, Stefano Marsi, Sergio Carrato and Giovanni Ramponi

Abstract Nowadays, the trend of the latest research in face recognition model shows that "the complex—the better" paradigm can be directly applied to these systems, whose accuracy effectively depends on both a large number of well-trained parameters and a complex functional structure. If this approach is sustainable for an offline processing on a consumer PC, it is far less appealing in the mobile environment, where processing power, as well as a high amount of onboard RAM could not be available. The *distillation* technique, applied on the cumbersome *dlib-resnet-v1* face recognition model results in a lighter version that, while maintaining a comparable accuracy, can achieve a faster processing rate (>10×) and a lower memory occupation (1/6). The final model has been implemented on a single board PC, also using a neural hardware accelerator.

39.1 Introduction

The face recognition problem has pushed more than any other research topic on Convolutional Neural Networks (CNNs), because the impact of human-like performances in this type of Artificial Intelligence is huge. Nowadays CNNs represent the key technology for reaching this objective and their apparent simplicity brought this kind of functionality in many mobile systems. Unfortunately, in the out-of-the lab environment high reliability is achieved only at the cost of a low processing rate, as a result of implementing a complex model; a viable option could be the use of an online cloud service, but the implications on privacy and reliability are evident. In a previous work [1], after having casted the problem as a "multi-class recognition in an open-set scenario", an open-source framework (Dlib) for face recognition has been identified and exhaustively tested. The resulting classification procedure can be carried out either with a shallow multi-layer perceptron (MLP) neural network (highest accuracy, short mandatory training phase), or with a simple distance metric

L. De Bortoli · F. Guzzi (✉) · S. Marsi · S. Carrato · G. Ramponi
DIA, Image Processing Laboratory (IPL), University of Trieste, Trieste, Italy
e-mail: francesco.guzzi@phd.units.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_39

(lower accuracy, insertion of identities in the database at runtime). All the classification algorithms we tested process the features provided by an open-source pretrained face-features extractor CNN (*dlib-resnet-v1*) [2] that, in conjunction with the subsequent classifiers, proved to be sufficiently discriminative. Besides this, while on a PC the presence of a CUDA-compatible GPU permits a reasonable processing rate of 5-10 fps, on a mobile hardware with an ARM CPU the average speed is in the order of roughly 0.5 fps (with Dlib compiled using ARM NEON [3] instructions), making a mobile use impractical. Another macroscopic problem of this pre-trained model is that it has been created within the Dlib framework. As a consequence, further modifications, fine-tuning and research, as well as a simple conversion of the model represent an unnecessarily difficult burden.

In face recognition, extracting general characteristics from the provided samples (during the supervised learning process) requires a more complex structure than the one needed for their actual representation (used during the inference). With this work we experimentally demonstrate knowledge transfer via *distillation* in a metric framework, and its actual implementation. The novelty of our contribution is threefold: (a) the entire feature vector is used, allowing theoretically for a blind swap of the oracle; (b) the entire framework is minimal, since it only requires the regression of the output target; (c) the input image width is reduced by half, permitting the recognition of small faces by design. By using the distilled model, within the entire frame processing procedure, face detection algorithms at HD resolution represent the most time consuming phase (roughly 1 s on mobile HW); frame preprocessing, face alignment and resizing are negligible in terms of computation time.

39.2 The Distillation Technique

The reduction of the time and memory complexity is a process that involves both the structure simplification and the parameter reduction; the sweet spot is given by a reduced set of parameters and a smart choice for the data processing flow that maintain the same level of accuracy as the original network. The form of compression used in this work [4–7] decorrelates the accuracy that a model achieve when performing a task, from its learned weights: what is really important to transfer (to distill) into a new model is the I/O relationship of the model itself, or the capacity to reveal the latent conditional distribution $p(\boldsymbol{T}|\boldsymbol{X})$ that relates the inputs $\boldsymbol{X}$ and the outputs $\boldsymbol{T}$. This capacity is called "*dark knowledge*" [6] and the act of transferring it from a slow but well-trained model (the teacher) to a student model is called "*knowledge distillation*" [6].

The training set for the distillation process, carried out as a supervised learning, is composed of the tuple $(\boldsymbol{X}, \boldsymbol{T})$, *input* and *corresponding target*. The distillation is carried out as a regression process, forcing the student network to provide the same descriptor generated by the teacher; in the case of an embedding network, this can be directly described in a distance metric framework, where a distance larger than the hypersphere radius of each cluster automatically flags a bad learning. This motivates

to choose as a loss metric the Euclidean Distance L_d [4] calculated between the target features vector $\boldsymbol{T}$ and the corresponding predicted descriptors vector $\boldsymbol{Y}$.

39.3 Distillation Experiments

The Dlib reference network (*dlib-resnet-v1*) is based on the *ResNet-34* [8] model which was modified by removing some layers and reducing the size of the filters by half [2]: it presents a 150 × 150 pixel RGB input, 29 convolutional layers and one fully-connected output layer with a 128D output, for a total of 5.58 M parameters.

As the base architecture for our distilled CNN, a newer architecture called "DenseNet" [9] has been selected, because it requires fewer parameters than a ResNet with the same accuracy. The core of this architecture are the so called "dense blocks", that consist in a sequence of bottlenecks and compression layers (DenseNet-BC). From the original DenseNet-121, four lighter version ("cuts") have been obtained, gradually halving both the number of dense blocks and the number of inner layers [4]. The network are denoted 'Net2.5', 'Net2.0', 'Net1.0' and 'Net0.5'. The number of parameters is equal to 3.94, 1.48, 0.38 and 0.12 M respectively. Similarly to Dlib, for all these generated networks the final layer is 128-D, while the input size has been reduced to 80 × 80 pixel (RGB). This resolution has been chosen observing that in our setup the smallest meaningful faces detected in an FHD video stream do not exceed 100 pixels at a couple of meters.

The training dataset of our distillation experiment is composed of a mixture of 250 k images taken from the Casia [10] and VGG [11] dataset: each image is pre-processed finding the face Region of Interest (ROI), aligning the detected face, and resizing it to a resolution of 80 × 80. The face detector and the face alignment procedures use the Dlib API [1]. The set of these generated images is then input to *dlib-resnet-v1* and the corresponding feature vector is saved, forming the tuple ($\boldsymbol{X}$, $\boldsymbol{T}$) that is consumed during the supervised learning of the student model. The best convergence has been reached removing the resulting average vector both from the target and the images and using Adam [12] as the optimizer. The training is relatively fast: 30 epochs on the training dataset (organized in batch of 128 images) suffice for a good convergence.

The test has been carried out on a completely different dataset, FaceScrub [13], that has been cleaned from mislabeled identities. Following the approach in [1] we designed a Multi Layer Perceptron (MLP) classifier composed of three fully-connected layers, of which the hidden layer presents 100 neurons. In order to maximise the classification reliability, for each CNN we have trained an ad hoc MLP classifier.

Each distillation experiment has been carried out within Keras [14], in order to simplify the deployment of our final model on Tensorflow-compatible hardware.

39.4 Simulation Results

The evaluation of the new CNNs has been accomplished comparing the resulting ROC curves with that of the original Dlib model. The system has to correctly identify faces of people in an ID database, without misclassifying the known identities and rejecting any other face (unknown ID). For this purpose one MLP for each model has been trained, using only samples belonging to the 'known' database; during the test phase, the dataset is composed of other images of the same ID group, plus an identical number of images of completely 'unknown' people taken at random from the remaining faces in the FaceScrub dataset.

The key for the rejection of the unknown identities is a *confidence index* (a form of normalized distance, defined in Eq. 39.1 that is used to decide on the reliability of the classifier decision: with an unknown ID a low confidence value is expected, whereas the opposite should happen for a known face.

$$C = \frac{d_1 - d_2}{d_1 - d_n} \tag{39.1}$$

where d_1, d_2 and d_n are all 'logit', i.e. the output of the latest layer of the MLP (before the SoftMax operator) respectively of the largest, the second-largest and the smallest value. The value of C is bounded between 0 and 1; by imposing a threshold for C it is possible to discriminate between known and unknown identities.

ROC curves are plotted calculating the True Positive Rate (TPR) and the False Positive Rate (FPR) as a function of the confidence index C, according to Eq. 39.2.

$$TPR = \frac{N_P}{N} \; ; \quad FPR = \frac{N_F}{N + F} \tag{39.2}$$

where N_p is the number of correctly classified samples (with C above the selected threshold of confidence) and N is the number of all known samples provided during the test; N_F is the number of misclassified samples (the number of known people that have been misclassified plus the number of the unknown people which are classified with a confidence index above the threshold, i.e. faces that have been erroneously classified as a known person) and F is the number of all the unknown samples.

During the training of the MLP, an increasing number of images (1, 2, 4, 10, 20, 40) has been used for each class (in the number N_c of 10, 20, 50). During the test, 70 samples of each known face have been used, while $N_c \times 70$ images of unknown people balance the testing-set. The entire procedure has been repeated 10 times in order to observe the average and standard deviation for each ROC curve.

Figure 39.1 shows the comparison of the four distilled CNNs with Dlib teacher, in the case of 10, 20, or 50 classes of known identities; a fixed number of 40 samples is used for the training. It is clear from the graph that the accuracy of the distilled models is very close to that of *dlib-resnet-v1*. An expected degradation is observable when the complexity of the network is reduced (less compressed networks perform better). Figure 39.1 exposes also a counterintuitive behavior, showing that for 10

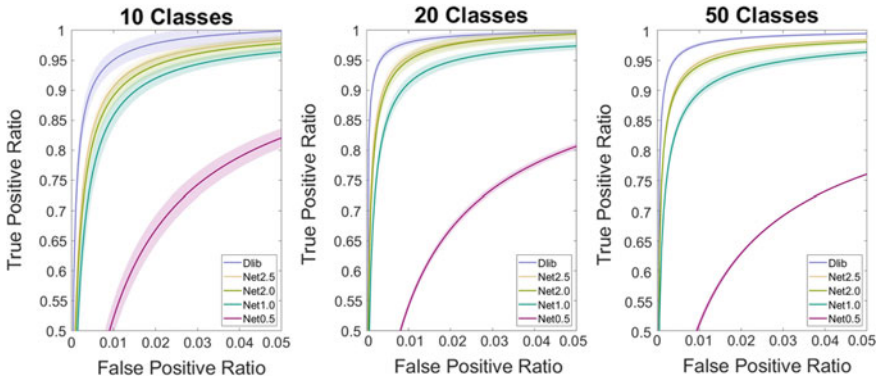


Fig. 39.1 ROC curve: comparing distilled network with the teacher Dlib network, in the case of 10, 20, 50 classes. Each graph is highly zoomed on the top left corner of the ROC space

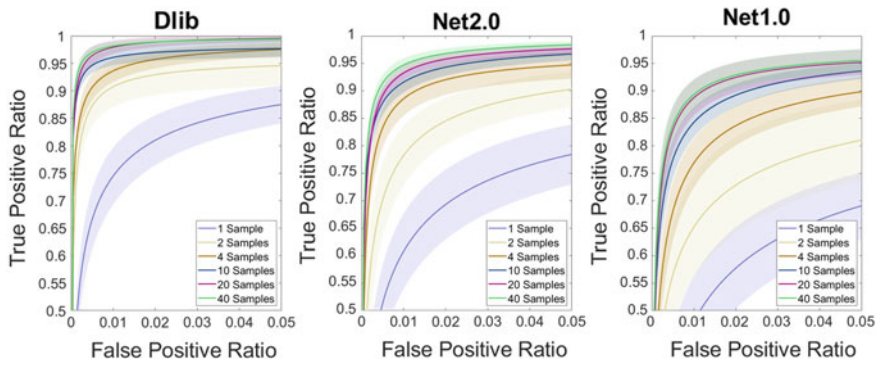


Fig. 39.2 ROC curve: performance variation observed using different training set sizes for the two best distilled network and the teacher Dlib

classes the performances are slightly lower than for 50: we suppose this is due to the combined action of how the network populates the embedding output space and of the threshold action on the evaluated confidence. Further testing is needed.

Figure 39.2 shows a comparison on the effect of the number of samples used to train the ad hoc classifier for Net2.0 and Net1.0; in this case the number of target classes is set to 50. It can be noted that, if necessary, even with less than 10 samples an MLP classifier can be trained effectively also on these new CNNs.

39.5 Implementation

The single board PC Odroid XU-4 has been selected as the reference mobile platform. On this hardware Dlib can exploit the CPU only. From the four variant of distilled

Table 39.1 Summary of average inference times

Implementation	CPU	TensorFlow Lite on CPU		Intel Movidius NCS	
CNN	Dlib	Net1.0	Net2.0	Net1.0	Net2.0
Inference time (ms)	816	63	195	50	67

network, again Net1.0 and Net2.0 have been ported to this mobile hardware; after having converted them to the TensorFlow Protocol Buffers format, two strategies of implementation have been examined: the first one uses TensorFlow Lite [15], while the second one exploits the Intel Movidius Neural Stick accelerator [16] as the target device. The latest incarnation of the Intel API for the hardware accelerator is called OpenVino [17] and allows for an easy deployment of trained models on many Intel heterogeneous devices (CPUs, GPUs, FPGAs, VPUs). The software development has been made easier than in the past because OpenCV now encapsulates the Deep Learning module of the OpenVino toolkit. While TFlite models running on CPU typically failback to the FP32 datatype, Intel Movidius support FP16 datatype only, thus making a quantization necessary. Even though the presence of this phase, the generated features remain well contained in the per-id-hypersphere. The inference time has been measured over 1000 inference cycles, also taking into account the required alignment process. Table 39.1 shows the measured time for each configuration. Using Intel Movidius, and considering 'Net2.0' as a reference model (the network with the best ROC), the required time is about 8% of the original Dlib processing time.

39.6 Conclusion

This paper presents a workflow that can be used to distill the knowledge of an expert oracle to a lighter CNN structure, that can be targeted to embedded devices. Through the teacher-student approach, the training of the new models can be reduced to a regression problem in which the convergence is reached in a relatively short time using a limited and unlabeled dataset of faces. The distilled TensorFlow model can run on an embedded CPU or with a HW accelerator. For our example facial recognition application we highlight a strategy to obtain a new CNN with inference time reduced by an order of magnitude, an accuracy comparable to the initial CNN, and a memory consumption reduced by 6 times.

Acknowledgements The support of the University of Trieste—FRA projects and of a fund in memory of Angelo Soranzo (1939–2012) is gratefully acknowledged.

References

1. Marsi S, et al (2018) A face recognition system using off-the-shelf feature extractors and an ad-hoc classifier. In: Saponara S, De Gloria A (eds) Applications in electronics pervading industry, environment and society, Lecture notes in electrical engineering, vol 550
2. Dlib API, http://blog.dlib.net/2017/02/high-quality-face-recognition-with-deep.html. Seen 6, 2019
3. ARM Neon, https://developer.arm.com/architectures/instruction-sets/simd-isas/neon. Seen 6, 2019
4. Guzzi F et al (2019) Distillation of a CNN for a high accuracy mobile face recognition system. In: Electronics and microelectronics (MIPRO) conference, Opatija, Croatia, 20–24 May 2019
5. Hinton G et al (2014) Distilling the knowledge in a neural network. In: Conference on neural information processing systems (NIPS), Montreal, Canada, 8–13 December 2014
6. Ba J et al (2014) Do deep nets really need to be deep? In: Advances in neural information processing systems 27, pp 2654–2662
7. Luo P et al (2016) Face model compression by distilling knowledge from neurons. In: Proceedings of AAAI-16, Hyatt Regency, Phoenix, Arizona (USA), 12–17 February 2016
8. He K et al (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
9. Huang G et al (2017) Densely connected convolutional networks. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 4700–4708
10. Yi D et al (2014) Learning face representation from scratch. https://arxiv.org/abs/1411.7923
11. Parkhi OM et al (2015) Deep face recognition. In: British machine vision conference, Swansea, UK, 7–10 September 2015
12. Kingma DP (2014) Adam: a method for stochastic optimization. https://arxiv.org/abs/1412.6980
13. Ng H-W et al (2014) A data-driven approach to cleaning large face datasets. In: Proceedings of IEEE international conference on image processing (ICIP), Paris, France, 27–30 Oct 2014
14. Chollet F et al. https://keras.io
15. TensorFlow Lite. https://www.tensorflow.org/lite. Seen 6 2019
16. Intel NCS. https://software.intel.com/en-us/neural-compute-stick. Seen 6, 2019
17. Intel OpenVINO Toolkit. https://software.intel.com/en-us/openvino-toolkit. Seen 6, 2019

Chapter 40
Fine-Grain Traffic Control for Smart Intersections

Jessica Bellitto, Valentina Schenone, Francesco Bellotti, Riccardo Berta and Alessandro De Gloria

Abstract As connected and, even more, autonomous vehicles are expected to bring significant novelties in the future road traffic patterns, we have investigated the control of a specific, yet very common topology, such as the intersection between two 2-lane roads. We have addressed the issue with a novel, fine-grain control approach, and proposed an adaptive prioritization algorithm which weights length of the queue and arrival order for each lane. From an Uppaal simulation, we deduce that the second factor looks more important, at higher arrival rates. Compared to a fixed Round-robin schedule, our algorithm achieves quite a better performance, especially at high traffic volumes, also with inhomogeneous traffic flow cases. In order to guarantee robustness to our design, we made a model checking analysis, considering safety and liveness requirements.

Keywords Smart intersections · Adaptive traffic light · Timed cyber-physical systems · Safety requirements · Uppaal · Intelligent transportation systems

40.1 Introduction

Intersections between two 2-lane roads is a very common topology, ever more addressed, overall in Europe, with roundabouts. But this solution requires significant territory space, and is cognitively engaging for drivers. Some other solutions based on traffic-light typically limit the turn possibilities, inevitably increasing trip times.

Considering fully connected and autonomous vehicles opens a completely new context, because of the high control and responsiveness of the vehicles. Literature has

J. Bellitto · V. Schenone · F. Bellotti · R. Berta (✉) · A. De Gloria
DITEN, Università degli Studi di Genova, Via Opera Pia 11/a, 16145 Genoa, Italy
e-mail: Riccardo.berta@unige.it; Berta@elios.unige.it

F. Bellotti
e-mail: Franz@elios.unige.it

A. De Gloria
e-mail: ADG@elios.unige.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_40

proposed a variety of solutions addressing crossroads with a huge number of lanes. But common topologies (like the one presented above) may have ad hoc simpler, yet effective solutions For instance, fine grain control could be enabled, which avoids traffic light phases, but manages single vehicles in each lane at each iteration.

This paper presents a performance analysis of an adaptive algorithm for smart intersections managed through fine-grain traffic control. Given the safety relevance of the application, we employed a requirement-based design approach [1], using Uppaal, an integrated tool environment for modeling, validation and verification of real-time systems modeled as networks of timed automata, extended with data types [2].

40.2 Related Work

Smart intersections are getting increasing focus in research on intelligent transportation systems. [3] implemented an adaptive traffic light control algorithm and simulated it in 4-way intersections with 12 lanes. The algorithm selects a set of phases where each phase is a set of concurrent lanes and picks the phase with the highest priority. The algorithm prioritizes lanes based on arrival times. Our approach is different, as we focus on a topology (4 incoming lanes with all directions available as leaving lanes), which is typically implemented through roundabouts. The number of lanes is much smaller (4 entering and 4 leaving), but every vehicle can take any direction. So, we do not consider phases. Finally, our controller schedules one vehicle per lane at each iteration. Differently form [3], which considers that all jobs need 1 unit of time to complete, we consider different intersection traversal times.

Younis and Moayeri [4] propose a novel framework dynamic intersection control relying on a sensor network to collect traffic data and includes novel protocols to handle congestion and facilitate more efficient traffic flow. Results show optimization in terms of traffic throughput, vehicle waiting time, and waiting line length. In [5], intersection control is formulated as a mixed-integer linear program (MILP) scheduling problem, and is solved by IBM CPLEX optimization package. A customized traffic microsimulation environment is developed to compare two baseline scenarios. Simulations take into account both autonomous-only and mixed traffic scenarios.

Bani Younes and Boukerche [6] presents a largest density first lane schedule, which is used to set the phases of each traffic light cycle. The work was later extended to a net of intelligent traffic lights [7], with the aim to guarantee high traffic fluency for the arterial flows in open-network scenarios. In the algorithm, simulated using NS-2, each traffic light uses the ratio between the traffic density of the competing traffic flows and the saturation density (i.e., saturation factor).

Saeed and Elhadef [8] proposed a performance evaluation of a distributed Internet of Vehicle-based intersection traffic control protocol using traffic and network simulators (Sumo, VEINS, OMNet++). Thamilselvam et al. [9] used Uppaal Stratego to study intelligent traffic light coordination (green wave).

40.3 The Model

Figure 40.1 shows the topology of the crossroad targeted by our work, with 4 incoming and 4 leaving lanes. Our model schedules at most one vehicle at a time per lane, thus implementing a fine-grain control. Intersection traversal time depends on the actual origin and destinations of the interested vehicle(s). As a baseline, we have implemented a trivial round-robin controller, with schedules the lanes according to a fixed clock-wise iteration. We have designed an adaptive algorithm which considers as priority parameters the queue length, the order of arrival and the expected traversal time, according to (1).

$$P(I) = \alpha * Q(I) + \beta * O(I) + \gamma * T(I) \tag{1}$$

where P is the priority of the lane, Q is its queue length, O is the arrival order, and T is the intersection traversal time of the first vehicle in the lane. α, β and γ are weights to be tuned.

Given the safety-related nature of the application, we modeled it using Uppaal. We defined four main templates. The Timer (Fig. 40.2a) implements the time basis,

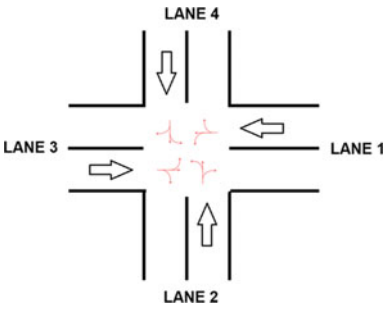


Fig. 40.1 Target intersection configuration

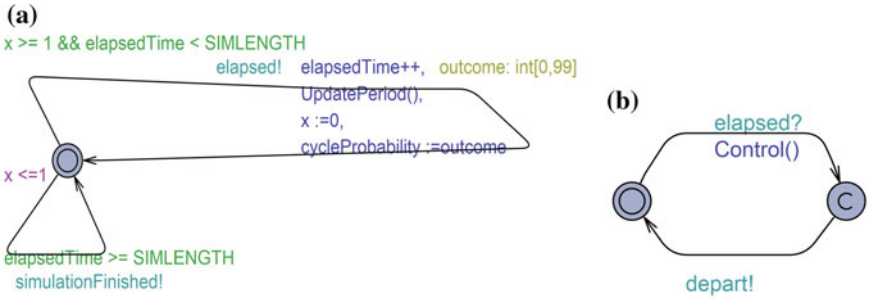


Fig. 40.2 The Timer (**a**) and Controller (**b**) Uppaal templates

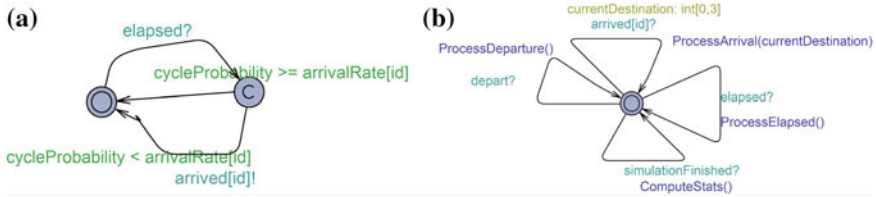


Fig. 40.3 ArrivalManager (**a**) and Lane (**b**) Uppaal templates

defines the random variable for each cycle, updates the simulated arrival rates according to the test plot and awakes the Controller (Fig. 40.2b). The ArrivalManager (Fig. 40.3a) generates the vehicles, whose arrivals and departures are managed by the Lane (Fig. 40.3b). Direction conflicts are taken in consideration by the Controllers, so that up to 4 vehicles may simultaneously pass during an iteration of vehicle passages.

40.4 Performance Analysis

The first step for the adaptive algorithm consists in tuning the α, β and γ parameters. To this end, we made a set of simulations, with 8 different values and considering different arrival rates (homogeneous), in a 10 min. time window. Results in Table 40.1 show that the best performance is achieved with $\alpha = 1$, $\beta = 1$, $\gamma = 1$. Differences are small for low arrival rates, and grow with them.

We set an experiment simulating a 100 min. with a traffic peak (0.5 vehicle/s. per lane from min. 40 to 75, 0.25 vehicle/s. per lane from min. 20 to 40, 0.1 in the other cases). Results are reported in Figs. 40.4 and 40.5, considering the two cases of homogenous and inhomogeneous arrival rates (one main road and one secondary road with halved rates). We can see that the fully adaptive algorithm achieves quite a better performance, especially at high traffic volumes (up to 40% delay reduction).

Table 40.1 Tuning of the algorithm performance (delay in seconds), with different arrival rates

Parameters			Arrival rate (per lane) [vehi/s]		
α	β	γ	0.1	0.25	0.5
0	0	1	0,9	3,6	5,5
0	1	0	1,5	2,4	6,2
0	1	1	1,4	3,3	5,4
1	0	0	1,4	3,7	5,4
1	0	1	1	3,0	5,1
1	1	0	1,1	3,4	6,0
1	1	1	0,8	2,2	4,8

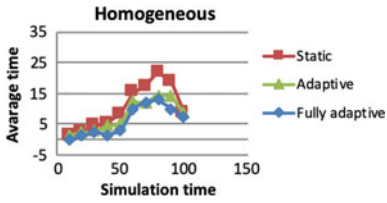


Fig. 40.4 Performance of the adaptive algorithm compared to the fixed scheduling in the homogeneous flows case

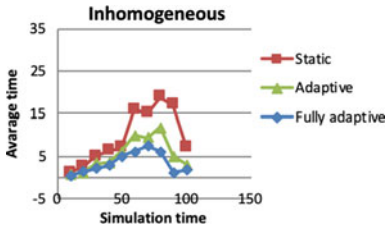


Fig. 40.5 Performance of the adaptive algorithm compared to the fixed scheduling in the not homogeneous flows case

Performance improvements are even higher in the inhomogeneous case (67%). While the adaptive case considers only the priority of a single lane, the fully adaptive case considers the priority of all the compatible concurrent lane combinations.

40.5 Model Checking

In a requirement-based system design, we defined a set of requirements to be satisfied by our model. Requirements concern safety and liveness properties. Particularly, we verified that up to four vehicles could pass at the same time and that, finally the total number of vehicles passing through the intersection is equal to the number of arrived vehicles. Because of the complexity of the model, we got an out of memory exception (with 26 GB of virtual memory) when verifying absence of deadlock and of vehicle crashes. Nevertheless, early violations of this safety constraints allowed us to spot a couple of bugs in the definition of the compatible concurrent trajectories.

40.6 Conclusions and Future Work

As connected and, even more, autonomous vehicles are expected to bring significant novelties in the future road traffic patterns, we have investigated the control of a specific, yet very common intersection topology. We have addressed the issue with a novel, fine-grain control approach, and proposed an adaptive prioritization algorithm which weights length of the queue, arrival order, and intersection traversal time of the first vehicle in each lane. From an Uppaal simulation, we deduce that the third factor looks more important. Compared to a fixed Round-robin schedule, our algorithm

achieves quite a better performance, especially at high traffic volumes (up to 40% delay reduction in the homogeneous traffic flow case, 67% in the inhomogeneous case). In order to guarantee robustness to our design, we made a successful model checking analysis, considering safety and liveness requirements. We believe that, in automated driving scenarios, this solution could reduce the need for roundabouts.

In modeling, we made some important limiting assumptions. Intersection traversal job times are fixed, even if we should have reduced the impact of this approximation, as our algorithm considers a single vehicle per lane, not platoons (e.g., as in [3]). Moreover, all vehicles are modeled as having the same length and dynamic responses. More important, we have considered a scenario with autonomous vehicles only, and completely ignored human factors and the possible presence of pedestrians or other vulnerable road users, which will need to be very carefully considered. In any case, more accurate simulations are needed to better characterize the dynamic behavior of the system and assess its improvement with respect to other optimizations.

References

1. Alur R (2015) Principles of cyber-physical systems. The MIT Press, Cambridge
2. http://www.uppaal.org/
3. Pandit K, Ghosal D, Zhang HM, Chuah C (2013) Adaptive traffic signal control with vehicular ad hoc networks. IEEE Trans Veh Technol 62(4):1459–1471
4. Younis O, Moayeri N (2017) Employing cyber-physical systems: dynamic traffic light control at road intersections. IEEE Internet Things J 4(6):2286–2296
5. Fayazi SA, Vahidi A (2018) Mixed-integer linear programming for optimal scheduling of autonomous vehicle intersection crossing. IEEE Trans Intell Veh 3(3):287–299
6. Bani Younes M, Boukerche A (2014) An intelligent traffic light scheduling algorithm through VANETs. In: 39th Annual IEEE conference on local computer networks workshops, Edmonton, AB, pp 637–642
7. Bani Younes M, Boukerche A (2016) Intelligent traffic light controlling algorithms using vehicular networks. IEEE Trans Veh Technol 65(8):5887–5899
8. Saeed I, Elhadef M (2018) Performance evaluation of an IoV-based intersection traffic control approach0 In: 2018 IEEE congress on cybermatics
9. Thamilselvam B, Kalyanasundaram S, Rao MP (2019) Coordinated intelligent traffic lights using Uppaal Stratego. In: 2019 11th International conference on communication systems and networks (COMSNETS), Bengaluru, India, pp 789–794

Chapter 41
A Graph Signal Processing Technique for Vibration Analysis with Clustered Sensor Networks

Federica Zonzini, Alberto Girolami, Davide Brunelli, Nicola Testoni, Alessandro Marzani and Luca De Marchi

Abstract The modal analysis of large structures, because of spatial and electrical constraints, generally requires cluster-based networks of sensors. In such solutions, dedicated procedures are required to reconstruct the global mode shapes of vibration starting from the local mode shapes computed on individual groups of sensors. Commonly adopted strategies are based on overlapped schemes, in which at least one sensing position is shared among neighbour clusters. In this paper, a non-overlapping monitoring approach is proposed. It relies on the intrinsic capability of graph signal processing to encode structural connectivity on edge weights and exploits the maximization of the global graph signal smoothness to define the best set of scaling factors between adjacent networks. Experiments on a pinned-pinned steel beam in condition of free vibrations proved that the proposed method is consistent with respect to numerical predictions, showing great potential for distributed monitoring of complex structures.

Keywords Graph signal processing · Cluster-based modal analysis · Mode shape assembly

F. Zonzini (✉) · N. Testoni · A. Marzani · L. De Marchi
ARCES—Advanced Research Center of Electronic Systems, 40136 Bologna, Italy
e-mail: federica.zonzini@unibo.it

N. Testoni
e-mail: nicola.testoni@unibo.it

A. Marzani
e-mail: alessandro.marzani@unibo.it

L. De Marchi
e-mail: l.demarchi@unibo.it

A. Girolami
DEI, University of Bologna, 40136 Bologna, Italy
e-mail: alberto.girolami2@unibo.it

D. Brunelli
DII, University of Trento, 38123 Trento, Italy
e-mail: davide.brunelli@unitn.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_41

41.1 Introduction

Operational Modal Analysis (OMA) is commonly applied to inspect the dynamic behaviour of structures, spanning from civil engineering to industrial applications [1]. The extraction of modal parameters, such as natural frequencies and mode shapes, is complicated in large scale monitoring scenarios, where the huge amount of data combined with the intrinsic structural complexity requires advanced and versatile solutions.

In such a context, clustered sensor networks, thanks to their capability to easy adapt to the geometric characteristics of the inspected structure, have been gradually considered as viable solutions to reduce the computational and energy budget associated to the gathering of sensor data and their transmission to a central processing unit. Nevertheless, this network architectural approach implies the development of dedicated post-processing methods to assemble the locally extracted modal information.

With reference to mode shape reconstruction, after modal coordinates have been obtained for each group of sensors, an optimal set of scaling factors between adjacent clusters must be computed. State-of-the-art solutions are based on overlapped sensor configurations, therefore at least one sampling location is shared among neighboring clusters. In [2], three covariance-driven methods were compared for modal shapes merging, showing similar satisfying performances in reconstructing vertical and lateral bending modes of bridges. Similarly, a least-squares minimization algorithm was implemented in [3] to assemble the modal coordinates of a bi-dimensional fan-shaped slab. Alternatively, a joint state space model was proposed in [4] to combine modal information from overlapping network configurations. All the above mentioned approaches suffer from some drawbacks, the most important of them concerning the increase in power consumption and computational efforts inherently related to the presence of superimposed sampling locations.

In this paper, a novel strategy based on non-overlapping clusters of sensors is proposed. Taking advantage of the Graph Signal Processing (GSP) techniques, the connections between the modal parameters extracted by different clusters are dealt with by purposely defining edge weights between adjacent sensors and then by maximizing the global graph signal smoothness. Beyond the obvious reduction in the number of sensors to be employed and the consequent energy saving, such a technique clearly encompasses some other electrical advantages. In detail, while considering large or even harsh environments, sometimes it might be difficult to install overlapped clusters due to physical or communication limitations (i.e. maximum distance between the closest devices, admitted connectivity ranges, geometrical obstacles). In addition, there are also some computational benefits associated to the minimization of data dimension while preserving the accuracy of the measurements. The implemented mode shape assembly algorithm was experimentally tested on a steel beam instrumented by means of clustered and irregularly spaced accelerometers. The results show satisfactory accuracy performances and perfect coherence with respect to the numerical predictions.

41.2 Graph-Defined Mode Shape Assembling

The analysis of signals defined on graphs has been gaining increasing attention due to its capability of modeling inherent patterns coded in the acquired data as similarities between adjacent vertices on a graph [5, 6]. Several application fields have recently benefited from this emerging signal representation, including smart cities, traffic networks and environmental processes [7]. Furthermore, a number of mathematical techniques have been developed, including the Graph Fourier Transform (GFT) and the Graph Laplacian (GL) operators, which can be used to transpose classical spectral characterization methods in equivalent tools for the vertex-frequency domain [8].

A graph is a mathematical entity described by a set of vertices connected by edges, whose Algebraic representation is expressed through the Adjacency and Degree matrices [5]. The weighted Adjacency matrix W expresses the vertex connectivity between two generic nodes n and m by means of a correspondent edge weight w_{nm}. Conversely, each entry of the Degree matrix D is given by the sum of all the weights incident on a specific vertex. The eigendecomposition of the graph Laplacian operator $L = D - W$ is an extremely useful tool to extract meaningful information from graph signals. In particular, it can be seen as the graph counterpart of the second-order derivative operator. Besides, a Fourier-like transform has been developed for graph spectral characterization, which consists of projecting graph signals on the Laplacian eigenvectors. The eigenvalues of the Laplacian matrix are also inherently related to the global graph signal smoothness of a generic function f sampled on the graph vertices:

$$\lambda = \frac{1}{2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} w_{nm} (f(n) - f(m))^2 = f^T L f \tag{41.1}$$

which quantifies the cumulative energy of signal changes sensed at different vertices [9].

41.2.1 Graph-Based Mode Shape Assembling

In vibration-based structural monitoring, spatially varying modal coordinates can be mapped as values on the vertices of an undirected arbitrary graph. Once a specific sampling grid has been deployed, edge weights can be defined as the inverse of the sampling points' spatial distance. In this context, no specific requirement about sensor density is additionally required apart from having the minimum cluster-size compliant to the number of modes to be investigated [10]. Given the quasi-sinusoidal dynamic regime typical of civil structures, which corresponds to smooth modal curves independently from the nature of the exciting force, the developed GSP technique iteratively tries to maximize the global graph signal smoothness introduced in Eq. (41.1) by correspondingly adapting a scaling factor α_c for each cluster, where subscript $c = 1, \ldots, N_c$ identifies one of the N_c subsets of sensors. The implemented

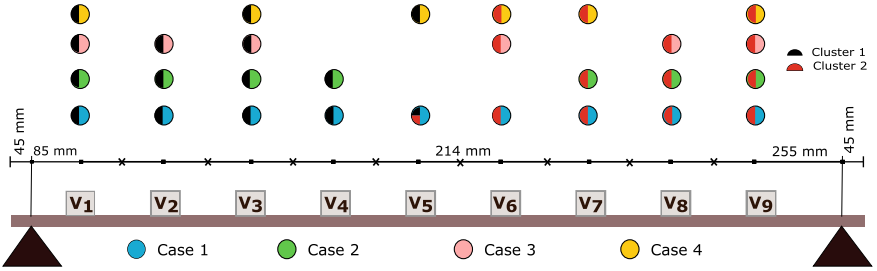


Fig. 41.1 Experimental setup with pinpointed sampling positions

algorithm comprises the following steps. During the starting phase, (*i*) a vector consisting of unitary scaling values is considered as the initial guess. Then, after the currently assembled mode shapes have been normalized (ii), the fitness function λ is computed (iii) according to (41.1). In particular, some of the graph data processing procedures from GSPBOX [11] were exploited. Finally, a prediction phase (iv) updates the scaling coefficients. More specifically, the values α_{k+1} predicted at iteration k are computed as $\alpha_{k+1} = \alpha_k - r_k \nabla f(\alpha_k)$, in which r_k and ∇f respectively represent the updating ratio and the gradient operator. Steps (ii–iv) are repeated until a convergence criterion is met, which is intended in the current approach as a smoothness variation between subsequent iterations inferior to a predefined threshold ϵ.

41.3 Experimental Validation

The effectiveness of the implemented graph-based mode shape assembly algorithm is tested on an instrumented steel beam, which was left to vibrate (free-vibration) after an initial stimulus. An extensive description of the geometric and physical properties of the structure, together with a detailed illustration of the employed electronic equipment, can be found in [12]. In particular, the circuitry consisted of low-cost tri-axial MEMS accelerometers capable of transmitting real-time data in a strictly synchronized manner by means of a CAN bus, each of them embedding an STMicroelectronics STM32L433 microcontroller unit.

Clusters of sensors were modelled on an undirected path graph of non homogeneous dimensions, the vertices of which holding modal coordinates extracted with conventional mode shape-extraction methods. As already discussed in [13], both classical Time or Frequency Domain Decomposition (TDD/FDD) methods and the unsupervised Second Order Blind Identification (SOBI) approach can be applied for this purpose. Considering that the predicted first three natural frequencies of vibration of the beam were below 50 Hz, a sampling frequency f_s 100 Hz was used; accordingly, clusters comprising at least three sensor nodes were used. Nine sampling positions were uniformly distributed along the beam length at a spatial step of 214 mm.

Four different configurations of two clustered networks were considered with various inter and intra-cluster distances between sensors. The sensor-to-cluster assignment adopted in each considered case is depicted in Fig. 41.1, from which it can be inferred that all the configurations except one (case 1) are non-overlapping. A maximum variation $\epsilon = 10^{-4}$ in successive evaluations of the fitness function was empirically estimated to be sufficient to achieve the best trade-off between the resulting modal accuracy and the convergence velocity.

To numerically quantify the level of superposition between theoretically predicted and graph-assembled modal curves, the Modal Assurance Criterion (MAC) [13] was computed, providing the modal correspondence indexes summarized in Table 41.1. Such quantities may range from 0 to 100, the latter value meaning a perfect recovery. An example of graph-combined mode shapes (ϕ_i), $i = 1, 2, 3$, is drawn in Fig. 41.2, where raw modal coordinates are extracted through the SOBI technique starting from sensing positions of case 4. Independently from the spatial distribution of the sensors and the adopted clustering scheme, making use of GSP tools a proper graph topology can be derived. As it can be observed, results yield to an almost perfect fitting between graph-assembled curves and numerical expectations, proved by a MAC value always above 95% (see Table 41.1). Additionally, it can be concluded that the sensor distribution and their relative distances seem not to affect the overall quality of the modal shape estimated for each specific mode under investigation. It is also worth noting that the performance of the proposed algorithm attains high scores with supervised (FDD and TDD) and unsupervised (SOBI) modal inspection methods. Furthermore, the number of iterations necessary to meet the convergence condition was always less than 15, thus limiting the required computational effort.

Table 41.1 MAC percentages between experimental and graph assembled mode shapes from overlapped and disjoint cluster network

	Case 1			Case 2			Case 3			Case 4		
	ϕ_1	ϕ_2	ϕ_3	ϕ_1	ϕ_2	ϕ_3	ϕ_1	ϕ_2	ϕ_3	ϕ_1	ϕ_2	ϕ_3
FDD	95.87	99.62	99.22	99.61	99.87	99.36	97.03	99.73	99.74	99.70	99.07	98.93
TDD	99.87	99.41	99.62	99.81	99.77	99.70	96.73	99.87	99.46	99.85	98.82	99.66
SOBI	95.29	99.77	99.34	99.79	99.94	99.43	97.47	99.86	99.55	99.83	99.24	99.09

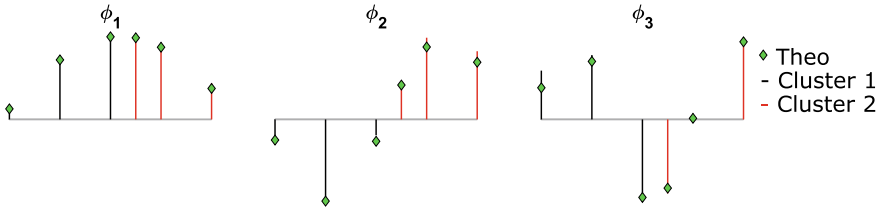


Fig. 41.2 Graph-assembled mode shapes at sensing locations chosen for case 4 exploiting SOBI modal reconstruction technique

41.4 Conclusions

This paper proposes a new approach for mode shape assembly of vibrating structures based on clustered sensor networks. Exploiting the advantages of graph signal domain to account for the underlying connectivity, the described method appears to be a powerful strategy to overcome the current limitation of state-of-the-art overlapped solutions. Different sampling grids were tested on the array of 9 sensors installed on a vibrating steel beam, assessing the robustness of the developed processing scheme in different spatial configurations. The consistency of the obtained results corroborates the possibility to deploy accelerometer sensor networks in large and complex civil structures. Future developments will address the validation of the proposed data fusion method in setups including damaged scenarios, to verify that the proposed approach does not affect the damage detection performance. Concurrently, denser sensor networks will be considered, allowing for a computational evaluation (e.g. convergence time, required processing resources) of the method under more complicated situations.

Acknowledgements This work has been partially funded by INAIL within the framework BRIC/2016 ID = 15, SMARTBENCH project.

References

1. Rainieri C, Fabbrocino G (2008) Operational modal analysis: overview and applications. Strategies for reduction of the seismic risk, pp 29–44
2. Döhler M, Reynders E, Magalhaes F, Mevel L, De Roeck G, Cunha A (2011) Pre- and post-identification merging for multi-setup OMA with covariance-driven SSI. In Dynamics of bridges, vol 5. Springer, New York, NY, pp 57–70
3. Au SK (2011) Assembling mode shapes by least squares. Mech Syst Signal Process 25(1):163–179
4. Cara FJ, Juan J, Alarcón E (2014) Estimating the modal parameters from multiple measurement setups using a joint state space model. Mech Syst Signal Process 43(1–2):171–191
5. Shuman DI, Narang SK, Frossard P, Ortega A, Vandergheynst P (2012) The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. arXiv preprint arXiv:1211.0053
6. Jabłoński I (2017) Graph signal processing in applications to sensor networks, smart grids, and smart cities. IEEE Sens J 17(23):7659–7666
7. Ortega A, Frossard P, Kovačević J, Moura JM, Vandergheynst P (2018) Graph signal processing: overview, challenges, and applications. Proc IEEE 106(5):808–828
8. Shuman DI, Ricaud B, Vandergheynst P (2016) Vertex-frequency analysis on graphs. Appl Comput Harmon Anal 40(2):260–291
9. Daković M, Stanković L, Sejdić E (2019) Local smoothness of graph signals. Math Probl Eng 2019
10. Liu X, Cao J, Lai S, Yang C, Wu H, Xu YL (2011) Energy efficient clustering for WSN-based structural health monitoring. In: 2011 proceedings IEEE INFOCOM, Apr. IEEE, pp 2768–2776
11. Perraudin N, Paratte J, Shuman D, Martin L, Kalofolias V, Vandergheynst P, Hammond DK (2014) GSPBOX: a toolbox for signal processing on graphs. arXiv preprint arXiv:1408.5781

12. Girolami A, Zonzini F, De Marchi L, Brunelli D, Benini L (2018). Modal analysis of structures with low-cost embedded systems. In: 2018 IEEE international symposium on circuits and systems (ISCAS), May. IEEE, pp 1–4
13. Testoni N, Aguzzi C, Arditi V, Zonzini F, De Marchi L, Marzani A, Cinotti TS (2018) A sensor network with embedded data processing and data-to-cloud capabilities for vibration-based real-time SHM. J Sens 2018

Chapter 42
Guided Waves Direction of Arrival Estimation Based on Calibrated Multiresolution Wavelet Analysis

Michelangelo Maria Malatesta, Nicola Testoni, Alessandro Marzani and Luca De Marchi

Abstract Damages produced by impacts can compromise structural integrity. Precise localization of the damage is fundamental to improve structural monitoring systems accuracy and reliability. A method based on the estimation of elastic waves Direction of Arrival (DoA) in plate-like structures by means of a DFT-Based Continuous Wavelet Transform (CWT) decomposition is proposed. To tackle the dispersive behaviour of guided waves, simultaneous multiband signal filtering in the Wavelet domain is performed. Subsequently, the cross-correlation method is applied to each computed scale to evaluate the Difference Time of Arrival (DToA). Finally, DoA is extracted applying an averaging procedure across scales to the arc-tangent of the ratio between the DToA among specific active areas of the sensors. This approach has been experimentally validated through measurements on an aluminum plate, after a calibration stage was performed.

Keywords Guided waves · Localization algorithm · Lamb waves · Continuous wavelet transform · PZT transducers · Calibration

42.1 Introduction

In the last few decades, Guided Waves (GWs) inspection of plate-like structures (Lamb waves) emerged as a promising Non-Destructive Evaluation (NDE) methodology. The possible applications of such kind of analysis range from aerospace, to

M. M. Malatesta (✉) · N. Testoni · A. Marzani · L. De Marchi
Adevanced Research Center on Electronic Systems (ARCES), 40136 Bologna, Italy
e-mail: michelange.malatest2@unibo.it

N. Testoni
e-mail: nicola.testoni@unibo.it

A. Marzani
e-mail: alessandro.marzani@unibo.it

L. De Marchi
e-mail: l.demarchi@unibo.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_42

marine, and civil industries, to foster the life-cycle cost reduction and the safety improvement of sensorized structures [1]. Two methods, *active-passive* and *passive-only*, are usually adopted. In the former, active piezoelectric transducers are used to generate the GWs, whereas passive piezoelectric transducers are employed for wave detection. In the latter approach, a passive sensor network, which continuously acquires data samples, is exploited [2]. Conventional passive GWs inspections usually allow for the detection and localization of damages such as crack formation or external impacts [3]. In particular, localization algorithms based on hyperbolic positioning are the most exploited methodologies due to their low computational cost and simple implementation. Unfortunately, the resolution and robustness of these approaches are limited by the dispersive behaviour of waves propagating within the structure. This inhibits a precise estimation of the Difference in Time of Arrival (DToA) of the incident wave among the active areas of the sensors. To tackle this problem, a signal processing procedure for impact localization entirely performed in the wavelet domain is proposed. By means of multiband signal filtering in the time-scale plane, the decomposition CWT coefficients of the acquired signals are extracted. Afterwards, the DToA is obtained by simply cross-correlating the coefficients. Due to the particular disposition of the PZT transducers on the structure, the Direction of arrival (DoA) is finally extracted by geometric calculations. The carried out procedure exploits the theoretical background described in [4, 5], taking into account the different piezoelectric topology of the new PZT cluster. Moreover, the enhancement of the final DoA estimation is achieved by a novel calibration method.

42.2 Sensor Disposition and Angle Definition

An innovative ad hoc PZT transducer designed with three active areas has been used in this work. The transducer is made of a cluster of three closely-located PZT elements placed to form an equilateral triangle: as such we call this disposition *equilateral*. Three signals are provided by the cluster, one for each active area S_1, S_2 and S_3. The distance between the PZTs centroids is defined as d. P is the point of impact which occurs at (x_p,y_p) and R_0 is the distance between P and the centroid of S_1. Distances D_1 and D_2 are defined as the distances the waves travel between S_1 and S_2, S_3 respectively, as shown in Fig. 42.1. If the far field approximation $R_0 \gg d$ is valid, following the derivation presented in [6], appropriately adapted to the equilateral configuration, the angle of arrival of the waveform generated by the impact can be written as:

$$\overline{\theta} \simeq \operatorname{atan}\left(\sqrt{3}\,\frac{1-{}^{D_1}/_{D_2}}{1+{}^{D_1}/_{D_2}}\right) \tag{42.1}$$

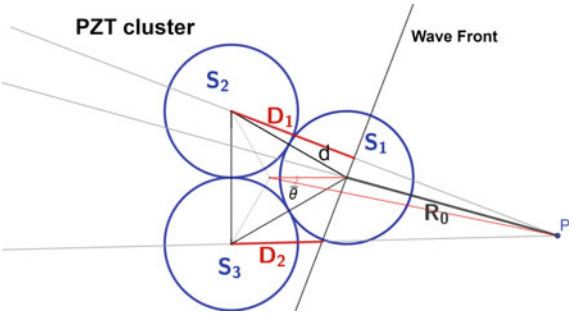


Fig. 42.1 Active elements disposition: the sensors active areas form an equilateral triangle. The GW travels from the right to the left of the figure

Defining as $\Delta t_{1,2}$ and $\Delta t_{1,3}$ the time intervals in which the wave travels the D_1 and D_2 paths respectively, i.e the so called DToA, the following equations held:

$$\Delta t_{1,2}(\nu) = \frac{D_1}{v_g(\overline{\theta}, \nu)} \qquad \Delta t_{1,3}(\nu) = \frac{D_2}{v_g(\overline{\theta}, \nu)} \tag{42.2}$$

where $v_g(\overline{\theta}, \nu)$ is the group velocity of the Lamb wave which impinges on the transducer; the frequency (ν) dependence which mathematically explains the dispersion of the considered wave mode is highlighted. Combining Eq. (42.1) with (42.2), the following result yields:

$$\overline{\theta} \simeq \text{atan}\left(\sqrt{3}\,\frac{1 - {}^{\Delta t_{1,2}}/_{\Delta t_{1,3}}}{1 + {}^{\Delta t_{1,2}}/_{\Delta t_{1,3}}}\right) \tag{42.3}$$

which, properly rotated of 30° for a more convenient coordinate reference system, becomes:

$$\overline{\theta} \simeq \text{atan}\left(\frac{1}{\sqrt{3}} - \frac{2}{\sqrt{3}}\frac{\Delta t_{1,2}}{\Delta t_{1,3}}\right) \tag{42.4}$$

It is also noteworthy that Eq. (42.3) can be easily reformulated for the specific case presented in [6]: in this case a rotation of the coordinate system reference of 45° and an angle θ_x between active areas of 90° must be considered.

42.3 DToA Estimation in CWT Domain

The DoA estimation is strictly related to the DToA evaluation, according to Eq. (42.4). Thus, a precise calculation of the DToA is fundamental in order to achieve the final goal. Because of the dispersive and multimodal behaviour of waves in plates, $\Delta t_{j,k}(\nu)$ depends on group velocity. Thus, the application of the *cross-correlation* method to the raw acquired signals is not strictly possible. The technique hereby proposed exploits a time-frequency decomposition in the CWT domain followed by

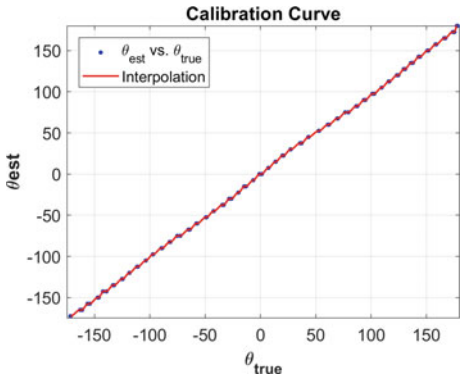


Fig. 42.2 Calibration curve

an isofrequential analysis to localize the signal both in the time and in the frequency domain. Let S_i and S_j be the ith and jth active area of the transducer, and $s_i(t)$ and $s_j(t)$ the acquired signals by S_i and S_j, respectively. The CWT coefficients are defined as:

$$W_i(\psi; a, b) = \int_{-\infty}^{+\infty} s_i(t)\Psi^*_{a,b}(t)\, dt \tag{42.5}$$

$$W_j(\psi; a, b) = \int_{-\infty}^{+\infty} s_j(t)\Psi^*_{a,b}(t)\, dt \tag{42.6}$$

where $\Psi_{a,b}(t) = |a|^{-\frac{1}{2}}\Psi\frac{t-b}{a}$ is the decomposition wavelet and $a \in \mathbb{R}$ is the scale parameter. In this domain, the wave components travelling at different velocities can be easily separated and filtered by means of the energy distribution analysis of the signal in the scale-frequency plane. Cross-correlation is then computed in the CWT domain for each scale (or the corresponding frequency range), i.e. by varying the scale parameter a, as shown in the following equation.

$$C_{i,j}(a, t) = (W_i * W_j)(a, t) = \int_{-\infty}^{+\infty} W_i^*(\Psi, a, b)W_j(\Psi, a, t + b)\, db \tag{42.7}$$

As a consequence, the DToA $\Delta t_{i,j}$ can be defined as

$$\Delta t_{i,j}(a) = \arg\max_t(C_{i,j}(a, t)) \tag{42.8}$$

Then, by applying Eq. (42.8) for each a parameter value, the DoA of the guided mode can be accurately estimated by means of an averaging procedure [4, 5].

42.4 Experimental Results

As a case of study, we tested the proposed algorithm to locate impacts in an aluminium 1050 A square plate 1000 mm × 1000 mm and 3 mm thick. The PZT cluster, made by three closely-located active areas with a diameter of 10 mm each, was attached at the centre of the plate. GWs were generated by hitting the plate with a metallic screw along a circumference of radius $R_0 \simeq 20\,\text{cm} \gg d$ centred at the transducer position in order to satisfy the far field approximation, as described in Sect. 42.2. The circle was divided into 48 angular intervals, each one 7.5° wide. The three piezoelectric signals were acquired by means of a Tektronix 3014 oscilloscope at a sampling frequency of 10 MHz, with the reference trigger on S_1 piezo signal. At first, a calibration procedure was implemented to achieve more precise and reliable results. Two cycles of impacts were generated on each angle of the quantized circumference, for a total of 96 training impacts. Therefore, the DoA of the impacts were estimated for each of them. Without any filtering or pre-processing, the maximum absolute error provided in this phase by the algorithm was around 5° with an average error of about 2°, revealing a remarkable accuracy along the entire angular range. The estimated angles were subsequently associated to the real DoA by a Cubic Spline Interpolation (CSI) in order to determine the calibration curve. An exhaustive mathematical treatment of the CSI is presented in [7]. The calibration curve obtained is depicted in Fig. 42.2. Afterwards, another dataset of 21 impacts was acquired. The impacts were produced randomly along the quantized angles of the circumference in order to test the DoA estimation algorithm and the calibration procedure. The modalities, the instrumentation and the setup exploited is identical to the previous experiment. The DoA and its relative error were computed for each impact. In particular, the average error measured was around 1.4°, with a maximum error of 4.8°. The comparable values in relation to the errors on the DoAs estimated during the calibration procedure reveal the consistency and the repeatability of the measurements. Furthermore, applying the calibration curve to the DoA angles estimated, the error decreases. In particular, the average error from 1.4° drops to 0.0517°, and the maximum error decreases from 4.8° to 0.8°. Figure 42.3 shows in detail the error trend. As expected, the proposed localization algorithm, in conjunction with an appropriate calibration procedure, is able to provide a reliable and very accurate estimation of the angle of arrival, with a resolution of less than one degree. It is also worthy to notice that the DoA is estimated in just 0.13 s, due to the DFT-based Wavelet approach. In fact, the CWT formulation as the inverse Fourier transform of a product of Fourier transforms is exploited, enabling the user to use the computationally-efficient fft and ifft algorithms to reduce the cost of computing convolutions.

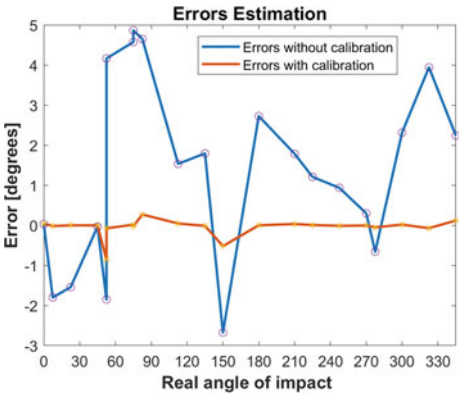


Fig. 42.3 Errors of the DoA estimation before and after the calibration procedure

42.5 Conclusion

In this work, a novel method to extract the DoA of Lamb waves generated by impacts in plate-like structures is proposed. The algorithm exploits a procedure carried out in the CWT domain in order to completely localize the signals in the time-scale domain. By cross-correlating the signals related to the same event, the DToA can be determined and used to locate the wave source according to the piezoelectric transducer configuration. Accurate and reliable results are shown through experimental tests, further enhanced by an appropriate calibration procedure. The errors achieved by the system are significantly less than 1°, revealing that the discussed method is especially suitable when high precision SHM localization is required.

Acknowledgements This research has been funded by INAIL within the framework BRIC/2016 ID = 15, SMARTBENCH project.

References

1. Ebrahimkhanlou A, Dubuc B, Salamone S (2016) Damage localization in metallic plate structures using edge-reflected lamb waves. Smart Mater Struct 25(8):085035
2. De Marchi L, Marzani A, Speciale N, Viola E (2011) A passive monitoring technique based on dispersion compensation to locate impacts in plate-like structures. Smart Mater Struct 20(3):035021
3. Kundu T, Das S, Jata K (2009) Detection of the point of impact on a stiffened plate by the acoustic emission technique. Smart Mater Struct 18(3):035006
4. Garofalo A, Testoni N, Marzani A, De Marchi L (2016) Wavelet-based Lamb waves direction of arrival estimation in passive monitoring techniques. In: IEEE international ultrasonics symposium (IUS). IEEE Press, Tours (FR), pp 1–4
5. Garofalo A, Testoni N, Marzani A, De Marchi L (2017) Multiresolution wavelet analysis to estimate Lamb waves direction of arrival in passive monitoring techniques. In: IEEE workshop on environmental, energy, and structural monitoring systems (EESMS). IEEE Press, Milan (IT), pp 1–6

6. Kundu T, Nakatani H, Takeda N (2012) Acoustic source localization in anisotropic plates. Ultrasonics 52(6):740–746
7. Marsdem M (1974) Cubic spline interpolation of continuous functions. J Approx Theory 10:103–111

Chapter 43
High-Frame-Rate Ultrasound Color Flow Imaging Based on an Open Scanner

Francesco Guidi, Enrico Boni, Alessandro Dallai, Valentino Meacci and Piero Tortoli

Abstract Standard scanned Color Flow Imaging (CFI) is a common blood flow visualization modality. Despite being introduced more than 30 years ago, this technique is still hampered by the conflicting requirements for either a good image quality or a high frame rate. In fact, good image qualities can only be obtained for frame rates between 10 and 20 Hz, which are unsuitable to show dynamically evolving events. This paper presents a high frame rate imaging modality that, once integrated with CFI, allows to overcome the above limitation. Results characterized by improved quality matched to the capability of properly tracking dynamically evolving flow rates are shown.

43.1 Introduction

Color flow imaging (CFI) is an ultrasound (US) technique [1] capable of producing B-Mode images in which the areas interested by blood flow are colored according to the instantaneous velocity and direction of red blood cells. If the flow is directed toward or away from the probe, red or blue colors are correspondingly used.

Coloring of US images is done line-by-line [2]. A beam focused on the current line is transmitted NP times (with NP, "packet size", typically between 6 and 16) from the active elements of a linear array probe. During the reception interval, the echoes received by each probe element are "beamformed" (i.e. properly delayed before being summed together). The mean Doppler frequency of the NP echo-samples beamformed for each depth is estimated through the autocorrelation approach, and the corresponding pixel is accordingly colored. Of course, the higher NP the more robust the frequency estimate, but also the lower the CFI frame rate (FR). For example, the time needed to form one CFI image of NL = 64 lines at pulse repetition frequency (PRF) of 10 kHz and NP = 10 is $64 \times 100 \times 10\ \mu s$, which yields ≈ 15 frames per

F. Guidi (✉) · E. Boni · A. Dallai · V. Meacci · P. Tortoli
Microsystems Design (MSD) Laboratory, Dipartimento di Ingegneria dell'Informazione, Università Degli Studi di Firenze, Firenze, Italy
e-mail: francesco.guidi@unifi.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_43

second (fps), neglecting the time needed to generate the B-mode background. Such a FR may not be sufficient to track rapid blood flow changes in the heart or in main human arteries such as the carotid.

The transmission of plane waves [3] may solve such problem, since the full region of interest (ROI) is insonified with a single transmission and the FR can thus be automatically increased by a factor NL [4]. In addition, such increase is so high that a larger packet size (up to 64 and more) becomes acceptable, with a corresponding image quality improvement. Such results are achieved provided the echoes received by the probe elements can be simultaneously beamformed along all the lines of interest. The term "parallel beamforming" indicates the capability of creating multiple image lines after a single transmission (TX) event. Of course, the larger the NL the higher the needed parallel beamforming speed.

Furthermore, the NL beamformed lines must be, at the same time, processed according to the color Doppler strategy. Such processing includes, besides the autocorrelation, high-pass filtering for clutter removal, the evaluation of proper criteria to avoid that residual small tissue movements are detected as blood movements, as well as temporal and spatial filtering of the final colored images. Performing all such processing in real time for multiple lines at high frame rates is quite challenging.

Parallel beamforming for high FR CFI is currently used in only one clinical system [5] and in commercial open scanners [6]. However, in both cases, the acquired raw data are beamformed and processed off-line. The goal of this paper is presenting a real-time high FR CFI system obtained by suitably using the programmable resources available on the ULA-OP 256 open scanner [7, 8]. It is shown that parallel beamforming is achieved by a special organization of the front-end FPGAs, while color Doppler processing is performed by one on-board multi-core DSP. With respect to the work presented in [9] better exploitation of ULA-OP 256 hardware permitted to obtain increased performance in terms of PRF, packet size range and global image quality.

43.2 Real-Time CFI Implementation

ULA-OP 256 is an open scanner developed by the MSD Laboratory of the University of Florence [7]. This research instrument has been designed with a modular approach, according to which eight front-end (FE) boards manage up to 256 channels connected to an equal number of transducer elements. The active FE boards are interconnected in a ring by a Serial RapidIO (SRIO) link with a bandwidth of 80 Gbit/s—full duplex.

Every FE board transmits, receives and elaborates 32 ultrasound signals. Each channel is equipped with an independent arbitrary waveform generator, capable of producing high voltage (up to 200 Vpp) signals at up to 20 MHz frequency. On the receiving path, 4 Analog Front Ends (AFEs) embed 32 low noise amplifiers followed by 32 analog to digital converters working at about 80 MSPS. These feed an FPGA which is in charge of beamforming (see Sect. 43.2.2). Beamformed data

are demodulated and low-pass filtered by an on-board DSP and finally sent to the Master Control (MC) board, which is here in charge of performing the high frame rate CFI algorithm.

43.2.1 Plane Wave Transmission

The 128 central elements of the LA533 linear array probe (Esaote SpA, Italy) were simultaneously excited at programmable PRF to alternate the transmission of NP CFI pulses and of 7 B-mode pulses. CFI pulses were 5-cycle Hanning tapered sinusoidal bursts at 6 MHz center frequency. B-mode pulses, were 3-cycle sinusoidal pulses at 9 MHz. The 7 B-Mode PWs were transmitted at steering angles of −7.5°, −5°, −2.5°, 0°, 2.5°, 5°, 7.5° before being coherently compounded [10]. The ROI was thus fully insonified during each Pulse Repetition Interval (PRI = 1/PRF).

43.2.2 Serial-Parallel Beamforming

After each TX event, the echo data were beamformed over a programmable depth range of the ROI, thanks to the serial-parallel beamformer (BF) implemented in the FE FPGA [11, 12]. This includes a Dual Port Memory (DPM) capable of storing up to 16384 words, each of 384-bit (12-bit per sample for 32 channels), and 4 parallel BFs, see Fig. 43.1. The serial-parallel BF implemented on FPGAs of the ARRIA V GX family (Altera, San Jose, CA, USA) are embedded on the FE boards.

The DPM stores the echoes digitized by 4 AFEs. The DPM is divided in two buffers (8192 words each); while one buffer is storing the echo-data of the current PRI, the second one is read to permit processing data acquired in the previous PRI. The BFs process multiple times the same data to focus the received echoes along corresponding multiple directions. Each BF works at 200 MHz and is composed of delay blocks, one per channel, which align the received data before they are coherently summed. A Memory controller embedded in the FPGAs manages an external memory in which the delays are stored during the system initialization. The serial-parallel beamformer implemented on FPGAs of the ARRIA V GX family (Altera, San Jose, CA, USA) embedded on FE boards can produce up to about 470 MSPS.

It exploits about 31% of the adaptive logic modules, and 24% of the logic registers present in the FE FPGA. The memory is allocated in the 10 kbit memory blocks (M10K), which are used at 71%.

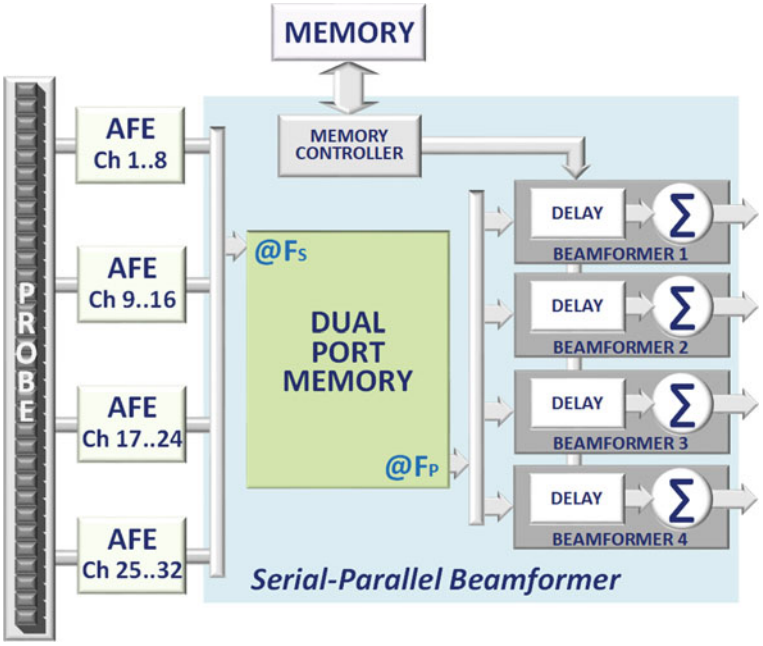


Fig. 43.1 Serial-parallel beamformer implemented on the research ultrasound scanner ULAOP-256

43.2.3 CFI Processing

The data produced by the serial-parallel BF are processed by the 8-cores DSP TMS320C6678 (Texas Instruments, TX, USA) mounted on the MC board. The CFI is here implemented using 1 core to manage the beamforming operation, 1 core as the master processing unit and the remaining 6 cores as the slave processing units.

The beamforming core is in charge of performing the last beamforming stage of the system. It sums together the samples from all the FE boards to produce a single data stream and stores it in external DDR memories, which act as temporary circular buffers.

The master core mainly acts as a supervisor and schedules tasks to the other cores. Every time a block of fresh samples is ready from the beamforming core, the Master core initiates a transfer of data from the DDR memories to the slave cores that are ready to process new tasks, and instructs them with appropriate processing parameters of the CFI algorithm. Once a slave core terminates its task, it notifies the master core that a new column of processed pixels is ready. The master core initially transfers the pixels to its internal memory, freeing resources in the slave core, which is accounted for subsequent scheduling. It then processes the pixels of multiple columns through a 3×3 median filter, and sends the result to the PC, where the CFI frames are displayed on a screen superimposing the B-Mode layer. All data transfers from and to any core are operated by DMA channels (Fig. 43.2).

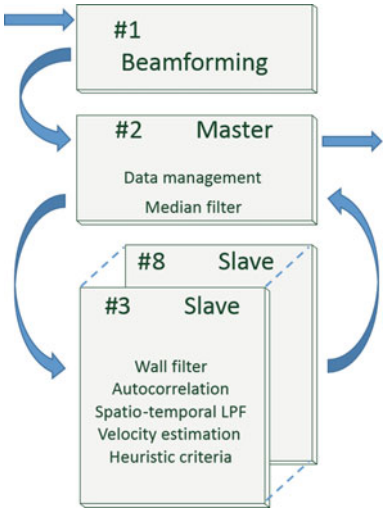


Fig. 43.2 DSP cores load distribution

Each slave core takes care of processing a vertical line of points. It receives a block of complex demodulated samples collected over $8 \div 64$ PRIs, depending on NP. The core calculates the incoming signal power and processes the samples through the wall filter, composed by a 4th order IIR high-pass filter. It afterwards calculates the autocorrelation and variance of the signal. The autocorrelation output is fed into a spatial filter made up of a 3×3 matrix that combines adjacent depths and lines to obtain smoothened images. Since different lines are processed in distinct cores, a particular state machine enables each core to directly collect, from other slave cores, the samples processed by the latter ones up to this stage. The line processed in the current core is thus combined with its two adjacent lines, and injected, first, into the spatial filter and then into the persistence filter, which is an IIR filter working across slow time. Finally, the slave core calculates the power and phase of the filtered autocorrelation, which are proportional to the intensity and the flow velocity respectively.

The computed phase is directly used to form a color map, while the power is non-linearly combined with other available parameters, to generate a pixel transparency mask, used to combine the final CFM and B-mode maps.

43.3 Results and Discussion

The CFI approach has been tested with a standard flow-phantom (ATS 524) connected to a peristaltic pump and a reservoir, in a closed loop containing a blood mimicking fluid (ATS 707).

The first experiment was conducted in stationary flow conditions. No temporal filters were used to maintain maximum temporal resolution. Figure 43.3 shows

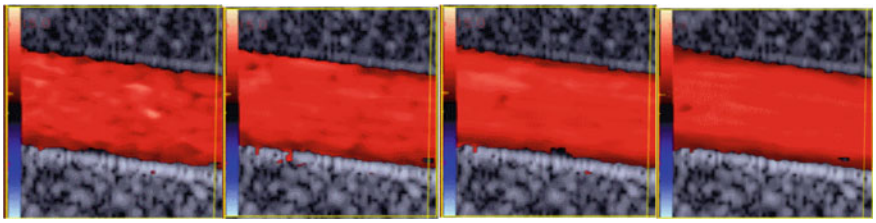


Fig. 43.3 Screenshots of ULAOP SW showing the output of real-time CFM for a steady flow and different packet lengths. From left to right: NP = 8, 16, 32, 64. The system PRF was 3 kHz, producing an output frame rate = 200, 130, 77, 42 [frames/s] respectively

samples of instantaneous images obtained with different packet sizes, maintaining PRF = 3 kHz in all cases. Good results are visible for all situations: the highest frame-rate (200 frames/s) was obtained with NP = 8 and the best quality for NP $\geq$ 32, when the frame rate is still high, corresponding to 77 (NP = 32) and 42 (NP = 64) frames/s respectively.

In the second experiment, a dynamic flow was used. The dynamic behavior was monitored by alternating a standard CFI method (based on the transmission of NP focused beams for each image line) and the new CFI (based on PW TX). In this way, it was possible to compare the results provided in the two conditions.

In Fig. 43.4, four snapshots obtained by the conventional CFI method are presented. The FR was 14 frames/s, which is a standard value in these cases. The two figures on the right highlight bizarre conditions, in which velocities in opposite directions are apparently present in two sections of the tube having the same distance from the walls.

Figure 43.5 shows four snapshots obtained with the CFI PW method for the same flow conditions. Here NP was set to 16 and the PRF was 2.5 kHz. The final output frame rate, including both CFI and B-mode, was 108 frames/s. All the lines in each frame are intrinsically coherent because produced by the serial-parallel beamforming of the same echoes. The two frames on the right still show different velocities, but now the difference appears between the center and the vessel walls, consistent with the expected dynamics of a pulsed flow in a partially elastic flow-phantom.

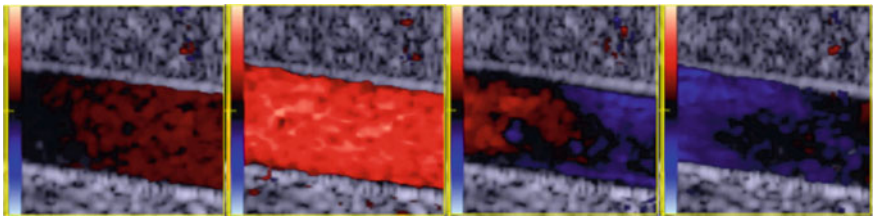


Fig. 43.4 Screenshots showing 4 frames of the real-time conventional scanned CFI, 2 before and 2 after the velocity peak. These were obtained with NP = 8

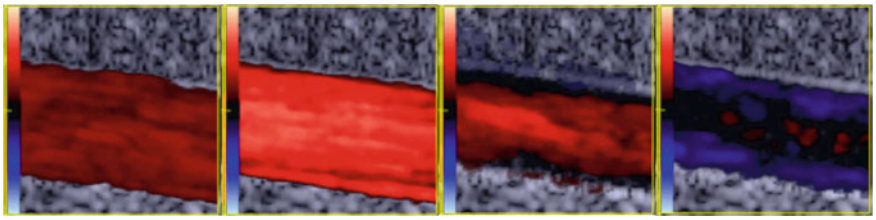


Fig. 43.5 Screenshots showing 4 frames of the real-time CFM PW, 2 before and 2 after the velocity peak. NP = 16, PRF = 2.5 kHz, no temporal filter was introduced to maintain high temporal resolution

Table 43.1 Frame rates achieved by alternating CFI and B-Mode (center) and by using only the CFI Mode (right), evaluated for different packet sizes

NP	B-mode + CFI (fps)	CFI software (fps)
8	267	403
16	174	222
32	103	116
64	56	57

The system was finally tested at PRF = 4 kHz to evaluate the FRs obtainable at different packet sizes. The results are reported in the Table, which shows the FR values obtained by mixing CFI and B-Mode (column 2) and by only operating in PW CFI Mode (column 3). As reported in the right column, up to 403 frames/s could be processed (NP = 8) (Table 43.1).

43.4 Conclusion

In this paper, a CFI method based on the transmission of PWs has been presented. The experimental results highlight that, thanks to the increased frames availability, it is possible to use longer packet-sizes to perform more reliable speed estimation, maintaining a final frame rate always very high. Moreover, thanks to the intrinsic frame coherence due to the PW imaging scheme, the output frames do not show artifacts during fast events as seen in standard CFM methods.

References

1. Kasai C, Namekawa K (1985) Real-time two-dimensional blood flow imaging using an autocorrelation technique. IEEE Ultrason Symp 1985:953–958
2. Evans DH, Jensen JA, Nielsen MB (2011) Ultrasonic colour doppler imaging. Interface Focus 1(4):490–502

3. Tanter M, Fink M (2014) Ultrafast imaging in biomedical ultrasound. IEEE Trans Ultrason Ferroelectr Freq Control 61(1):102–119
4. Bercoff J et al (2011) Ultrafast compound doppler imaging: providing full blood flow characterization. IEEE Trans Ultrason Ferroelectr Freq Control 58(1):134–147
5. Supersonic Imagine, "UltraFast^{TM} Imaging." Online www.supersonicimagine.com/Aixplorer-R. 06 Dec 2016
6. Ekroll IK, Swillens A, Segers P, Dahl T, Torp H, Lovstakken L (2013) Simultaneous quantification of flow and tissue velocities based on multi-angle plane wave imaging. IEEE Trans Ultrason Ferroelectr Freq Control 60(4):727–738
7. Boni E et al (2016) ULA-OP 256: A 256-channel open scanner for development and real-time implementation of new ultrasound methods. IEEE Trans Ultrason Ferroelectr Freq Control 63:1488–1495
8. Boni E, Yu ACH, Freear S, Jensen JA, Tortoli P (2018) Ultrasound open platforms for next-generation imaging technique development. IEEE Trans Ultrason Ferroelectr Freq Control 65(7):1078–1092
9. Guidi F, Dallai A, Boni E, Ramalli A, Tortoli P (2016) Implementation of color-flow plane-wave imaging in real-time. IEEE Int Ultrason Symp (IUS) 2016:1–4
10. Berson M, Roncin A, Pourcelot L (1981) Compound scanning with an electrically steered beam. Ultrason Imaging 3(3):303–308
11. Boni E et al (2017) Architecture of an ultrasound system for continuous real-time high frame rate imaging. IEEE Trans Ultrason Ferroelectr Freq Control 64:1276–1284
12. Meacci V et al (2019) FPGA-based multi cycle parallel architecture for real-time processing in ultrasound applications. In: International conference on applications in electronics pervading industry, environment and society, pp 295–301

Part IX
Vehicular, Robotic and Energy Electronic Systems

Chapter 44
Empowering Deafblind Communication Capabilities by Means of AI-Based Body Parts Tracking and Remotely Controlled Robotic Arm for Sign Language Speakers

Silvia Panicacci, Gianluca Giuffrida, Luca Baldanzi, Luca Massari, Giuseppe Terruso, Martina Zalteri, Mariangela Filosa, Giovanni Tonietti, Calogero Maria Oddo and Luca Fanucci

Abstract Deafblind people face remarkable challenges in communicating, because of their severe disability. The only way to interact with other people is the usage of the tactile sign language, which consists in understanding the sign language putting their hands on the signer's hands. But this approach works only when the signers are in the same place. The aim of this project is to reduce the gap between deafblind people and the other ones, giving them the capability to communicate remotely. By collecting

S. Panicacci · G. Giuffrida (✉) · L. Baldanzi · L. Fanucci
Department of Information Engineering, University of Pisa, Via G. Caruso 16, 56122 Pisa, Italy
e-mail: gianluca.giuffrida@phd.unipi.it

S. Panicacci
e-mail: silvia.panicacci@phd.unipi.it

L. Baldanzi
e-mail: luca.baldanzi@ing.unipi.it

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

L. Massari · G. Terruso · M. Zalteri · M. Filosa · G. Tonietti · C. M. Oddo
Sant'Anna School of Advanced Studies, The BioRobotics Institute, Via R. Piaggio 34, 56025 Pontedera, Italy
e-mail: luca.massari@santannapisa.it

G. Terruso
e-mail: giuseppe.terruso@santannapisa.it

M. Filosa
e-mail: mariangela.filosa@santannapisa.it

G. Tonietti
e-mail: giovanni.tonietti@santannapisa.it

C. M. Oddo
e-mail: calogero.oddo@santannapisa.it

L. Fanucci
CINI, National Interuniversity Consortium for Informatics, Via Ariosto 25, 00185 Rome, Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_44

images with two cameras, the signer's body is tracked with a deep neural network. The extracted coordinates of the body parts (chest, shoulders, elbows, wrists, palms and fingers) are used to move one or more robotic arms. The deafblind person can put his hands on the robots to understand the message delivered by the person on the other side. The entire system is based on a cloud architecture.

44.1 Introduction

One of the main characteristics of human beings is the ability and the will to communicate with other people, thanks to a shared language. Typically, the language is a spoken language, with acoustic-vocal modalities. Unfortunately, deaf people cannot hear any sounds. For this reason, they use another kind of language, a gestural-visive one, the Sign Language (SL). Deaf people can then communicate making gestures, moving the hands and doing facial expressions. Since they cannot hear words or see gestures, deafblind people cannot use either voice or sign language. Their language is then tactile, the so-called Tactile Sign Language (TSL). It is based on the SL, but the receiver's hands are placed on the ones of the signer, in order to follow the signs made. It works fine when the deafblind is in the same place of the signer, but deafblind people cannot communicate remotely, like other people do with phone or video calls, because they need to touch the other person's hands. This limitation due to their severe disability can lead to isolation and depression [1]. Nowadays, from 0.2 to 3.3% of the world population is deafblind [2].

The idea of the PARLOMA project is to reduce the gap between deafblind people and the others, in order to decrease the cases of isolation and depression, giving them the possibility to communicate remotely. This is done tracking the movements of a person in front of two cameras and reproducing these gestures on a robotic arm, which can be touched by the deafblind to understand the message delivered remotely.

44.2 System Architecture

In order to let the deafblind people communicate in a remote way without external tools, there is the need of two cameras. The first one is used to capture the frames for the identification of the pose of the upper part of the body (chest, shoulders, elbows and wrists) and the second camera is used to detect the position of the hands, and of a robotic arm, which reproduces the gestures done in front of the cameras. Because of the complexity to reach this aim, the whole system is based on a client-server cloud architecture, with the possibility of the connection of more clients on the camera side and more clients on the robotic arm side (multi-client system). In the simplest configuration, the system is composed by three entities: a client which manages the two cameras (Camera Client, CC); the server, which receives the images from the CC and computes the complete pose of the person (Elaboration Unit, EU);

a client which has the task of moving the robotic arm (Robot Client, RC). When more CCs and more RCs are connected to the elaboration unit, each CC can choose with whom to communicate, selecting one RC, like in a phone call, or more ones, simulating a radio scenario. The multi-client scenario is managed by using a simple database, composed by three tables: the CCs table and the RCs table, which contain the same fields (name, IP, port and a boolean indicating if it is online or not), and the communication table, which is used to store the relationships between the CCs and the RCs. Because of their tasks, both the CCs and the RCs can be common laptops, while the EU must have at least one GPU.

44.2.1 Camera Client

The Camera Client is the part of the system that collects the frames from the two cameras and sends them to the EU. The cameras are placed like in Fig. 44.1. The camera in front of the signer is an Orbbec Astra Pro [3]. It is used to collect the frames that will be used by the EU for the identification of the coordinates of the joints of the upper part of the body. Each frame is composed by two images with 640 × 480 @ 30 fps resolution, an RGB image and a depth image. Both they are needed to have a 3D vision and to let the robotic arm move correctly in the space, after all the elaborations. Since the depth image is noisy, some filters are applied. First, the depth image is dilated, with an elliptic kernel, increasing the size of the objects and deleting some shadows [4]. After that, the image is convolved with a low-pass filter (gaussian blur), in order to reduce the noise and smooth the image itself [5]. The two images are sent through Internet, so it is needed to reduce their dimensions: the RGB image is compressed in jpg, the filtered depth image is compressed in gz. The

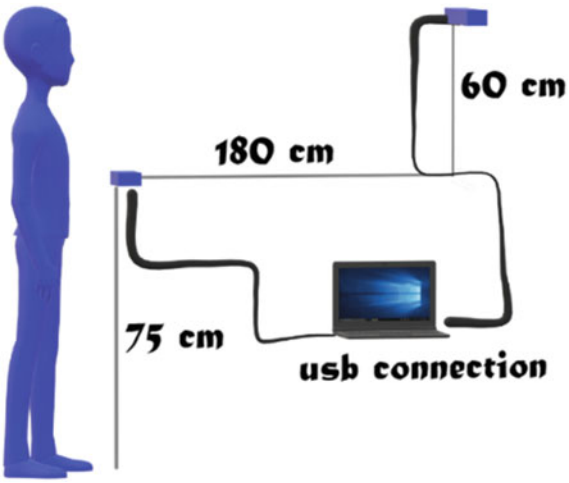


Fig. 44.1 Camera client pose configuration

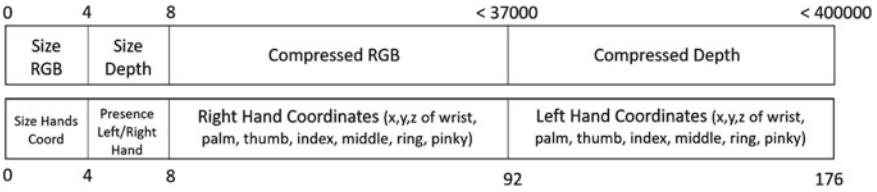


Fig. 44.2 Structure of the packet sent by the camera client to the elaboration unit

camera placed under the signer's hands is a LeapMotion [6]. Since it already has a hand tracking algorithm, the coordinates (x, y, z) of the hands with respect to the LeapMotion are directly extracted in the CC and sent to the EU.

The packets sent to the EU are then composed by the size of the compressed RGB image, the size of the compressed depth image, the compressed RGB image, the compressed depth image, the size of the coordinates of the hands, the presence of the two hands and the coordinates of the hands (x, y, z of the wrists, the palms and the 5 fingers of each hand, for a total of 42 integers for each detected hand). The structure of the packet is shown in Fig. 44.2. For simplicity, the single packet is split into two parts: the upper one regards the images coming from the Orbbec camera and the lower one shows the data from the LeapMotion. Furthermore, to avoid that the packets can be stolen by *man in the middle*, an encrypted socket is used to communicate with the EU.

44.2.2 Elaboration Unit

The Elaboration Unit represents the core of the entire application. Here the packets are taken and elaborated in order to obtain the key-points of the body. It is divided in two parts: the connection management and the GPU server.

The connection management takes all the packets coming from the CC and associates them to the selected RC. The association is handled exploiting the communication table which represents the connection between the IP addresses and the ports of the Robot-Camera clients.

The GPU server receives the packets, extracts the data (RGB, depth and hands points) and decompresses the images (RGB and depth) in order to obtain the original frames again. When RGB and depth images are available, the first one is feed to a Deep Neural Network (DNN) called Open Pose (explained in Sect. 44.3), which analyses the RGB image and gives as output 7 key-points. These points represent the coordinates (x, y) of each joint in the camera frame, where (0, 0) is on the upper-left edge of the frame. But they are not enough to control the robot: the robotic arm needs the three spatial coordinates (x, y, z). Since each pixel of the depth image represents the distance between the camera and the objects in the scene, it is possible to extract the z, mapping the coordinates (x, y) given by the DNN on the depth image.

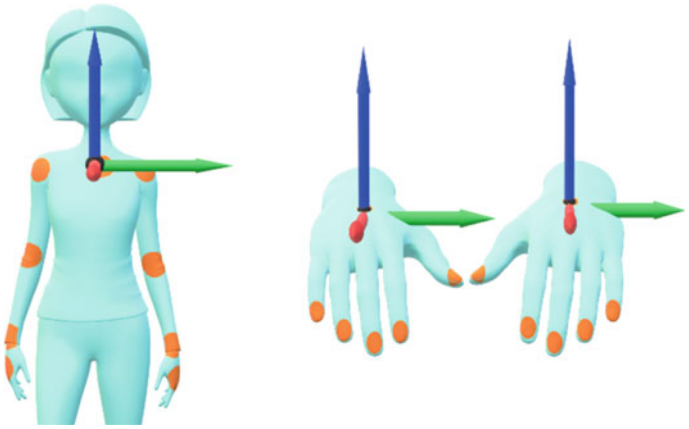


Fig. 44.3 Reference systems: on the left, the body parts which refer to the chest; on the right, the hand points which refer to the palms

Before sending the key points to the RC, it is necessary to convert both the body and the hands coordinates in the same reference system. We chose the chest as the origin for the body points (shoulders, elbows, wrists and palms) and each palm as the origin for every hand's finger points. The three reference systems are shown in Fig. 44.3. Since the DNN does not identify the palms, which are instead pinpointed by the LeapMotion, the palms are linked to the rest of the body by exploiting the coordinates of the wrists taken by the network and the ones given by the LeapMotion.

At the end of this step, all the points are in the right reference system and the new set of coordinates (in millimetres) is sent to the RC. The packet is then composed by 54 integers, 3 (x, y, z) for each joint, in the following order: left arm, right arm, left hand and right hand, where each arm is composed by shoulder, elbow, wrist and palm w.r.t the chest and each hand is composed by thumb, index, middle, ring and pinky w.r.t its palm.

44.2.3 Robot Client

The Robot Client converts the set of coordinates received from the EU in joint variables and gives them as input to the robotic arms. Since each joint is identified by three coordinates, it is easy to control the robots in the space. If there are not both the right and the left robotic arms, the user can select which arm should move.

Fig. 44.4 Rendering of the artificial arm. The black parts represent the structural shell made of hard plastic while the white parts represent the soft artificial skin

44.2.4 Robotic Arm

The robotic arm, shaped to resemble the human arm, is divided in three main modules: upper, representing the shoulder and composed by two revolute joints; middle, representing the bicep and composed by two revolute joints; lower, representing the forearm and the hand and composed by three revolute joints. Thus, the robotic arm consists of seven Degree of Motion (DoM) arranged in a humanoid kinematic tree, with spherical joints for shoulder and wrists, as shown in Fig. 44.4. In order to enable a smooth and safe interaction between human and robot, we conceived design choices such as integration of back-drivable servomotors coupled with serial elastic transmission, development of low mechanical inertia structure and realization and integration of a soft sensorized artificial skin. Through such a skin, we were able to retrieve critical information about the contact between user and robotic arm. We designed and developed large-area skins embedding optical sensors to solve both the magnitude and the localization of an applied load onto the skin surface using a customized neural network. The movement of the arm is based on seven motors: two linear actuators to handle two out of three DoM of the wrist, and five rotary actuators for shoulder and elbow joints and for the last DoM of the wrist. The structural shell of the robotic arm has been printed using a 3D printer and hard plastic filament.

44.3 Pose Estimation

The pose estimation is a re-elaboration of the Open Pose DNN [7]. There are several implementations of the Open Pose network: Python, OpenCL, Unity or CUDA. All these make use of the GPU acceleration, to decrease the inference time and to maintain the real-time. It detects up to 16 points exploiting a 2D image and it returns their (x, y) coordinates. The coordinates are computed starting from the upper-left edge of the image and they are normalized in order to allow images with different resolutions. For our purpose, the body parts needed to communicate are only 7 (chest, shoulders, elbows and wrists). Thus, the network was slightly changed to obtain a more fluid and higher speed algorithm, since it is important that all the images are processed before the next packet arrives to reduce lag. In particular, we simplified the network removing all the parts of the lower body (e.g. legs and hips) and of the

face (e.g. ears and nose) and adding a sorting in the output. In this way, it is possible to understand if some parts have not been well recognized and which ones are. This allows EU to send a message to the CC to inform the user of the wrong position. In addition, if some parts are not recognized, the protocol features that the packet is not sent to the RC: the loss of a frame at the speed of 30 fps does not lead to a wrong movement of the robotic arms.

The results of the DNN are very promising, both for the accuracy and for the inference time: it always recognizes all the joints if they are in the RGB image and an inference lasts about 30 ms, serving each client without communication delay.

44.4 Conclusion

Deafblind people are affected by a severe disability that binds them to communicate only with other people in the same room with the tactile sign language and that does not allow them to communicate with other people remotely. This can lead to social gap and isolation, which can cause depression. The combination of an AI-based body parts tracking algorithm and remotely controlled robotic arms gives them the possibility to communicate remotely, increasing their social relationships. The AI-based system tracks the sign language movements of a person behind two low-cost cameras. Because of the precision and the speed of the inference, these movements are reproduced correctly and in real-time by one or more robotic arms, positioned in a different place with respect to the signer. So, the deafblind person can lay his hands on the robots and understand the message delivered by the signer remotely, in the same way they do when communicating with another person in the same room with the tactile sign language.

Acknowledgements This study was funded in part by the Italian Ministry of Education, Universities and Research within the "Smart Cities and Social Innovation Under 30" program through the PARLOMA Project (SIN_00132).

References

1. The PARLOMA project. http://www.asp-poli.it/smain-innovative-small-appliances-accessories/
2. About deafblindness. https://www.deafblindinformation.org.au/about-deafblindness/
3. Orbbec Astra Pro. https://orbbec3d.com/product-astra-pro/
4. Morphological operations. https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_imgproc/py_morphologival_ops/py_morphological_ops.html
5. Filtering. https://docs.opencv.org/3.1.0/d4/d13/tutorial_py_filtering.html
6. LeapMotion. https://www.leapmotion.com/
7. Cao Z, Hidalgo G, Simon T, Sheikh Y (2018) OpenPose: realtime multi-person 2D pose estimation using part affinity field

Chapter 45
Project VELA, Upgrades and Simulation Models of the UNIFI Autonomous Sail Drone

Enrico Boni, Marco Montagni and Luca Pugi

Abstract Autonomous Sail Drone represent an interesting research topic but also a potential solution for different applications mainly related to the monitoring of large marine areas or freshwater basins. In this work authors from University of Florence presents the current evolution of UNIFI sail drone. Current vehicle development is focused on many activities. In this work authors have focused their attention on current development of the drone with a particular attention to design of energy management system which plays a key role in maintaining vehicle efficiency and reliability.

45.1 Introduction: Brief Description of the UNIFI SAIL DRONE

UNIFI sail drone has been gradually developed from a preliminary design developed during the master engineering thesis of one of the authors [1]. Proposed layout is aligned to innovative solutions currently proposed in literature [2]. A simplified scheme and a brief description of the current evolution of the system are described in Fig. 45.1 and Table 45.1: the vehicle is compose by a composite hull (fiberglass and carbon), which is propelled by a sail whose design, as visible in Fig. 45.1 has been gradually improved in the last two years. Electric energy needed to manage on board electronics, payload and to actuate sail and rudder surfaces is provided by

E. Boni · M. Montagni
DINFO—Department of Information Engineering, University of Florence, Via di Santa Marta 3, 50139 Florence, Italy
e-mail: enrico.boni@unifi.it

M. Montagni
e-mail: marco.montagni@unifi.it

L. Pugi (✉)
DIEF—Department of Industrial Engineering, University of Florence, Via di Santa Marta 3, 50139 Florence, Italy
e-mail: luca.pugi@unifi.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_45

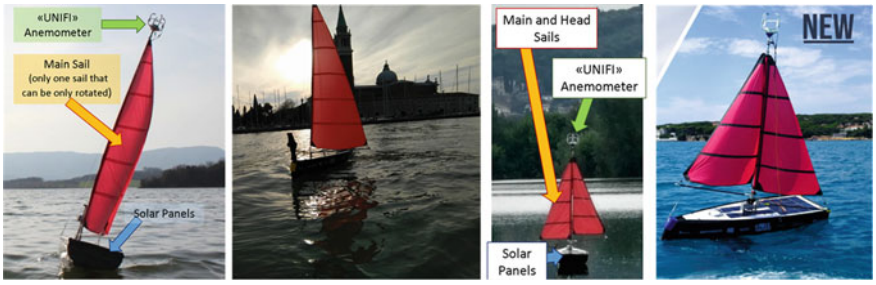


Fig. 45.1 Brief schematic and evolution of the UNIFI sail drone

Table 45.1 Main features and specifications of UNIFI sail drone

Specification	Value	Note
Length	Less than 2 m	Easy transportation
Weight	10–20 kg (payload)	Manual handling
Sail	Conventional sail system	Lighter, cheaper, efficient
Sail actuation	commercial servo systems	Cheap assembly and maintenance
Sail control	Full sail setting	Sail settings according different weather cond
Elec. control units	Cheap microcontrollers	Cheap assembly
Anemometer	Low cost automotive components	Cheap assembly and maint. Easy customization
Energy Man.	Solar panels and backup batteries (48 h of operations)	Night functionality, robust against weather or system degradation
Autonomy	Virtually unlimited	
Use mode	Autonomous/rem. Control	Easy calibration with remote control
Final Prod. cost	1000–5000 €	Preliminary raw evaluation

solar panels in order to assure a near to null environmental impact and an almost unlimited autonomy of the vehicle. Since the vehicle must be able to operate even at night when solar energy is not available, an on-board storage system, charged by solar panels, is used to provide a stable power source. Vehicle is controlled by the simple navigation logic described in Fig. 45.2:

- A high-level trajectory planner (details Fig. 45.2a, b) is used to generate a mission profile congruent to the navigation between imposed waypoints. Mission can be modified according the intervention of vision-based obstacle identification and avoidance algorithms (described in Fig. 45.2c).
- Once trajectory is planned, an inner navigation loop (described in Fig. 45.2d) regulates the orientation of sails in order to assure an optimal incidence of the incoming wind, while the rudder is used to further correct vehicle trajectory. Wind direction

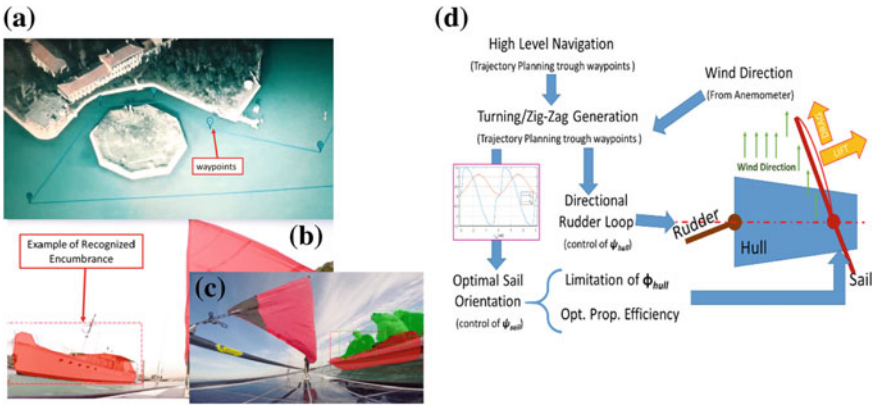


Fig. 45.2 **a–d** Upper and lower control layer of the UNIFI sail drone

and intensity is measured through an internally designed ultrasound anemometer [3]. Since it's not possible to travel against the wind the system provides a feedback to the upper control layer to modify desired trajectory in order to reach the next waypoint with a zig-zag pattern that allows to the boat to sail upwind [4].

Authors are also installing some further sensors to evaluate water and air quality of inspected areas (both physical and chemical parameters are considered). The continuous increase of complexity of installed on board systems are boosting the requirements in terms of required energy and probably in a next future of needed encumbrances on the boat so for this reason authors have to implement an improvements of the current management systems in order to increase the amount of generated energy, the efficiency of conversion and storage processes, also trying to reduce the encumbrance of some important components such as batteries.

45.2 New Solar Panel Design

The original solar panel (Fig. 45.3) was positioned almost on the rear of the boat and divided into two part. Standard solar cells where employed, and the total installed power was 48 W nominal (24 W each). The new solar panel design (Fig. 45.3) includes the use of high-efficiency next-generation SunPower flexible solar cells.

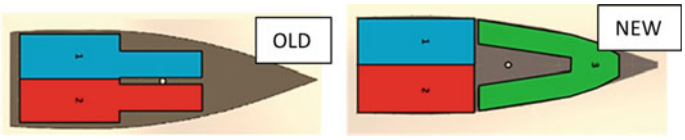


Fig. 45.3 Position on the first prototype (old) and in the new one (new)

Three solar panels are placed on the top surface of the boat. Two panels are installed on the rear side, one on the left side, the other on the right side. The third panel is on the front side. Each panel has a nominal power of 33 W, they are composed of a series of 20 half cells, each one providing 0.55 V and 3 A at maximum power point. The nominal panel voltage is 11 V. Dividing the solar panel in 3 regions allows to maximize the available solar energy, thanks to the fact that at least one full panel is guaranteed being not covered by the sail shadow in all conditions.

45.3 $LiFePO_4$ Battery Pack

The original dual battery pack was composed by two Lead-Gel 12 V batteries, with a capacity of 12 Ah each. The new dual battery pack is composed by two independent series of 4 $LiFePO_4$ cells, each one with a capacity of 20 Ah. The nominal voltage of the battery pack is 13.2 V, with a minimum of 10 V (2.5 V per cell) and a maximum of 14.4 V (3.6 V per cell). The minimum and maximum working voltage of the cells was chosen within the extreme values allowable for this chemistry (2.45–3.65 V) to preserve the battery life. Since the cells in a series can slowly reach different charging states, due to microscopical differences from cell to cell, to avoid damaging the battery we designed a battery balancing circuit. The circuit is based on the BQ76920 chip from Texas Instruments. The chip automatically measures cell voltages and when one reaches the maximum programmed value (3.6 V) a discharging circuit, in parallel with the cell, is activated. A digital interface allows an external microcontroller to monitor cells status and manually override the protection algorithm. Moreover, the microcontroller can communicate with the central unit of the autonomous sailboat, thus providing vital information related to the battery pack to the main controller. The two battery strings are OR-ed with two Schottky diodes. This ensure additional power supply reliability regarding the possible failure of one half of the battery pack.

45.4 High Efficiency MPPT Battery Charger

Each panel is equipped with a separate MMPT battery charger, to better cover different shading conditions and to increase the overall reliability of the system. Panels #1 and #3 are used to recharge one half of the battery pack, while panel #2 is connected to the second half of the battery pack. The old MPPT charger was based on a simple constant panel-voltage circuit. While the circuit is very simple, it actually doesn't fit the use case. In fact, different shading conditions, due to the sail position and boat orientation, easily bring different (lower) MPP voltages for the panel. Moreover, the high temperatures that the panels can reach during a sunny day forced the use of a low MPP voltage, thus lowering the maximal power extraction in all

other conditions. The new MPPT charger is based on the LT8490 chip from Linear Technologies. The chip is a buck-boost switching regulator battery charger. The internal logic provides automatic continuous maximum power point tracking with a perturbe-and-observe algorithm. The panel is also scanned periodically to avoid settling on a local maximum power point for long periods of time, in the case of non-uniform panel illumination. The circuit design involved careful optimization of all the components in order to achieve the maximum efficiency. Panel input specifications, based on the new solar panels, were: 7–11 V input voltage and 3.5 A maximum input current. Battery output was configured as 9–14.8 V, 4.3 A maximum output current.

Based on the input/output specifications, the buck-boost switching cell (Fig. 45.4) was optimized. We started choosing a 170 kHz switching frequency: a compromise between reducing the switching losses and keeping the inductor small. Then, an inductor was selected, considering the lower bound of 7 μH due to the combination of switching frequency, min/max input/output voltage and current values:

$$L_{(MIN)} = \frac{V_{IN(MIN)} \cdot \left(\frac{DC_{(MAX,M3,BOOST)}}{100\%}\right)}{2 \cdot f \cdot \left(\frac{V_{RESNSE(MAX,BOOST,MAX)}}{R_{SENSE}} - \frac{I_{OUT(MAX)} \cdot V_{OUT(MAX)}}{V_{IN(MIN)}}\right)} H$$

The selected component (Wurth Electronics 74436411500) has 15 μH inductance (to keep some margin with the minimum requested inductance) and a DC resistance of 1.3 mΩ. We evaluated the maximum power dissipation due to resistive losses to be 0.13 W. The four MOSFETs contribute to power losses by both resistive losses due to channel resistance and switching losses due to input/output capacitance. Considering all the loss contributions, a TPWR8503NL from Toshiba was selected, with 1 mΩ $R_{DS(ON)}$ and very low gate charge and input/output capacitance. The estimated maximum total losses were 1.22 W for the four MOSFETs. A PCB was designed ad assembled following all the above specifications. To evaluate the circuit efficiency,

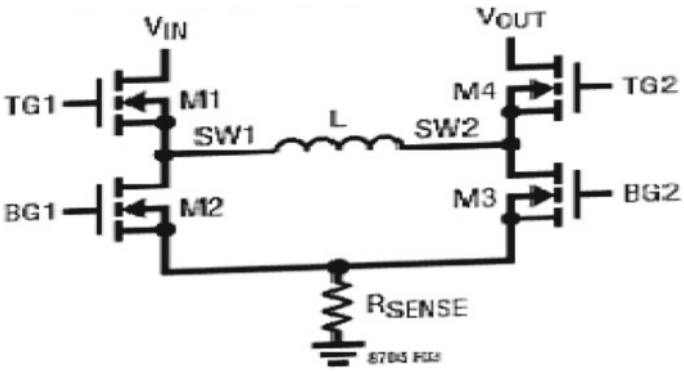


Fig. 45.4 Buck-boost switching cell topology

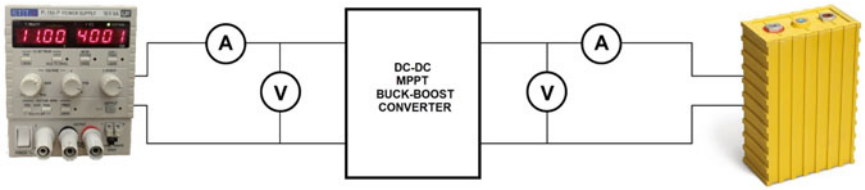


Fig. 45.5 Measurement setup to evaluate circuit efficiency

the MPPT system was supplied with a current limited power supply, and the output was connected to a constant voltage load. The power supply was swept between 7 and 11 V and for each voltage the current limit was swept between 0.2 and 4 A. The output voltage was set to 13.2 V. Input and output voltages, input and output currents were measured with four Peaktech 3440 multimeters connected to a data collecting PC (Fig. 45.5). The corresponding input and output power and thus efficiency was calculated for each setpoint.

45.5 Results and Final Conclusions

The maximum input power of the designed MPPT charger is 38.5 W (11 V, 3.5 A), while the maximum estimated switching losses are 1.35 W (0.13 + 1.22 W). The predicted maximum efficiency is thus 96.5%. 1160 efficiency measurement points where extracted with the measurement setup described in Sect. 45.2. Figure 45.6 shows the actual efficiency measurements results. From the graphs we found that the circuit reach the maximum efficiency of 96.2% at input values of 10.2 V, 3.5 A. This well matches with the predicted values. Regarding the battery pack, for the old lead-acid cells we estimated an 82% charging/discharging efficiency (total energy

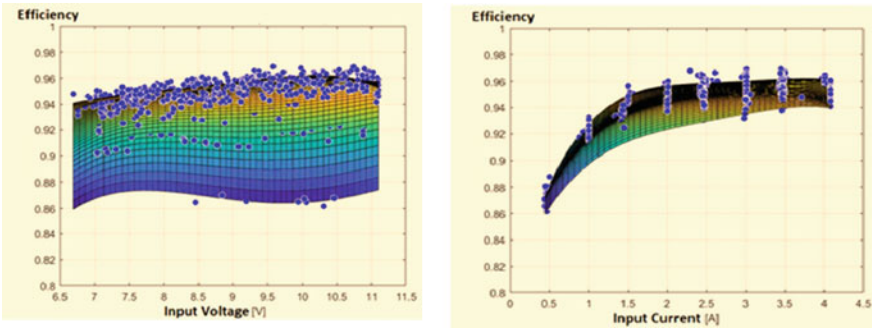


Fig. 45.6 Measurement results, ploted against input voltage (left) and input current (right). The shaded surface is the least square interpolant over the measured points (blue dots), colors are referred to efficiency values

Table 45.2 Performance comparison of energy storage and management system

Design layout	Panel power (W)	Derated panel power due to 30° inclination (W)	Derating by MPPT efficiency	Derating by battery chemistry charging efficiency	Energy stored in 6 h of full sun (Wh)
Old	48	32	21 W (67%)	17.5 W (Pb, 82%)	105
New	99	66	64 W (96.2%)	63 W ($LiFePO_4$, 98%)	378

extracted from the pack vs total energy provided to the pack), while for the new $LiFePO_4$ pack we estimate a 98% charge/discharge efficiency. Table 45.2 reports the incremental reduction with respect to the nominal panel power due to three main reduction factors: sun incidence angle (and average shading), MPPT circuit efficiency, battery charge/discharge efficiency. The last column shows how much energy can realistically be stored in a sunny day considering those factors. As we can see the new system allows to store 378 Wh, while the old one 105 Wh. The battery balancing chip included in the battery packs has a negligible impact on the overall system efficiency, being its current consumption, when fully active, less than 200 μA. These performances are superior respect to power management systems of sail robots in literature [5]. This feature is very interesting especially not only for the proposed applications but more generally for management and balancing of batteries for other applications, such as example electric road vehicles [6] or other kind of autonomous underwater drones [7].

Acknowledgements Authors wish to thank Fondazione Cassa di Risparmio di Firenze that have granted the funding of VELA Project.

References

1. Allotta B, Pugi L, Massai T, Boni E, Guidi F, Montagni M (2017) Design and calibration of an innovative ultrasonic, arduino based anemometer. In: Conference proceedings—2017 17th IEEE international conference on environment and electrical engineering and 2017 1st IEEE industrial and commercial power systems Europe, EEEIC. I and CPS Europe 2017, art. no. 7977450. https://doi.org/10.1109/eeeic.2017.7977450
2. Alves JC, Cruz NA (2008) FAST—an autonomous sailing platform for oceanographic missions. In: OCEANS 2008, Quebec City, QC, p 1–7. https://doi.org/10.1109/oceans.2008.5152114
3. Pugi L, Allotta B, Boni E, Guidi F, Montagni M, Massai T (2018) Integrated design and testing of an anemometer for autonomous sail drones. J Dyn Syst Meas Control Trans ASME 140 (5):055001. https://doi.org/10.1115/1.4037840
4. Sauzé C, Neal M (2011) Long term power management in sailing robots. In: OCEANS 2011, IEEE, Santander, p 1–8. https://doi.org/10.1109/oceans-spain.2011.6003406

5. Boni E, Montagni M, Pugi L (2019) Autonomous sail surface boats, design and testing results of the MOUNTAINS prototype. In: Lecture Notes in Electrical Engineering, vol 550. pp 453–459 (9783030119720). https://doi.org/10.1007/978-3-030-11973-7_54
6. Pugi L, Grasso F, Pratesi M, Cipriani M, Bartolomei A (2017) Design and preliminary performance evaluation of a four wheeled vehicle with degraded adhesion conditions. Int J Electr Hybrid Veh 9(1):1–32. https://doi.org/10.1504/ijehv.2017.08281
7. Pugi L, Pagliai M, Allotta B (2018) A robust propulsion layout for underwater vehicles with enhanced manoeuvrability and reliability features. Proc Inst Mech Eng Part M: J Eng Marit Environ 232(3):358–376. https://doi.org/10.1177/1475090217696569

Chapter 46
DC-Link Capacitor Sizing Method for a Wireless Power Transfer Circuit to Be Used in Drone Opportunity Charging

Andrea Carloni, Federico Baronti, Roberto Di Rienzo, Roberto Roncella and Roberto Saletti

Abstract Resonant-coupled inductive Wireless Power Transfer systems are very appealing as opportunity charging systems for drone applications. Drones are compact systems in which weight and size are critical constraints, so the on-board electronics and the battery must be as small and light as possible. The paper presents the LTSpice simulation analysis of a circuit on the WPT secondary side that uses the intrinsic inductance of the Li-poly battery and only an external capacitor as filter of the full-wave bridge rectifier that typically constitutes the DC-link. The analysis shows the trade-off between the power delivered to the battery and the capacitor size. It results that it can be found a capacitor value that maximizes the power transfer to the battery at the expense of a non-optimal transfer efficiency and increased ripple in the battery current. That value sets the *LC*-filter resonant frequency close to the double of the excitation frequency of the WPT system.

46.1 Introduction

The good mid-range power transfer capabilities achievable with resonant-coupled inductive Wireless Power Transfer (WPT) systems have led to many battery opportunity charging implementations for application to flying devices such as drones [1]. A basic WPT system consists of two magnetically coupled resonant circuits called primary and secondary circuits. Usually, the primary circuit design is straightforward, as it is not limited by weight and volume constraints. It is powered by an external energy source and employs a D-class amplifier [2]. Instead, the secondary circuit design is critical as it is located on board the drone, where weight is a major

A. Carloni · F. Baronti · R. Di Rienzo · R. Roncella · R. Saletti (✉)
Dipartimento di Ingegneria dell'Informazione, University of Pisa, via Girolamo Caruso, 16, 56122 Pisa, Italy
e-mail: r.saletti@iet.unipi.it

A. Carloni
e-mail: andrea.carloni@ing.unipi.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_46

issue. It is mainly made by a bridge rectifier that supplies the battery to be recharged [3]. Depending on the application type and the quality of the WPT system, a DC-DC converter and/or filters are present in the secondary side to rectify the transmitted AC power and to control the charge of the battery. The authors reach and track the maximum charging efficiency using a DC-DC converter with variable duty-cycle in [4]. However, this architecture might be not affordable for drones where the overall dimensions and weight are important constraints. Therefore, a simpler solution based on an *LC*-filter may be preferable. The DC-link *LC*-filter architecture used in [5] is safely sized by choosing the capacitor value *C* large enough to obtain a constant output DC voltage on it. However, this assumption may lead to overestimate the capacitor value and to add useless size and load to the drone. The capacitor and inductor sizes are related to the maximum current and voltage values that they can withstand, so these factors can be critical when the power requirement of the opportunity charger becomes large. Moreover, as WPT systems usually work with resonant frequencies from tens of kHz to MHz [3], the available capacity values of commercial capacitors are much less than those that work at lower frequencies. The aim of this paper is to investigate how the sizing of the filtering capacitor influences the power transfer and the efficiency of a generic WPT, by means of LTSpice time-domain simulations. In addition, as a generic Li-ion battery shows an intrinsic inductive component in the WPT frequency range as shown in [6], our idea is to eliminate the external passive inductor and to use the battery itself as the inductive component of the *LC*-filter to achieve a further size and load saving. The final goal of the paper is to find the best trade-off between the size of the *LC*-filter capacitor and the power delivered to the battery.

46.2 Methodology

Figure 46.1 shows the equivalent circuits for the series-series (SS) WPT architecture investigated in this paper. The secondary circuit consists of the diode rectifier bridge

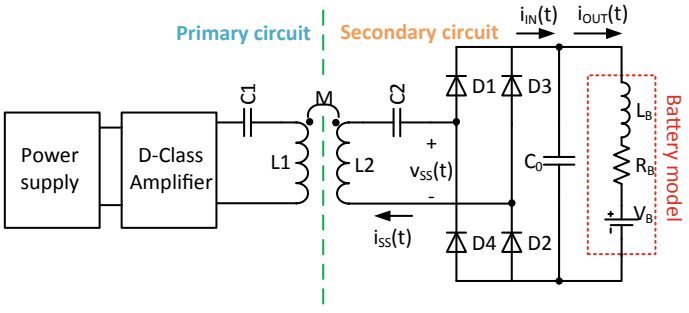


Fig. 46.1 Equivalent circuit of the SS WPT architecture proposed in this paper

and only the passive component C_0, whereas R_B and L_B are the intrinsic parameter of the Randal model [6] of a generic Li-ion battery valid at the frequencies of interest. Let us determine the frequency response of the *LC*-filter consisting of C_0 and the parasitic inductance of the battery.

46.2.1 LC-*Filter Frequency Response*

As described in [7], the SS architecture fixes the current in the secondary circuit. Therefore, the frequency response of the *LC*-filter showed in Fig. 46.1 can be evaluated by considering the current $i_{IN}(t)$ coming from the rectifier bridge as input, and the current $i_{OUT}(t)$ that flows in the Li-ion battery as output. The circuit behaves like a second order low-pass filter, the Laplace domain response of which is well known and shown in (46.1), together with the discriminant Δ of the polynomial [8]. The value of Δ determines the position of the poles of the filter. The value of C_0 that makes $\Delta = 0$, i.e. C^* reported in (46.2), sets the limit between real and complex conjugate poles. If C_0 is lower than C^*, the filter has complex conjugates poles.

$$\frac{i_{OUT}(s)}{i_{IN}(s)} = \frac{1}{s^2 L_B C_0 + s R_B C_0 + 1} \quad \Delta = C_0\left[R_B^2 C_0 - 4L_B\right] \tag{46.1}$$

$$C^* = \frac{4L_B}{R_B^2} \tag{46.2}$$

By defining, f_0 as the resonant frequency and ξ as the damping factor of the filter as expressed in (46.3),

$$f_0 = \frac{1}{2\pi\sqrt{L_B C_0}}, \quad \xi = \frac{R_B}{2}\sqrt{\frac{C_0}{L_B}}, \tag{46.3}$$

Equation (46.1) can finally be written as in (46.4)

$$\frac{i_{OUT}(s)}{i_{IN}(s)} = \frac{1}{\frac{1}{4\pi^2 f_0^2}s^2 + \frac{\xi}{\pi f_0}s + 1}. \tag{46.4}$$

When the damping factor approaches to 0, the time-domain response will show an increased oscillating behaviour.

46.2.2 Time-Domain Simulation Analysis

The time-domain response of the secondary circuit in Fig. 46.1 was evaluated as a function of the circuit parameters by means of the LTSpice electrical simulator. Since the SS WPT architecture behaves like a current generator [7], we represent the circuit up to the bridge rectifier as a sinusoidal current generator $i_{ss}(t)$, with 25 A amplitude and 150 kHz frequency. These values resemble those commonly used in medium power applications such as drones [2, 9]. The diodes are modelled as ideal switches with a voltage drop V_γ of 640 mV. The battery voltage is fixed at 22 V, whereas L_B and R_B were extracted from a real Li-ion battery as it will described in the following subsection. The step directive in LTSpice allows us to perform a parametric simulation, where the value of C_0 is logarithmically swept between 50 nF and 500 μF. Finally, the total power P_B transferred to the battery and the input-output efficiency η are evaluated. Let us note that P_B is the active power transferred to the battery, being defined as the average of the product between the battery electromotive force V_B and the battery current $i_{OUT}(t)$ in Fig. 46.1. Moreover, η is the ratio between P_B and the power at the bridge rectifier input.

46.2.3 Battery Parameter Experimental Extraction

The intrinsic resistance and inductance of a real Li-ion battery specific for drone applications were measured by performing the Electrochemical Impedance Spectroscopy (EIS) of a TA-15C-16000-6S1P-EC5 battery. It consists of 6 cells in series, with 22.2 V nominal voltage and 16 Ah capacity. The spectroscopy test was performed by means of a Gamry Reference 3000 [10] set in galvanostatic mode. The instrument sets a 0.1 AC-current on the single cell and measures the cell voltage between 1 Hz and 500 kHz in ten points per decade. The extracted Bode diagrams of the impedance were fitted by the Gamry Echem Analyst software [11] by which the intrinsic resistance and inductance of the six cells of the battery were derived.

46.3 Results and Discussion

The resistance values derived as above show an average value of 29.957 mΩ with a standard deviation of 76.085 μΩ, whereas the average inductance is 217.3 nH, with 10.721 nH standard deviation. Thus, the total battery resistance R_B and inductance L_B are 179.4 mΩ and 1.304 μH, respectively. These values were used by the simulator to perform the analysis described in Sect. 46.2.2. Furthermore, the value C^* defined in (46.2) is 160.5 μF. The power transferred to the battery P_B and the efficiency η obtained by the circuit simulations when the capacitance C_0 is the parameter are shown in Fig. 46.2.

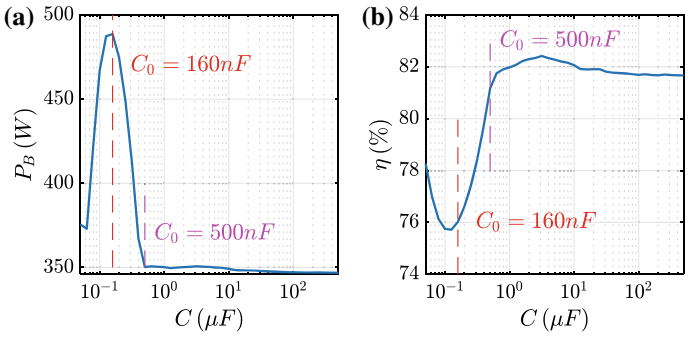


Fig. 46.2 **a** Total power transferred to the battery; **b** power efficiency

It results that the power delivered to the battery and the circuit efficiency significantly depend on the value of C_0 and thus on the resonant frequency f_0 of the *LC*-filter. Considering the power graph in Fig. 46.2a, it is possible to determine three intervals, where the circuit has three different behaviors. Figure 46.3 shows the current and voltage waveforms at the rectifier input for three C_0 values, (a), (b) and (c),

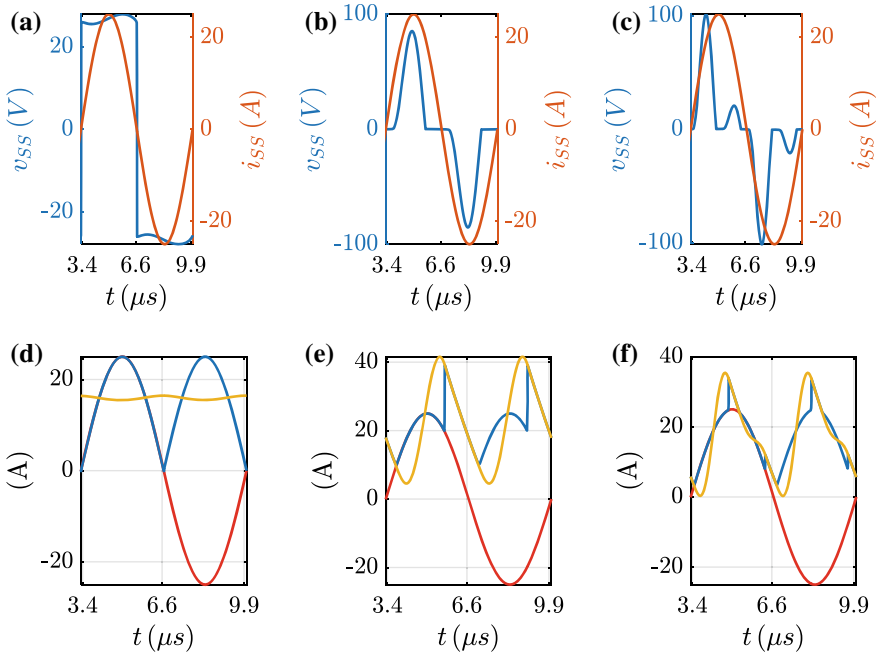


Fig. 46.3 Voltage $v_{SS}(t)$ and current $i_{SS}(t)$ waveforms at the input of the bridge rectifier, with C_0 = 5 µF (**a**), C_0 = 126 nF (**b**) and C_0 = 50 nF (**c**). Bridge rectifier input current $i_{ss}(t)$ (red line), bridge rectifier output current $i_{IN}(t)$ (blue line) and battery current $i_{OUT}(t)$ (yellow line) for C_0 = 5 µF (**d**), C_0 = 126 nF (**e**) and C_0 = 50 nF (**f**)

respectively, each one representing a particular behavior. Figure 46.3 also shows the current waveforms at the rectifier input (red), rectifier output (blue) and in the battery (yellow), for the same three C_0 values, in (d), (e) and (f), respectively.

For capacity values higher than 500 nF and thus for resonant frequencies lower than 197 kHz, the power is almost constant at 350 W. The bridge rectifier works in continuous mode in this capacity interval. Its input voltage $v_{SS}(t)$ resembles a square wave as shown in Fig. 46.3a; the diodes D3, D4 and D1, D2 work as rectifying couples, and the filter provides good low-pass effect on the battery current (see Fig. 46.3d).

Instead, the power delivered to the battery grows for C_0 values between 500 nF and about 160 nF, i.e. for f_0 between 197 and 349 kHz, as it can be seen in Fig. 46.2a. We note that the bridge rectifier now works in discontinuous mode. There are time intervals in each semi-period of Fig. 46.3b, where $v_{SS}(t)$ is fixed at $-2\ V_\gamma$, because all the diodes of the bridge simultaneously conduct. However, a beneficial effect is that $v_{SS}(t)$ is more like a sine wave than before, and the active power delivered to the load is higher with respect to the previous case. As C_0 approaches 160 nF, the phase angle between the fundamental components of $v_{SS}(t)$ and $i_{SS}(t)$ reduces itself, and the power delivered increases. This is a very appealing behavior, particularly for opportunity charging, where the goal is to deliver the maximum power possible for a limited amount of time. The drawback is found in the reduced filtering action, as the battery current becomes more oscillating (see Fig. 46.3e). However, this fact does not affect the battery ageing as demonstrated in [12].

Then, the power starts to decrease when the capacity becomes lower than 160 nF. Finally, for capacity lower than 50 nF and resonant frequency above 624 kHz, the waveforms in Fig. 46.3c and the battery current in Fig. 46.3f exhibit consistent overshoots and undershoots. This is a region to avoid, as the damping coefficient drops below than 0.02. The efficiency profile in Fig. 46.2b can also be divided in three sections with capacity intervals similar to the previous ones. The efficiency is quite constant at 80% for C_0 higher than 500 nF. Here, the $i_{OUT}(t)$ ripple is very low and produces a negligible loss on the resistance of the battery. For capacity lower than 500 nF, the efficiency starts to decrease, because of the increased power losses on the battery resistance due to the higher $i_{OUT}(t)$ values and on the diodes due to the discontinuous conduction.

46.4 Conclusions

Sizing C_0 of the output filter of a WPT in order to obtain a stable output brings the designer to oversize its value, adding useless size and load to the drone. The paper shows that the filter can be reduced to a single capacitor, as the inductance can be provided by the battery itself, reducing the on-board circuitry. The power delivered to the battery and the process efficiency were evaluated as a function of the C_0 value. As in opportunity charging the aim is to maximize the power delivered to the battery, it has been shown that choosing a value of C_0 that fixes the resonant frequency of the

LC-filter near the double of the excitation frequency of the WPT system leads to the maximum power transfer. The drawback is a reduced filtering effect on the battery current and a non-optimal value of the efficiency in the power transfer. Instead, if the goal is to maximize the efficiency, a value of C_0 that sets the resonant frequency of the *LC*-filter close to the WPT excitation frequency leads to the maximum efficiency in power transfer.

References

1. Lu M, Bagheri M, James AP, Phung T (2018) Wireless charging techniques for UAVs: a review, reconceptualization, and extension. IEEE Access 6:29865–29884
2. Campi T, Cruciani S, Feliziani M (2018) Wireless power transfer technology applied to an autonomous electric UAV with a small secondary coil. Energies 11(2)
3. Zhang Z, Pang H, Georgiadis A, Cecati C (2019) Wireless power transfer—an overview. IEEE Trans Ind Electron 66(2):1044–1058
4. Zhong WX, Hui SYR (2015) Maximum energy efficiency tracking for wireless power transfer systems. IEEE Trans Power Electron 30(7):4025–4034
5. Liu X, Wang T, Yang X, Jin N, Tang H (2017) Analysis and design of a wireless power transfer system with dual active bridges. Energies 10(10):1–20
6. Amanor-Boadu JM, Abouzied MA, Sanchez-Sinencio E (2018) An efficient and fast Li-ion battery charging system using energy harvesting or conventional sources. IEEE Trans Ind Electron 65(9):7383–7394
7. Zhang W, Mi CC (2016) Compensation topologies of high-power wireless power transfer systems. IEEE Trans Veh Technol 65(6):4768–4778
8. Seborg DE, Mellichamp DA, Edgar TF, Doyle FJ (2010) Process dynamics and control. Wiley, p 81
9. Campi T, Dionisi F, Cruciani S, De Santis V, Feliziani M, Maradei F (2016) Magnetic field levels in drones equipped with wireless power transfer technology. Asia-Pac Int Symp Electromagn Compat APEMC, 01:544–547
10. Gamry Reference 3000 (Online). Available: https://www.gamry.com/potentiostats/reference-3000/
11. Gamry Echem Analyst (Online). Available: https://www.gamry.com/application-notes/software-scripting/
12. De Breucker S, Engelen K, D'hulst R, Driesen J (2013) Impact of current ripple on Li-ion battery ageing. World Electr Veh J 6:0532

Chapter 47
Distributed Video Antifire Surveillance System Based on IoT Embedded Computing Nodes

Alessio Gagliardi and Sergio Saponara

Abstract This paper shows the design and the implementation of a distributed video antifire surveillance system based on Raspberry Pi embedded computing board and RPi Camera, able to detect the smoke and trigger autonomously a fire alarm. These smart cameras will be placed in different areas under surveillance, connected together according to an IoT-scheme via wired (e.g. ethernet) or wireless (e.g. Wi-Fi) links, and accessible to several users via web browser. A centralized web interface node shows the video stream of each camera in real time, while a video processing algorithm is responsible for the smoke identification and for the decision making of a fire alarm. Furthermore, the system is able to auto record the video in case of fire alarm. Target applications are distributed smoke/fire alarms in smart cities or smart transport systems or smart factories.

Keywords IoT (Internet of Things) · Distributed smoke/fire alarm systems · Embedded video processing · Raspberry pi

47.1 Introduction

Fire is an undesirable event that causes every year billions of dollars in damage to property and the environment. Fire and smoke can be detected at the state of art by installing smoke/fire detector nodes that typically exploit ionization and photometry. Through these mechanisms, they can identify the presence of certain particles and trigger a fire alarm. Although the technologies are becoming affordable, these state of art systems have the drawback to react slowly in large areas and they cannot be installed in open spaces. Closed circuit television systems (CCTV) and cameras instead, are already installed in factory buildings, city streets and public transportation for surveillance purpose. Exploiting an already existing video infrastructure allows

A. Gagliardi (✉) · S. Saponara
Department of Information Engineering, University of Pisa, Via G. Caruso 16, 56122 Pisa, Italy
e-mail: alessio.gagliardi@phd.unipi.it

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

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_47

to reduce the purchase/installation costs of additional add-on products, increasing only the complexity of the video processing algorithm used to identify the smoke, since it appears before the flames and represents the first event of a fire hazard. Albeit there are available commercial and open-source solutions that act as Surveillance Monitoring Systems, they mainly are used for security purpose since they offer poor integration in the event of fire. Often computer vision algorithm such as motion detection, are embedded in the system like in MotionEye [1] or RPi Cam Web Interface [2], specifically designed for Linux OS and Raspberry Pi board. Other solutions like [3–6] consist only in the implementation of the fire-smoke detection algorithm with the lack of integration of a centralized system for surveillance. The main objective of this paper is to show the design and implementation of a video antifire surveillance system using several independent Raspberry Pi embedded computing nodes and Pi Cameras, able to detect the smoke and trigger autonomously a fire alarm. Each node will be a smart camera with IoT connection capabilities. These smart cameras will be placed in different areas under surveillance while a central node provides the monitoring of such areas from a centralized point of view, through a web user interface. A centralized video processing algorithm is responsible for the smoke identification and for the decision making of a fire alarm. Furthermore, the proposed distributed system meets a minimum security standard by being password protected with different user permission. The paper is organized as the following: Sect. 47.2 describes the System Architecture while Sects. 47.3 and 47.4 present the Hardware and Software Implementation respectively. Section 47.5 discusses about the experimental results and it is followed by the conclusion in Sect. 47.6.

47.2 Distributed Video Antifire Surveillance System Architecture

The architecture of the proposed system is shown in Fig. 47.1.

The system is composed by multiple Raspberry Pi streaming video feed to a central hub that allows access from any devices connected to the network such as a computer, tablets or smartphone. Each camera is connected to just one Raspberry Pi board, which represents a smart node in the system architecture. Although it is possible to connect multiple cameras to the same Raspberry Pi board, the network topology in Fig. 47.1 guarantees the best performance in terms of stability and maintaining of a fluid frame rate during the video feeds. All the smart nodes of the system are interconnected through a router using a static IP address. The router directs the traffic and video streaming of each node and serves as networking device. The main interface of the smart antifire surveillance system resides on the Central Hub. This would mean that the entire system can be accessed through a web interface provided by the Central Hub via web browser.

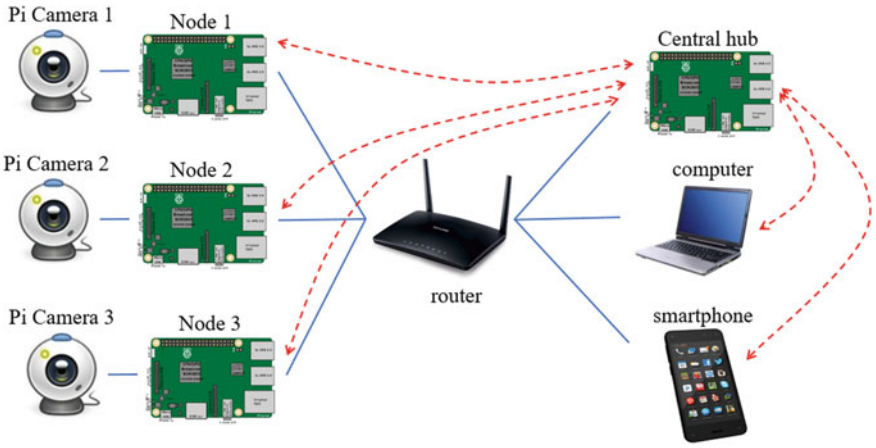


Fig. 47.1 Smart surveillance antifire system architecture

47.3 Hardware Implementation

The nodes in the network in Fig. 47.1 consists of several Raspberry Pi 3 model B units, a low cost, low power single board embedded computer. The board is equipped with a Broadcom BCM2836, a System on Chip (SoC) including a 1.2 GHz 64-bit quad-core ARM Cortex-A53 processor, 512 KB of cache L2, 1 GB of DDR2 RAM, Video Core IV GPU, 4 USB 2.0 ports, on-board WiFi @2.4 GHz 802.11n, Bluetooth 4.1 Low Energy, 40 GPIO pins and many other features [7]. The camera module used in the implementation is a PI Camera Board v1.3 able to deliver a 5 MP resolution image, or 1080p HD video recording at 30 fps [8]. The Pi Camera plugs directly into the CSI (Camera Serial Interface) connector of the Raspberry Camera board. We used a common Netgear DGN2200v3 as router network device.

47.4 Software Implementation

The backbone of the software implementation is the Smoke Detection algorithm, developed in MATLAB 2019. The technique involves several processing steps as motion detection (based on a Kalman Filter), colour segmentation, feature extraction, bounding box extraction, and alarm generation. The details of the algorithm are discussed in [9]. The MATLAB functions of the algorithm are converted in C code as library, transferred on the Raspberry Pi running Raspbian OS and compiled using GCC compiler (already built in) to create a shared library. Then the function is wrapped by using ctypes library [10] that provides compatible data types and functions compliant in Python. Python was preferred as main language for coding due to its simplicity and its strong integration with Raspberry Pi boards and its

components. In fact, the web application is written in Python using Flask [11], a web application framework based on Werkzeug WSGI (Web Server Gateway Interface) toolkit and Jinja2 template engine. The WSGI is a specification for a universal interface between the web server and the web applications, and its toolkit implements requests, response object and utility functions. Jinja2 is a template engine that combined with a certain data source (HTML template, relational database, XML files, etc.) permits to render dynamic web pages. The monitoring system makes use of three main templates: login.html, home.html and index.html. Each template is bounded in some specific Python functions that associate respectively three main URL: http://central-hub-IP/login, http://central-hub-IP/home, http://node-IP/, where "*central-hub-IP*" represents the IP address of the Central Node, while "*node-IP*" represents the IP address of the smart node. The login page, as well as the other webpages, are based on Semantic UI [12] framework, which permits to build fast and concise HTML, along with a complete mobile responsive user experience. The system is built for a multi-user usage, so a database layer is added to the application. Unfortunately, Flask does not have any database support out of the box. So that, we used SQLAlchemy library [13], which is able to manage and query our relational SQLite [14] database built into the central hub Raspberry Pi. The structure of the database is very simple with just one table called "*users*" with three columns: one called *id* that act as primary key, one called *username* and one called *password.* Obviously, the tuples of the database represent a user who is allowed to access to the Antifire System. OpenCV 3.4 [15] is responsible for the management of the Pi camera.

47.5 Experimental Results

To access the program from a smartphone or a computer it is necessary to open a web browser such as Google Chrome and navigate to http://central-hub-IP/login/ (Fig. 47.2).

Once the user is authenticated by inserting valid credentials in the login format, he/she is redirected to the home page (http://central-hub-IP/home) showing the main dashboard (Fig. 47.3). The interface displays all the Camera Nodes available in the network feeding independent video stream. Our test was made by using four Raspberry Pi 3 nodes: one acting as central hub and three as smart camera node. By clicking on the video camera stream the user is then redirected to the Web Interface showing a bigger video screen and multiple option for recording video, saving snapshot and for setting the parameters of the camera (Fig. 47.4). Potential smoke in the camera stream will be identified and inserted into bounding boxes as expected by the smoke detection algorithm [9]. A fire alarm will be generated if the smoke exceeds some time/volume threshold. The video stream will be recorded automatically once the alarm is triggered, and a red circle will appear on the top left of such video stream. Media files of recorded video and snapshot will be stored and accessible in a local drive on each node. Since it was impossible to reproduce a real fire scenario

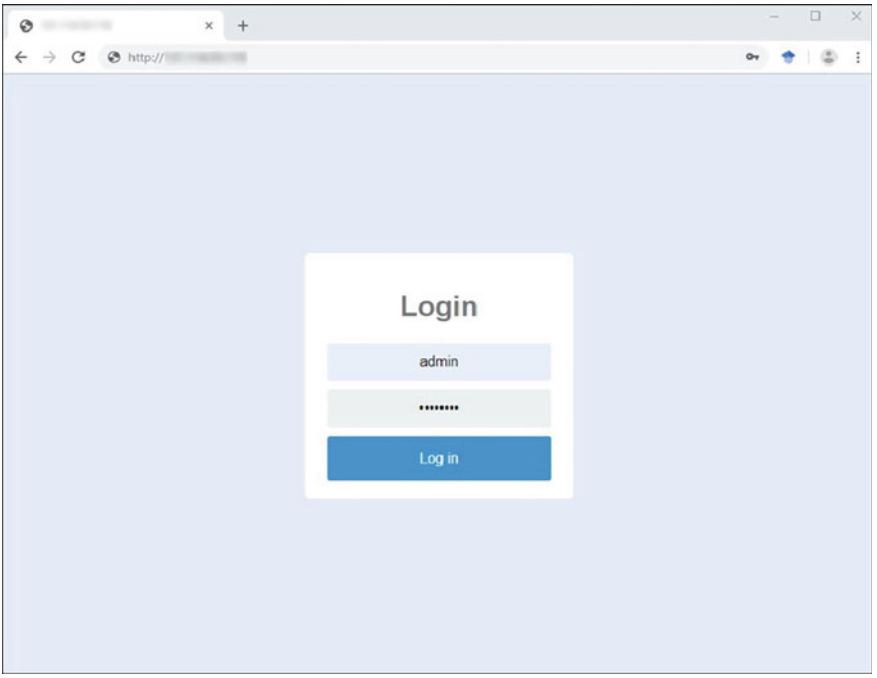


Fig. 47.2 Login webpage

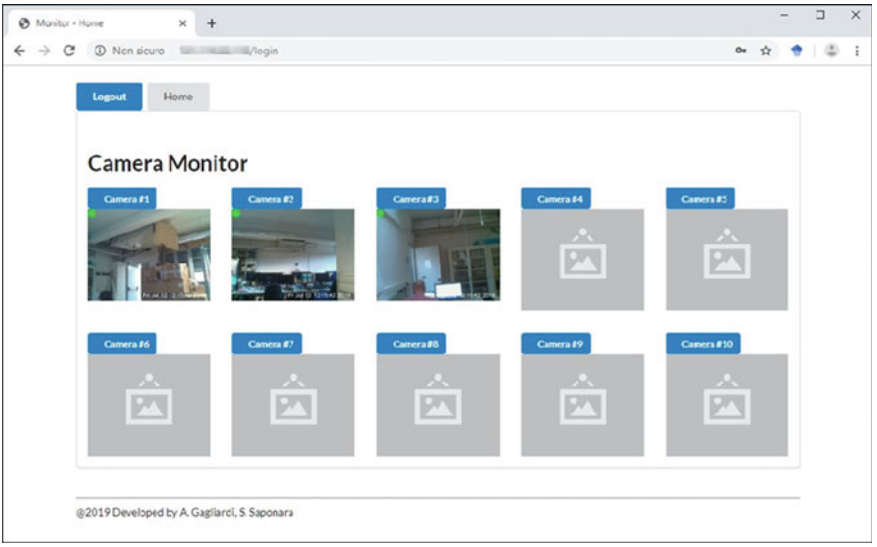


Fig. 47.3 Central hub camera dashboard

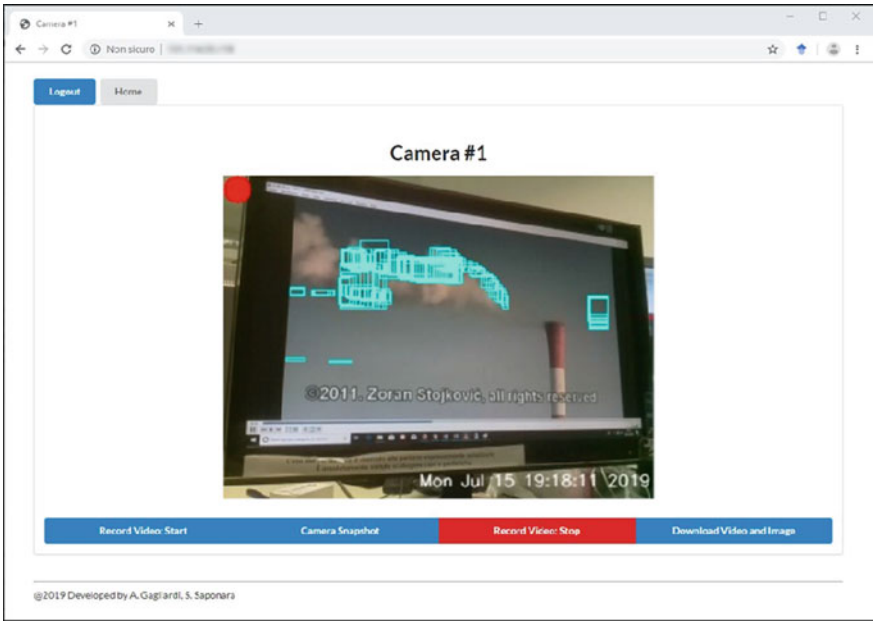


Fig. 47.4 Smart node web interface

in our laboratory, the system was tested on a video playback displaying smoke on the PC monitor and caught in real time by one of the RPi camera node as shown in Fig. 47.4. In fact, the web interface shows the bounding box around the smoke properly detected and a fire alarm is triggered in few seconds as expected.

47.6 Conclusion and Future Work

This paper has proposed a low cost implementation of a distributed Smart Antifire Surveillance System based on Raspberry Pi embedded platform. The system takes advantages of an existing Video Smoke Detection algorithm, proposed by authors, to detect smoke in real-time from several cameras distributed in different areas. Every smart node can feed a central hub that works as collector of the video live stream while a web Interface permits the monitoring of such cameras, to save video and to modify camera parameters. A simple test was carried out using four Raspberry Pi 3 proving the feasibility of the whole system. The communication is done over HTTP so it is not encrypted and it is vulnerable to man-in-the-middle and eavesdropping attacks. A future improvement would be implementing an HTTPS communication to protect the authenticity of the webpage/web-interface, to secure accounts and to keep private user communication and identity. Another possible improvement would be an automatic notification via email to the users as result of a fire detection alarm.

References

1. Crisan C (2017) MotionEyeOs 1 Jan 2017 (Online). Available: https://github.com/ccrisan/motioneyeos/wiki
2. RPi-Cam-Web-Interface 29 Jan 2016 (Online). Available: https://elinux.org/RPi-Cam-Web-Interface
3. Vijayalakshmi S, Muruganand S (2017) Smoke detection in video images using background subtraction method for early fire alarm system. In: 2nd international conference on communication and electronics systems (ICCES), Coimbatore, pp 167–171
4. Tao C, Zhang J, Wang P (2016) Smoke detection based on deep convolutional neural networks. In: International conference on industrial informatics-computing technology, intelligent technology, industrial information integration (ICIICII), pp 150–153
5. Muhammad K et al (2018) Convolutional neural networks based fire detection in surveillance videos. IEEE Access 6
6. Filonenko A et al (2018) Fast smoke detection for video surveillance using CUDA. IEEE Trans Ind Inform 14(2):725–733
7. Raspberry Pi Foundation (Online). Available: https://www.raspberrypi.org/
8. RPI Camera Module v.1.3 (Online). Available: https://www.raspberrypi.org/documentation/hardware/camera/
9. Saponara S, Gagliardi A (2019) AdViSED: Advanced Video SmokE Detection for real-time measurements in antifire surveillance indoor and outdoor systems. IEEE Trans Instrum Meas
10. ctypes—A Foreign function for Python—Documentation (Online). Available: https://docs.python.org/3/library/ctypes.html
11. Flask—Web development, one drop at time—Documentation (Online). Available: https://flask.palletsprojects.com/en/1.0.x/
12. Semantic UI—Documentation (Online). Available: https://semantic-ui.com/
13. SQLiteAlchemy—The Python SQL Toolkit and Object Relational Mapper—Documentation (Online). Available: https://www.sqlalchemy.org/
14. SQLite—Documentation (Online). Available: https://www.sqlite.org/docs.html
15. OpenCV 3.4.0—Open Source Computer Vision—Documentation (Online). Available: https://docs.opencv.org/3.4.0/

Chapter 48
Integrated Simulation Environment for Co-design/Verification of Mechanic, Electronic and Control of Automotive E-Drives: The Smart-Latch Case Study

Emanuele Abbatessa, Davide Dente and Sergio Saponara

Abstract With reference to SW-controlled mechatronic units for the new generation of electrified and assisted vehicles, this work proposes and validate a methodology to simulate together its three main subsystems: electronic HW components (passive and actives, both integrated circuits and board-level components), algorithms and relevant SW implementation running on a Microcontroller unit, mechanical part. With the support of MAGNA, worldwide leader in the production of automotive components, particularly for door systems, we have considered the Smart Latch as case study. The Smart Latch is a new, SW-controlled, mechatronic doors latch. The proposed methodology allows the creation of a digital virtual design and verification environment suited both in design phase for multi-domain component specification (HW, SW, mechanics) or for diagnostic/verification in case of faults.

Keywords SW-defined mechatronics · Model/Hardware-in-the-Loop · Smart Latch

48.1 Introduction

Due to the increasing demands of the modern economy, the development of new products must reach new levels concerning the complexity and the implemented intelligence, while saving resources and reducing the time needed for design and production. Such products often are interdisciplinary systems and they are also called mechatronic systems, referring to the synergistic integration of SW, electronics and mechanics. Hence, mechatronics is interdisciplinary and was defined by Harashima F. as "the synergistic combination of mechanical and electrical engineering, computer science, and information technology, which includes control systems as well as numerical methods used to design products with built-in intelligence".

E. Abbatessa · S. Saponara (✉)
Department of Information Engineering, University of Pisa, Pisa, Italy
e-mail: sergio.saponara@unipi.it

D. Dente
MAGNA Mechatronics, Guasticce (LI), Italy

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_48

To satisfy the requirement of saving time and cost for design and production, model-based design techniques exist, which allow simulating and optimizing the behavior of the designed structure in conditions similar to the real ones. After the virtual development of the product, the system must be implemented in real world and tested in real life conditions in order to validate simulation results. The testing procedure implies validation of different parts of the product like control algorithms, sensor systems, actuation systems, electronic boards. This step can be very challenging and time consuming without using adequate tools and implies a need for methods and tools which can assist in the various processes involved in realizing such products. Application of computer aided engineering (CAE) capabilities, together with custom engineering strategies within the company, is a way that is being implemented in a lot of mechatronic companies. The goal is the usability of all tools and models throughout the whole process of design from the first conceptual ideas to its use for troubleshooting of systems that are already in revenue service. This article shows and tests a methodology to simulate in an integrated environment all 3 subsystems of a mechatronic system: electronic HW components (passive and actives, integrated circuits and board-level components), algorithms and relevant SW implementation running on a Microcontroller unit, mechanical part.

With the support of MAGNA, leader in automotive door systems, we have considered the "Smart Latch" as case study [1–3]. It is a new SW-defined mechatronic product and it is a good system for applying the methodology presented below, since it is based on the interaction among mechanics, electronics and control algorithms (SW running on a microcontroller). Hereafter, Sect. 48.2 reviews the Smart Latch architecture and the limits of state of art commercially-available design and verification methodologies for SW-defined mechatronics. Section 48.3 presents the integrated simulation/verification environment. Simulation results and comparison to experimental measurements are discussed in Sect. 48.4. Section 48.5 deals with conclusion and state-of-art comparison.

48.2 Review of Smart Latch Architecture and State of Art Design Flow

48.2.1 The Smart-Latch New Concept

The "Smart Latch" developed by MAGNA creates a new paradigm for side-door latches by "the first fully electronic side door latch in the market". The industry first application of the latch is on the BMW i8 and it has been selected by other global automotive manufacturers for future vehicle programs. The features of "Smart Latch" are: it removes all mechanical latch system components and eliminates the need for cables, rods and moving handles in the door; significant weight savings compared to mechanical latches; reduced number of components; flexibility to be used in any type of car or truck; improved safety and sound quality.

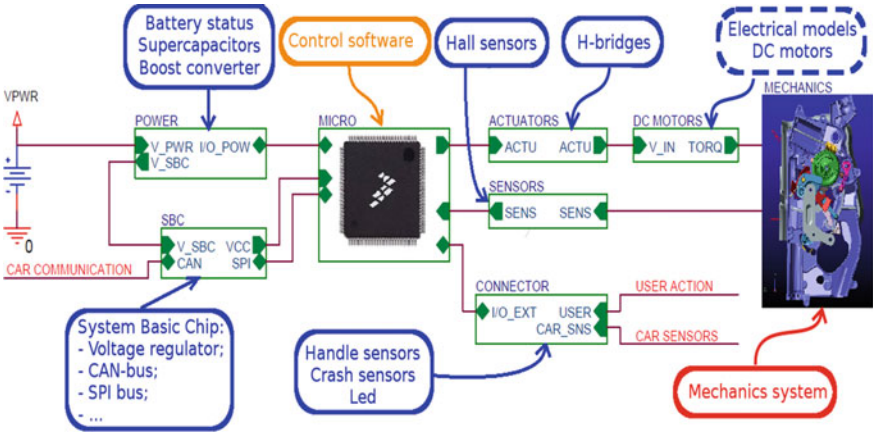


Fig. 48.1 Smart Latch architecture

The "Smart Latch" system includes an on-board ECU (Electronic Control Unit) that has power backup capabilities and generates signal to drive motors for a soft close function for automatic door cinching. The "Smart Latch" provides additional functions like connection with a car's network, diagnostic, passive entry, crash detection and post-crash safety. Figure 48.1 shows a diagram blocks of whole "Smart Latch" system. This picture highlights the subsystems that describe the model inside the "Smart Latch". The "Smart Latch" has HW blocks (pointed by blu arrows in Fig. 48.1), that ensure interfacing the microcontroller core to power supply, motors and sensors. The microcontroller of the Smart Latch runs a state machine that implements the control algorithm (orange arrow in Fig. 48.1). Sensors and motors interact with the mechanical part (red arrow in Fig. 48.1) to accomplish latch door functions.

48.2.2 Limits of State of Art Design and Verification Methodologies

Today many SW houses aim at developing tools to aid the engineers in developing new mechatronics systems. In the following, we consider the main 3 tools in the market.

Simulink Simscape: It provides an environment for modeling and simulating physical systems spanning mechanical, electrical, hydraulic, and other physical domains. It provides fundamental building blocks from these domains that you can assemble into models of physical components, such as electric motors, inverting op-amps, hydraulic valves, and ratchet mechanisms. Because Simscape components use physical connections, the models match the structure of the system under development. Simscape models can be used to develop control systems and test system-level performance. The libraries can be extended using the MATLAB

based Simscape language, which enables text-based authoring of physical modeling components, domains, and libraries. Models can be parameterized using MATLAB variables and expressions. To deploy developed models to other simulation environments, including hardware-in-the-loop (HIL) systems, Simscape supports C-code generation.

Altair Activate: Activate SW is a tool for rapidly modeling and simulating products as multi-disciplinary systems in the form of 1D models (signal-based or physical block diagrams), optionally coupled to 3D models. Some Altair Activate features are:

- Block Diagrams; Control System Design: Providing a natural modeling approach for developing smart systems involving sensors, actuators, feedback, and built-in logic;
- Mix Signal-based and Physical Modeling in Same Diagram: Leveraging the power of predefined Modelica libraries for modeling common Mechanical, Electrical, and Thermal physical components;
- Connections to Other Altair Tools: Enabling true multi-disciplinary system simulation via model exchange or co-simulation with **MotionSolve** for controlled multi-body dynamics, with Flux for controlled motor dynamics models, etc.;
- Typically much Faster than 3D Simulations: Relying on high model abstraction levels enables more product-performance insight earlier and rapid design exploration;
- Support for Functional Mockup Interface (FMI): Including Functional Mockup Units (FMU) enables model exchange or co-simulation connections to non-Altair tools which also support the FMI standard.

MotionSolve performs 3D multi-body system simulations to predict the dynamic response and optimize the performance of products that move. By considering realistic motion-induced loads and environmental effects, designers can be confident that their products, when made and operated, will perform reliably, meet durability requirements, and not vibrate excessively or fail from fatigue. MotionSolve facilitates multi-disciplinary collaboration across product development teams since it enables combined simulation of subsystems for mechanical plants together with electrical/electronic ones.

Synopsys SaberRD: SaberRD [4, 5] is an intuitive, integrated environment for designing and analyzing power electronic systems and multi-domain physical systems. With the proven Saber simulation technology at its core, SaberRD combines ease of use with the power to handle today's complex electrical power problems, allowing engineers to explore design performance, optimize robustness and assure system reliability for a broad range of generation, conversion and distribution applications. SaberRD's true multidomain physical modeling capability and unmatched analysis capabilities provide engineers with a virtual prototyping platform that supports complete system design. With an intuitive and flexible user interface for casual and expert users alike, SaberRD accelerates design for engineering organizations in automotive, aerospace, defense and industrial power. SaberRD promises quick virtual prototyping of complex power electronic systems. Its main features are: (i)

integrated design environment: Schematic design, mixed-signal multi-domain circuit simulation, waveform analysis and automatic report generation capabilities; (ii) built-in design flow: a modern and streamlined interface guides the user to results, stepping through the key step of a simulation-based workflow, including design, modeling, simulation and analysis; (iii) proven in production over 25 years.

The above SW tools permit the design of a mechatronics system but we aim to obtain a model that is as accurate as possible. This means that we'd like to take in account simulations of multibody dynamics, electronics and control system dynamics. Hence, we search an environment that allows in a simple way to obtain results by the co-simulation of 3 different models. The difference of use the methodology presented below is that we make use of other SW and tools, and the advantage is that we have three different solver that exchange information and perform the simulation. This way we can use more accurate SW tools to model the subsystems of a mechatronic product.

48.3 Integrated Simulation/Verification Environment

A "Smart Latch" system, being a multidisciplinary mechatronic unit, needs tools that permit the design of subsystems models belonging to 3 main disciplines, see Fig. 48.2:

- Electronics: used to model the HW interfaces among microcontroller, sensors and mechanical actuation;
- Mechanics: used to model the gears and levers actuated from electrical motor inside door cinching;
- Informatics: used to model the control SW running on microcontroller.

Although the CAD tools presented in Sect. 48.2.2 permit to model systems belonging to the three disciplines, we have to take in consideration the real exigence of the company, MAGNA in this case. This means that to realize the model of "Smart Latch" we first consider the possibility of use the SW that just helps MAGNA to design new components. Hence, as shown in Fig. 48.2, we have (i) Electronics subsystems model designed by "OrCAD Capture CIS"; (ii) Mechanics subsystems model designed by "ADAMS MSC"; (iii) Control software model designed by "Simulink Stateflow". The difference among the SW described in the previous section is that the SW tools used in this work are specific for each discipline and are presented hereafter.

Simulink Stateflow: It is developed by MathWorks and it is a control logic tool used to model reactive systems via state machines and flow charts within a Simulink model. Stateflow uses a variant of the finite-state machine notation established by David Harel, enabling the representation of hierarchy, parallelism and history within a state chart. Stateflow also provides state transition tables and truth tables. Stateflow permits to obtain the C code than can load on a microcontroller.

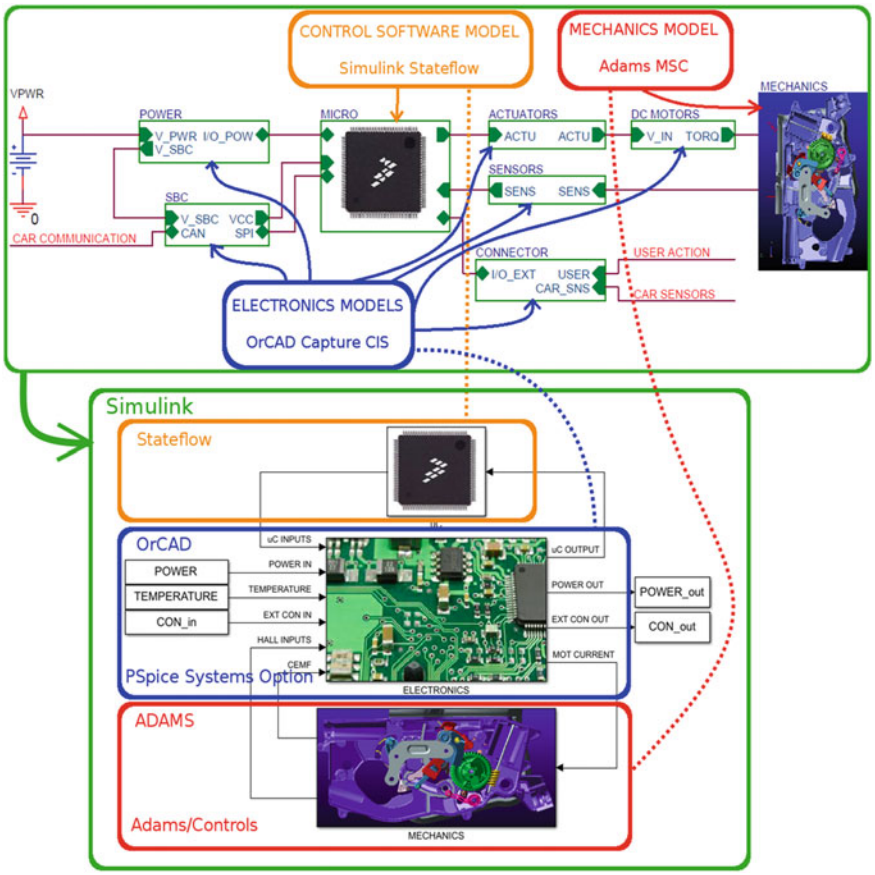


Fig. 48.2 Integrated simulation environment

OrCAD Capture CIS: OrCAD Systems Corporation was a SW company that made OrCAD, a proprietary tool suite used primarily for electronic design automation (EDA). The SW is used mainly by electronic design engineers and technicians to create electronic schematics, perform mixed-signal simulation and electronic prints for manufacturing printed circuit boards. **OrCAD** Capture is a schematic capture application, and part of the OrCAD circuit design suite. Capture does not contain built-in simulation features, but exports netlist data to the simulator, OrCAD EE. Capture can also export a HW description of the circuit schematic to Verilog or VHDL, and netlists to circuit board designers such as OrCAD Layout, Allegro, and others. **Capture** includes a component information system (**CIS**), that links component package footprint data or simulation behavior data, with the circuit symbol in the schematic. Capture includes a Tcl/Tk scripting functionality that allows users to write scripts, that allow customization and automation. Any task performed via the GUI may be automated by scripts. **OrCAD EE PSpice** is a **SPICE** circuit

simulator application for simulation and verification of analog and mixed-signal circuits. PSpice is an acronym for Personal Simulation Program with Integrated Circuit Emphasis. OrCAD EE typically runs simulations for circuits defined in OrCAD Capture, and can optionally integrate with MATLAB/Simulink, using the Simulink to PSpice Interface (SLPS recently became OrCAD PSpice Systems Option). OrCAD Capture and PSpice Designer together provide a complete circuit simulation and verification solution with schematic entry, native analog, mixed signal, and analysis engines. The OrCAD PSpice-Simulink integration, **OrCAD PSpice Systems Option** provides co-simulation and helps verify system level behavior. A circuit to be analyzed is described by a circuit description file, which is processed by PSpice and executed as a simulation.

ADAMS MSC: ADAMS is acronym of Automated Dynamic Analysis of Mechanical Systems and is a multibody dynamics simulation SW [6] equipped with Fortran/C++ numerical solvers. Adams has been proved as very essential to VPD (Virtual Prototype Development) through reducing product time to market and product development costs. ADAMS provides some basic modules: Adams/View; Adams/Solver; Adams/Postprocessor. Several additional modules sold separately are available for extended functionality, for example: Vibration analysis through ADAMS/Vibration includes mode shape analysis; SISO and MIMO closed loop control system modeling and simulation is available through **ADAMS/Controls**; simulate flexible links, via Adams/ViewFlex and/or Adams/Flex. Its approach to flexible body modelling is that of modal analysis which uses a modal neutral Adams/Controls is very well integrated into MathWorks Simulink by some S-functions. A closed loop between Simulink and Adams/Controls makes simulation of non LTI systems very simple. A non-linear time variable model of plant is modeled within ADAMS/Controls and its behavior is reported to Simulink via Named pipe or TCP/IP communication as feedback, whereby analyzed by some controller within Simulink and through some actuators act upon ADAMS/Controls plant in the same communication scheme. Also through control export mechanism, ADAMS/Control can provide MATLAB's Control System Toolbox with a state space model of system under study to be used further for design of controller. Adams also supports importing a compiled DLL version of Simulink models built using Simulink Coder. Functional Mock-up Interface has been supported. It is an open standard interface intended for coupling tools from different vendors for Model Exchange and Co-simulation. Adams is highly integrated with Actran frequency-domain solver for chained simulation analyses of moving mechanisms such as gearbox run-up and impact noise studies, such as door latch mechanisms.

Either OrCad and Adams permit the importing of their models in Simulink. This means that with the tool "PSpice Systems Option" we can import the OrcCAD model of electronics system in "Simulink" and we can make the same to import the ADAMS model of mechanics system with "Adams/Controls". In addition, the use of "Stateflow", that is already a "Simulink" tool suggest the choice of this last software as main software to develop the model of a mechatronics systems. The advantage of this approach is that we can exploit the models already developed by MAGNA for electronics, mechanics and control SW. In this way we avoid the copy model

errors that can occur when we translate model between two SW based on different languages.

To exploit OrCAD model developed from MAGNA in co-simulation with Simulink, we have added to OrCAD model voltage generators to simulate output pins of microcontroller and added circuit equivalent of DC motors to take in account the effect of counter EMF. At this point we have to consider that equivalent circuit of DC motors models, corresponding at block pointed by the dotted shape in Fig. 48.1, it is necessary if we want obtain a system that takes in account the interaction between electronics and mechanics. In fact, the torque developed by a motor is proportional to the current that goes through the inductance of electrical model of DC motors and the counter EMF is proportional to the rotor angular velocity.

To take in account the sensor mounted on the mechanics some new library parts have developed and added to OrCAD model. Connected the sensor developed parts to other voltage generators we obtain the inputs for the electronics model that depend by position of sensors in the mechanics model. To exploit ADAMS model, instead, we added some state variables before export model to Matlab.

After this modification of OrCAD and ADAMS models and applying the tools presented before we can obtain a single integrated environment that incorporates all subsystems used to design a mechatronic system. Figure 48.3 shows results of the model.

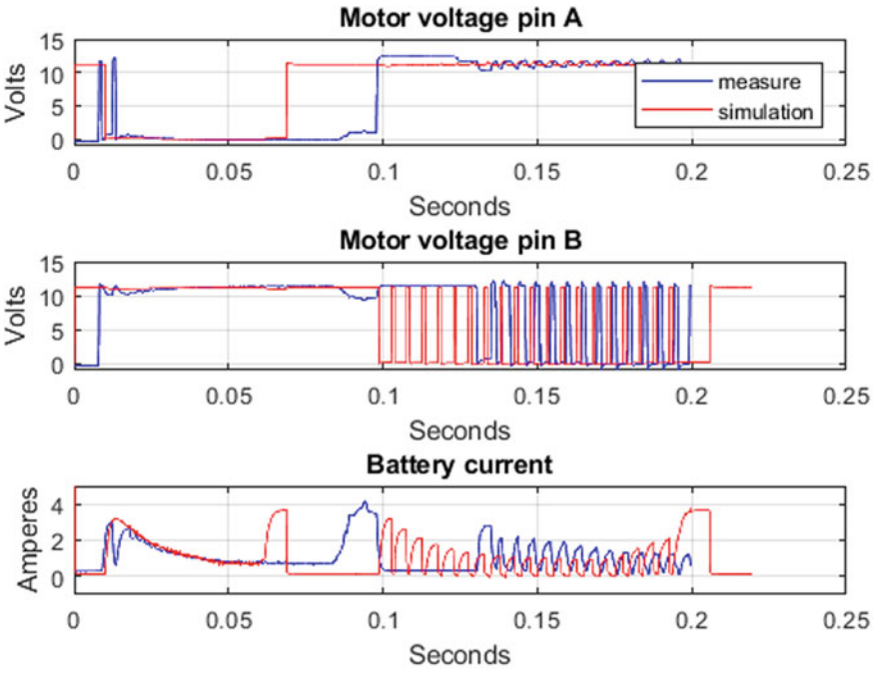


Fig. 48.3 Power release action

48.4 Simulation and Experimental Measurements Results

To validate the methodology explained we compare simulation results with measures of a real "Smart Latch". The door latch system has two main functions, the release of cinching system, that permits the door opening and the reset of it, for the door closing. These functions start after the command arrived by the pushing of handle switch and are performed by the actuation of DC motor into "Smart Latch" and by the sensing of two hall sensors. For safety reason, a more important requirement of the "Smart Latch" is that it must works in the absence of power coming from the main car battery.

Our principal aim is testing of interaction between mechanic and electronic models. To improve the performance of simulation and taking into account the previous presented "Smart Latch" behavior, we have used a simple "Stateflow" chart to model the behavior of control SW. Then, in the model of "Smart Latch", the control SW does not coincide with the one actually present on the real "Smart Latch". The mechanics and electronics, on the other hand, are very accurately modeled as we have developed models starting from the models used by MAGNA. With this assumption we have simulated the "power release" and "power reset" of "Smart Latch" considering two different conditions:

Boost off: fully charge of battery; Boost on: battery complete discharged.

From a control point of view, the "Stateflow" chart used in the model is done ad hoc to perform the two functions before presented. For this reason, the following tests show only a part of the real capacity of this methodology and some inevitable difference are waited between measures and simulations results.

48.4.1 Boost off Test

In the first test type the boost converter is not running because the battery is fully charged. To replicate this condition in the proposed model we have set the power supply to 12 V and we have disabled the boost converter. To compare model and real "Smart Latch" we have measured the voltage of signals used to drive the power release DC motor (pin A and pin B) and current absorbed from battery by Smart Latch (Battery current). Real system measures are reported in Fig. 48.3 with blue lines, while model system results are red lines in the same figure. With the analysis of graphics in Fig. 48.3 we can observe the difference of driving signals, that are constant during power release action, instead, are PWM during power reset action.

The first part corresponds to power release action and in the first two graphics of Fig. 48.3 we observe some difference between real and modeled command used to drive the DC motors. This translate into the difference between the two data type that we see in the first part of the "Battery current" graphic in Fig. 48.3. In the same graphics we highlight the difference of stalling time between the two kinds of data, probably due to the unmodeled friction torque and unmodeled command. The second

part corresponds to power reset action and as we can see in the first two graphics of Fig. 48.3, the power reset action signals used to drive the motor are PWM but real Smart Latch makes use of a variable duty cycle, instead in the model we have supposed a constant duty cycle.

48.4.2 Boost on Test

In the second test type the battery is fully discharged, but the converter is running. To replicate this condition in the proposed model, we have set the power supply to 0 V and we have enabled the boost converter. To compare model and real "Smart Latch" we have measured the voltage of signals used to drive the power release DC motor (pin A and pin B), current absorbed from release DC motor (motor current) and the voltage generate by the boost converter of Smart Latch (V PROT). Real system measures and model systems results are in Figs. 48.4 and 48.5 with blue lines and red lines, respectively.

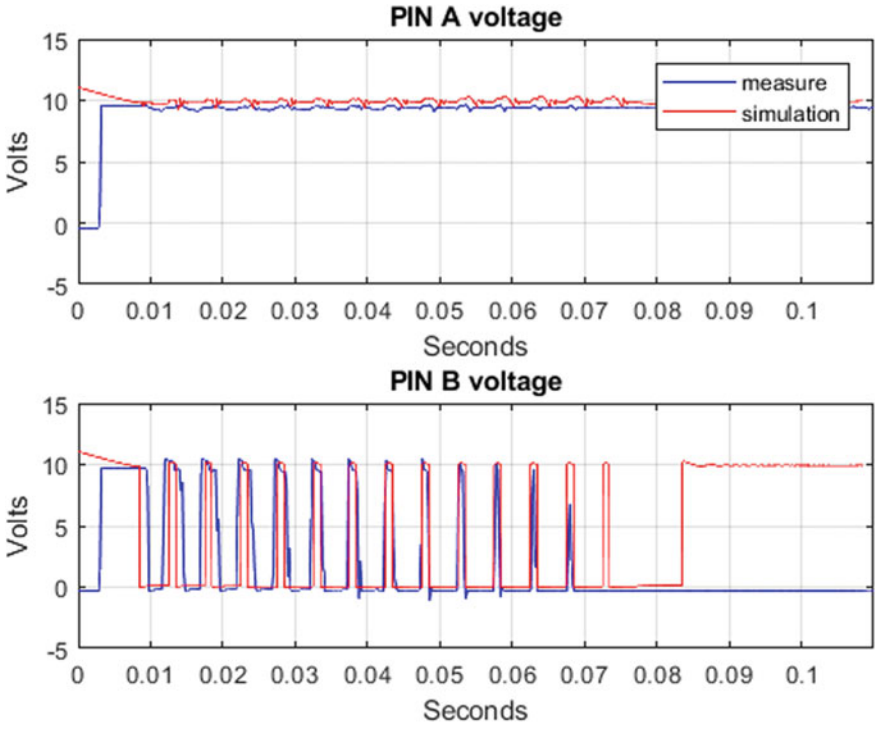


Fig. 48.4 Boost on inputs

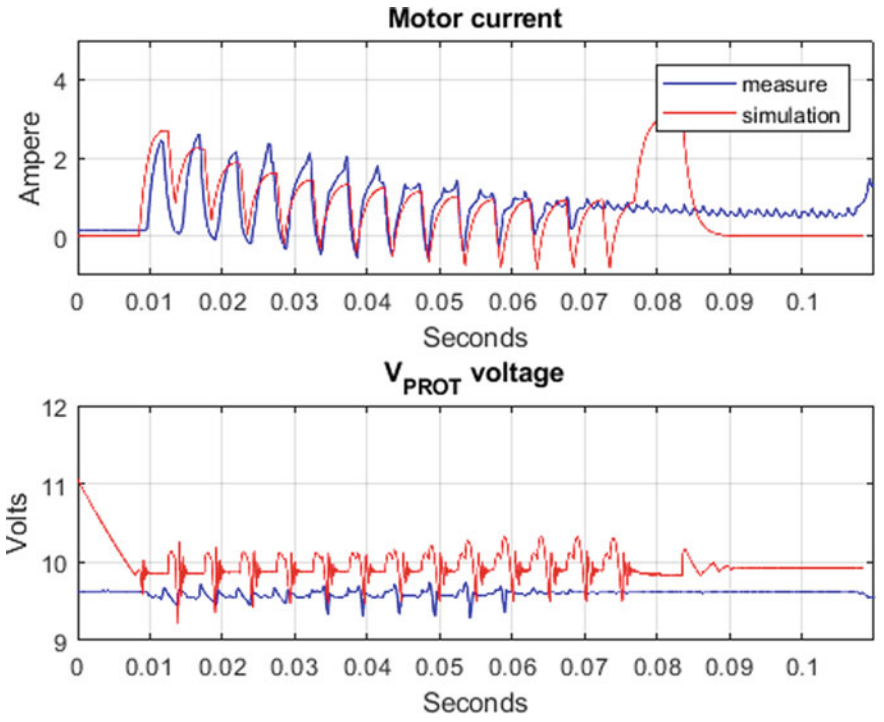


Fig. 48.5 Boost on outputs

48.5 Conclusion and Comparison to the State-of-Art

After the comparison between signals obtained with the proposed model and the real measures made on a Smart Latch, we can conclude that:

- Motor stalling time difference is probably due to the frictions present in the real mechanics.
- Graphics of current show that the signals extracted form simulations are accurate since the average error versus real measurements is limited and hence the model obtained is a good approximation of reality.
- The time taken to simulate 250 ms of a real SW-defined mechatronic component is about 7 h (CPU i7-4500U @ 1.80 GHz, 2.40 GHz and 4 GB ram), but we must consider that, with a single simulation, we obtain information regarding the behavior of all three subsystems.
- It is believed that the developed model is a starting point to obtain an even more accurate model to explore in detail also the integration of the real control software present on Smart Latch.
- For a qualitatively better study, we could add comparisons between measurements and simulations of other parts of the "Smart Latch".

– Simulation proofs validity of methodology to simulate a mechatronics system and shows that we can use this model to predict the behavior of real Smart Latch.

According to the achieve results, the proposed methodology can be a strategical choice for a company because, with the use of software like Simscape, Activate, and Saber, engineers can make only models that have a higher level of abstraction with respect to models made by software like OrCAD, Adams and State-flow, that instead work on a specific discipline and permit more complete and accurate designs and simulations. Moreover, for each discipline the engineers that make variations on a model can see directly the effect on the other subsystems, without affecting the work of other team. We can conclude that the methodology shown in this paper is the right compromise between flexibility and completeness of the models and allows to obtain reliable results that can be used by the company to design new products.

References

1. Saponara S, Bove A, Baronti F, Saletti R, Roncella R, Dente D, Leonardi E, Marlia M, Taviani C (2013) Thermal, electric and durability characterization of supercaps for energy back-up of automotive ECU. In: 2013 IEEE international symposium on industrial electronics, IEEE
2. Saponara S, Saletti R, Fanucci L, Roncella R, Marlia M, Taviani C (2014) Supercap-based energy back-up system for automotive electronic control units. Lecture notes in electrical engineering, vol. 289, pp 1–12
3. Saponara S (2016) An actuator control unit for safety-critical mechatronic applications with embedded energy storage backup. Energies 9(3)
4. Goodwin W et al (2008) Designing automotive subsystems using virtual manufacturing and distributed computing. SAE World Congress, Detroit, US
5. Gerket T et al (2008) Model based design of robust vehicle power networks. SAE World Congress, Detroit, US
6. Schiehlen W Multibody systems handbook. Springer, ISBN 978-3-642-50997-1

Chapter 49
Spice Model of Photovoltaic Panel for Electronic System Design

Mirco Muttillo, Tullio de Rubeis, Dario Ambrosini and Giuseppe Ferri

Abstract The aim of this work is to propose a Spice model of photovoltaic panel for electronic system design. The model is based on R_p-model of PV cell and implements the open-circuit voltage and short-circuit current variations from temperature and solar irradiation. The model was implemented on the LTSpice software characterized by comparing the System Advisor Model (SAM) software and MATLAB models with a commercial panel. The results of IV and PV curves are here reported.

49.1 Introduction

The simulation and the models of photovoltaic modules allow to characterize their behavior and to find the maximum power point with variations of solar irradiation and temperature. This kind of simulation is also important for the analysis and design of the electronic circuits that exploit them as a power source [1–7] or for Smart modules [8]. However, considering the recent development of research on energy harvesting systems, MPPT (maximum power point tracking) techniques are being integrated for the development of circuits with energy consumption of a few mW. In these cases the electronic interface circuit with the panel must be carefully designed to increase efficiency [9]. In general, the use of circuit simulators for commercial panels poses problems for photovoltaic generator models in creating voltamperometric (I-V) characteristic. In [10–16], several models of Orcad-Pspice or LTSpice are implemented.

In this paper a Spice model of photovoltaic panel for electronic system design was presented. The model, based on R_p-model of PV cell with five input parameters, implements the open-circuit voltage and short-circuit current variation based on solar irradiation and temperature. A commercial panel was chosen from SAM software

M. Muttillo (✉) · T. de Rubeis · D. Ambrosini · G. Ferri
Department of Industrial and Information Engineering and Economics, University of L'Aquila, L'Aquila, Italy
e-mail: mirco.muttillo@graduate.univaq.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_49

database and the results were compared with the I-V and P-V curves of SAM and MATLAB models.

49.2 Equivalent Circuit of PV Cells

The operation of a photovoltaic module can be represented by the equivalent circuit shown in Fig. 49.1, which model is the so-called single diode Rp-model of PV cell.

The solar cell behaves like a simple p-n junction diode in the absence of solar radiation. The solar cell behaviour is not very different from that of a diode, so it can be described by the following Shockley equation [17]:

$$I_D = I_0\left[e^{\frac{qV}{akT}} - 1\right] \tag{49.1}$$

In Eq. (49.1), I_D is the diode conduction current, a is the ideality factor of the diode and I_0 represents the saturation current. Furthermore, k is the Boltzmann constant (1.380653 × 10−23 J/K), q is the absolute constant value of electron charge (1.60217646 × 10−19 C) and T is the junction temperature (K).

The model of the real behaviour of the cell is described in Eq. (49.2), which includes the R_s series resistance and R_p parallel resistance (called shunt resistance). The first term represents the internal losses of the cell while the second one describes the leakage currents [16]. Equation (49.2) presents five parameters: I_{pv}, a, I_0, R_s and R_p.

$$I = I_{pv} - I_0\left[e^{\left(\frac{q}{akT}\right)(V+R_sI)} - 1\right] - \frac{V + R_sI}{R_p} \tag{49.2}$$

Photovoltaic panels have voltage and current variations that depend on temperature and solar irradiation. In the datasheets of the panels, two coefficients, K_i and K_v, sized in %/°C, consider these variations. The first parameter K_i represents the temperature coefficient of I_{SC} (short-circuit current) while K_v is the temperature

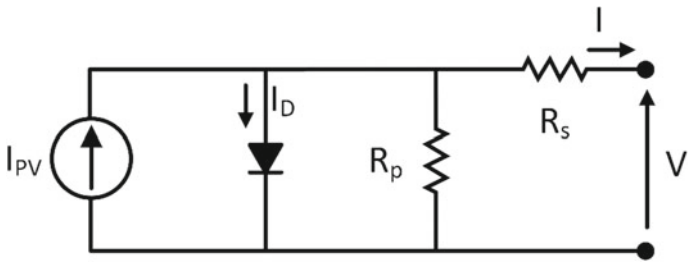


Fig. 49.1 Equivalent circuit of the single diode R_p-model of PV cell

coefficient of V_{OC} (open-circuit voltage). These coefficients are the percentage variation of open-circuit voltage and short-circuit current with respect to the ambient temperature (25 °C) and change according to the type of panel. These changes in temperature and solar irradiation are reported in Eqs. (49.3) and (49.4) that express V_{OC} and I_{SC} [17], respectively.

$$V_{OC} = V_{OC,STC}[1 + K_v(T - T_{STC})] + a\frac{kT}{q}\ln\frac{G}{G_{STC}} \tag{49.3}$$

$$I_{SC} = I_{SC,STC}\frac{G}{G_{STC}}[1 + K_i(T - T_{STC})] \tag{49.4}$$

In Eqs. (49.3) and (49.4) the open-circuit voltage $V_{OC,STC}$ and short-circuit current $I_{SC,STC}$ are the values reported by the manufacturer measured in standard test conditions (STC) with ambient temperature T_{STC} = 25 °C and solar irradiation G_{STC} = 1000 W/m^2.

49.3 PSpice Model

The model was implemented in the LTSpice software [18] using the scheme shown in Fig. 49.1 with a current generator, a diode and two resistors. The diode model must be scaled as shown in [19], due to variations in voltage and current with respect to temperature and solar irradiation. From Eq. (49.1) we can derive the new value of the ideality factor a of the solar cell using open-circuit condition with $V = V_{OC}$, $I_D = I_{SC}$ and T = 300 K [17]. The new value of a should be placed in the Spice model of the diode as parameter N. Nevertheless, it is not enough to scale the diode Spice model making it dependent on temperature and solar irradiation. Indeed, the only parameters for the level 1 diode Spice model that determine the variation of the current I_D with respect to the temperature are *XTI* (Saturation-current temperature exponent) and *EG* (Energy gap), in addition to the temperature Spice parameter T [20]. These parameters must be multiplied by N by a quantity equal to their default value. The following Eqs. (49.5) and (49.6) show the new parameters to be included in the diode model in LTSpice.

$$N = a = \frac{\frac{q}{kT}V_{OC}}{\ln\frac{I_{SC}}{I_0}} \tag{49.5}$$

$$EG = 1.11N, \quad XTI = 3N \tag{49.6}$$

Compared expressions [19], the values returned by Eqs. (49.3) and (49.4) can be considered for V_{OC} and I_{SC} in (49.5). Indeed, thanks to the use of these equations, it is possible to have a variation of the diode current dependent on temperature and solar irradiation.

Table 49.1 Parameters of the "Pythagoras Solar Midi PVGU Windows" panel present in SAM software in STC

Parameter	Value
P_{MAX}	20.286 W
V_{MAX}	16.1 V
I_{MAX}	1.3 A
V_{OC}	19.4 V
I_{SC}	1.4 A
K_v	−0.322%/°C
K_i	0.140%/°C
a	0.795311
I_0	3.37×10^{-11}
R_S	0.714915 Ω
R_P	633.18 Ω

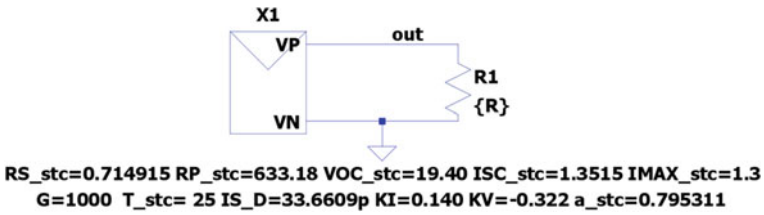


Fig. 49.2 LTSpice sub-circuit instance of the proposed model

To compare the results of the proposed Spice model, a panel present in the database of the SAM (System Advisor Model) software by NREL [21] is used. The panel chosen is named "Pythagoras Solar Midi PVGU Windows", a Mono-c-Si panel with parameters shown in Table 49.1.

Figure 49.2 shows the LTSpice sub-circuit instance of the proposed model. In this instance, it is possible to give accurate parameters for the simulation. For the simulation of the I-V characteristic, a variable resistance load was used.

49.4 Results and Discussion

The LTSpice simulation returns the I-V and P-V characteristics of the analyzed panel shown in Fig. 49.3. The simulation results have been compared with the SAM software data and a common MATLAB model used [22]. Simulations of all the models have been done in STC (1000 W/m^2, 25 °C).

Further results of the simulations are shown in Table 49.2. The relative error with respect to the SAM for the proposed Spice model is lower than that of the MATLAB

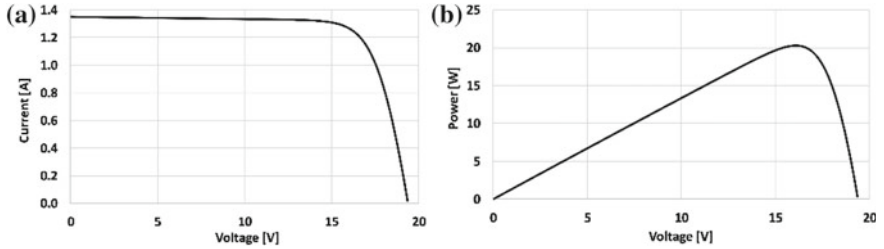


Fig. 49.3 Spice simulation results of the proposed Spice model: **a** IV characteristic; **b** PV characteristic

Table 49.2 The relative error of the proposed Spice model with respect to MATLAB model and SAM model

	P_{max} [W]	V_{oc} [V]	I_{sc} [A]	$P_{max,\varepsilon}$ [%]	$V_{oc,\varepsilon}$ [%]	$I_{sc,\varepsilon}$ [%]
SAM	20.286	19.400	1.350	–	–	–
SPICE	20.265	19.356	1.350	0.103	0.228	0.002
MATLAB	20.808	19.399	1.361	−2.571	0.008	−0.807

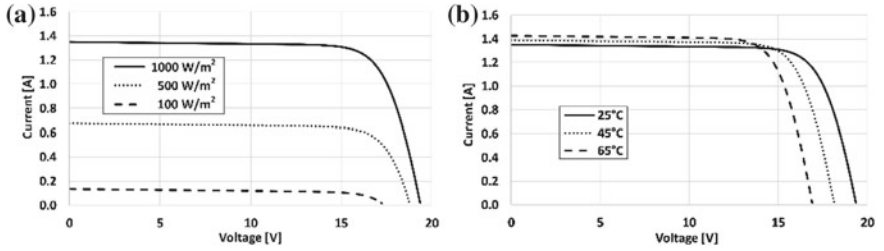


Fig. 49.4 Spice simulation results with a variation of the solar irradiance (**a**) and temperature (**b**)

model for maximum power and short-circuit current. The MATLAB model shows a lower relative error than the proposed Spice model for the open-circuit voltage.

Furthermore, Spice simulations carried out varying the solar irradiation and temperature, thanks to the use of Eqs. (49.3) and (49.4), are shown in Fig. 49.4. For the solar irradiation and temperature, the values of 1000, 500, 100 W/m^2 and 25, 45, 65 °C respectively were chosen.

49.5 Conclusion

In this work a Spice model of photovoltaic panel for electronic system design was presented. The model is based on R_p-model of PV cell with five input parameters. The model implements the equations for the variation of the open-circuit voltage

and short-circuit current as a function of irradiation and temperature. A 20 W panel was chosen and the results were compared with the I-V and P-V curves of the SAM and MATLAB models. The results show that the proposed model is better on some values than the MATLAB model. Future developments will be focused on the use of the model for the design of circuits that use solar panels as a power source.

References

1. Gabriele T, Pantoli L, Stornelli V, Chiulli D, Muttillo M (2015) Smart power management system for home appliances and wellness based on wireless sensors network and mobile technology. In: 2015 XVIII AISEM annual conference
2. Pantoli L, Barile G, Leoni A, Muttillo M, Stornelli V (2018) A novel electronic interface for micromachined si-based photomultipliers. Micromachines 9:507
3. Orsetti C, Muttillo M, Parente F, Pantoli L, Stornelli V, Ferri G (2016) Reliable and inexpensive solar irradiance measurement system design. Procedia Eng 168:1767–1770
4. Pantoli L, Muttillo M, Ferri G, Stornelli V, Alaggio R, Vettori D, Chinzari L, Chinzari F (2019) Electronic system for structural and environmental building monitoring. Lecture notes in electrical engineering, pp 481–488
5. Fusacchia P, Muttillo M, Leoni A, Pantoli L, Parente F, Stornelli V, Ferri G (2016) A low cost fully integrable in a standard CMOS technology portable system for the assessment of wind conditions. Procedia Eng 168:1024–1027
6. Leoni A, Stornelli V, Pantoli L (2018) A low-cost portable spherical directional anemometer for fixed points measurement. Sens Actuators, A 280:543–551
7. Smarra F, Jain A, de Rubeis T, Ambrosini D, D'Innocenzo A, Mangharam R (2018) Data-driven model predictive control using random forests for building energy optimization and climate control. Appl Energy 226:1252–1272
8. Baka M, Catthoor F, Soudris D (2016) Near-static shading exploration for smart photovoltaic module topologies based on snake-like configurations. ACM Trans Embed Comput Syst 15:1–21
9. Leoni A, Stornelli V, Ferri G, Errico V, Ricci M, Pallotti A, Saggio G (2018) A human body powered sensory glove system based on multisource energy harvester. In: 2018 14th conference on Ph.D. research in microelectronics and electronics (PRIME)
10. Afghan S, Almusawi H, Geza H (2017) Simulating the electrical characteristics of a photovoltaic cell based on a single-diode equivalent circuit model. MATEC Web Conf 126:03002
11. Devasia A, Kurinec S (2011) Teaching solar cell I-V characteristics using SPICE. Am J Phys 79:1232–1239
12. Jiang Y, Qahouq J, Orabi M (2011) Matlab/Pspice hybrid simulation modeling of solar PV cell/module. In: 2011 twenty-sixth annual ieee applied power electronics conference and exposition (APEC)
13. Gadjeva E, Hristov M (2015) Generalized SPICE model of photovoltaic modules. In: 2015 22nd international conference mixed design of integrated circuits & systems (MIXDES)
14. Abdulkadir M, Samosir A, Yatim A (2013) Modeling and simulation of a solar photovoltaic system, its dynamics and transient characteristics in LABVIEW. Int J Power Electr Drive Syst (IJPEDS) 3
15. Dondi D, Brunelli D, Benini L, Pavan P, Bertacchini A, Larcher L (2007) Photovoltaic cell modeling for solar energy powered sensor networks. In: 2007 2nd international workshop on advances in sensors and interface
16. Diaz-Bernabe J, Morales-Acevedo A (2015) Photovoltaic module simulator implemented in SPICE and Simulink. In: 2015 12th international conference on electrical engineering, computing science and automatic Control (CCE)

17. Chin V, Salam Z, Ishaque K (2015) Cell modelling and model parameters estimation techniques for photovoltaic simulator application: a review. Appl Energy 154:500–519
18. LTspice|Design Center|Analog Devices. https://www.analog.com/en/design-center/design-tools-and-calculators/ltspice-simulator.html
19. Intusoft Newsletter (2005). http://intusoft.com/nlhtm/nl78.htm#The_Solar_Cell_SPICE_Model. Available online on 6 Sept 2019
20. Diode Model (PN-Junction Diode Model). http://literature.cdn.keysight.com/litweb/pdf/ads2008/ccnld/ads2008/Diode_Model_(PN-Junction_Diode_Model).html
21. Home—System Advisor Model (SAM). https://sam.nrel.gov/
22. Implement PV array modules—Simulink—MathWorks Benelux. https://nl.mathworks.com/help/physmod/sps/powersys/ref/pvarray.html

Chapter 50
Exhaustive Modeling of Electric Vehicle Dynamics, Powertrain and Energy Storage/Conversion for Electrical Component Sizing and Diagnostic

Gaia Fiore, Lucian Mihet-Popa and Sergio Saponara

Abstract Electric Vehicles (EVs) will play a major role in meetings Europe's need for clean and efficient mobility. The development of new simulation tools, functionalities and methods integrated with the controlled development of a vehicle-centralized controller will also be part of the future solutions for the next generation of EVs. To improve the safety analysis and reduction costs, the solutions will be based on flexible user-friendly interfaces and specialized software tools. This work presents an exhaustive modeling of EV dynamics, powertrain and energy storage/conversion. The simulation model is useful for both electrical component sizing at designed time and on-board diagnostic to check component aging. The aim is to model the transient response of the system while preserving the simplicity and feasibility of simulation. The design of an EV requires, among others, the development and optimization of a complete electric powertrain system, including the longitudinal car, battery system components, power electronics, electric machine and control system. The paper presents the modelling and implementation of an entire powertrain system of EVs to describe the EV dynamics with respect to mechanical and electrical system components. Mathematical models based on equations and equivalent circuits are developed and implemented in MATLAB-Simulink and further study for predicting the final vehicle driving performance is performed.

Keywords Electric vehicle (EV) · Electronic systems for EV · Powertrain models

50.1 Introduction

Electrification of vehicle powertrain is one of the main trends in current vehicle development. EVs offer an increased efficiency (energy savings) through better fuel

G. Fiore (✉) · S. Saponara
Department of Information Engineering, University of Pisa, Pisa, Italy
e-mail: gaia.fiore91@gmail.com

L. Mihet-Popa
Faculty of Engineering, Oestfold University College, Halden, Norway

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_50

economy, reducing the emissions/pollution in the same time. Decarbonized electricity would become the dominant form of energy supply, posing challenges and opportunities for economic growth and climate policy. The use of renewable energy as a power source for EVs involves an important step in reducing greenhouse gas (GHG) emissions [1]. EVs to compete in a highly dynamic world market [2] first need to become comparable with conventional cars on several attributes, such as price, range and size. The main obstacles for future EV customers are well known: long charging times, short range and high purchase price [3]. Furthermore, the speed of public charging is often expected to be similar to conventional refueling. For this reason, research and political interest in public charging focus more and more on fast charging options with higher power rates. The EV system is an integration of many sub-components such as energy storage, power converters and electric motors. Each component must meet certain requirements. Specifically, an electric motor with high power and torque densities is desirable. EVs ranges from ultralight vehicles (e.g. Renault Twizzy) to electric race cars and minibuses with a few kW power up to hundreds of kW. As the heart of EVs, lithium-ion battery has gained priority due to its excellent characteristics [4] such as high cell voltage, high energy density, low self-discharge rate. A requirement is a correct dimensioning between vehicle performance (e.g. speed, acceleration, autonomy) and dimensioning of components (e.g. battery, power converter, electric machine). In this work we present the design of all the electric/electronic and control components of an electric vehicle, including energy storage (based on lithium-ion batteries), power conversion considering energy recovery and recharging capacity (DC/DC bi-directional converter), and the implementation with both 3-phase electric motors, e.g. AC Permanent-Magnet Synchronous Machine (PMSM), and DC motor, the former for propulsion and the latter for ancillary loads. For this purpose, we provide a MATLAB-Simulink complete model to simulate all the conversion, energy storage and driving model of an electric vehicle. The model is useful in the diagnostic phase as well as to validate the correct sizing of the electrical/electronic architecture. The model is parametric and can be scaled to different vehicle configurations, battery pack, motor, covering different scenarios. Hereafter, Sect. 50.2 presents the organization of the model divided into subsections and related controls. Section 50.3 provides a configuration of parameters, and examples of co-simulation between the electrical subsystem and the mechanical/dynamic performance of the vehicle. Section 50.4 shows how to change parameters to determine the correct configurations of electrical components, according to the characteristics of the vehicle, as well as possible use for diagnostics. Conclusions and discussions for future work are given in Sect. 50.5.

50.2 Materials and Methods

In this work a conventional architecture (Fig. 50.1) is chosen [5, 6].

The forces acting on a vehicle moving up a grade includes tire rolling resistance, aerodynamic drag, and uphill resistance. The traction force of a vehicle can be

described by Eq. (50.1), where F_t is the traction force, α is the angle of the driving surface, M is the mass of the vehicle, V is the velocity of the vehicle, a is the acceleration of the vehicle, g is the free fall acceleration, ρ is the air density of dry air, C_{rr} is the tire rolling resistance coefficient, C_d is the aerodynamic drag coefficient and A_f is the front area. Table 50.1 shows the specification of the vehicle dynamic model considered for simulation results in Sect. 50.3. The values in Table 50.1 refer to a light-duty vehicle (e.g. a 3-wheel electric scooter) but the model is parametric and can be applied to any vehicle. Indeed, for the simulations in Sect. 50.4 the values in Table 50.1 have been rescaled for a 450 kg electric vehicle, like the Renault Twizzy.

$$F_t = Ma + Mg\sin(\alpha) + Mg\cos(\alpha)C_{rr} + \frac{1}{2}\rho C_d A_f V \tag{50.1}$$

Due to its high-power density and high efficiency the PMSM motor-type is selected as propulsion system for the vehicle. The electric machine is divided into an electric part and a mechanical part. In the dq reference frame [7], the electrical part for the d-axis voltage V_d, the q-axis voltage V_q of the PMSM are expressed as:

$$V_d = R_s i_d + L_d \frac{di_d}{dt} - \omega_r L_q \tag{50.2}$$

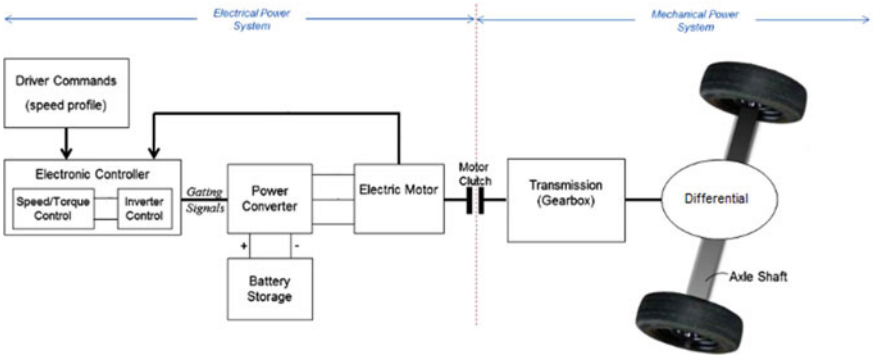


Fig. 50.1 Block diagram of a conventional powertrain

Table 50.1 EV dynamic parameter

Parameter	Value	Unit
Vehicle mass (M)	150	kg
Front area of the vehicle (A_f)	1.18	m^2
Wheel radius (r)	0.3	m
Coefficient of aerodynamic drag (C_d)	0.19	–
Air density (ρ)	1.2	kg/m^2
Rolling resistance coefficient (C_{rr})	0.0048	–

Table 50.2 PMSM parameters for two configurations, PMSM 1 and PMSM 2, used for simulations in Sects. 50.3 and 50.4, respectively

Parameter	PMSM 1	PMSM 2	Unit
Rated speed (ω_n)	4500	1500	rpm
Rated power (P_n)	9.4	15	kW
Number of poles (p)	6	3	–
Rated torque (T_n)	20	95.5	Nm
q axis inductance (L_q)	47.2×10^{-6}	0.05	H
d axis inductance (L_d)	28.7×10^{-6}	0.05	H
Flux linkage (λ_d)	9.71×10^{-3}	0.01	Wb
Stator winding resistance (R_s)	0.00962	3.3	Ω

$$V_q = R_s i_q + L_q \frac{di_q}{dt} + \omega_r L_d i_d + \omega_{rd} \tag{50.3}$$

where, R_s is the stator winding resistance, and ω_r is the electrical angular speed of the rotor, L_d and L_q denote the dq-axes inductance components, and λ_d is the flux linkage. The electromagnetic torque of the PMSM Tm and the mechanical dynamics are given by Eqs. (50.4) and (50.5), respectively

$$T_m = 3/4p\left[\lambda i_q + \left(L_d - L_q\right) i_d i_q\right] \tag{50.4}$$

$$J_m \frac{d\omega_m}{dx} = T_m - T_L - B_m \omega_m, \omega_m = 2/p\omega_r \tag{50.5}$$

where, ω_m is the mechanical angular speed of the rotor, J_m, B_m, and T_L are the moment of inertia, viscous friction coefficient, and load torque, respectively. The mechanism of regenerative braking is used in this EV [8]. The control strategy that suits best for the PMSM is Indirect Field Oriented Control (FOC). The torque produced by the PMSM is controlled indirectly, by monitoring the stator current i_s. The reference currents, i_{sd}* and i_{sq}*, are obtained from the Maximum Torque Per Ampere (MTPA) control strategies [9]. For the control algorithm Proportional-Integrator (PI) regulators are chosen. A generic model of Lithium-ion battery according to Shepherd's [10] model has been chosen from the MATLAB graphical editor Simulink and experimented in this work. The battery is chosen with maximum rated capacity of 20 Ah and a nominal voltage of 215 V. A MATLAB function is also included to adjust the correct operation of the battery in the right ranges. Powering the vehicle with batteries has the benefit of recapturing the braking energy loss along with zero emissions. Recapturing the braking energy requires bidirectional DC/DC converters [11]. The parameters of the Simulink model built for the PMSM drive are reported in Table 50.2. To demonstrate the model scalability two set of values for two different PMSM, i.e. PMSM 1 [12] and PMSM 2 [13], are reported and used in Sect. 50.3 and Sect. 50.4 respectively.

A dual Half-bridge topology with zero-voltage-switching (ZVS) in either direction of the power flow is implemented. The DC-DC bidirectional converter is utilized to convert the 215 V battery voltage to 700 V DC-link voltage and vice versa. In this model the magnitude and the direction of the power flow are controlled by the phase shift angle between low voltage side (LVS) and high voltage side (HVS) φ, setting the duty cycle at 50% and the $T_s = 5 \times 10^{-5}$ s. The control variable φ is given as output by the PI controller that take as input the difference between the reference and actual DC-link voltage. As a part of the control objective, the DC-link must be maintained at a constant value (this is a parametric value, set at 700 V for the simulations in Sect. 50.4). The EV has a large amount of electrical loads which should be supplied by the battery. These loads are either due to safety, e.g., light, wipers, horn and/or comfort, e.g., radio, heating, air conditioning. These loads are not constant and are supplied with low voltage (e.g. 5, 12 V). A DC buck converter is necessary to connect the high voltage bus with the DC car loads. Both the converter and the DC brush motor are provided by the Simscape library in MATLAB-simulink.

50.3 MATLAB Model of Electric Vehicle Powertrain

The EV was designed using library models of MATLAB-Simulink, and mathematical equation. The simulation model is composed by battery storage, buck converter, electronic controller as PWM block, and DC-brush motor designed with blocks provided by Simscape. In addition, the vehicle dynamic model, PMSM, bidirectional DC/DC converter and all the controller are realized with model equation based. The analyzed scenario presents the vehicle accelerating up to a steady speed along a slope followed by a period of descent during which electrical power is returned to the battery. In Fig. 50.2 the behavior of the EV subject to driver inputs and environmental conditions is reported. The vehicle accelerates until the driver maintains constant speed. As the driver applies the brakes, the vehicle slows down to zero speed. In Fig. 50.2b the torque produced by the PMSM motor in the EV is shown. During the first half of the simulation the motor accelerates the vehicle to the commanded speed and then continues to apply torque to push the vehicle up a hill. During the second half of the simulation, the motor works as a generator as shown by the change in sign of motor torque. The behavior of the battery in response to the vehicle is presented in Fig. 50.2c. The PMSM responds adequately both in motor and regenerative mode. The battery is discharged when the PMSM works as a motor and it is recharged when it works as a generator, while it does not undergo changes when the vehicle is stationary.

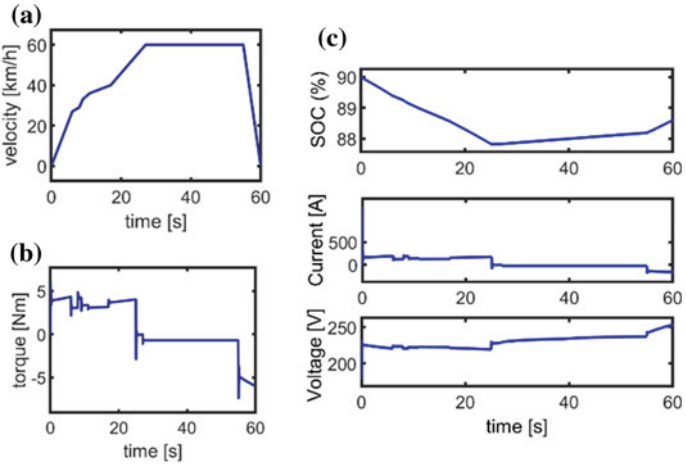


Fig. 50.2 **a** Drive cycle, the vehicle speed increases from 0 to 60 km/h and remains constant until the vehicle stops. The vehicle goes up a slope for the first half of time and then a descent for the remaining time. **b** Electromechanical torque. **c** State of charge, current and voltage of the lithium-battery subsystem

50.4 Model Application to Different Scenarios

The model described allows to test the vehicle in different scenarios and with different components. Each element of the model (motor, battery storage, dc-bus, converters, auxiliary load and so on) can be configured for specific scenarios, changing the parameters. Another test is shown to verify the validity of the model, to determine correct configurations of electrical components according to the characteristics of the vehicle, as well as possible use for diagnostics. The same scenario is used for the speed profile. The road angle α is set to zero. The ultralight vehicle Twizzy is considered in this simulation with a mass of 450 kg and a PMSM of 15 kW whose parameters are shown in Table 50.2. Figure 50.3a reports the electromechanical torque behavior. Figure 50.3b illustrates the traces for battery characteristics, the state of charge (SOC), after the vehicle drives for 60 s, is decreased by 1.5% of its initial value. The EV dynamics including tractive force, speed/acceleration in the vehicle can be simply monitored.

50.5 Conclusion

EVs contain different electrical, mechanical and electrochemical components in powertrain systems. These components are modelled mathematically and simulated on a MATLAB-Simulink. Every component modelled in the current paper is first validated standalone: energy storage system, power conversion considering energy recovery

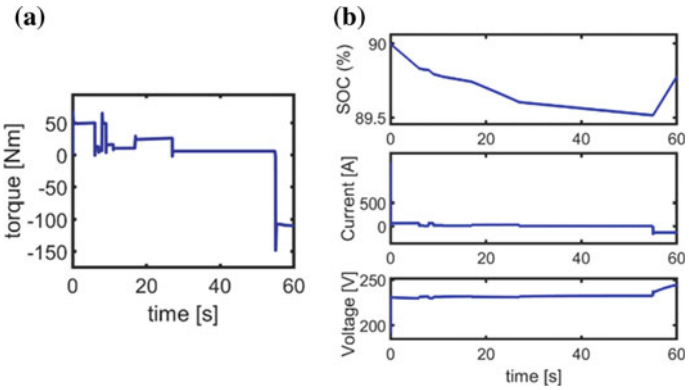


Fig. 50.3 **a** EV electromechanical torque, **b** State of charge, current and voltage of the battery

and recharging capacity, electric motors PMSM and DC motor, and the electric vehicle dynamics. The model is useful in the diagnostic phase as well as to validate the correct sizing of the electrical/electronic architecture. The model is parametric and can be scaled to different vehicle configurations, battery pack, motor, covering different scenarios. For this purpose, two different configurations have been considered with different PMSM motors and different light-duty vehicles. Our results suggest that the proposed model is a complete dynamic model for an electric vehicle powertrain, including all main subsystems.

References

1. Morgadinho L, Oliveira C, Martinho A (2015) A qualitative study about perceptions of European automotive sector's contribution to lower greenhouse gas emissions. J Clean Prod 106:644–653
2. van Vliet OPR, Kruithof T, Turkenburg WC, Faaij APC (2010) Techno-economic comparison of series hybrid, plug-in hybrid, fuel cell and regular cars. J Power Sour 195(19):6570–6585
3. Høyer KG (2008) The history of alternative fuels in transportation: the case of electric and hybrid cars. Util Policy 16(2):63–71
4. Saw LH, Ye Y, Tay AAO (2016) Integration issues of lithium-ion battery into electric vehicles battery pack. J Clean Prod 113:1032–1045
5. Park G, Lee S, Jin S, Kwak S (2014) Integrated modeling and analysis of dynamics for electric vehicle powertrains. Expert Syst Appl 41(5):2595–2607
6. Ehsani M, Gao Y, Gay SE, Emadi A (2005) Modern electric, hybrid electric, and fuel cell vehicles
7. Mondal D, Chakrabarti A, Sengupta A (2014) Power system small signal stability analysis and control
8. Adib A, Dhaouadi R (2017) Modeling and analysis of a regenerative braking system with a battery-supercapacitor energy storage. In: 2017 7th international conference on modeling, simulation, and applied optimization (ICMSAO), pp 1–6
9. Boldea I, Nasar SA (1900) ELECTRIC drives

10. Shepherd CM (1965) Design of primary and secondary cells II. An equation describing battery discharge. J Electrochem Soc 112(7):657–664
11. Mihet Popa L, Saponara S (2018) Toward green vehicles digitalization for the next generation of connected and electrified transport systems. Energies 11(11):3124
12. Captain C (2009) Torque control in field weakening mode. Aalborg University M.Sc. thesis, PED4-1038
13. Dini P, Saponara S (2019) Cogging torque reduction in brushless motors by a nonlinear control technique. Energies 12(11)

Part X
IoT and Integrated Circuits

Chapter 51
Analysis of 3-D MPPT for RF Harvesting

Michele Caselli and Andrea Boni

Abstract We discuss the issues arising in the design of RF harvesters for ultra low-power environments. The 3-D MPPT approach in [1] is the only one taking into account the presence of variable output load. Its architecture and performance are compared with other state-of-the-art MPPT implementations.

51.1 Introduction

Energy harvesting is a physical process aimed at collecting energy from the environment to power or recharge an accumulator whenever possible. This technique plays a fundamental role in the development and full exploitation of the Internet of Things (IoT) emerging world, that is characterized by a huge number of connected devices that must be fully autonomous. Among the possible energy sources, radiofrequency electromagnetic field (RF field) is an attractive option since it does not require any movement or friction nor a thermal gradient, and it is available in both indoor and outdoor environments [1]. On the other hand, the RF source is ambient dependent, uncontrollable, and unpredictable. The a priori study of the availability of the power spectral density (PSD), carried out in the final device location, would be the base for the design of a dedicated RF harvester. However, the huge number of devices potentially involved in IoT environments requires a more flexible approach. Therefore, several techniques for searching and tracking the point of maximum transferred power have been proposed in literature (MPPT techniques). In this paper, we discuss some of the aspects characterizing MPPT for RF harvesters. Moreover, the 3-D MPPT approach reported in [1] is analyzed and compared with other reported MPPT implementations.

M. Caselli (✉) · A. Boni
Department of Engineering and Architecture, University of Parma, Parma, Italy
e-mail: michele.caselli@unipr.it

A. Boni
e-mail: andrea.boni@unipr.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_51

51.2 RF Harvester: Model and Considerations

An RF harvester is made by few elements: an antenna used to collect RF energy available in the environment, modelled as a sinusoidal power source P_{AV} with a series resistance R_{ANT}, an RF rectifier circuit for the AC-DC conversion represented as a double bipole, and a storage device C_H with a load in parallel R_{LH}, Fig. 51.1.

Additionally, a DC-DC converter can be located at the output of the rectifier to control the power transfer in the storage element. L_M in Fig. 51.1 represents the element for the matching between the antenna and the input impedance of the rectifier. The R-L-C series circuit composed by the antenna radiation resistance R_{ANT}, L_M, and the harvester input impedance Z_{IN} has a major impact on the overall harvesting efficiency. At the resonance frequency $f_{LC} = 1/(2\pi\sqrt{L_M \cdot C_{IN}})$, this resonant circuit has a quality factor:

$$Q_{LC} = \frac{1}{[R_{ANT} + R_{IN}]} \cdot \frac{1}{2\pi f_{LC} C_{IN}} \tag{51.1}$$

where C_{IN} and R_{IN} are the input capacitance and resistance of the rectifier. Q_{LC} provides voltage amplification and is strictly related with the sensitivity S_{IN} of the harvester circuit. The non-linear behavior of the RF rectifier generates a lower bound for the peak input voltage leading to the minimum input power able to activate the system:

$$S_{IN} = \frac{(V_{ID})^2 \cdot R_{IN}}{2 \cdot Q_{LC}^2 \cdot [R_{ANT} + R_{IN}]^2} \tag{51.2}$$

where V_{ID} is the internal voltage drop due to the rectifying devices.

In case of harvesting in environments without any dedicated source or of far field energy scavenging, it is necessary to achieve a high Q factor to keep S_{IN} as low as possible, despite the reduction of the bandwidth of the resonant filter ($\Delta f = f_{LC}/Q_{LC}$). On the contrary, in case of dedicated RF sources in near field a large amount of energy can be normally collected and the constraint on the Q factor can be relaxed.

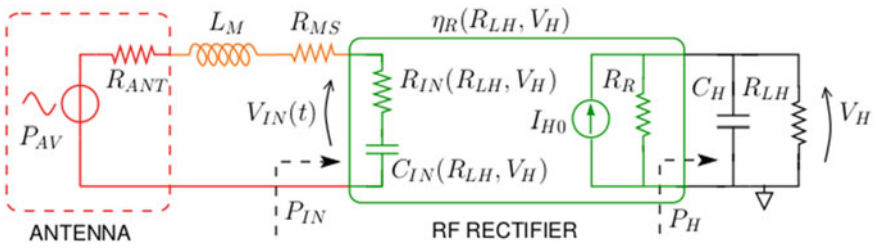


Fig. 51.1 Large signal model of an RF harvester circuit

The overall transfer efficiency η_{TOT} is given by the product of the cascade of the efficiencies of the circuits involved in the harvesting process. In this perspective, the efficiency of every stage has to be maximized for the best energy collection.

The rectification efficiency η_R largely depends on the rectifier architecture that can realize both voltage conversion and multiplication. Fully passive rectifier architectures, such as those adopted in [1–3], do not require any energy from the battery, at the cost of power losses either on the rectifying devices or to apply techniques of self-polarization. These circuits are those most suitable for energy scavenging where RF available energy is limited and the battery must be preserved. Active architectures for harvesters with dedicated sources and large energy availability have also been proposed to maximize η_R at the cost of power from the battery [4].

51.2.1 Input Impedance Variation

Achieving a high Q-factor is fundamental in RF energy harvesting, since it minimizes the input sensitivity S_{IN}. On the other hand, high values of Q_{LC} reduce the filter bandwidth B_{LC}. This can have detrimental effects in RF harvesting applications where the power distribution in band is not known a priori, cutting out frequencies that could be potentially exploited. To avoid this trade-off, a variation of the input capacitance can be applied to obtain high quality factors and exploit the whole band [5]. A bank of capacitors or of varactors, giving an additional capacitance C_{TUN}, can be added in parallel with the input terminals of the rectifier to vary the input reactance and obtain a shift of the resonance frequency at $f_{LC1} = 1/(2\pi\sqrt{L_M \cdot (C_{IN} + C_{TUN})}$. A large increase of the operative band can be obtained with a limited cost in terms of occupied area and a slight decrease of Q_{LC} by increasing C_{TUN}.

According to Eqs. (51.1) and (51.2), an increase of the series inductance can shift up f_{LC}, thus yielding high Q_{LC}. Despite these remarkable features, this option should be exerted with care for general purpose and low cost RF harvesters, due to the complexity and cost (in terms of area and implementation option) of the integration of high Q inductors in the standard CMOS process [6]. Finally, an increase of Q_{LC} can be obtained by means of the reduction of antenna resistance R_{ANT}. However, decreasing the value of R_{ANT} below 10 Ω makes the antenna design quite challenging.

51.2.2 MPPT Power Transfer Verification and Algorithm

As discussed in the previous section, a specific parameter can be varied to perturb the steady state (e.g. C_{IN}) and increase the overall system efficiency implementing a Hill Climbing approach. To perform the MPPT it is necessary to choose at least one monitor parameter providing the feedback signal for the control system and closing the control loop. The systems in [5, 7], for example, vary the input capacitance by means of a bank of capacitor, whereas [2, 3, 8] modify their rectifier structures still

operating on C_{IN}. All these circuits monitor the output voltage for a specific value of the output resistance. The maximum power of the transferred power is obtained maximizing V_H and therefore $P_H = V_H^2/R_{LH}$. For this purpose, V_H is stored on a capacitor and then compared with the V_H of the previous algorithm step.

Alternative approaches are proposed in [9, 10] where the rectifier in coupled with a DC-DC converter. The former aims at maximizing the input current of the DC-DC converter considering constant its output voltage and acting on the switching frequency f_S of the converter. The delivered power is hence calculated as $P_H = V_H \bullet I_H$. Despite the originality, this system does not consider at all the effect of the output load on any parameter of the rectifier and it does not seem effective for ultra low-power RF harvesters. Martins and Serdijn [10] still operates on f_S but considering the effects of I_H in particular on the R_{IN} of the rectifier and it maximizes the converted power maximizing V_H. In all the mentioned methods, only one parameter is considered, thereby simplifying the control strategy and speeding up the MPPT algorithm. Differently from these approaches, the harvester in [1] computes the maximum power considering both the output voltage and the output load of the rectifier.

51.3 3-D MPPT for a Real Maximum Power Transfer

The system in [1] adopts an alternative approach to MPPT for ultra low-power RF harvesters and it is designed for adaptation to mutable RF environments in space and time. The search space of the point of maximum transferred power is three-dimensional since it considers the input capacitance C_{IN} for the shifting of f_{LC}, the output voltage V_H, and the output resistance R_{LH}. By means of the last two parameters, the real maximum power for RF harvesters with non-constant output load can be computed.

51.3.1 RF Harvester Model

In [1] a Full-Wave Mirror Stacked rectifier with threshold voltage cancellation has been chosen for the minimum value of the input capacitance, leading to high Q-factors and very low sensitivity S_{IN} [11]. A mathematical model of the rectifier developed in MATLAB and post-layout simulations with back-annotated parasitic extraction demonstrate that the point of maximum transferred power moves in a 3-D space, function of the additional capacitance C_{TUN}, V_H, and R_{LH}. Figure 51.2 shows the maximum values of P_H (R_{LH}) on V_H and P_{AV}, with the best C_{TUN}, obtained from a MATLAB simulation that includes the model of the rectifier and the computation of the reflection coefficient Γ_L. In the algorithm C_{TUN}, V_H, R_{LH}, and P_{AV} are varied to find the best matching capacitance value C_{TUN} and the maximum P_H for given f_S, R_{ANT}, and L_M. It is worth to notice that the point of maximum power transfer is not always at the maximum V_H.

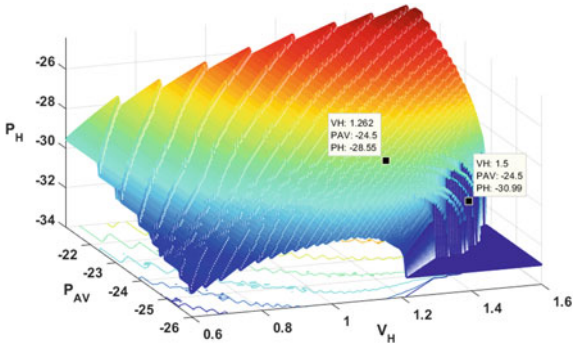


Fig. 51.2 Maximum P_H (R_{LH}) versus P_{AV} and V_H. f_S = 950 MHz, L_M = 90 nH, R_{ANT} = 12 Ω

The limitation of the maximum V_H is a matter of fact in real harvesters and the result shown in Fig. 51.2 suggests to split the graph in two sections: one for high P_{AV} and one for low P_{AV}. If the V_H and the operative P_{AV} ranges are defined a priori, for high value of RF available power the maximum delivered power is located at the maximum V_H, whereas for low values of available power, where the MPPT is most necessary, in order to find the maximum P_H also R_{LH} must be evaluated.

51.3.2 3-D MPPT Circuit Design

Based on the theoretical results, an RF harvester including an MPPT capable to sweep the 3-D space has been designed in ST 65 nm CMOS technology. The system (Fig. 51.3, left) is composed by the modelled rectifier, a bank of capacitors, a power meter with a variable load, and a finite state machine (FSM). The bank of binary scaled capacitors is used to sweep the resonance frequency. The power meter (Fig. 51.3, right) is connected at the output of the rectifier. The circuit embeds a programmable load resistance, replacing R_{LH}, made by binary scaled resistors and bypass switches driven by the n_R-bit word B_R. A shifted down version of the rectifier output voltage

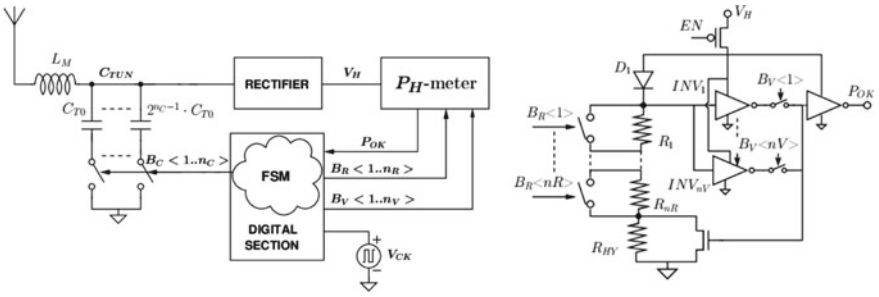


Fig. 51.3 Left: schematic of the MPPT system for RF harvesting. Right: power meter with variable load and voltage threshold [1]

V_H is compared with the threshold voltages V_{TH_N} of a bank of inverters with different aspect ratios. Only one inverter at the time is selected for the comparison by means of the n_V-bit word B_V.

The dedicated FSM advances the MPPT algorithm by means of the feedback signal P_{OK}, obtained from the comparison of V_{TH_N} and V_H. The value of the power delivered by the rectifier is computed by means of a look-up table. The correct computation of the delivered output power assumes the linearity of the threshold voltage steps. The FSM actuates a *perturb and observe* algorithm operating on the input capacitance and monitoring the different configurations of the V_H–R_{LH} pair. The evaluation starts from the upper limit of the chosen band with $C_{T0} = 0$.

The load resistance is progressively decreased sweeping B_R until V_H falls below V_{TH} and P_{OK} goes to zero. Here the resonance frequency is shifted down by means of the bits n_C until V_H rises again over V_{TH}, P_{OK} goes to logic one, and the sweep of the load resistance resumes. Finally, if the maximum computable power has not been achieved, V_{TH} is modified and the nested loops are repeated.

51.4 Comparison with State-of-the-Art and Discussion

Simulation results [1] demonstrate the capability of the proposed system to deal with multiple tones at different frequencies and to choose the correct point of maximum delivered power at given quantization levels for the controlling bit words B_R, B_C, B_V.

From the comparison of the designed system with several state-of-the-art RF harvesters equipped with MPPT reported in Table 51.1, some remarks can be offered. Harvester implementations for both far field without any dedicated RF source [1–3, 7, 10] and near field with dedicated sources operating on a specific frequency [5] can exploit the impedance matching variation to enlarge the system bandwidth, although the latter category obtains a limited improvement of performance. The 3-D MPPT system in [1] is the only one capable to compute the delivered power taking into account both V_H and R_{LH}. The MPPT implementations in [2, 3, 5, 7, 8]

Table 51.1 Comparison of the state-of-the-art RF harvesters including MPPT

	[1]	[2]	[3]	[5]	[7]	[10]
Tech. node (nm)	65	180	90	180	130	180
Monitor param.	V_H-I_{LOAD}	V_H	V_H	V_H	V_H	V_H
Sweep. param.	C_{IN}-V_H-R_{LH}	C_{IN}	C_{IN}	C_{IN}	C_{IN}	C_{IN}
f_{LC} [MHz]	760–960	2400	868	13.56	860–960	405
S_{IN} [dBm]	−30	−22	−26.3	NA	−20	−27
I_{CC}	<50 nA	NA	<30 nA	<400 nA	<2.3 μA	<20 nA
Sim/Meas	Sim	Sim	Meas	Sim	Sim	Sim

search the maximum power point monitoring only the output voltage of the rectifier V_H. However, this approach is accurate only for RF harvesters with constant and defined R_{LH}. The information about the best output configuration (V_H-R_{LH}) can be particularly useful for RF harvesters cascading an input control DC-DC converter at the output of the rectifier [12, 13]. The control strategy of the DC-DC converter can be tuned to operate at the desired average V_H, whereas the average load resistance of the rectifier is defined by the incoming power P_{IN} [14]. Most approaches in literature propose the periodic verification of the point of maximum transferred power due to the variability of the RF field. In order to limit power consumption, the control section is normally powered down at the end of the algorithm and turned again on with a small duty cycle. Little information is provided in literature about the power consumption of the MPPT systems even if this is crucial given the low RF power possibly available [14]. Currently, the best reported values are in the order of few tens of nanoamperes [1, 3, 10] thanks to the low MPPT duty cycle and the minimal architecture. Moreover, to cope with low power environments high sensitivity rectifier architectures are required [10, 11].

References

1. Caselli M, Boni A (2019) 3-D Maximum power point searching and tracking for ultra low power RF energy harvesters. In: IEEE SMACD
2. Zeng Z et al (2016) A WLAN 2.4-GHz RF energy harvesting system with reconfigurable rectifier for wireless sensor network. In: IEEE ISCAS
3. Stoopman M et al (2013) Self-calibrating RF energy harvester generating 1 V at −26.3 dBm. In: IEEE symposium on VLSI
4. Wang SH et al (2018) The design of CMOS 13.56 MHz high efficiency 1x/3x 1.99 V/6.29 V active rectifier for implantable neuromodulation systems. In: IEEE ISCAS
5. Gosselin A et al (2017) A CMOS automatic tuning system to maximize remote powering efficiency. In: IEEE ISCAS
6. Abouzied MA et al (2017) A fully integrated reconfigurable self-startup RF energy-harvesting system with storage capability. IEEE J Solid-State Circ 52(3)
7. Bakhtiar AS et al (2010) An RF power harvesting system with input-tuning for long-range RFID tags. In: IEEE ISCAS
8. Xia L et al (2014) 0.56 V, −20 dBm RF-powered, multi-node wireless body area network system-on-a-chip with harvesting-efficiency tracking loop. IEEE J Solid-State Circ 49(6)
9. Hua X, Harjani R (2018) A 5 μW–5mW input power range, 0–3.5 V output voltage range RF energy harvester with power-estimator-enhanced MPPT controller. In: 2018 IEEE custom integrated circuits conference (CICC)
10. Martins GC, Serdijn WA (2018) An RF energy harvester with MPPT operating across a wide range of available input power. In: 2018 IEEE international symposium on circuits and systems (ISCAS)
11. Nakamoto H et al A passive UHF RF identification CMOS tag IC using ferroelectric RAM in 0.35–μm technology. IEEE J Solid-State Circ 42(1)
12. Wang J et al (2017) 900 MHz RF energy harvesting system in 40 nm CMOS technology with efficiency peaking at 47% and higher than 30% over a 22 dB wide input power range. In: ESSCIRC 2017-43rd IEEE European solid state circuits conference

13. AEM40940 (2018) e-Peas Sem. https://e-peas.com/products/energy-harvesting/rf/aem40940/
14. Caselli M et al (2019) Design and analysis of an integrated RF energy harvester for ultra low-power environments. Int J Circ Theor Appl 47(7)

Chapter 52
Analysis and Simulation of a PLL Architecture Towards a Fully Integrated 65 nm Solution for the New Spacefibre Standard

Marco Mestice, Bruno Neri and Sergio Saponara

Abstract This paper presents the modeling and design activity of a PLL (Phase-Locked Loop) architecture to generate the clock reference for the new ESA Spacefibre standard for on-board satellite communications up to 6.25 Gbps. Starting from a 6.25 GHz VCO rad-hard design, integrated in 65 nm technology within an IMEC-University of Pisa collaboration, this work presents a PLL architecture including configurable integer divider, down to a reference signal of 156.25 MHz, phase-frequency detector, charge pump and passive loop filter. Modeling and simulation analysis, carried out in Keysight ADS environment, show that a fully integrated solution can be achieved with a 6 MHz low-pass PLL loop filter whose passive devices can be integrated on chip with an area of about 4600 μm^2. The PLL phase noise performance are in line with that of the original VCO, and for the stability a gain and phase margins of 86 dB and 50° are achieved. PLL lock time is about 555 ns. A preliminary circuit for the charge pump implementation is also proposed.

Keywords PLL (Phase-Locked Loop) · SpaceFibre communications · VCO (Voltage controlled Oscillator) · Loop-filter · Charge-pump · Space electronics

52.1 Introduction

The new Spacefibre standard for on-board satellite communication up to 6.25 Gbps has been recently released by ESA [1]. A key block for its implementation is the clock reference generator, which should be tolerant to SEE (Single event effects) and TID (Total ionization dose), and able to sustain up to 6.25 GHz, as well as its

M. Mestice (✉) · B. Neri · S. Saponara
Department of Information Engineering, University of Pisa, Pisa, Italy
e-mail: marco@mestice.it

B. Neri
e-mail: bruno.neri@iet.unipi.it

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

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_52

divided frequencies by 2 and by 4. In aerospace applications, the TID level to be sustained is 300 krad. Similar needs for an integrated clock reference characterize High Energy Physics (HEP) experiments at CERN, where the upgrade of LHC will exploit modules in 65 nm such as the RD53 [2] sensor front-end. RD53 targets 1.2 Gbps/module and some hundreds of Mrad as TID. From a temperature point-of-view the aerospace application field requires supporting −40 to 150 °C, while in HEP applications the temperature is not critical.

Reaching 6.25 GHz in worst case PVT (process-voltage-temperature) and rad-hard conditions means working at even higher frequencies in normal conditions. Due to ITAR-free issues, a European design solution is needed, available as hard-macro. Unfortunately, state of art of rad hard designs made-in-Europe is limited to PLL below 6 GHz in nominal condition (even lower in worst case). For example, only 2.56 GHz are achieved on [3, 4]. A rad-hard LC-tank VCO design, integrated in 65 nm technology through IMEC, is part of a design collaboration between University of Pisa and IMEC in [5]. This work aims at modeling and simulating the whole circuitry around the PLL, including configurable integer divider, phase-frequency detector (PFD), charge pump (CP) and passive loop filter. As rad-hard quartz-based reference oscillator a 150 MHz solution has been selected. Modeling and simulation analysis were carried out in both Simulink and Keysight ADS environments, achieving similar results.

Hereafter, Sect. 52.2 presents the PLL architecture and a preliminary circuit schematic design in cadence of the charge pump and passive loop filter, Sect. 52.3 shows the ADS model and results and the Simulink model. Section 52.4 shows the achieved PLL architecture sizing results. Conclusions are drawn in Sect. 52.5.

52.2 PLL Architecture and Charge-Pump Preliminary Schematic Design

52.2.1 PLL Architecture

The PLL's architecture is shown in Fig. 52.1. It is composed of a PFD (Phase Frequency Detector), that provides two digital signals (UP and DOWN) depending on the phase and frequency difference between the input signals, a CP (Charge Pump), that converts UP and DOWN signals in a current (positive or negative), a passive Loop Filter, a VCO, that generates the output signal of the PLL, and an integer N frequency divider that provides one of the inputs of the PFD.

The main target frequency that have been chosen for this work (6.25 GHz) is obtained from a reference frequency of 156.25 MHz thanks to an integer divider with $N = 40$. The other frequencies (3.125 and 1.5625 GHz) could be obtained by divisions by 2 and by 4 of the main frequency without adding other components to the architecture.

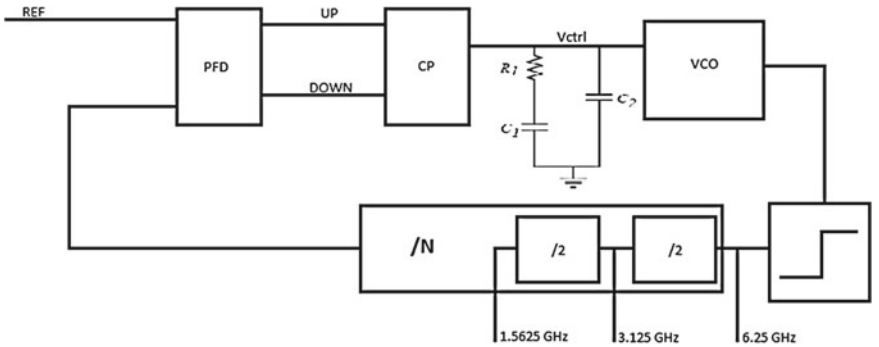


Fig. 52.1 PLL architecture

The VCO has a tuning range of 0–1.2 V and the supply voltage for the PLL is 1.2 V and presents a phase noise <−100 dBc/Hz at 1 MHz. The target phase noise for this work has been chosen to be <−80 dBc/Hz at 1 MHz, in line with that of the VCO and better than what required in the Spacefibre standard.

The target architecture of the PFD, shown in Fig. 52.2, consists of two D-FF with a logic 1 at the inputs. The edges of the reference clock and of the divided clock from the VCO force the signal UP and DOWN respectively to a logic 1. When both UP and DOWN are active, the internal feedback chain resets the D-FF, forcing the two signals to a logic 0. The delay of the reset chain has to be carefully chosen to avoid the dead zone problem.

The loop filter is the component that determines the bandwidth and stability of the PLL. It consists of two capacitors (C1 and C2) and one resistor (R1), see Fig. 52.1. In its simplest form with only one capacitor (C1), it would bring instability and, for this reason, a resistor (R1) is added. The second capacitor (C2) is added to reduce spurious tones due to the current mismatch, caused by the not ideal charge pump, and has to be at most C1/5. C1 and R1 are chosen to reach a loop bandwidth of 6 MHz. This bandwidth provides a good tradeoff between low-noise performance and integrability of the filter. A complete integrated filter is preferred since all the

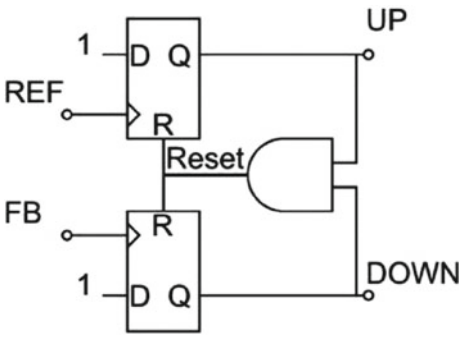


Fig. 52.2 Target PFD architecture

problems deriving from exiting the chip are avoided. Actually, higher loop bandwidth provides higher filtering of VCO's and loop filter's noise with smaller capacitors, while lower bandwidth provides higher filtering of charge pump's and reference's noise, but larger capacitors. Furthermore, higher bandwidth provides faster lock time. For these reasons, an I_{cp} of 40 μA has been chosen and, given the selected loop filter bandwidth of 6 MHz and the need of having both resistor and capacitors integrable on-chip, the following values for the passive devices have been defined: 8 pF for C1, 12 kΩ for R1, 1 pF for C2.

52.2.2 *Charge-Pump Preliminary Schematic Design in Cadence*

A preliminary charge pump's schematic design has been done in Cadence for two different architecture shown in Figs. 52.3 and 52.4.

The first one in Fig. 52.3 is a simpler architecture but presents some disadvantages: first, it behaves as a current source with low output impedance; second, the switches are directly connected to the output node, influencing it with charge injection and

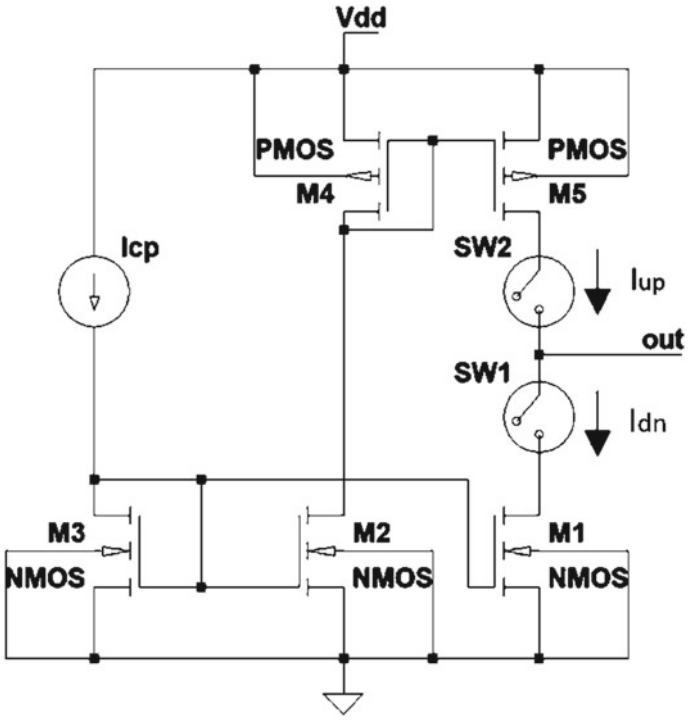


Fig. 52.3 Charge pump first architecture

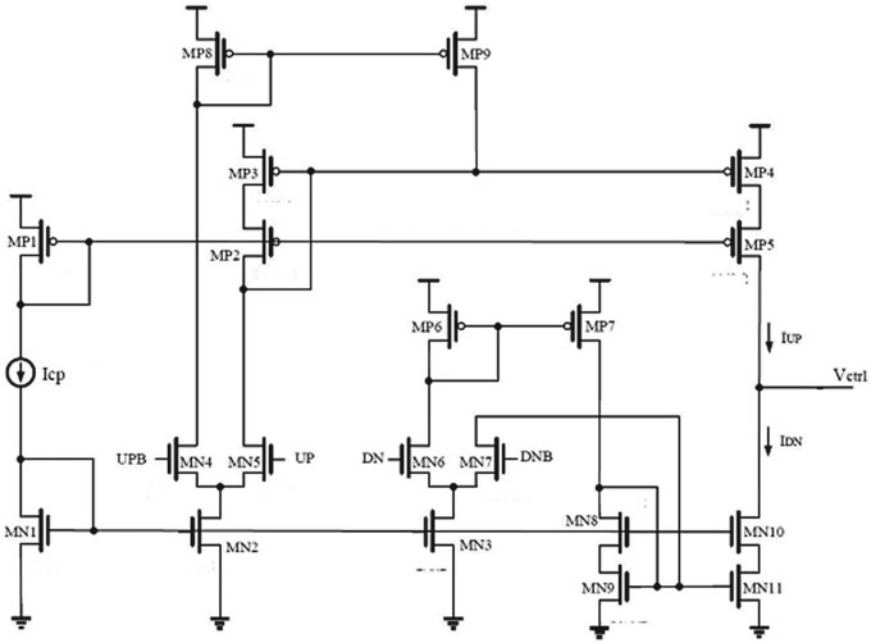


Fig. 52.4 Charge pump second architecture

clock feed through effects; third, M1 and M5 spends some time in linear region when SW1 and SW2 are enabled [6].

The second charge-pump circuit in Fig. 52.4 uses as UP/DOWN signals a differential pair (UP, UPB and DN, DNB). This charge-pump circuit when compared to the one in Fig. 52.3 shows higher output impedance, less effects on the output node due to switching activity but it is a more complex architecture. This circuit solution has been derived from [7]. In this work, with respect to [7] different bias signals to MN2/MN3 and MN8/MN10 have been provided.

The first charge-pump architecture has been dimensioned with no minimum length for all mirror's transistors to enhance the output impedance, while the switches are low Vt transistors with minimum length and quite large width to reduce the voltage drop on them. To evaluate the output impedance of the circuits in Figs. 52.3 and 52.4 voltage sources have been applied at the output of the circuits and the relevant Iup, Idn currents have been measured. The results are shown in Fig. 52.5 for the charge-pump of Fig. 52.3 and in Fig. 52.6 for the charge-pump of Fig. 52.4. In Fig. 52.5 on the left the UP (red) and DOWN (black) currents are shown as a function of the output voltage, while in Fig. 52.5 on the right their difference (black) and the derivative of the difference (red) are shown. As expected this solution presents a quite low output impedance. The second architecture has been sized with no minimum length for mirror's transistors, while the differential pairs are quite small. Results are shown

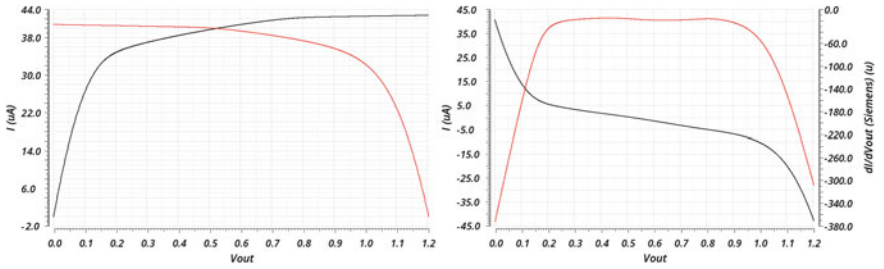


Fig. 52.5 Currents as function of Vout of the charge-pump architecture of Fig. 52.3

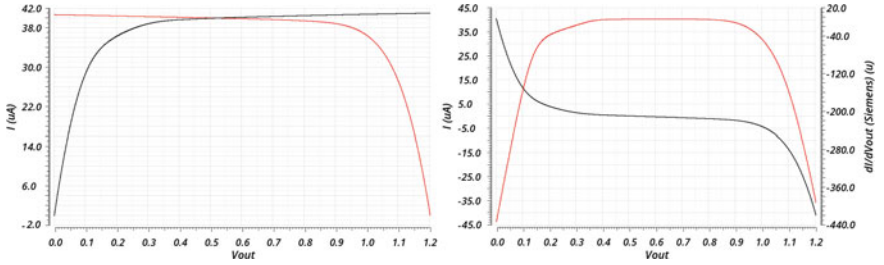


Fig. 52.6 Currents as function of Vout of the charge-pump architecture of Fig. 52.4

in Fig. 52.6. It shows an higher output impedance for the charge-pump circuit of Fig. 52.4 versus that of Fig. 52.3, with nearly the same output range.

52.3 PLL Modeling and Simulation Results

52.3.1 PLL Modeling and Simulation Results in ADS

Firstly, the PLL has been modeled in phase domain to simulate and analyze the behavior in terms of stability and bandwidth. Closed and open loop PLL models are shown in Fig. 52.7. The PFD plus charge-pump and the divider blocks are linearized models with constant gains of $I_{cp}/2\pi$ and 1/N, while the VCO behaves like an integrator. The total transfer functions are those in Eq. (52.1) in open loop and Eq. (52.2) in closed loop:

$$H_{ol} = \frac{I_{cp}}{2\pi} Z(s) \frac{K_{vco}}{s} \frac{1}{N} \tag{52.1}$$

$$H_{cl} = \frac{\frac{I_{cp}}{2\pi} Z(s) \frac{K_{vco}}{s}}{1 + \frac{I_{cp}}{2\pi} Z(s) \frac{K_{vco}}{s} \frac{1}{N}} \tag{52.2}$$

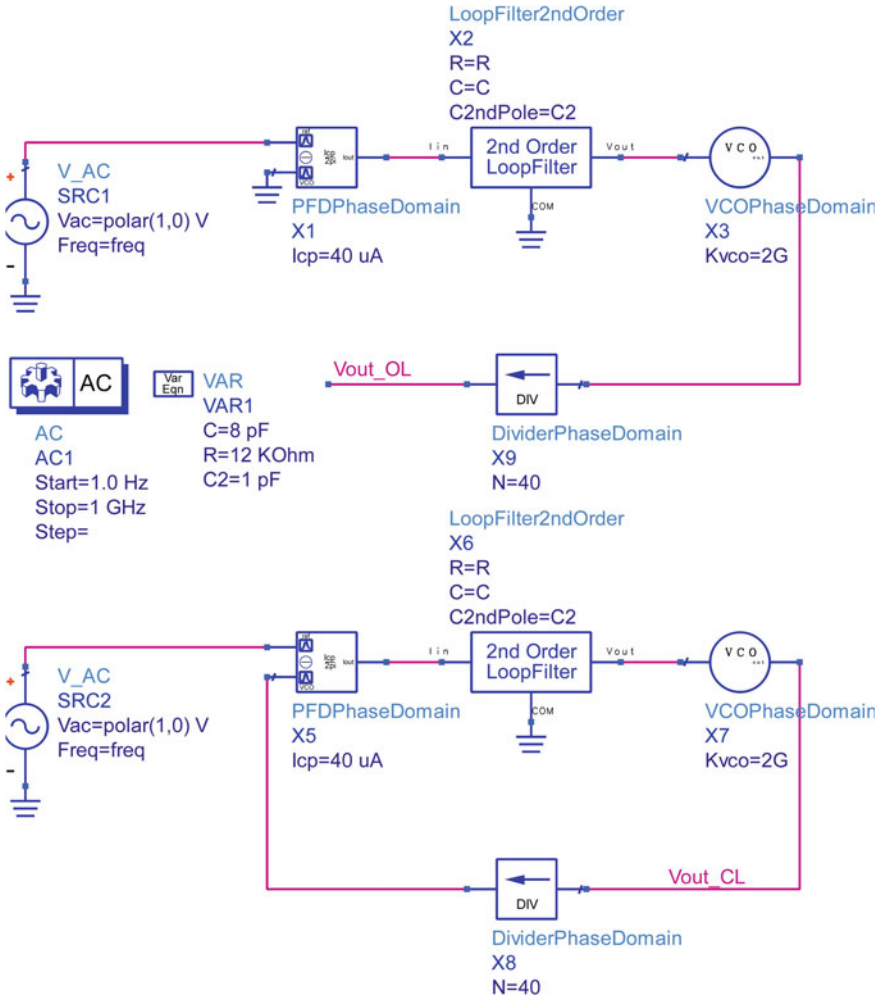


Fig. 52.7 Open and closed loop models, phase domain

$$\begin{aligned} &\mathrm{I_{cp}} = 40\ \mu\mathrm{A}, \quad \mathrm{N} = 40, \quad \mathrm{K_{vco}} = 12.57 \times 10^9\ \mathrm{rad}/(V * s), \\ &Z(s) = \frac{1}{s(C1 + C2)} \frac{1 + sR1C1}{1 + sR1\frac{C1C2}{C1+C2}} \end{aligned} \tag{52.3}$$

An AC simulation has been done and the results are shown in Figs. 52.8 and 52.9 for the open and closed loop transfer functions. In Fig. 52.10 the step response is shown for an input phase step of 1°. The zero introduced by R1 in the loop filter of Fig. 52.1 stabilizes the loop, while C2 tends to reduce the stability and, for this

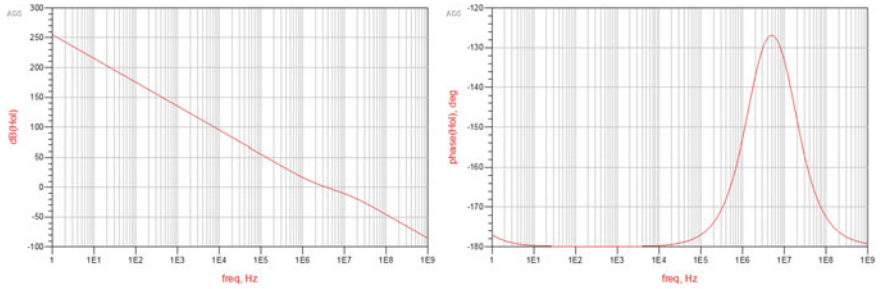


Fig. 52.8 Open loop frequency response

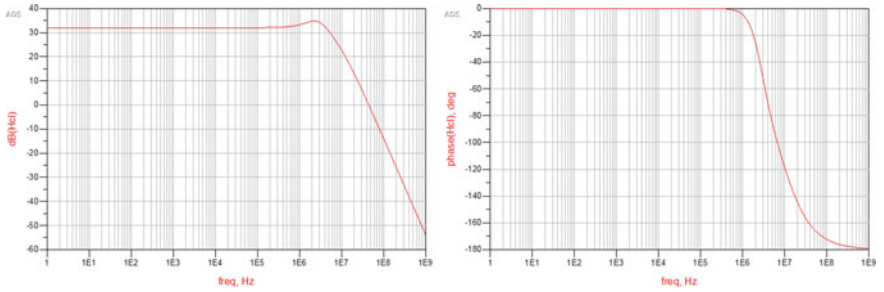
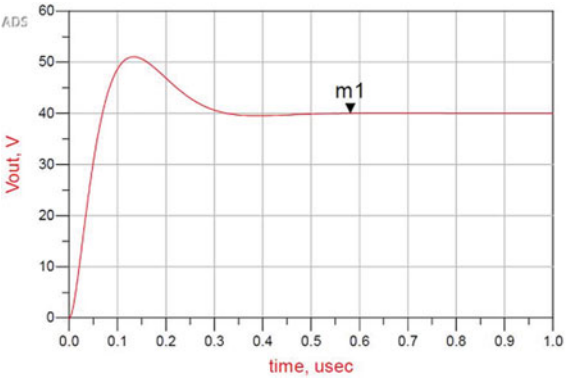


Fig. 52.9 Closed loop frequency response

Fig. 52.10 Step response



reason, its value is chosen to maximize the phase margin. The unity gain frequency is 3.31 MHz and the phase margin is 50.9°. From the closed loop analysis the bandwidth is 5.37 MHz.

Secondly, a PLL's model in time and frequency domain has been done to analyze the lock time of the system and noise performances, as is shown in Fig. 52.11. To achieve this goal, an envelope simulation has been done with both open loop and

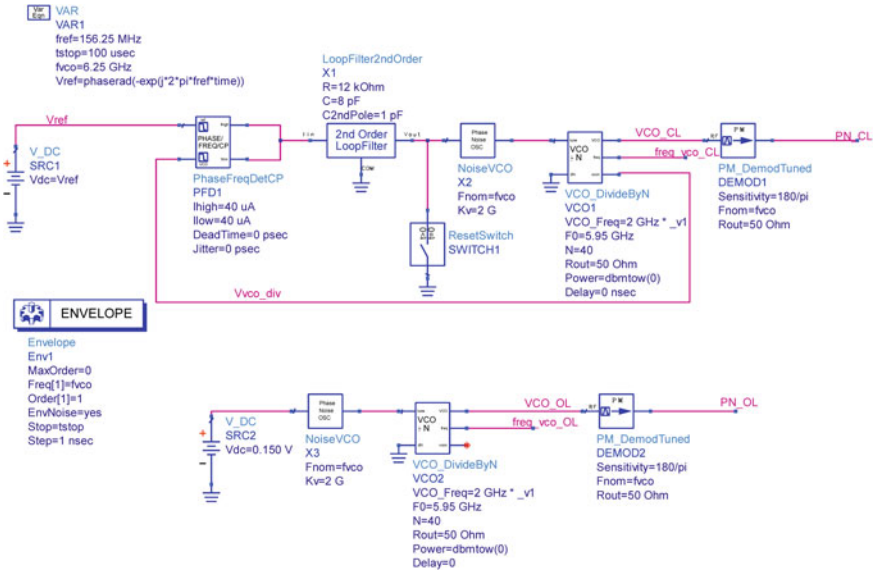
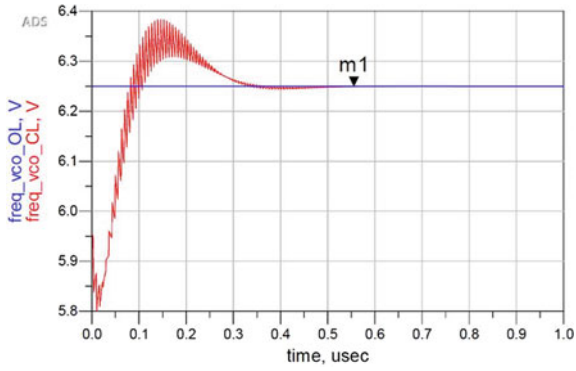


Fig. 52.11 Open and closed loop models

closed loop models to compare the two systems. The models of the *VCO_DivideByN* and the *PhaseFreqDetCP* are noise-free, as well as the reference source *SRC1*. Therefore, the block *NoiseVCO* has been added to insert the VCO phase noise in the analysis. This block, starting from a piecewise linear curve approximation in frequency domain of the VCO's phase noise, provides the equivalent noise on the control voltage of the oscillator. Instead, ADS' noise models have been used for the loop filter.

In Fig. 52.12 the lock transient is shown in terms of frequency. From this analysis a lock time of 555.6 ns results considering a locking error below 0.01%. As expected, during the transient, peak frequencies are present due to the resistor in the loop filter. These peaks are not present when the PLL is in locked state because in this model

Fig. 52.12 Lock time



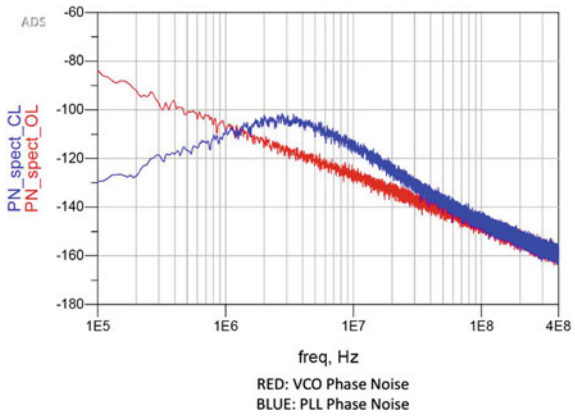


Fig. 52.13 Noise characterization in ADS

the charge pump is ideal, but not ideality has to be considered and, therefore, the second capacitor has been added in the loop filter. This capacitor has been sized so that the previous results in terms of loop bandwidth, unity gain and phase margins are kept roughly the same. In Fig. 52.13 the noise analysis' results are shown in dBc/Hz. In Fig. 52.13, the target phase noise (<−80 dBc/Hz) is achieved with a good margin, but reference phase noise and charge pump noise were not considered. These contributions will be added in the on-going activities. The loop filter's noise contribution is predominant at mid frequencies (bandpass response), while VCO's contribution prevails at higher frequencies (highpass response).

52.3.2 PLL Modeling and Simulation Results in Simulink

The same analysis has been done in Simulink using the models provided by the *Mixed Signal Blockset*. The achieved results are similar to those of ADS' model and the VCO's Simulink model follows the real VCO behaviour in terms of noise. In Fig. 52.14 the VCO phase noise in orange, target, and the PLL resulting phase noise in blue are shown.

52.4 Achieved Results

The achieved results from the analysis that has been done are summarized in Table 52.1.

An estimation of the occupied area has been done for the loop filter considering the 65 nm TSMC process design kit, resulting in a total area of about 4600 μm^2. C1 should occupy, as MIM capacitor, about 4000 μm^2, C2, MIM capacitor as well, 500 μm^2 and R1, N Well under OD resistor, 102.5 μm^2.

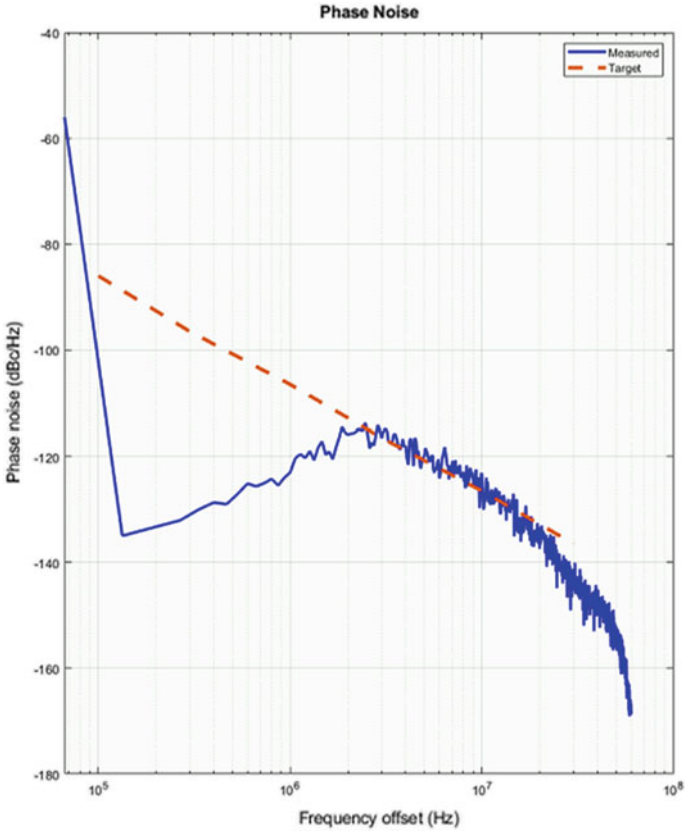


Fig. 52.14 Noise characterization Simulink

Table 52.1 Achieved results from the PLL models in Simulink and ADS

	Simulink model	ADS model
Phase margin, open loop	52°	50.892°
Bandwidth, closed loop	6 MHz	5.37 MHz
Lock time, closed loop	573.44 ns	555.6 ns
Phase noise, closed loop	<−110 dBc/Hz	<−100 dBc/Hz

52.5 Conclusions and Future Work

The paper has presented the modeling and design activity in both Simulink and ADS CAD environments of a PLL architecture to generate the clock reference for the new ESA Spacefibre standard. This standard allows on-board satellite communications up to 6.25 Gbps. Starting from a 6.25 GHz VCO rad-hard design, integrated in 65 nm technology, this work presents a PLL architecture including configurable integer

divider, down to a reference signal of 156.25 MHz, phase-frequency detector, charge pump and passive loop filter. A fully integrated PLL solution in 65 nm can be achieved with a 6 MHz low-pass PLL loop filter whose passive devices can be integrated on chip with an area of 4600 μm^2. PLL lock time is about 555 ns. The PLL phase noise performance are in line with that of the VCO, and for the stability a gain and phase margins of 86 dB and 50° are achieved. A preliminary circuit for the charge pump implementation is also proposed. Next steps are refining the charge pump circuit and implementing the blocks for PFD and clock divider.

References

1. European Cooperation for Space Standardization (2019) ECSS-E-ST-50–11C, SpaceFibre—very high-speed serial link
2. De Maria N et al (2016) Recent progress of RD53 collaboration towards next generation pixel read-out chip for HL-LHC. J Instrum 11
3. Monda D (2019) Analysis and design of a radiation hardened voltage controlled oscillator for space applications in 65 nm CMOS technology. Master Thesis, University of Pisa
4. Prinzie J, Christiansen J, Moreira P, Steyaert M, Leroux P (2018) A 2.56-GHz SEU radiation hard LC-tank VCO for high-speed communication links in 65-nm CMOS technology. IEEE Trans Nucl Sci 65(1)
5. Prinzie J, Christiansen J, Moreira P, Steyaert M, Leroux P (2017) Comparison of a 65 nm CMOS Ring- and LC-oscillator based PLL in terms of TID and SEU sensitivity. IEEE Trans Nucl Sci 64(1) 2017
6. Moustakas K, Siskos S (2015) Low voltage CMOS charge pump with excellent current matching based on a rail-to-rail current conveyor. In: IEEE 13th international new circuits and systems conference (NEWCAS)
7. Zhang C, Au T, Syrzycki M (2012) A high performance NMOS-switch high swing cascode charge pump for phase-locked loops. In: IEEE 55th international midwest symposium on circuits and systems (MWSCAS)

Chapter 53
Stability and Startup of Non Linear Loop Circuits

Francesca Cucchi, Stefano Di Pascoli and Giuseppe Iannaccone

Abstract The reliable analysis of DC operating point in circuits with positive feedback topology is often challenging, and frequently performed with ad hoc methods. These techniques are often error prone and lead to the frequent use of sub-optimal or unnecessary additional circuits for the stabilization or determination of the operating point (startup circuits). We present a simple and reliable technique for the determination of "stable" circuit solutions, that is based on the use of available circuit simulators and hence takes advantage of accurate device models. The method has been experimentally validated on a self-biasing current generator fabricated with a standard 0.18 μm CMOS process.

Keywords Self biasing · Operation point · Analog circuits

53.1 Introduction

In the realm of electronic circuits containing active devices, the determination of the operating point is a basic step of the design process. It is one of the few engineering techniques requiring the solution of an inherently non-linear physical system. Since non-linear systems cannot generally be solved in closed form, the electronic designer has to resort to approximate solutions, numerical analysis tools or, sometimes, clever ad hoc tricks. In fact, this intrinsic non-linearity is seldom a problem, since most circuits are *designed* to have an operating point that can be easily determined.

However, some applications demand the use of circuits for which the computation of the operating point is non trivial. The typical case is a circuit with a positive feedback such as the well known Eccles-Jordan flip-flop. These circuits can have a few operating points, some of which "unstable". Due to the mentioned non-linearity, the analysis of these circuits can be challenging; furthermore, in this case commonly

F. Cucchi · S. Di Pascoli (✉) · G. Iannaccone
Dipartimento di Ingegneria dell'Informazione, University of Pisa,
Via Caruso 16, 56122 Pisa, Italy
e-mail: stefano.di.pascoli@unipi.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_53

used circuit simulators, such as SPICE, often provide unreliable information, since they can converge to the "unstable" solution.

General methods have been developed for the non-linear analysis of active circuits [1–3], but are generally too abstract, provide poor physical insight on circuit operation, and are of little help to the circuit designer. As a consequence, non-linear circuits are usually analysed with simple pencil and paper methods [4]. These calculations are constrained to the use of crude first-level device models, which can lead to grossly approximated solutions, missed solutions and also to spurious solutions. Another common way to investigate the stability properties of circuits is the use of (time consuming) transient simulations, but these can also provide unreliable information in case of circuits with widely separated time constants (ill-conditioned systems). In order to overcome these shortcomings, we propose a method that is able to find the operating points and the stability properties of many commonly used non-linear feedback circuits.

53.2 Problem Definition

A non-linear time-independent circuit (i.e., without capacitors and inductors) can be described with a system of equations $F(\mathbf{x}) = \mathbf{0}$, where the vector $\mathbf{x}$ is composed by node voltages and/or branch currents. The system can have an unknown number of solutions $\mathbf{x_i}$. Most circuits have only one solution, but circuits with more than one solution are well known. Eccles-Jordan circuits generally have three solutions, one of which is "unstable".

We must note that even the "stability" of the solution is not a well-defined concept. Solutions of time-independent circuits cannot be "stable" or "unstable". Indeed, unstable solution are not solutions at all. A formal definition of "stable solution" can be found in [5]: a solution of $F(\mathbf{x}) = \mathbf{0}$ is *potentially stable* if it is possible to build—adding capacitors between nodes and inductors in series to the branches of the given circuit—an augmented circuit which is *robustly stable* in the time domain. *Robustly stable* means that the stability is not compromised by the addition of another set of sufficiently small capacitors and inductors to the given circuits (i.e. the values of the first set of capacitors and inductors must not be critical). Solutions which are not potentially stable are unstable.

Many non-linear circuits with more than one solution are based on a positive-feedback loop topology, like, for example, self-biased current generators, in which two current-controlled current generators are connected back-to-back in a positive-feedback loop. We will take this circuit as an example for illustrating the method (Fig. 53.1a).

Transistors M3 and M4 form a linear current mirror, duplicating the current fed into the drain of M4 (I_{in_um}) onto the drain of M3 (I_{out_um}). This current mirror provides a linear relationship between its input and output:

$$I_{out\_um} = k_{um} I_{in\_um}, \tag{53.1}$$

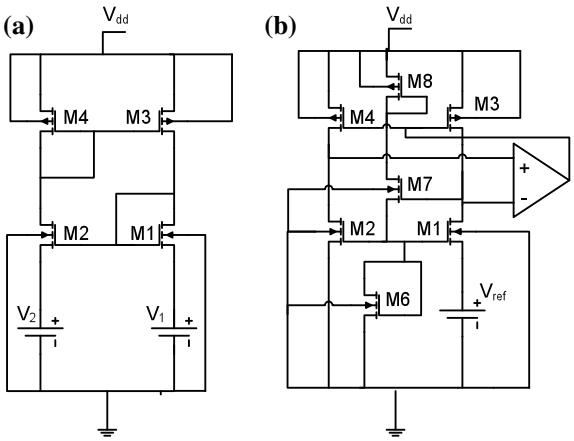


Fig. 53.1 Self-biased current generator: simplified proof of concept circuit (**a**) and complete circuit (**b**); M6, M7 and M8 are needed to set the bias point of M1 (a native transistor with negative threshold voltage); the operational amplifier imposes $V_{DS_{M3}} = V_{DS_{M4}}$ improving the accuracy of the upper current mirror; since in the complete circuit $V_2 = 0$ no generator is connected in series with M2

where k_{um} depends on the geometry of M3 and M4. On the other hand, the lower mirror (M1, M2, V1, and V2) provides a nonlinear relationship between the input current (the drain current of M1, I_{in_lm}) and the output (the drain current of M2 I_{out_lm}):

$$I_{out\_lm} = f(I_{in\_lm}). \tag{53.2}$$

The ratio of the input to the output current k_{lm} depends on the input current. At equilibrium we must have

$$k_{um} = 1/k_{lm}. \tag{53.3}$$

If k_{lm} is a monotonic function of the input the (53.3) can be satisfied for a single set of circuit currents. However, as [4] points out, both mirrors of the circuit provide zero current when fed with a zero input and hence another equilibrium point exists, with all null currents (where k_{lm} is undefined). For this reason most designers of self-biased current generators include a startup circuit which forces the circuit to the desired solution, avoiding the zero-current one [6–8].

However, the above discussion is oversimplified. Simulating the circuit (with a UMC .18 μm CMOS technology, and with identically sized M3 and M4) we find that if $\beta_1 > \beta_2$ and $V_1 > V_2$, where $\beta_i = \mu C_{ox} W_i/L_i$ (W_i and L_i are transistor width and length, μ is carrier mobility and C_{ox} is the gate oxide capacitance per unit area) are referred to transistors M_i, the circuit undergoes a transient ending at the equilibrium point with non-zero currents. Hence, no startup circuit seems required. Instead, if $\beta_1 < \beta_2$ and $V_1 < V_2$ the circuit never settles in the equilibrium point suggested by Eq. (53.3), and no startup circuit can help. For the other possible configurations ($\beta_1 < \beta_2$ and $V_1 > V_2$; $\beta_1 > \beta_2$ and $V_1 < V_2$) Eq. (53.3) is never verified and no equilibrium point is possible.

53.3 Proposed Solution

To solve this problem we developed a technique that provides valuable information on the equilibrium points of nonlinear circuit. If we can consider a nonlinear circuit as a closed loop of nonlinear blocks (Fig. 53.2a), we can cut open the loop and insert the circuitry shown in Fig. 53.2b. Even if the method can be adapted to cuts in any branch, we will discuss only the most useful case, when the current flowing in the severed branch is non zero. The case of zero current is indeed simpler, but less general. The independent current source sends in the circuit a test current I_t which gives rise to a voltage V_p across its terminals. The voltage-controlled generator imposes the same voltage V_p to node B, the other end of the cut loop. Obviously, when the current I_v sinked by the voltage generator is equal to I_t, the original uncut circuit is in equilibrium. The two sides of the cut could be directly connected without altering the branch currents and the node voltages. Hence if we plot I_v versus I_t, equilibrium points can be identified as the intersections between the $I_v(I_t)$ curve and the $I_v = I_t$ line. In addition, the derivative $\partial I_v/\partial I_t = \lambda$ at the equilibrium point enables us to determine the stability of the equilibrium point.

Let us call R_t the differential resistance seen by the I_t generator: if the test current increases by ΔI_t, the voltage V_p increases by $\Delta V_p = R_t \Delta I_t$. The current I_v, instead, increases by $\Delta I_v = \lambda \Delta I_t$. Since the nodes A and B are at the same voltage, we connect them and redraw the circuit as in Fig. 53.2c. The total differential resistance seen between nodes $A \equiv B$ and ground (as shown in Fig. 53.2c) can be written as:

$$R_d = \frac{\Delta V_p}{\Delta I_{tot}} = \frac{R_t \Delta I_t}{\Delta I_t - \lambda \Delta I_t} = \frac{R_t}{1 - \lambda} \tag{53.4}$$

where ΔI_{tot} is indicated in Fig. 53.2c. From (53.4) we can conclude that if $\lambda > 1$ this solution is unstable. Let us underline that we assumed $R_t > 0$, which is the typical situation is practical circuits, but the method can in theory be easily generalized to any initial sign of R_t. Furthermore, λ is the small-signal DC loop gain, and hence the fact that values in excess of 1 lead to instability is well known.

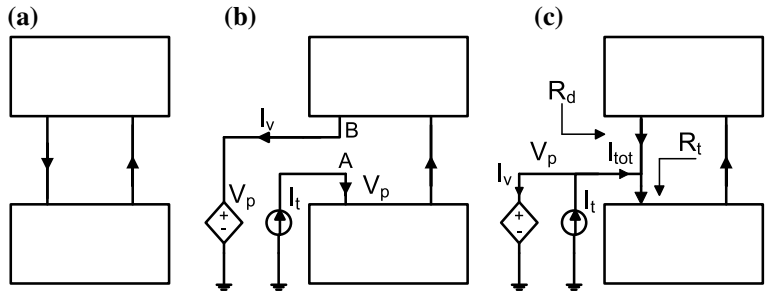


Fig. 53.2 Non-linear loop analysis

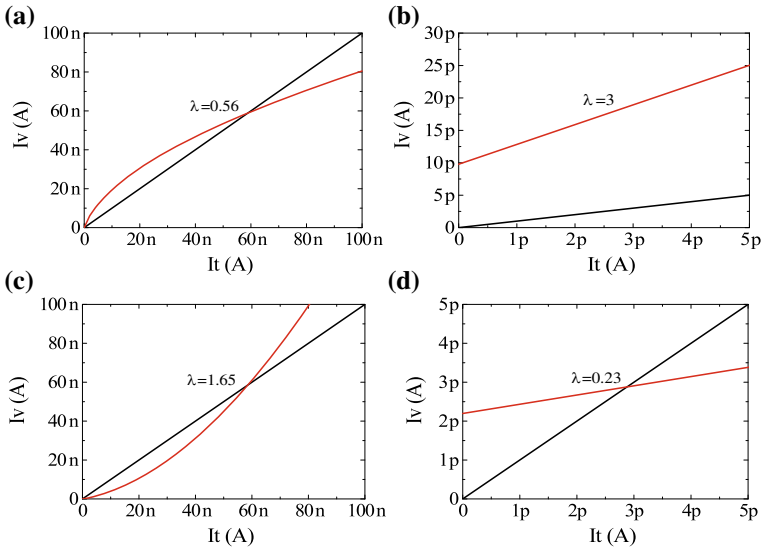


Fig. 53.3 SPECTRE dc sweep of circuit of Fig. 53.1, cut at the drain of M3: current generator to gate-drain of M1, voltage generator to M3 drain. $\beta_1 > \beta_2$ and $V_1 > V_2$ (**a**); particular of low current region (**b**); $\beta_1 < \beta_2$ and $V_1 < V_2$ (**c**); particular of low current region (**d**) (λ is the derivative of the current at the intersection; the black straight lines are $I_V = I_t$, while the red lines show the simulation results)

Hence, the practical application of the method consists of cutting open a loop, inserting the proper generators and performing a DC simulation of the circuit with an input current sweep. The analysis of circuit Fig. 53.1a (for which is $R_t > 0$) leads to the results of Fig. 53.3a and b, which show that a single and stable operating point is obtained only for $\beta_1 > \beta_2$ and $V_1 > V_2$. It is worth noticing that in this case no equilibrium point exist at $I_t = 0$ and hence no startup circuitry is needed. Figure 53.3c and d shows instead that for $\beta_1 < \beta_2$ and $V_1 < V_2$ the solution is unstable, and another stable solution is present for very small currents. Therefore, with the use a circuit simulator equipped with accurate device models we can learn that often some pencil-and-paper results, such as the zero-current stable solution, can indeed be artifacts due to the use of too simplistic device models.

Furthermore, this approach provides valuable physical insights on the circuit. Since the $I_v(I_t)$ relationship provided by the simulations can be interpreted as the input-output characteristic of an amplifier, a designer can usually devise modifications to the circuit which can modify it in a foreseeable manner. Hence, the above analysis not only can provide evidence of bias or stability problems, but is also a tool for their solution.

The circuit of Fig. 53.1b has been designed and fabricated, using native transistors (with threshold voltage < 0) for M1 and M2. In this version of the circuit M1 was not diode-connected and a proper bias circuit was added in order to bias M1 in

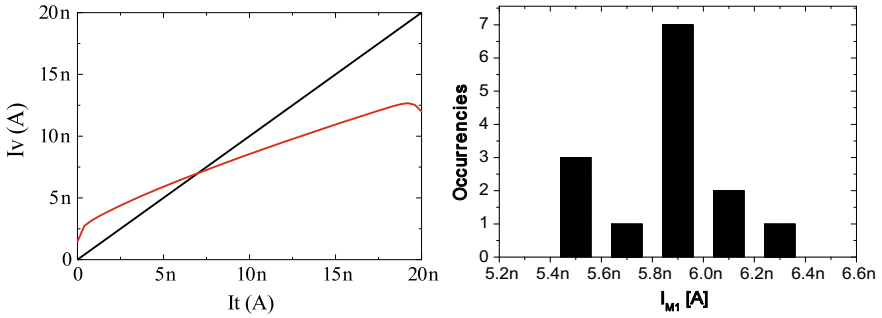


Fig. 53.4 I_v versus I_t for the complete circuit of Fig. 53.1b (left) and I_{M1} distribution in 14 samples of Fig. 53.1b circuit (right)

saturation. V_1 and V_2 were set to 335 mV and 0, respectively. Using the proposed method, we obtained the results of Fig. 53.4 (left). The current in M1 is about 7 nA, and the operating point is stable. This is confirmed by measurements on 15 samples realized in a 0.18 µm UMC CMOS technology. Figure 53.4 (right) shows the current distribution in 14 working samples; the mean current is 5.85 nA ($\sigma = 0.24$ nA) and no start up problems were observed.

Acknowledgements This work has been partially supported by the Electronic Components and Systems for European Leadership Joint Undertaking and by the Italian Ministry of Education, University and Research (MIUR) under grant agreement No. 737434 (CONNECT).

References

1. Chua L, Green D (1976) A qualitative analysis of the behavior of dynamic nonlinear networks: stability of autonomous networks. IEEE Trans Circuits Syst 23:355–379
2. Green M, Wilson A (1995) An algorithm for identifying unstable operating points using SPICE. IEEE Trans Comput-Aided Des Integr Circuits Syst 14:360–370
3. Gajani GS, Brambilla A, Premoli A (2008) Numerical determination of possible multiple DC solutions of nonlinear circuits. IEEE Trans Circuits Syst 55:1074–1083
4. Sedra A, Smith K (1997) Microelectronic circuits. Oxford University Press, New York
5. Green M, Willson A (1992) How to identify unstable dc operating point. IEEE Trans Circuits Syst 39:820–832
6. Guo J, Leung KN (2012) A CMOS voltage regulator for passive RFID tag ICs. Int J Circuit Theory Appl 40:329–340
7. Liang CJ, Chung CC, Lin H (2011) A low-voltage band-gap reference circuit with second-order analyses. Int J Circuit Theory Appl 39:1247–1256
8. Tsitouras A, Plessas F, Birbas M, Kikidis J, Kalivas G (2012) A sub-1V supply CMOS voltage reference generator. Int J Circuit Theory Appl 40:745–758

Chapter 54
IoT Ubiquitous Edge Engine Implementation on the Raspberry PI

Ahmad Kobeissi, Riccardo Berta, Francesco Bellotti and Alessandro De Gloria

Abstract In the Internet of Things (IoT) ecosystem, sensors and actuators represent the edge that is the source of data. The amount of data being generated by edge devices is exploding. Storage and processing of all the data in the cloud has become too slow and costly to meet the requirements of the end user. Edge computing presents a substantial solution through facilitating the processing of device data closer to the source. However, computing data from various and different sources is a formidable challenge for edge programming. This abstract presents lab experiments for testing versions of a multi-purpose generic edge engine on open-hardware edge devices, specifically the Raspbersry Pi 3 as a test bed with a standard Operating System (OS) and the STMxx as an MCU with Real-Time Operating System (RTOS).

54.1 Introduction

The edge denotes the layer closest to the physical world that we are interested in sensing. Other than mobile devices, edge devices are considered low-end computing systems due to their limited computational power abilities. The edge engine that we propose in this paper is designed to provide feasible edge computing capabilities on low-end IoT edge devices. It is Ubiquitous in the way that it can be adopted for different use case implementations. In this paper, part 2 discusses related work. Part

A. Kobeissi · R. Berta (✉) · F. Bellotti · A. De Gloria
DITEN, Università degli Studi di Genova, Via Opera Pia 11/a, 16145 Genova, Italy
e-mail: Berta@elios.unige.it

A. Kobeissi
e-mail: Ahmad.Kobeissi@elios.unige.it

F. Bellotti
e-mail: Franz@elios.unige.it

A. De Gloria
e-mail: AD@elios.unige.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_54

3 presents the architecture and implementation of the edge engine. Then, in part 4, we present the test methodology and discuss the results. Finally, we conclude with future work.

54.2 State-Of-The-Art

Edge computing presents great opportunities to achieve ubiquitous computation in the Internet ecosystem. It is proposed to overcome the intrinsic challenges of computing at the cloud side. Edge computing offers gathering more sensory data, reducing the response time, freeing up network bandwidth, and ultimately reducing the workload on the cloud.

Recently, AWS offered serverless functions called Lambda@Edge [1] in a pay-per-computation billing scheme. Content delivery through the Amazon CloudFront can be customized as well as compute resources and execution time. Azure IoT Edge [2] offers deployment of models (built and trained in the cloud) on the edge. In case of intermittent connectivity, Azure IoT Edge device management automatically synchronizes the latest state of edge devices after they are reconnected to ensure seamless operability.

The deployment of IoT applications to distributed nodes is a tedious procedure. In [3], a proposed approach is presented where the IoT application can be modeled in one place, where after modeling; the different pieces of application are annotated with location information. Based on this annotation, the application is decomposed into fragments that are deployed to corresponding individual compute nodes, automatically generating code to remotely connect the application fragments to other application fragments on other compute nodes in the edge or in the cloud. In addressing the domain-diversity aspect in data sharing in IoT, [4] proposes a cross-domain, secure, and feasible data sharing scheme in cooperative edge computing. To ensure the data's safety achieve data's fine-grained access, the scheme employs CP-ABE as an encryption mechanism for data privacy (Fig. 54.1).

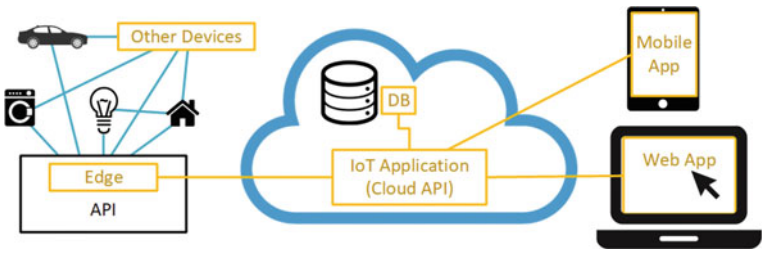


Fig. 54.1 IoT end-to-end ecosystem

54.3 Edge Engine

The main requirement of edge processors is real-time computing from continual input in small time-periods. Computations such as aggregation, filtering, processing and other form of data manipulation must keep up with the input flow of raw data. Another requirement of edge processors is the backup and storage of important data to the cloud. These requirements are suitable for limited bandwidth on the communication channels available for IoT node connections (I2C, BLE, Wifi) as well as internet connection to the cloud. Furthermore, for better data management and structural organization of edge devices, a common edge hub for multiple sensory nodes works better than connecting each node directly to the cloud.

We developed a ubiquitous engine that runs through the Express.js framework (programmed in NodeJs [5]) that enables the edge device to act as an IoT hub between sensors/actuators and the cloud. Figure 54.2 shows the block diagram of the framework where the edge hub plays a central role.

The edge engine, when run, goes through two main stages: Initialization and run loop. The complete flowchart representation is shown in Fig. 54.3.

In the initialization stage, the engine is set up. First, by logging in to the cloud through user credentials. The login is an HTTP POST request to the URL of the cloud server: https://api.atmosphere.tools. The request body would contain the user credentials (username/password) which are provided to the operator by the database administrators. In case of successful request, the server returns (in an HTTP response) a Json Web Token (JWT) for use in further http requests to the cloud for a time frame of 30 min, then it must be renewed. Second, by downloading the edge script by specifying the script Id. The script contains all the necessary information for the edge engine to run according to the edge device operator's intent, since the operator is responsible of writing the script and storing it in the database beforehand.

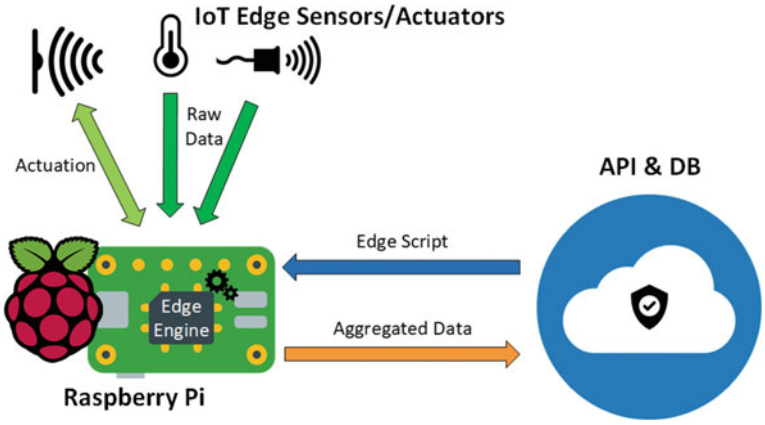


Fig. 54.2 Block diagram showing the edge engine within the IoT ecosystem

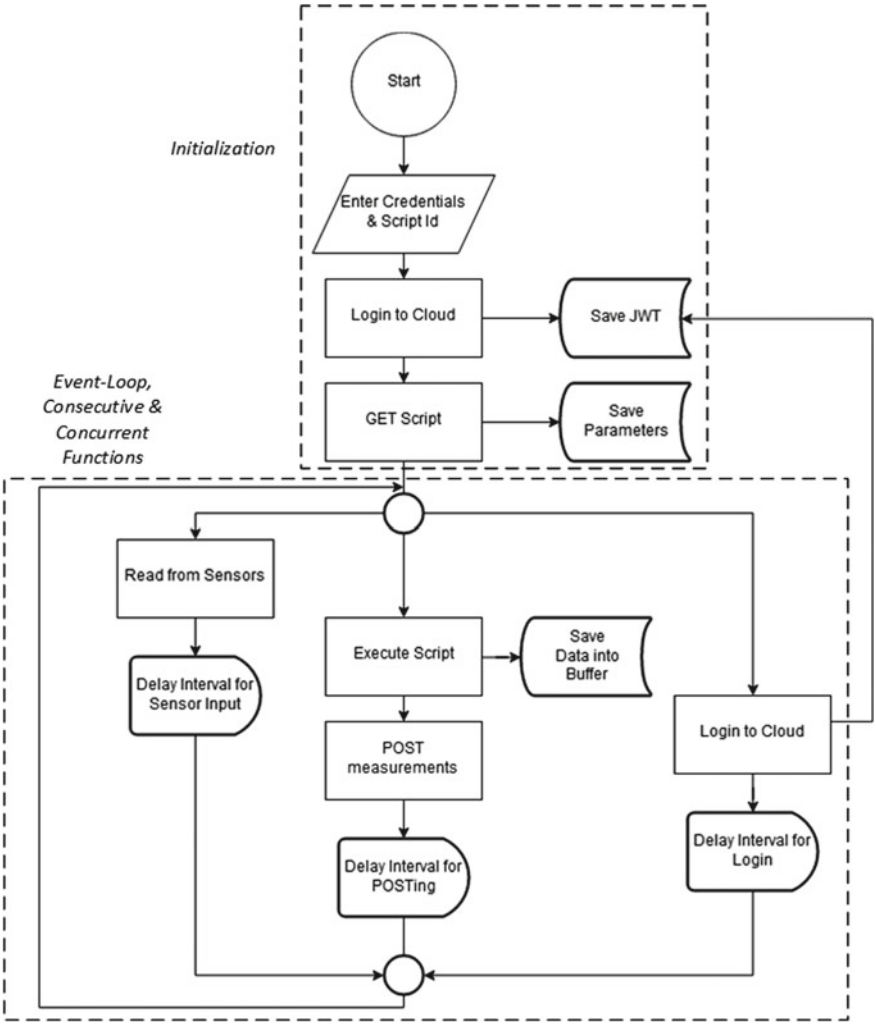


Fig. 54.3 Flowchart representing the coding of the edge engine in Node.js

The information consists of edge descriptor information (tags, http method, features, device properties, etc.), delay intervals for looping processes, and computations (operations) parameters. A sample of the edge script is viewed in Fig. 54.4 along with labels indicating the different information contained in the script.

After a successfull initialization, the edge engine enters a recursive stage call run loop. The run loop exploits the event loop mechanism within Nodejs to run three different processes consecutively and concurrently. The first to execute is reading from densors, which check the connected ports for input data which it saves in and

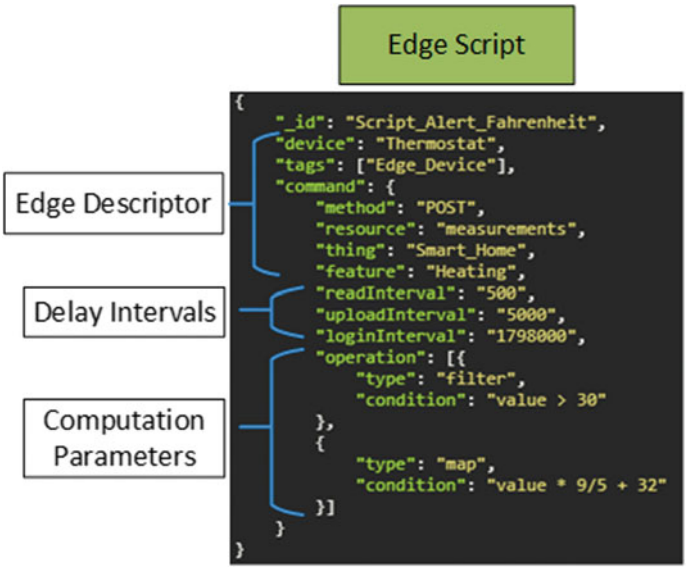


Fig. 54.4 A sample edge script for a 2-stream edge device

input buffer. The next process is dependant on the first, thus it executes immediately afterwards. This processor applies computation operations on the input buffer, replacing it's contents with aggregated data row for row.

The computation operations are specified within the edge script which specifies the type and required condition/parameters. In the example of Fig. 54.4, the script specifies two operations in order: a filter keeping the values that exceeds 30, and a map function to apply a transformation of the initial value.

The example is a demonstration of a temperature monitoring application where only high values are recorded then transformed from Celsius to Fahrenheit. These operations are supported by the 'array.prototype' JavaScript constructor which contains up to 30 operational methods. The last process in the run loop is a login just like the one performed in the initialization. A new login is required every 30 min to re-aquire a valid JWT for the measurements upload to be allowed to the cloud. The intervals for each of the three processes in the run loop can be set within the script, if the operator fails to do so, the engine will load default values for each interval.

We implemented backup scenarios for certain common events that occurs at the edge. One scenario is the offline more (disconnectivity from the Internet) or intermittent connectivity issues. In this case, we activate local storage of the aggregated buffer untill the connection is re-established or the memory if ull, in which the engine starts replacing the oldest of the records. Another scenario is the case of incomplete scripts or even no script at all. The engine has default values for essential parameters to run, where raw data are uploaded to the cloud as is. The final case scenario is

corrupt data handling. Data can get corrupted upon certain types of operations, the engine detects and de-activates computing by recording raw data immediately to the cloud.

54.4 Results and Discussion

We designed a lab experiment to test the performance and parametric limits of the edge engine deployment on a Raspberry Pi 3 b [6]. The experiment was designed to simulate a smart home IoT environment. It included up to 16 sensors, wired connected to the GPIO port of the Raspberry Pi. Those sensors are 4 dual temperature and pressure sensors, 4 switch sensors, 3 photodetectors, 3 passive infra red (PIR) sensors, 1 humidity sensor, and 1 moisture sensor. These sensors have different polling rates, with the fastest at 100 Hz frequency reached by the PIR sensor. That indicated that the minimum delay that still captures a change in measurements from the sensors is 10 ms. In the experiments, we ran 7 different edge scripts. Each script specifying different delay parameter for input reading from sensors and output writing to the cloud, both at the same time keeping the ration between input and output streams 10x. The scripts ran the same number of consecutive operation at four, since the change in this parameter had little to no effect on the Raspberry Pi's load. The experiment result in two main observation as presented in Fig. 54.5. The CPU usage reached it's maximum at 90% with 4 threads running on the 4-core CPU at the minimum limit of possible input stream delay at 1 ms. The typical delay of 10 ms for input stream corresponded to 27% CPU usage with 3 running threads. Such usage is acceptable considering the number of input streams (16) and computations (4) running a 100 times per second. The other observation, which concerned the memory usage was unexpected. The test recorded a decline in memory usage in regards to higher output stream delays. One explanation for this observation is the cashe management mechanism within the Raspbian OS, which keeps the buffers that were cleared by the engine saved for a while. So, the more buffer gets cleared by the engine in a smaller timeframe, the more buffers the OS is cashing.

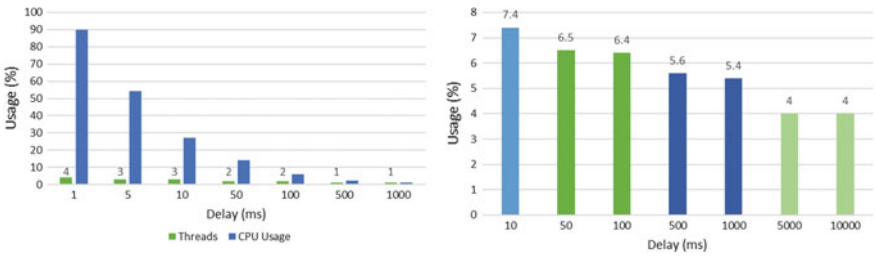


Fig. 54.5 Test observations of CPU and memory usages with respect to delay changes

The presented results proves the feasibility and usability of the edge engine architecture on low-end computing devices. As a test case, the Raspberry Pi performed rather well under extreme parametric conditions.

54.5 Conclusion and Future Work

In this paper, we presented a ubiquitous IoT edge engine implementation for low-end computing devices such as microcontrollers. We performed a lab experiment to test the engine's performance on a Raspberry Pi unit connected to 16 sensors. Results came in support of the ability of such devices to perform remarkable computations (100 * 16 *4) per second within acceptable hardware usage. In future work, we look forward to perform similar experiments on RTOS-based microcontrollers like the STM32 and Arduino. A possible addition to the supported operations at the edge is lite machine learning algorithms in both the supervised and unsupervised categories.

Acknowledgements The heading should be treated as a 3rd level heading and should not be assigned a number.

References

1. Amazon Web Service (2019) AWS Lambda@Edge. https://aws.amazon.com/lambda/edge/
2. Azure IoT Edge (2019) Microsoft. https://azure.microsoft.com/en-in/services/iot-edge/
3. Jain R, Tata S (2017) Cloud to edge: distributed deployment of process-aware IoT applications. In: 2017 IEEE international conference on edge computing (EDGE), Honolulu, HI, pp 182–189
4. Fan K, Pan Q, Wang J, Liu T, Li H, Yang Y (2018) Cross-domain based data sharing scheme in cooperative edge computing. In: 2018 IEEE international conference on edge computing (EDGE), San Francisco, CA, pp 87–92
5. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) Edge computing: vision and challenges. IEEE Internet Things J 3(5):637–646
6. Raspberry Pi 3 model b. https://www.raspberrypi.org/products/raspberry-pi-3-model-b/

Chapter 55
Non-intrusive Load Monitoring on the Edge of the Network: A Smart Measurement Node

Hugo Wöhrl and Davide Brunelli

Abstract To efficiently reduce energy usage in buildings, it is necessary to understand how energy is consumed today. Non-intrusive load monitoring (NILM) is a promising approach where appliance level load profiles can be extracted from an agglomerated single-point measurement using statistical or machine-learning methodology. Moving NILM to the edge of the network holds many advantages like reduced operation cost and decreased power consumption while minimizing privacy concerns. In this paper, we present a NILM hardware that can apply real-time NILM on the edge of the network on an ultra-low power AI-optimized microcontroller.

Keywords Non-intrusive load monitoring · Smart meter · Power efficiency · Energy disaggregation · Blind source separation problem

55.1 Introduction

Every record heat summer shows the need for action against the ongoing waste of resources, their intrinsic release of climate affecting gasses and their heating of the atmosphere. To effectively combat the misuse of power consumption, it is essential to get a clearer image of how energy is consumed in today's life, as [1] pointed out, households play a crucial role here. A very promising concept to understand electrical as well as thermal power consumption or water usage is Non-Intrusive Load Monitoring, where consumption is measured on a single point, and appliance level power consumption is extracted from an agglomerated signal. This practice, compared to equipping single consumers, cuts installation cost dramatically and therefore makes scaling at a city scope possible. As a key component of smart cities Non-Intrusive

H. Wöhrl
Department of Electronics and Computer Science, Technical University of Berlin, Straße Des 17. Juni 135, 10623 Berlin, Germany
e-mail: hugowoehrl@mailbox.tu-berlin.de

D. Brunelli (✉)
Department of Industrial Engineering, University of Trento, Via Sommarive 9, 38123 Trento, Italy
e-mail: davide.brunelli@unitn.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_55

Load Monitoring holds the chance to efficiently measure and visualize not only private household power consumption data but also buildings of the commercial sector as offices, malls or factory sites. Also, for the Industry 4.0 in-depth power analysis plays a major role and holds many possibilities [2]. Through pattern detection anomalies in power consumption can be detected in real-time [3, 4] and machines can be maintained before a major fault occurs, and thus further consequences eluded.

Non-intrusive Load Monitoring describes the task of disaggregation power consumptions of single appliances from an agglomerated mains power measurement. From the machine learning point of view, this is considered a single-channel blind source separation problem, where multiple sources need to be extracted from one combined measurement. George W. Hart founded the field of energy disaggregation in the 1980s and published 1992 the seminal paper for Non-intrusive Load Monitoring [5], where he introduced different NILM scenarios and implemented first disaggregation algorithms based on low-frequency features at a sampling rate of 1 Hz. In 2015, along with an overall rising interest in the field of machine learning the topic of NILM gained a boost in popularity, resulting in various publications combining different classification methods and features. This can generally be distinguished into two different approaches. One of which is using low frequency data and e.g. machine learning methods as Kelly in 2015 with the first application of Neural Networks to NILM [6]. The other one is deploying richer features of higher sampled measurements as for example S. Gupta [7] by using high frequent electromagnetic interference features. The most significant advantage of the low-frequency approach is its applicability in commercially available smart meters; this method still has some shortcomings as requiring big amounts of labeled data while still facing accuracy challenges [8]. The higher frequency approach is due to its increased hardware cost generally less investigated, but as e.g. Bernard [9] and Gupta [7] have shown, a richer feature set can significantly improve existing NILM algorithms, allowing them to classify more complex as well as similar loads. It furthermore permits new usage scenarios like anomaly detection.

In this paper, we present a smart measurement node that uses sampling rates up to 10 kHz and outperforms previous prototypes [10, 11], even with harvesting capability [12]. Therefore, it can be flexibly deployed in various environments allowing a combination of low and high-frequency features.

55.2 System Description

To measure the power intake of the respective building, the measurement node is connected to the in-house mains power supply, which also serves for the power supply of the board. For the current measurement, different analog interfaces can be selected via a multiplexer. The analog interfaces are described in more detail in Sect. 55.2.1. Two microcontrollers are deployed to handle the data stream and classification, this comes with the advantage of a fast implementation of the training phase on one microcontroller while allowing to have an optimized real-time classification phase

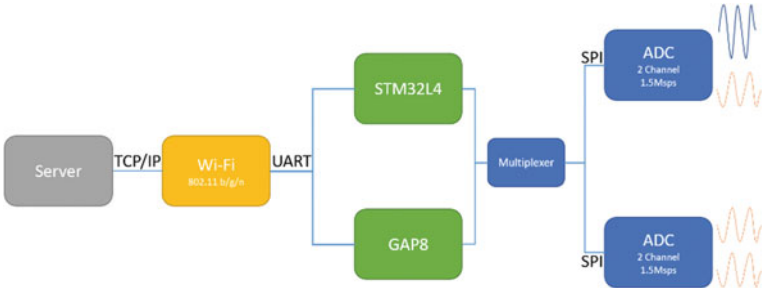


Fig. 55.1 Schematic overview of the system setup

on the other microcontroller, Sect. 55.2.2 explains this in detail. The measurement data then gets streamed via the onboard Wi-Fi Module (Sect. 55.2.3) to a TCP/IP Server, which stores the data and trains the classification model. After training the model gets retransferred to the microcontroller, which then is ready to do online classification (Fig. 55.1).

55.2.1 Analog Frontend

To acquire the voltage and current measurements, we deploy two interfaces, each driven by a 3MSps ADC capable of sampling simultaneously two differential channels at 1.5Msps. Their low current intake makes a power-efficient operation possible. The SPI Busses of the ADCs are connected to a multiplexer which routs one SPI connection to the microcontrollers. The first interface measures the grid voltage via a voltage divider on one channel and the consumed mains current via a shunt on the other channel. The second interface offers another option for the current measurement by having a hall-effect current sensor connected on one channel and a Rogowski Coil on the other. While the second interface already handles only isolated signals, the first interface is directly connected to mains voltage and therefore has a digital isolator between the ADC and multiplexer to prevent AC voltages on the logical voltage level in case of failure. The installed STM32 microcontroller theoretically would allow operation at the maximum sampling rate, making the PCB potentially usable for further applications as Non-intrusive load monitoring on electromagnetic interference [7] or power quality analyzers.

In our implementation, the analog frontend gets sampled by creating a steady pulse with a frequency of 10 kHz to trigger the conversion. After the trigger, the data is fed into the microcontroller using a DMA, where it is stored in two ping pong buffers.

55.2.2 Dual Microcontroller Concept

The heart of the system consists of two microcontrollers, of which one is active, and one is idle at a time. One microcontroller is an ultra-low-power RISC-V processor developed by Greenwave Technologies named GAP8, while the other is an ultra-low-power Cortex-M4 microcontroller from the STM32-F410 family (Fig. 55.2). The first contains a multicore processor with a cluster of 8 cores which offers enough computing power to do near real-time classification on the chip. This allows to move the classification to the edge of the network and so cuts power consumption drastically since recorded data can be processed locally instead of in the cloud. Furthermore, an application on the network edge cuts operation cost by making maintenance of a server expendable and reduces privacy concerns by storing most data locally.

The STM32-M4 microcontroller has the necessary interfaces to communicate between the different submodules of the device and is therefore used in the training phase where data needs to be fetched from the ADCs and passed to the Wi-Fi Module. It also comes with an ultra-low power intake, making a power-efficient operation possible. As a processor of the ARM architecture it is highly flexible and can be programmed also to fit different applications, or compression algorithms [13].

In the training phase the STM32 microcontroller sends the measured data to a server which learns the classification model in the next step. Afterwards, the extracted model gets transferred to the GAP8 microcontroller, which then does the classification of newly recorded data.

55.2.3 Wi-Fi Module

To establish the connection to the server, we use a 2.4 GHz 802.11 b/g/n Wi-Fi module with an integrated MCU. Other standards, such as Bluetooth, have not enough bandwidth [14], and, on the other side, the Wi-Fi module makes the implementation considerably easy and reduces programming effort while increasing flexibility. The Wi-Fi module is connected via UART to the STM32 microcontroller at a baud rate of 1Mbps, which in optimal circumstances allows conversions at up to 35KSps. The UART communication is implemented with a DMA that streams the buffers away as soon as they are full.

55.3 Evaluation

To evaluate the operation of the sub-modules firstly, the throughput of the Wi-Fi connection was tested. A program was written for the STM32 microcontroller that opens a TCP/IP Server and then sends dummy data to the server. This resulted in a net bandwidth of approximately 100 kB/s, which would allow sampling rates up to 20 kHz.

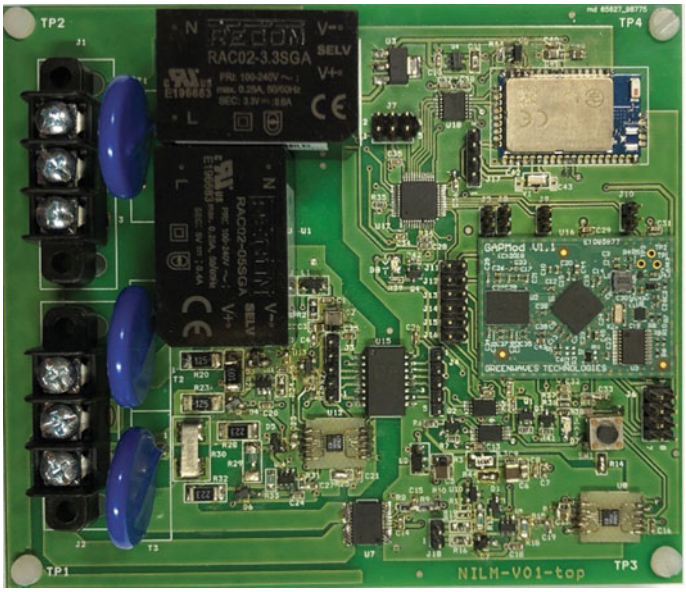


Fig. 55.2 Hardware layout of the measurement node

In a second test, the STM32 was to sample data at 10 kHz from the mains power with different appliances. This test showed promising results in respect to a NILM operation since differences in spectra even for very similar devices can be clearly seen by analyzing their harmonics (Fig. 55.3).

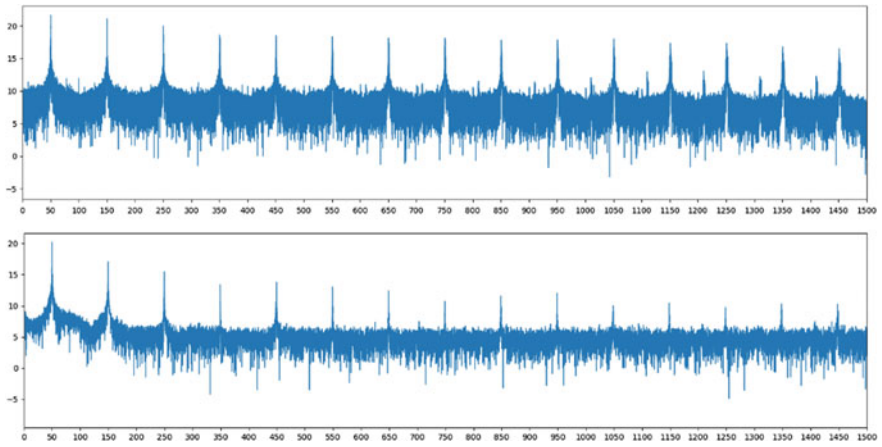


Fig. 55.3 FFT spectra of current intake of two notebook power supplies in idle operation

References

1. Armel KC, Gupta A, Shrimali G, Albert A (2013) Is disaggregation the holy grail of energy efficiency? The case of electricity. Energy Policy 52:213–234
2. Rossi M, Rizzon L, Fait M, Passerone R, Brunelli D (2014) Energy neutral wireless sensing for server farms monitoring. IEEE J Emerg Sel Top Circ and Syst 4(3):324–334
3. Nardello M, Rossi M, Brunelli D (2017) A low-cost smart sensor for non intrusive load monitoring applications. In: 2017 IEEE 26th international symposium on industrial electronics (ISIE), Edinburgh, pp 1362–1368
4. Nardello M, Rossi M, Brunelli D (2017) An innovative cost-effective smart meter with embedded non intrusive load monitoring. In: 2017 IEEE PES innovative smart grid technologies conference Europe (ISGT-Europe), Torino, pp 1–6
5. Hart GW (1992) Nonintrusive appliance load monitoring. Proc IEEE 80(12):1870–1891
6. Kelly J, Knottenbelt W (2015) Neural NILM: deep neural networks applied to energy disaggregation. In: Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments, pp 55–64
7. Gupta S, Reynolds MS, Patel SN (2010) ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home. In: Proceedings of the 12th ACM international conference on ubiquitous computing, pp 139–148
8. Kelly D (2016) Disaggregation of domestic smart meter energy data
9. Bernard T Non-intrusive load monitoring (NILM): combining multiple distinct electrical features and unsupervised machine learning techniques
10. Porcarelli D, Brunelli D, Benini L (2014) Clamp-and-Forget: a self-sustainable non-invasive wireless sensor node for smart metering applications. Microelectron J 45(12):1671–1678
11. Balsamo D, Porcarelli D, Benini L, Davide B (2013) A new non-invasive voltage measurement method for wireless analysis of electrical parameters and power quality. In: SENSORS, IEEE, Baltimore, MD, pp 1–4
12. Porcarelli D, Brunelli D, Benini L (2012) Characterization of lithium-ion capacitors for low-power energy neutral wireless sensor networks. In: 2012 ninth international conference on networked sensing (INSS), Antwerp, pp 1–4
13. Brunelli D, Caione C (2015) Sparse recovery optimization in wireless sensor networks with a sub-nyquist sampling rate. Sensors (Switzerland) 15 (7):16654–16673
14. Negri L, Sami M, Macii D, Terranegra A (2004) FSM-based power modeling of wireless protocols: the case of Bluetooth. In: Proceedings of the 2004 international symposium on low power electronics and design (IEEE Cat. No.04TH8758), Newport Beach, CA, USA, pp 369–374

Chapter 56
Design of a SpaceFibre High-Speed Satellite Interface ASIC

Pietro Nannipieri, Gianmarco Dinelli, Luca Dello Sterpaio, Antonino Marino and Luca Fanucci

Abstract In the last few years, data rate requirement in on-board data handling for space missions has continuously grown, due to the presence of high resolution instruments. This lead the European Space Agency to start working on a new communication standard named SpaceFibre. It is able to fulfil a data rate of 6.25 Gbit/s per communication lane (up to 16 communication lanes). This work proposes the design of a SpaceFibre interface Application Specific Integrated Circuit. The block diagram of the system is presented, together with results in terms of area occupation and power consumption (excluding serialiser-deserialiser circuitry) after the synthesis on a 65 nm CMOS technology.

Keywords SpaceFibre · CODEC · 65 nm · ASIC · Logic synthesis · On-board data-handling · Satellite · High-speed

56.1 Introduction

SpaceFibre is a novel very high-speed serial link, designed specifically for space applications. It has been recently standardized by the European Cooperation for Space Standardisation (ECSS) [1]. It is thus adoptable as on-board data handling protocol

P. Nannipieri (✉) · G. Dinelli · L. Dello Sterpaio · A. Marino · L. Fanucci
Department of Information Engineering, University of Pisa, Pisa, Italy
e-mail: pietro.nannipieri@ing.unipi.it

G. Dinelli
e-mail: gianmarco.dinelli@ing.unipi.it

L. Dello Sterpaio
e-mail: luca.dellosterpaio@ing.unipi.it

A. Marino
e-mail: antonino.marino@ing.unipi.it

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

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_56

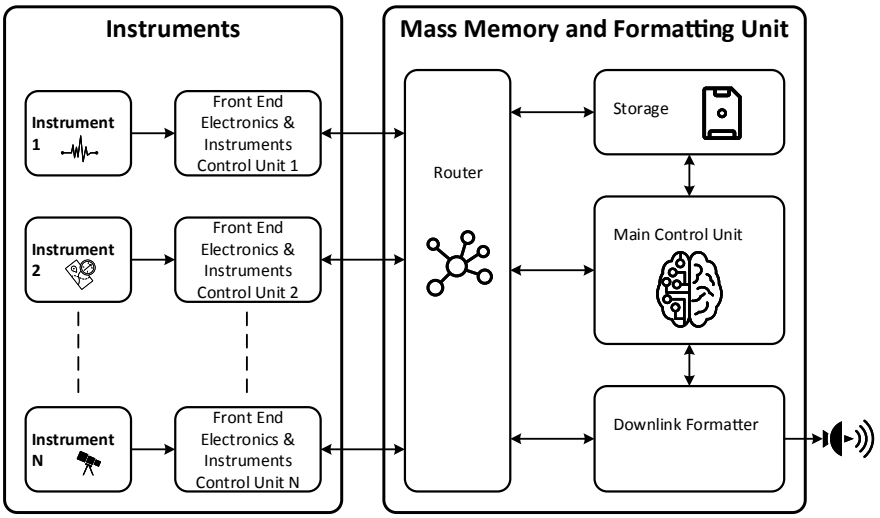


Fig. 56.1 Generic spacecraft network topology

in next generation space missions, from earth observation to telecom and science satellites (i.e. FLEX [2] and BIOMASS [3]) will mount Synthetic Aperture Radars (SARs) and hyper-spectral imagers which will require high speed on-board communication), where data rate requirement is particularly demanding. Different missions obviously have different on-board communication architectures. However in the following we present a scheme which describes a generic high-speed communication architecture for space applications.

In Fig. 56.1, it is possible to observe a schematic satellite on-board data handling topology. Generally, several instruments are hosted on the same spacecraft; each one produces a significant amount of data which will are then processed with front end electronics and sent to the Main Control Unit (MCU) of the experiment, where data are stored, processed or sent to the downlink. Redundancy is usually required (i.e. each link should be doubled). Moreover, each instrument may have a separate bus for communication to the MCU. It is known that constraints in space applications are particularly harsh, particularly in terms of radiation tolerance, fault tolerance, low power consumption, harness and data-rate. Such stringent requirements led to development of highly optimised solutions rather than adaptation of existing commercial product.

56.2 The SpaceFibre Protocol

SpaceFibre (SpFi) [4] is multi-Gigabit/s full stack on-board communication technology, whose design has been promoted by the European Space Agency (ESA). It has been developed specifically to support next generation satellite data-handling

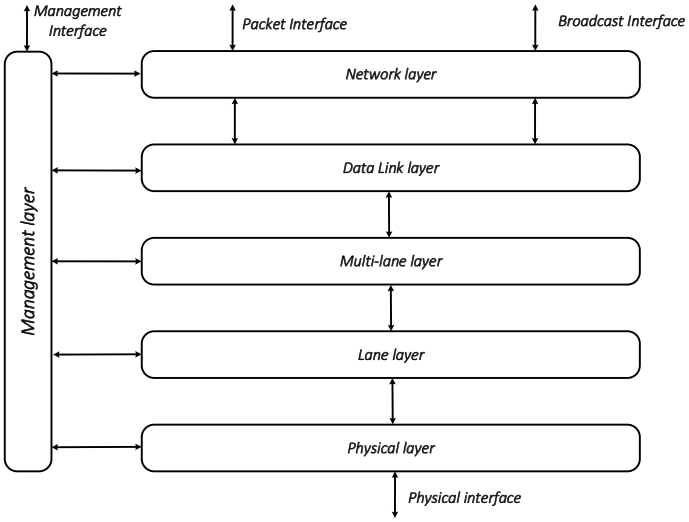


Fig. 56.2 SpaceFibre protocol stack

requirements. It is able to operate both on copper cables and fibre optic and supports data rate of 6.25 Gbit/s per communication lane. SpaceFibre includes built-in quality of service (QoS) and Fault Detection, Isolation and Recovery (FDIR) techniques, which provides system level benefits without requiring complex limiting software implementation. SpaceFibre is currently being integrated onto various FPGA technologies by IngeniArs [5], STAR-Dundee [4] and Chobam Gaisler [6]. The SpaceFibre protocol stack is shown in Fig. 56.2.

- **Network Layer** is responsible for packet transfer over the link or the network. This is an optional layer, see [7].
- **Data Link Layer** is responsible for the QoS, flow control and for resending information in case a temporary fault occurs over the link. It is also responsible for packaging data to be sent over the link and for broadcasting (and receiving) short messages across a SpaceFibre network.
- **Multi-lane Layer** is responsible for running several SpaceFibre lanes in parallel to provide higher data throughput. This is an optional layer.
- **Lane Layer** is responsible for lane initialisation error detection and re-initialisation. Symbols are encoded with 8b/10b encoding [8], with AC coupling of data signals.
- **Physical Layer** is responsible for serializing and de-serializing encoded symbols and for transmitting them over the physical link. It also recovers clock from the received data.
- **Management Layer** is responsible for the control and configuration of each layer.

A SpaceFibre link is composed of two differential lines, one for serial data transmission and one for serial data reception. The clock signal is transmitted together

with the data as symbols are processed with 8b10b encoding, providing a number of bit transitions sufficient to recover the clock from the incoming bit stream with a PLL. Attempts has been done in literature to design and build a SpaceFibre interface ASIC [9]. Unfortunately, no details are shared onto the technology node chosen, the circuit complexity and power consumption. Moreover, the work reported in [9] refers to a very early draft version of the standard (F3). Therefore, we identified the need to document an ASIC implementation of SpaceFibre codec fully compliant to the recently released final version of the standard, including indication of area and power consumption. In this work, the IngeniArs SpaceFibre IP has been used for the design of a SpaceFibre ASIC. In Sect. 56.3, the architecture of the synthesized circuit is shown, and in Sect. 56.4 synthesis results in terms of area occupation and power consumption, for the single SpaceFibre interface, is presented. Finally, in Sect. 56.5 conclusions are drawn.

56.3 The SpaceFibre CODEC

In Fig. 56.3, the block diagram of the system is shown. The blocks displayed are the following

- **SPFI Codec**: SpaceFibre single lane CODEC, implemented by IngeniArs. To have two independents communication lanes two separate single lane codecs have been included.
- **VC Switch**: as the SpaceFibre interface is wide due to the presence of several separated Virtual Channels, a multiplexer to reduce the number of I/O pins is necessary.
- **SPI Slave**: this block, which enables the device to be configured as SPI slave peripheral, is provided as external IP.
- **SERialiser DESerialiser (SERDES) Interface**: it is the lower part of SpaceFibre Lane layer. It comprehends 8b10b encoding/decoding, symbol and word synchronisation and elastic buffering.
- **HSSL**: High Speed Serial Link: this block is an IP technology dependent, which serialise and de-serialise input and output data streams. Please note that this block is the SERDES itself, which is usually a technology dependent block. Therefore it is not taken into account in the presented results.

The HDL of the CODEC itself has been tested and validated in previous work [5].

56.4 Synthesis Results

In this section, SpaceFibre CODEC synthesis results on a commercial 65 nm technology are presented. The tool used is Synopsys. These results are meant to be used as preliminary indications for future implementation of the SpaceFibre CODEC on a

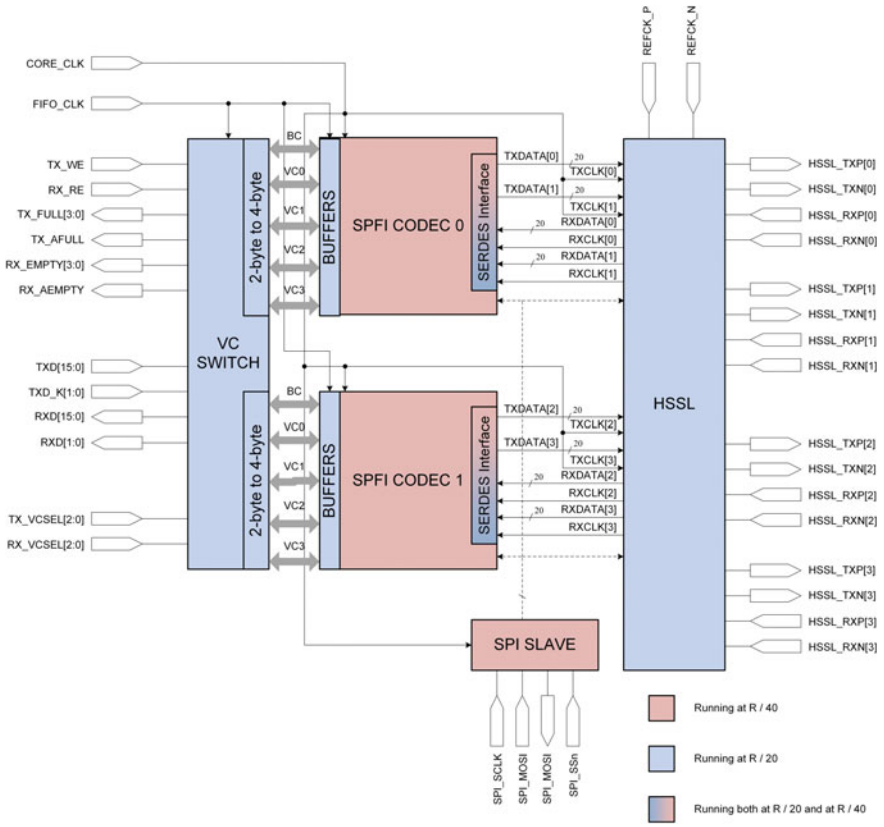


Fig. 56.3 SpaceFibre ASIC block diagram

rad-hardened silicon technology. The system has been synthesised in order to reach the operating clock frequency of 312.5 MHz in the lower sections of the CODEC (8b10b encoder/decoder, symbol synchronizer) where the data path is 2 symbols wide, and 156.25 MHz where the data path is 4 symbols wide (in the rest of the CODEC), which corresponds to a serial data rate of 6.25 Gbps (the fastest data rate reachable according to the standard). Area occupation is presented in Table 56.1 and the estimated power consumption is presented in Table 56.2. A NAND cell area of 3.12 μm is considered to compute the gate equivalent area.

Table 56.1 SpFi CODEC area occupation on a commercial 65 nm technology

Comb. area (μm^2)	Non-comb. area (μm^2)	Memory area (μm^2)	Total area (μm^2)	Total area (Kgate)
50,110	80,490	324,554	455,154	145,883

Table 56.2 SpFi CODEC power consumption on a commercial 65 nm technology

Switching power (mW)	Int power (mW)	Leak power (mW)	Total power (mW)
0.136	5.454	0.089	5.689

56.5 Conclusions

This work reports preliminary synthesis results of a SpaceFibre CODEC on a commercial 65 nm technology. Results are reported also in gate equivalent, in order to estimated SpFi CODEC area occupation also in other technologies. As the results refer to synthesis only, they do not take into account routing (both for area and power consumption estimation). The SpaceFibre codec requires 0.45 mm^2 of area and about 145 MGate equivalent. It consumes about 5.7 mW. Please note that these numbers do not take into account serialisation and de-serialisation.

Acknowledgements IngeniArs SpaceFibre technologies have been developed in the framework of the project SIMPLE (Spacefibre IMPLementation design & test Equipment). This project has received funding from the European Unions Horizon 2020 research and innovation programme under Grant Agreement No. 757038.

References

1. Space engineering—SpaceFibre—very high-speed serial link. European Cooperation for Space Standardisation, ECSS-E-ST-50-11C, May 2019
2. Rivera JP, Sabater N, Tenjo C, Vicen J, Alonso L, Moreno J (2014) Synthetic scene simulator for hyperspectral spaceborne passive optical sensors. Application to ESA's FLEX/sentinel-3 tandem mission. In: Proceeding of the 2014 6th workshop on hyperspectral image and signal processing: evolution in remote sensing (WHISPERS), Lausanne, SW, 24–27 June 2014
3. Toan TL et al (2018) The biomass mission: objectives and requirements. In: Proceeding of 2018 IEEE international geoscience and remote sensing symposium, Valencia, SP, 22–27 Jul 2018
4. Parkes S et al (2015) SpaceFibre: multi-gigabit/s interconnect for spacecraft on-board data handling. In: Proceeding of the IEEE aerospace conference, Big Sky, MT, USA, 2015, pp 1–8
5. Nannipieri P, Dinelli G, Davalle D, Fanucci L (2018) A SpaceFibre multi lane CODEC system on a chip: enabling technology for low cost satellite EGSE. In: 2018 14th conference on Ph.D. research in microelectronics and electronics (PRIME), Prague, 2018, pp 173–176
6. Siegle F, Habinc S, Both J (2016) SpaceFibre Port IP Core (GRSPFI): SpaceFibre, poster paper. In: Proceedings of the 7th international spacewire conference, Yokohama, Japan, 2016, pp 1–5
7. Leoni A, Nannipieri P, Fanucci L (2019) VHDL design of a SpaceFibre routing switch. IEICE Trans Fundam Electron Commun Comput Sci E102A(5):729–731
8. Nannipieri P, Davalle D, Fanucci L (2018) A novel parallel 8B/10B encoder: architecture and comparison with classical solution. IEICE Trans Fundam Electron Commun Comput Sci E101A(7):1120–1122
9. Villafranca AG, Ferrer A, McLaren D, McClements C, Parkes S (2015) VHiSSI: experimental SpaceFibre ASIC. In: European Space Agency (Special Publication) ESA SP, SP-732

Chapter 57
An FPGA Realization for Real-Time Depth Estimation in Image Sequences

Stefano Marsi, Sergio Carrato, Luca De Bortoli, Paolo Gallina, Francesco Guzzi and Giovanni Ramponi

Abstract This paper proposes a method for depth estimation in video sequences acquired by a monocular camera mounted on a mobile platform. The proposed algorithm is able to estimate in real time the relative distances of the objects in the field of view exploiting the parallax effect, provided the platform movement complies with a few constraints. The developed system is designed to operate at the input pixel cadence and is thus applicable to any video resolution. The final architecture, using operators no more complex than an adder and a memory that is just a fraction of a frame memory, can be realized in a low-cost FPGA.

Keywords Depth estimation · Monocular camera · Real time · FPGA · Fast matching

57.1 Introduction

Recent climatic upheavals lead to environmental catastrophes like the one that happened in northern Italy in Autumn 2018, when many thousands of centuries-old trees were uprooted during a storm in a wide and sparse mountain area. To help coping with these phenomena, a continuous monitoring of the territory is required, which in turn implies the availability of low-cost systems suitable to operate autonomously and capable to reach areas that are poorly accessible by normal vehicles or even on foot. Drones are a very interesting option [1] for this task. To increase their capability of autonomous flight they should be equipped with an economic and effective system able to estimate in a 2-D view the distance of the obstacles, in order both to analyze the underlying territory and to independently establish the most appropriate route, the exploration area and possibly the landing area.

Systems capable of detecting the distance of a target can adopt several types of sensors [2], often combined with dedicated illuminators. They can be based e.g. on

S. Marsi (✉) · S. Carrato · L. De Bortoli · P. Gallina · F. Guzzi · G. Ramponi
University of Trieste, Trieste, Italy
e-mail: marsi@units.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_57

laser beams, on infrared systems, or on ultrasonic sources; they differ in range, accuracy, sensitivity, resolution, and so on. *Time of Flight* cameras are devices equipped with an illuminator [3] and a special camera able to evaluate for each pixel the time taken by the emitted beam to be reflected back to the camera. These systems are generally complex and show important limitations. They are affected by other light sources and present a rather limited range.

Systems based on laser scanning can be easily used to deal with large distances [4], but are quite expensive, operate with difficulty if mounted on a moving acquisition system, and typically require some time to perform a complete 2-D acquisition.

Passive multi-camera systems are cheaper and more robust, but have a range limited to a few tens of meters; moreover, they are obviously more complex, expensive and heavier than a possible single-camera system. The latter, however, typically adopts quite complex algorithms [5–7], often based on neural networks, to provide the distance information; these algorithms may be unsuitable for real-time applications since they require massive computing resources and actually provide fairly approximate results.

To overcome these limitations, we have designed a simple and effective method that can operate in real time using just a single camera. The developed system, mounted on a moving platform (drone, aircraft, operating machine, ...) uses the principles of stereoscopic vision but takes advantage of the different acquisitions made by a single camera during the platform motion.

The depth estimation algorithm that we propose in this paper can be realized using a low-cost FPGA. The proposed architecture works at the cadence of the input pixels and is therefore independent of the resolution of the input images. Furthermore, the proposed design uses only operators no more complex than an adder and a memory as large as only a fraction of a frame memory.

57.2 The Algorithm

The method we developed to extract the distance information from the image sequence acquired by the camera relies on the following constraints:

- the optical axis of the camera should be orthogonal to the movement direction;
- the camera should move at an approximately constant speed;
- the direction of the motion should be close to the horizontal axis of the acquired frame.

Coarsely speaking, the drone should follow at a steady pace a straight path, orthogonal to the axis of the camera. It should be noted that the third constraint is not fundamental for the method, but, when verified, permits a major simplification in the implementation of the algorithm. Moreover, this constraint is not particularly binding and can be simply obtained by rotating the camera around its optical axis. Thanks to this particular configuration, the distance information can be inferred by the relative motion of the various objects in the scene. Taking advantage of the

parallax, in fact, the objects closest to the camera appear to move faster than those placed in the background; objects placed at infinite distance will appear practically still. Moreover, thanks to the third constraint, this motion is purely horizontal. The devised method estimates the apparent horizontal speed of the various parts of the scene and, from these estimates, their distance from the camera.

To evaluate the movements within a sequence of images, a technique commonly adopted in the literature is block matching. For our applications, however, we have some information about motion which can be exploited to simplify the search. Indeed, we already know that the motion will proceed mostly along the horizontal direction, with a known orientation and a limited entity.

The main idea is therefore to switch from a standard 2-D block matching to a 1-D matching of the vertical projections of the blocks. Two blocks match if the vectors of the averages of their columns are similar.

Considering two video frames having size $m \times n$, the first phase of the algorithm consists in sectioning the input images into horizontal slices of size $b \times n$, where b is the size of the chosen block. These slices can partially overlap. If o_v is the overlap amount in pixels, the total number of slices of the image will be

$$u = \left\lfloor \frac{m}{b - o_v} \right\rfloor \tag{57.1}$$

For each slice we then proceed to calculate all the vertical averages of the pixels:

$$p_i(j) = \frac{1}{b} \sum_{k=0}^{b-1} x_i(k, j) \tag{57.2}$$

where $p_i(j)$ are the projections of the pixels of slice i and $x_i(k, j)$ represents the pixel in position (k, j) in slice i. The projection is further segmented into a suitable number of blocks which may also partially overlap by o_h positions; then, the number of blocks available from each projection is

$$v = \left\lfloor \frac{n}{b - o_h} \right\rfloor \tag{57.3}$$

For each block l belonging to the slice i, the algorithm searches for the best matching block on the same slice in the second image; let $d_{i,l}$ be its shift with respect to the position of the reference block. To determine the best match we minimize the SAD (sum of absolute difference)

$$d_{i,l} = \min_d \sum_{k=0}^{b-1} \text{abs}[p_i(s + k) - p'_i(s + k + d)] \tag{57.4}$$

where

$$0 \leq l < v, \quad 0 \leq i < u, \quad s = l(b - o_h) \tag{57.5}$$

The adoption of a 1-D matching yields several advantages:

- the memory requirement is reduced by a factor approximately equal to $b - o_v$, since it is sufficient to keep only the projections data instead of the whole image;
- the matching operations are vastly simplified: working in 1-D requires to perform only b differences for each match instead of $b \times b$;
- small displacements of the camera in a direction orthogonal to the motion are automatically filtered: indeed, the adopted projection can be interpreted as a vertical low-pass filter that makes the system less sensitive to such variations.

Searching for the minimum in Eq. 57.4 would require to repeat the summation for a large range of possible d values. However, taking advantage of the hypothesis that the motion has a known direction and a limited magnitude, the search can be limited to values of $0 \leq d \leq d_M$, where d_M represents the expected maximum displacement. This is useful to reduce not only the processing time but also the rate of wrong matches, as we limit the search within the most probable area. The values of $d_{i,l}$ represent an estimate of the distances of the various portions of the images from the camera. To get rid of some sparse matching errors we employ a small 2-D median filter at the end of the process.

57.3 Architecture

The proposed method presents several features which can be fruitfully exploited in a real-time realization:

- it requires a very small memory;
- it can be realized using operators not more complex than an adder;
- it can be highly parallelized by organizing the operations into a pipeline; in this way the system can work at the input pixels rate, and therefore the implementation is suitable for any desired video resolution.

The system adopts as input a video stream which sequentially provides, line by line, all the pixels coded with an 8 bit gray levels representation. The processing, organized in a pipeline, can subdivided in the following steps:

- **Projection processing**: Using port A of a true dual port memory M_p, all vertical projections of the pixels blocks are calculated pixel by pixel loading the previous data from the memory, adding to them the present pixel, and re-storing the results in real time. When all the pixels constituting the slice have arrived, the memory address pointer moves to the next row and proceeds to evaluate the projections for the next slice.

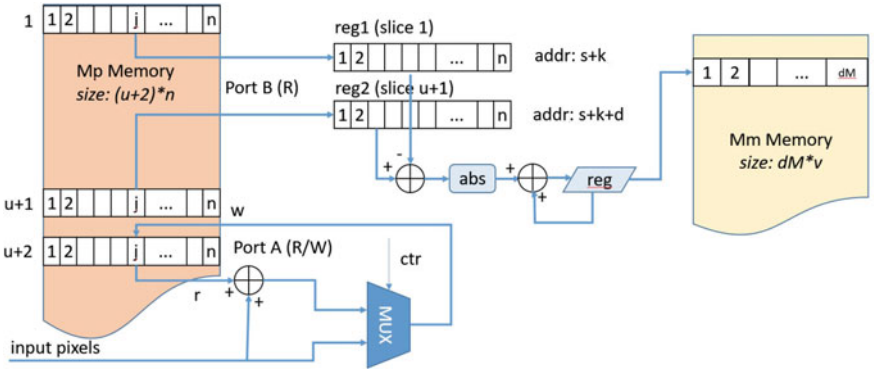


Fig. 57.1 Simplified block scheme (without the control system) for the first two algorithm steps: the projection processing and the matching evaluation

– **Matching**: Using the second port of the memory M_p (port B), the system accesses the data belonging to the area of interest and stores the projections values into two local buffers. From there, through appropriate pointers the system accesses the values of $p_i(s+k)$ and $p_i(s+k+d)$ (see Eq. 57.4) to determine the absolute value of the difference, and accumulates the results in a temporary register by varying k. When k reaches the value $b-1$, the data in the register is stored to the location (d, s) of a memory M_m.
  An approximate block scheme of these two steps is depicted in Fig. 57.1.
– **Minimum estimation**: By sequentially analyzing the data stored into the memory M_m, a dedicated circuit searches for the minimum in each row and determines its location. The location data, which represent the estimated displacement of the block, is stored in a M_D memory composed by a suitable number of shift registers with length v that feed the following median filter. The number of shift registers corresponds to the size of the median filter.
– **Median Filter**: To reduce local errors, the data present in the memory M_D are filtered through a suitable (typically 3×3) median filter realized through a systolic array of comparators [8, 9] before being supplied as output data.

Some overlap among the blocks is desirable, since it permits to improve the resolution by increasing u and v. Indeed, both the slices of pixels on which the projections are calculated and the blocks on which the matching is performed can overlap. This solution obviously requires more resources, but their organization can be easily planned by maintaining the same pipeline operation and simply replicating the structure described above.

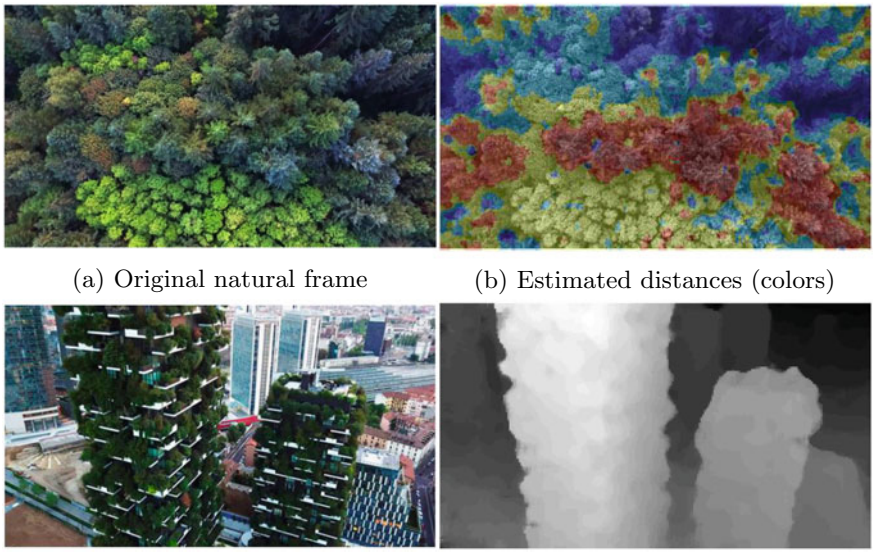
57.4 Simulation Results

The method has been tested using several sequences compliant with the constraints reported in Sect. 57.2. We show here two relevant results.

In Fig. 57.2a a drone flies above a forest and takes up a view of the underlying elements. The estimated distances are reported in false colours superimposed to the original frame (Fig. 57.2b).

By increasing the time delay between the pair of used frames, the distance can be estimated also for far away objects: in Fig. 57.2c the drone acquires a lateral view of the "Vertical Forest" skyscrapers in Milan: it can be noted from Fig. 57.2d (shown in gray levels for a better visualization) that the system is able to discriminate the distance not only of the elements in the foreground, but also those of the most distant ones such as the skyscrapers in the background. Both videos have been taken with HD 1080 resolution at 30 f/s; in the first sample we have used all the frames, while in the second one the time delay between the considered frames has been increased by a factor 10.

It should be noticed that in these first experiments only the relative distances of the objects have been assessed. However, using further information always available in practical cases (speed of the drone, data of the optics and from other sensors of the on-board camera), the distances can be determined in a quantitative way.



(a) Original natural frame (b) Estimated distances (colors)

(c) Original urban frame (d) Estimated distances (b/w)

Fig. 57.2 Frames extracted from sequences acquired by a drone flying horizontally and taking up an above (**a**, **b**) or side (**c**, **d**) view

The proposed system has been synthesized on a Cyclone V—5CSEMA5F31C6—FPGA. Considering a Full HD 1080p input video, the resource required are: 103 out of 397 available M10k RAM blocks (26%) and a logic utilization (in ALMs) about 2%. The maximum estimated frequency is about 115 MHz, which is compatible with most video standards.

References

1. Saracin CG, Dragos I, Chirila AI (2017) Powering aerial surveillance drones. In: 2017 10th international symposium on advanced topics in electrical engineering (ATEE), Mar 2017, pp 237–240
2. Schning J, Heidemann G (2016) Taxonomy of 3D sensors—a survey of state-of-the-art consumer 3D-reconstruction sensors and their field of applications. In: Proceedings of the 11th joint conference on computer vision, imaging and computer graphics theory and applications, vol 3: VISAPP (VISIGRAPP 2016). INSTICC, SciTePress, pp 192–197
3. Foix S, Alenya G, Torras C (2011) Lock-in time-of-flight (ToF) cameras: a survey. IEEE Sens J 11:1917–1926
4. Xharde R, Long B, Forbes D (2006) Accuracy and limitations of airborne LiDAR surveys in coastal environments. In: 2006 IEEE international symposium on geoscience and remote sensing, Jul 2006, pp 2412–2415
5. Duan X, Ye X, Li Y, Li H (2018) High quality depth estimation from monocular images based on depth prediction and enhancement sub-networks. In: 2018 IEEE international conference on multimedia and expo (ICME), Jul 2018, pp 1–6
6. Wang J, Liu H, Cong L, Xiahou Z, Wang L (2018) CNN-monofusion: online monocular dense reconstruction using learned depth from single view. In: 2018 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct), Oct 2018, pp 57–62
7. Zhang Z, Xu C, Yang J, Gao J, Cui Z (2018) Progressive hard-mining network for monocular depth estimation. IEEE Trans Image Process 27:3691–3702
8. Hu Y, Ji H (2009) Research on image median filtering algorithm and its FPGA implementation. In: 2009 WRI global congress on intelligent systems, vol 3, May 2009, pp 226–230
9. Cadenas J (2015) Pipelined median architecture. Electron Lett 51(24):1999–2001

Part XI
Digital Circuits and Systems

Chapter 58
Integration of a SpaceFibre IP Core with the LEON3 Microprocessor Through an AMBA AHB Bus

Gianmarco Dinelli, Gabriele Meoni, Pietro Nannipieri, Luca Dello Sterpaio, Antonino Marino and Luca Fanucci

Abstract Nowadays, requirements for satellite electronics are becoming more stringent due to the increasing complexity of space missions. In particular, data rate requirement is growing up due to the adoption of high-speed payloads such as Synthetic Aperture Radars and hyper-spectral imagers that overcome the capability of state-of-the-art on-board data handling system. The European Space Agency answered to this request introducing a new high-speed communication protocol, SpaceFibre. At the same time, data collected by high-speed interfaces may be processed on-board with specific hardware or general-purpose microprocessor such as the LEON3. The aim of this paper is to describe the integration of a SpaceFibre IP core in the Cohbam Gaisler GRLIB library, to integrate the functionalities offered by the SpaceFibre CODEC with the potential of the LEON3 microprocessor. Implementation results on a Xilinx Virtex-6 and an analysis of the performance of the SpaceFibre interface on an AMBA 2.0 AHB bus are presented.

Keywords SpaceFibre · Data handling · LEON3 · Satellite communication · FPGA

G. Dinelli (✉) · G. Meoni · P. Nannipieri · L. Dello Sterpaio · A. Marino · L. Fanucci
Department of Information Engineering, University of Pisa, Via G. Caruso 16, 56122 Pisa, Italy
e-mail: gianmarco.dinelli@ing.unipi.it

G. Meoni
e-mail: gabriele.meoni@ing.unipi.it

P. Nannipieri
e-mail: pietro.nannipieri@ing.unipi.it

L. Dello Sterpaio
e-mail: luca.dellosterpaio@ing.unipi.it

A. Marino
e-mail: antonino.marino@ing.unipi.it

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

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_58

58.1 Introduction

In the last years, data rate requirement for satellite on-board data handling systems continuously grew: for example, payloads such as Synthetic Aperture Radars (SARs) and hyper-spectral imagers require high-speed communication interfaces, able to transfer several Gb/s. The European Space Agency (ESA) answered to this request introducing a new protocol, SpaceFibre (SpFi) [1], a multi-Gb/s high-speed serial link that supports data rate up to 6.25 Gb/s per lane. SpaceFibre is a candidate to become the successor of the SpaceWire (SpW) protocol [2], which is the state-of-the-art for spacecraft on-board communication, supporting a maximum data rate of 200 Mb/s. SpFi is backward compatible with SpW at packet level and can operate on both copper cable and optical fibre. The protocol stack described in the SpFi standard [3] is composed of:

- **Network layer**: it is responsible for transferring data packets over a SpFi network. The Network layer is optional and can be omitted for point-to-point communication link.
- **Data Link layer**: it is responsible for the Fault Detection Isolation and Recovery (FDIR) mechanism, which resend a data packet in case a communication error occurs. It handles independent flows of information through Virtual Channels (VCs) and the broadcast service.
- **Multi-lane layer**: it allows to parallelize the communication up to 16 lanes. The Multi-lane layer is optional.
- **Lane layer**: it is responsible for establishing the communication between the two ends of the communication link. Data words are encoded/decoded using 8b/10b encoding/decoding.
- **Physical layer**: it is responsible for sending/receiving data over the physical link.

For more details about the architecture of protocol stack layers please refer the SpaceFibre standard [3].

The GRLIB Intellectual Property (IP) library by Cohbam Gaisler is a collection of IPs (i.e. Ethernet interface, memory controller, etc.) interconnected through an Advanced Microcontroller Bus Architecture (AMBA) 2.0 Advanced High-Performance Bus (AHB). It also includes the LEON3, a 32-bit Reduced-instruction-Set-Computing (RISC) processor compliant with the Scalable Processor ARChitecure (SPARC) V8, available under the GNU GPL license [4]. The LEON3 microprocessor exploits a SPARC V8 instruction set, and it has a 7-stage pipeline and a Floating-Point Unit (FPU) compliant with the IEE-745 FPU standard. The LEON3 processor is compatible with the AMBA 2.0 AHB bus interface and runs up to 125 MHz on FPGAs, guaranteeing 1.4 DMIPS/MHz. A fault-tolerant and Single Event Upset (SEU) proof version of the LEON3, the LEON3FT, is available, and its Single Event Effects (SEEs) performances has been evaluated [5].

The LEON3 processor was exploited in several ESA missions such as CHEOPS [6] and PROBA-3 [7] and National Aeronautics and Space Administration (NASA) missions such as Lunar Flashlight, INSPIRE and MarCO [8]. The LEON3 found

different applications inside the avionic system. Indeed, in many space systems it is adopted as the On-Board Computer (OBC) [9], whose functionalities and architecture are specified by Space AVionics Open Interface aRchitecture (SAVOIR) initiative [10]. In this case, the use of high throughput data links is necessary for data transmission from the payload to the platform, (i.e. science data that are necessary for the control of a platform) and for transmitting platform telemetry data, using payload telemetry hardware, as described in [10].

Even when it is used for different aims, applications may require a medium/high-speed data link for communicating with the LEON3 processor. For instance, in CHEOPS [6], the LEON3 is used to process sensory data and it is interfaced through the SpW protocol. In the CCSDS File Delivery Protocol (CFDP) IP Core present in ESA portfolio, a LEON3 processor is exploited to implement various IP functionalities, and a high-throughput data link (e.g. Ethernet, SpW) is used to realize the UnitData Transfer layer [11].

In view of that, the aim of this paper is to describe the integration of the IngeniArs SpaceFibre IP core [12] in the GRLIB IP library in order to exploit the high data rate capabilities offered by the SpFi standard, with the features of the state-of-the-art space qualified LEON3 microprocessor. In Sect. 58.2, a description of the system architecture is presented. In Sect. 58.3, implementation results for a Xilinx Virtex-6 are presented and discussed. Finally, in Sect. 58.4 conclusion are drawn.

58.2 System Architecture

The architecture of the proposed system is shown in Fig. 58.1.

The system is structured according to the Von Neumann architecture with instructions and data store in the same memory, a Double Data Rate 3 (DDR3) Synchronous Dynamic Randam Access (SRAM). The system is composed of:

- **LEON3 microprocessor**: it is the core of the system responsible for data elaboration.
- **MIG memory controller**: it allows the communication with an external DDR3 memory.
- **Interrupt controller**: it is responsible for handling interrupt requests.
- **JTAG**: it is used to download the bitstream. The bitstream is the file that contains the programming information of an FPGA.
- **SpaceFibre IP core**: it is the high-speed interface of the system. It is interfaced with the AHB bus through an AHB master interface. The Tx DMA receives data from the AHB bus and pass them to the SpFi codec. The Rx DMA passes the data received by the SpFi CODEC to the AHB bus. The SpFi is externally interfaced with a high-speed SERializer/DESerializer (SERDES).
- **SpaceFibre register file**: it is configurable through an Advanced Peripheral Bus (APB) slave interface. It configures DMAs and SpFi IP core operations.
- **APB/AHB bridge**: it allows to connect a APB peripheral to a AHB bus.

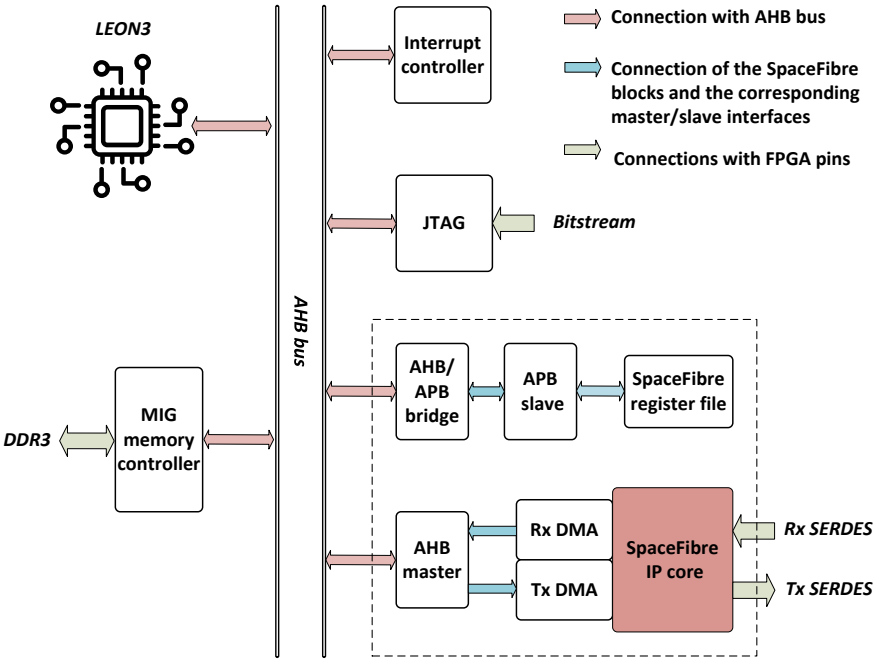


Fig. 58.1 Architecture of the GRLIB-based system with the SpaceFibre IP core

58.3 Implementation Results

The system described in Sect. 2 has been implemented on a Xilinx Virtex-6 ml605 Evaluation kit, mounting a XC6VLX240T-1FFG1156 FPGA. Implementation results are presented in terms of Look-Up-Tables (LUTs) (Virtex-6 family combinatorial logic is based on 6-input LUTs), Registers (Reg), Block RAM18 (18 kb block RAM) and block RAM36 (36 kb block RAM) necessary for the implementation of the proposed design. The percentage of used resources out of the total is also indicated. The LEON3 maximum frequency for the target FPGA is 80 MHz. SpFi CODEC target frequency is 62.5 MHz that guarantees a data-rate of 2.0 GHz (the protocol transmits/receives a 32-bit word per clock cycle to/from the bus). The implementation of the LEON3 requires also four Digital Signal Processors (DSPs) (not included in Table 58.1).

58.4 Performance Analysis

The integration of the SpaceFibre IP core in the Cohbam Gaisler GRLIB is intended to exploit the potential of the AMBA 2.0 AHB bus, which supports high-bandwidth

Table 58.1 Resource occupation for the GRLIB-based system and for the SpaceFibre IP core on a Xilinx Virtex-6 FPGA

	LUT	LUT (%)	Reg.	Reg. (%)	Block Ram18	Block Ram18 (%)	Block Ram36	Block Ram36 (%)
SpFi interface	5245	3	3629	1	12	1	0	0
Leon3/GRLIB peripherals	27842	18	16579	5	8	1	26	3
Total	33087	21	20208	6	20	2	26	3

operation. In the architecture described in Fig. 58.1, the LEON3 microprocessor and the SpaceFibre IP core compete for accessing the DDR3 memory, where both instructions and data are stored. Considering that the AHB bus transfers 32-bit data words at a clock frequency (f_{bus}) of 80 MHz, it offers a maximum bandwidth of 2.560 Gb/s. The data rate (δ) achievable on the bus (expressed in Gb/s) by a generic master interface can be calculated as in Eq. (1), considering f_{bus}, the master interface bus occupation (β) and the number of bits per word (n) transmitted on the AHB bus (32).

$$\delta = \beta \cdot f_{bus} \cdot n \tag{1}$$

The SpaceFibre IP core master interface has a data rate δ_{spfi} of 2 Gb/s. To transfer data at full speed its bus occupation β_{spfi} shall be as shown in Eq. (2):

$$\beta_{spfi} > \delta_{spfi}/(f_{bus} \cdot n) = 78.12\% \tag{2}$$

That means that for lower values of β_{spfi}, the SpaceFibre IP actual data rate is reduced owing to the limited availability of the AHB bus. On the other hand, if the IngeniArs SpaceWire IP core is considered [13], a maximum data rate δ_{spw} of 200 Mb/s is available. For this reason, to transfer data at full speed it requires a bus occupation β_{spw}, as shown in Eq. (3):

$$\beta_{spw} > \delta_{spw}/(f_{bus} \cdot n) = 7.81\% \tag{3}$$

These results suggest SpaceWire represents the bottleneck in the data transfer, when it is possible to guarantee a bus occupation higher than 7.81%. In view of that, in these cases SpaceFibre is a better choice, since it guarantees a higher communication throughput, since it allows to exploit the available capacitance of the AMBA 2.0 AHB bus.

58.5 Conclusions

In this paper, the integration of the IngeniArs SpaceFibre IP core on the Cohbam Gaisler GRLIB library is presented in order to combine SpFi high data rate capability with the state-of-the-art space qualified LEON3 microprocessor. This platform can represent an enabling technology for future high-speed elaboration system, involving the newest high-bandwidth spacecraft payloads. In particular, a system architecture interconnected through an AHB bus, including the LEON3 and the IngeniArs SpaceFibre IP core is described. Furthermore, implementation results of the proposed architecture on a Xilinx Virtex-6 are presented and an analysis of the SpFi achievable data rate on an AMBA 2.0 AHB bus is discussed.

Acknowledgements IngeniArs SpaceFibre technologies have been developed in the framework of the project SIMPLE (Sacefibre eIMPLementation design & test Equipment). This project has received funding from the European Union Horizon 2020 research and innovation programme under Grant Agreement No. 757038.

References

1. Parkes SM, McClements C, Mclaren D, Florit AF, Villafranca AG (2015) SpaceFibre: a multi-gigabit/s interconnect for spacecraft onboard data handling. In: 2015 IEEE aerospace conference
2. Parkes SM, Armbruster P (2005) SpaceWire: a spacecraft onboard network for real-time communications. In: 14th IEEE-NPSS real time conference, 2005
3. Space engineering—SpaceFibre—very high-speed serial link, ECSS-EST-50-11C-DIR1, May 2018
4. Cobham Gaisler AB documentation about LEON3 microprocessor and GRLIB. Available at: https://www.gaisler.com/index.php/downloads/leongrlib?task=view&id=156
5. Hafer C, Griffith S, Guertin S, Nagy J, Sievert F, Gaisler J, Habinc S (2009) LEON 3FT processor radiation effects data. In: 2009 IEEE radiation effects data workshop
6. CHEOPS mission. Available at: https://gsp.esa.int/documents/10192/46710/C4000108880ExS.pdf/30d1d3a8-bf14-4a00-b3c9-6f65cecd5ac7
7. PROBA-3 mission. Available at: https://earth.esa.int/web/eoportal/satellite-issions/p/proba-3
8. Imken T, Castillo-Rogez J, He Y, Baker J, Marinan A (2017) CubeSast flight system development for enabling deep space science. In: 2017 IEEE aerospace conference
9. Furano G, Menicucci A, Roadmap for on-board processing and data handling systems in space. In: Dependable multicore architectures at nanoscale. Springer, Cham, pp 253–281
10. Space AVionics Open Interface aRchitecture (SAVOIR). Available at: http://savoir.estec.esa.int/
11. Meoni G, Valverde A, Magistrati G, Fanucci L (2019) Estimating the downlink data-rate of a CCSDS file delivery protocol IP core. In: International conference on applications in electronics pervading industry, environment and society, (in press), 11–13 Sept 2019, Pisa (Italy)
12. Nannipieri P, Dinelli G, Davalle D, Fanucci L (2018) A SpaceFibre multi lane codec System on a Chip: enabling technology for low cost satellite EGSE. In: 2018 14th conference on Ph.D. research in microelectronics and electronics (PRIME)
13. SpaceWire CODEC IP core developed by IngeniArs. Available at: https://www.ingeniars.com/italiano/prodotti/spazio/sw-codec.html

Chapter 59
A RISC-V Fault-Tolerant Microcontroller Core Architecture Based on a Hardware Thread Full/Partial Protection and a Thread-Controlled Watch-Dog Timer

Luigi Blasi, Francesco Vigli, Abdallah Cheikh, Antonio Mastrandrea, Francesco Menichelli and Mauro Olivieri

Abstract The electronics devices that operate in the extreme space environment require a high grade of reliability in order to mitigate the effect of the ionizing particles. For COTS components this can be achieved using fault-tolerant design techniques which allow such design to fulfil the space mission requirements. This paper presents the design and the implementation of one of the Klessydra *F03x* microcontroller soft core family, called the *F03_mini*, which is a RISC-V RV32I compatible fault-tolerant architecture enhanced by a Hardware Thread (HART) full/partial protection and a thread-controlled Watch-Dog Timer module. The core architecture has been synthesized and implemented on an ARTIX-7 A35 FPGA and fault-injection by the meaning of a functional RTL simulation has been performed in order to evaluate the robustness to Single Event Effects (SEE). Experimental results are provided, illustrating the impact and the benefits obtained by the usage of the proposed TMR protection techniques as well as a thread-controlled Watch-Dog Timer.

Keywords Microcontroller core architecture · Fault-tolerance · RISC-V instruction set · Interleaved multithreading · Single event effects · Watch-dog timer

L. Blasi (✉) · F. Vigli · A. Cheikh · A. Mastrandrea · F. Menichelli · M. Olivieri
Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Via Eudossiana, 18, 00184 Rome, Italy
e-mail: luigi.blasi@uniroma1.it

A. Cheikh
e-mail: abdallah.cheikh@uniroma1.it

A. Mastrandrea
e-mail: antonio.mastrandrea@uniroma1.it

M. Olivieri
e-mail: mauro.olivieri@uniroma1.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_59

59.1 Introduction

The electronic devices that operate in the extreme space environment require a high grade of reliability in order to mitigate several effects of ionizing particles [1]. In our design, we considered only soft-errors (SE), such as Single Events Effects (SEE), as we focus on low clock speed (25 MHz) applications.

The usage of Commercial Off-the-Shelf (COTS) components as well as an open-source Instruction Set Architecture (ISA) allow a reduction in cost due to the low volume demand for aerospace applications. From this point of view, the growing interest for an extendable microprocessor Instruction Set Architectures (ISA) has led many companies to support the RISC-V open standard [2, 3].

Since this kind of components are not intrinsically protected at hardware level, a fault-tolerant architecture design is required in order to fulfil with the severe environment requirements as well as with resource availability [4, 5].

This paper describes the design and the implementation of a compact variant of the Klessydra *F03x* microcontroller soft core family (named *F03d* or *F03_mini*) which is a RISC-V RV32I compliant, fault-tolerant architecture enhanced by a TMR-based full/partial Hardware Thread (HART) protection and a Thread Controlled Watch-Dog Timer (TC-WDT) module.

In the following, Sect. 59.2 provides an overview of the core microarchitecture and its compatibility with the RV32I instruction set. Section 59.3 describes the proposed HART full/partial protection techniques, as well as the utilization of the dedicated TC-WDT. Section 59.4 reports experimental results about the FPGA implementations and the HDL fault-injection simulation, and finally Sect. 59.5 provides the conclusions.

59.2 The Klessydra Processing Core Family and RV32I Compliance

The Klessydra processing core family is a set of cores featuring full compliance with the RISC-V instruction set and intended to be integrated within the PULPino microcontroller platform [6]. To date, the Klessydra family [7] includes:

- a minimal gate count single-thread core, Klessydra *S0*;
- a set of multi-threaded low-end IoT-oriented cores, Klessydra *T0x*;
- a set multi-threaded fault-tolerant cores, Klessydra *F03x*, featuring different fault-tolerance techniques;
- a set of multi-threaded cores, Klessydra *T1x*, supporting vector processing acceleration for high-speed controllers in high-end IoT nodes [8].

The Klessydra core family features can be summarized as follows:

- Full compliance with the RISC-V architecture specification (instruction set, control and status registers, interrupt handling mechanism and calling convention);

- Compliance with the standard RISC-V compilation toolchain;
- Interleaved multi-threaded execution of RISC-V HARTs. In particular, each HART has its own Program Counter (PC), Control and Status Registers (CSR) and Registers File (RF) and every HART can send a software interrupt to another HART. A new instruction is fetched from a different PC at each clock cycle, according to the interleaved multi-threading scheme;
- Easy and standardized multi-threading programming interface;
- Core synthesis on FPGA (presently, Xilinx Series 7 implementations have been tested);
- Hardware compliance with the PULPino microprocessor platform, as a pin-to-pin compatible alternative of the PULPino RI5CY and Zero-riscy cores;
- Software compliance with the PULPino microprocessor platform, as compatible I/O memory map, interrupt handler memory map, program/data memory map.

We focus our discussion on the "mini" version of the Klessydra *F03x* core, which implements the 32-bit integer RISC-V machine mode instruction set, namely user-level RV32I base integer instruction set version 2.2 [2] and M-mode privileged instruction set version 1.10 [3].

59.3 The F03_Mini Fault-Tolerant Microarchitecture

In this section we discuss the architectural choices included in the *F03_mini* core in order to minimize area overhead required by fault-tolerance features. The core shares the same baseline architecture as the *T03x* [9, 10], on which classic Triple Modular Redundancy (TMR) has been applied (Fig. 59.1). As opposed to the Klessydra *F03a* core, featuring full TMR protection on all the HARTs supported by the hardware microarchitecture, the *F03_mini* introduces the following general characteristics in order to save hardware resources:

- Different degrees of error protection among the HARTs.
- Reduced set of Counter and Performance Registers.

All the CSRs and PCs are protected using TMR technique, while the non-critical Counter and Performance Registers (CPRs) are not protected at all, in order to reduce the usage of hardware resources.

Klessydra *F03_mini* supports the execution of 3 HARTs. The hardware microarchitecture features a fully-protected datapath for HART0 and a weakly-protected datapath architecture for HART1 and HART2. Actually, the Processing Pipeline (PP) is completely TMR-protected, while only HART0 has a TMR-protected register file. In this way it is possible to reduce the use of resources by reducing the reliability of two HARTs. A limited degree of protection is guaranteed on HART1 and HART2 by the introduction of the TC-WDT. Moreover, the user can implement software protection techniques to prevent processing failures on weak-protected HARTs, by exploiting the thread-communication features offered by the Klessydra architecture. From the

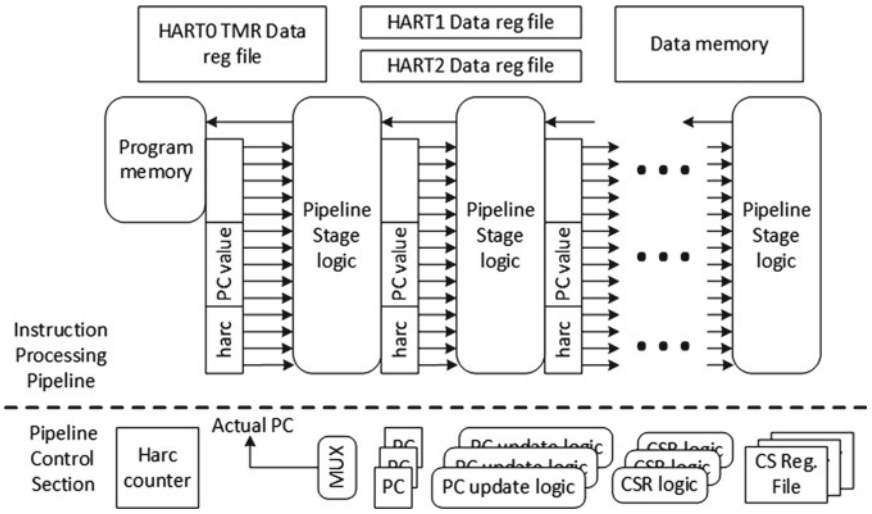


Fig. 59.1 *F03_mini* microarchitecture

application software point of view, HART0 will handle the mission critical tasks of the satellite, while HART1 and HART2 will handle non-critical tasks (Fig. 59.2).

The TC-WDT is a critical component for the correct behaviour of the microcontroller core to be used in the space environment, as it provides a limited degree fault-tolerance for weakly protected HARTs whenever a loss of control due to a SEE occurs within the application program flow.

The TC-WDT can be controlled only by HART0, i.e. the only one which has full TMR protection. In normal operation, i.e. in the absence of critical SEE on HART1 and HART2, all threads will send their reset request (RST_REQ) to the WDT before its timeout has elapsed. The reset command (RST_CMD) o the WDT can be sent only by HART0 once it has verified the correctness of the results for weak-protected HARTs (HART1 and HART2). All the requests and commands are performed with write/read access on memory mapped register (accessed by AMBA Peripheral Bus (APB) interface). The complete reset request sequence is described by the following points:

1. HART1 and HART2 send their reset requests (WDT_RST command).
2. The WDT enables the flags for HART1 and HART2 (in the WDT_CSR register).
3. HART0 checks periodically both flags in the WDT to check the correct behaviour of HART1 and HART2.
4. HART0 requests the WDT reset by the WDT_RST command (Fig. 59.3).

According to the above description, whenever a SEU causes a loss of control within the program flow of HART1 or HART2, HART0 will detect a mismatch when checking the WDT flags. In this case, a dedicated software routine will handle

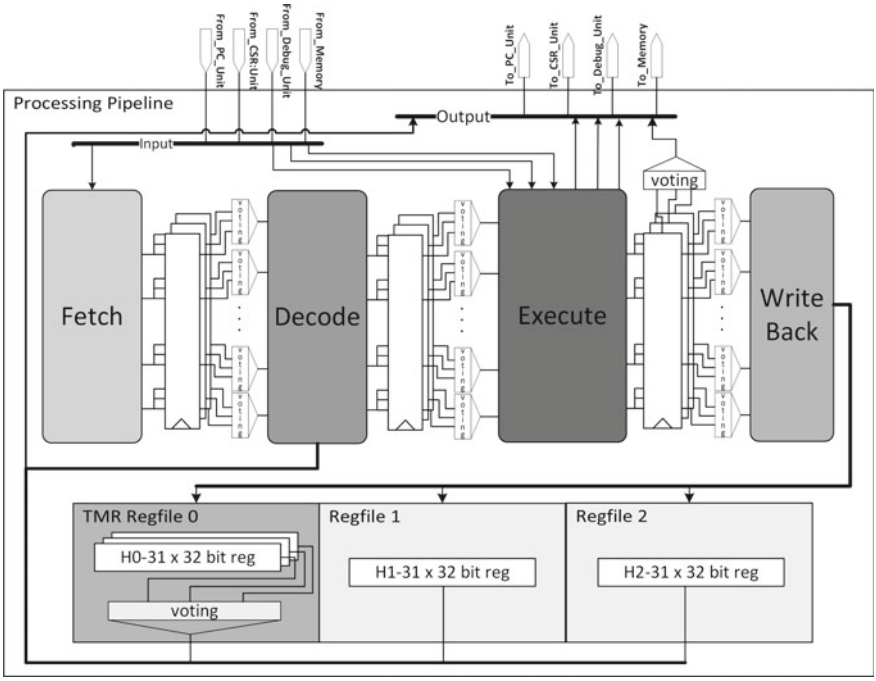


Fig. 59.2 Processing pipeline architecture

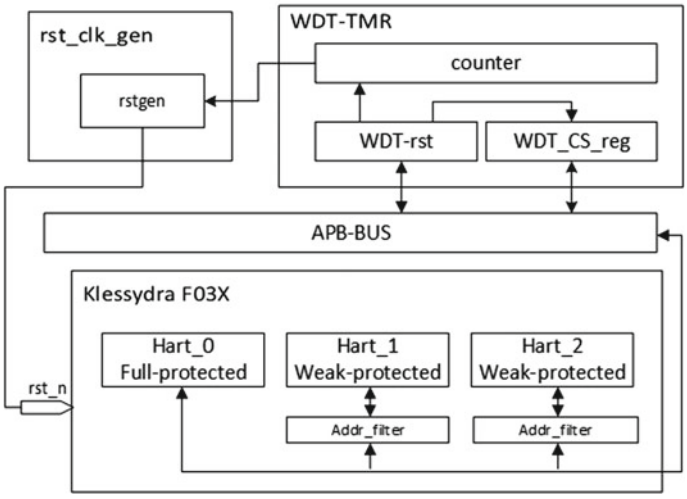


Fig. 59.3 Thread controlled WDT architecture

the error detection in the proper way, which is application dependent. A software support by means of dedicated error recovery routines is therefore required in order to ensure a reliable behaviour.

59.4 Experimental Results

Klessydra *F03_mini* has been coded in VHDL-2008 HDL language and implemented on a Xilinx ARTIX-7 xc7a35tftg256 device. Here we report the essential results related to area, speed and fault tolerance tests.

Table 59.1 provides results for the area usage, compared with the fully TMR-protected *F03a* version.

Table 59.2 reports a group of tests that have been executed to compare the performance between the *F03_mini* fault tolerant core and non fault-tolerant *T03x* core. We can see that the application of the TMR technique in the *F03_mini* cores does not reduce performance in terms of cycle count.

To verify the proposed fault tolerant features, we performed several HDL fault-injection simulation. The tests are based on TCL scripts which force random bit flip in each flip-flop inside the core with a rate up to *18 upset bits/µs*. Table 59.3 provides the results of fault-injection campaign.

Table 59.1 *F03x* versus *F03_mini* resources usage

Architecture	LUT	LUTRAM	FF	IO	BUFG	BRAM	MMCM
F03a	28704	1	20840	28	7	16	1
F03_mini	19465	1	14464	28	7	16	1

Table 59.2 *T03x* versus *F03_mini* performances test

Test name	F03_mini	T03x
testALU	123,131 cycles	123,135 cycles
testCSR	63,098 cycles	63,098 cycles
testIRQ	316,383 cycles	316,383 cycles
testException	43,949 cycles	43,949 cycles

Table 59.3 *F03_mini* fault-injection results for several upset rates @ 20 MHz

F03_mini	6 upset bits/100 µs	18 upset bits/10 µs	18 upset bits/µs
testCSR	6156 µs	6156 µs	6156 µs
testALU	3154 µs	3154 µs	3154 µs
testIRQ	15819 µs	15819 µs	15819 µs

59.5 Conclusions

We illustrated the fault-tolerant microarchitecture used for the implementation of a microcontroller core belonging to the Klessydra *F03x* processing core family, which is compliant with RISC-V integer 32-bit instruction set and with the widely known PULPino System-on-Chip platform. Performance analysis by the means of FPGA area usage and fault-injection by HDL simulation results was also reported, showing the trade-offs between different micro-architecture organizations (*F03x* and *F03_mini*). Future work will be focused on several fault- tolerant techniques based on the intrinsic interleaved multithreading architecture in order to enhance the fault-tolerant of the presented architecture. This work is a fundamental step towards the utilization of RISC-V RV32I microcontroller core as a payload for Nanosatellites (CubeSats, Picosats) which allow academic institutions and small companies to afford space mission research. The first launch of a satellite equipped with a reconfigurable computing sub-system based on *F03x* cores is expected in spring 2020.

References

1. ESA-ESTEC (2016) Space product assurance—techniques for radiation effects mitigation in ASICs and FPGAs handbook. ECSS-Q-HB-60-02A
2. RISC-V User-Level ISA Specification v2.2, Online: https://riscv.org/specifications/
3. RISC-V Privileged ISA Specification v1.10, Online: https://riscv.org/specifications/privileged-isa/
4. Gupta S, Gala N, Madhusudan G, Kamakoti V (2015) SHAKTI-F: a fault tolerant microprocessor architecture. In: IEEE 24th Asian test symposium
5. Blasi L, Mastrandrea A, Menichelli F, Olivieri M (2018) A space-rated soft IP-core compatible with the PIC® hardware architecture and instruction set. Adv Astronaut Sci 163:581–594
6. PULP Platform, Open hardware, the way it should be!, Online: https://pulp-platform.org/
7. Klessydra Processing Core Family Technical Manualv8, 2019. Online: http://github.com/klessydra
8. Cheikh A, Sordillo S, Mastrandrea A, Menichelli F, Olivieri M (2019) Efficient mathematic accelerator design coupled with an interleaved multi-threading RISC-V microprocessor. In: Applications in electronics pervading industry, environment and society. ApplePies
9. Cheikh A, Cerutti G, Mastrandrea A, Menichelli F, Olivieri M (2019) The microarchitecture of a multi-threaded RISC-V compliant processing core family for IoT end-nodes. In: Applications in electronics pervading industry, environment and society. ApplePies 2017. Lecture Notes in Electrical engineering, vol 512. Springer, Berlin
10. Olivieri M, Cheikh A, Cerutti G, Mastrandrea A, Menichelli F (2017) Investigation on the optimal pipeline organization in RISC-V multi-threaded soft processor cores. In: 2017 New Generation of CAS (NGCAS), Genova

Chapter 60
Estimating the Downlink Data-Rate of a CCSDS File Delivery Protocol IP Core

Gabriele Meoni, Alberto Valverde, Giorgio Magistrati and Luca Fanucci

Abstract The Consultative Committee for Space Data Systems File Delivery Protocol (CFDP) is a protocol designed for the transmission of files in space environment, characterized by frequent link disconnections and high transmission delay. This work presents the characterization of the CFDP IP Core included in the European Space Agency (ESA) IP portfolio in terms of downlink data-rate through a custom methodology. For the characterization of the acknowledged/unacknowledged transmission modes, the CFDP IP Core was implemented on board Virtex-5 and Virtex-6 Field Programmable Gate Arrays and tested by using the ESA Ground Segment CFDP reference software, acting as a secondary CFDP entity. The delivered CFDP packets were encapsulated in Unit Datagram Protocol (UDP) packets and transmitted through Ethernet protocol. Wireshark was used to measure the time for a file transmission. The presented methodology provides a way to estimate the IP Core maximum transmission throughput and to identify the architectural bottlenecks.

Keywords CFDP · Downlink throughput characterization · FPGA · LEON3 · Hardware/software · CCSDS · Space

G. Meoni (✉) · L. Fanucci
Information Engineering Department, University of Pisa, Pisa, Italy
e-mail: gabriele.meoni@ing.unipi.it

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

A. Valverde · G. Magistrati
European Space Agency, European Space Research and Technology Centre, Noordwijk, The Netherlands
e-mail: alberto.valverde@esa.int

G. Magistrati
e-mail: giorgio.magistrati@esa.int

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_60

60.1 Introduction

In 2007 the Consultative Committee for Space Data Systems (CCSDS) issued the CCSDS File Delivery Protocol (CFDP) in view of the necessity of a unique file delivery protocol that can transmit files in space environment, characterized by frequent link disconnections, limited availability of bandwidth and high transmission delays [2]. To address the necessity of a broad variety of missions, the CFDP protocol permits to exploit different communication protocols as *UnitData Transfer* (UT) layers and supports *Acknowledged* and *Unacknowledged single hop* (*class 1–class 2*) and *multi hop* (*class 3–class 4*) transmission modes [2].

Nevertheless, the improvements in resolution of onboard instruments, the increment of data storages and the availability of high throughput communication links allowed to increase the quantity of onboard data of space missions, which shall be transmitted to ground [6]. For such reason, research focused to optimize the CFDP protocol to increase the transmission throughput and to reduce the time requested for a file delivery [10]. In view of that, this work presents the characterization in terms of downlink data-rate of the CFDP VHDL Intellectual Property (IP) Core included on the European Space Agency (ESA) portfolio, through a dedicate methodology.

Such IP Core was designed by Braunschweig University [1], which also realized the first complete prototype implementing the CFDP IP Core on ML509 Virtex-5 Field Programmable Gate Array (FPGA)-development board [7]. Moreover, through system level simulations, Braunschweig University provided an estimation of the maximum transmission throughput of such implementation. A new prototype was realized onboard Virtex-6 ML605 board [8] to verify the design portability and to measure the increment in performances due to the use of a FPGA of the next generation. The presented methodology permitted to measure the prototypes downlink data-rate during in *class 1* and *class 2* modes and to delineate the bottlenecks of the architecture, confirming the results of the simulations performed by Braunschweig University.

60.2 Description of the CFDP Prototype

The architecture of the CFDP IP Core, which supports *class 1* and *class 2* transmission and reception, is shown in Fig. 60.1.

Such IP is realized according to a hardware/software codesign. In particular, the *CFDP hardware* realizes an accelerator to control and carry out the different CFDP transactions. The *CFDP software* is integrated in the Real-Time Executive for Multiprocessor Systems (RTEMS) [5], and it is organized in a layered structure composed of three parts: *CFDP Drivers* + *CFDP Entity API*, realizing a CFPD entity; *CFDP client*, which permits to implement different test scenarios. Furthermore, the outgoing and ingoing CFDP Protocol Data Units (PDUs) are encapsulated in User Data Protocol (UDP) packets and transmitted over Ethernet. The communication between

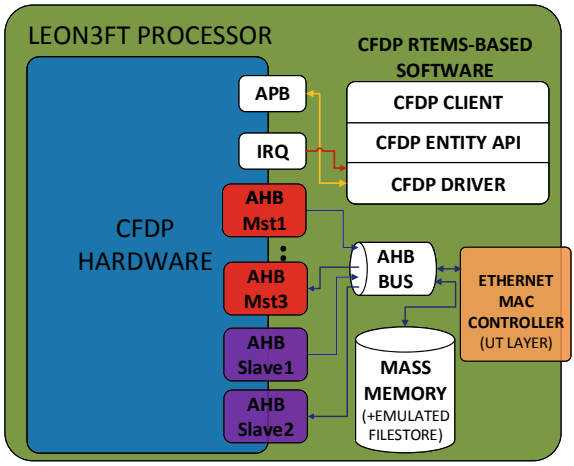


Fig. 60.1 Architecture of the CFDP IP Core

CFDPHardware and *LEON3FT* processor is performed through Advanced High-performance Bus (AHB) bus, used for data transmission, and Advanced Peripheral Bus (APB), exploited for the configuration of the hardware registers. Moreover, Interrupt ReQuests (IRQs) are generated by hardware to trigger the software in presence of particular events. Moreover, the IP Core includes other hardware peripherals, such as a *Mass Memory*, containing system software and the stored files (*VIRTUALIZED FILESTORE*) and a 100 Mbit-Ethernet Media Access Controller (MAC), which is used for the communication with other CFDP entities.

60.3 Characterization Methodology

60.3.1 Performance Characterization Set-Up

The set-up shown in Fig. 60.2 was used for the characterization of the CFDP implementations in terms of the transmission data-rate.

To test the Virtex-5 and Virtex-6 prototypes, the *ESA Ground Segment CFDP* software, provided by the *European Space Operations Centre* (ESOC), was used as a CFDP receiving entity [3]. To exploit its functionalities, the *ESA Ground Segment CFDP* software was run on a Personal Computer (PC) and linked to the prototypes on FPGAs by using a UDP over Ethernet approach, as specified in Sect. 60.2. To run the various tests, different *CFPD client* layers were executed on the LEON3FT processor, as described in Sect. 60.2, through *GRMON2* [4] software. Finally, to observe outgoing packets exchanged between the two CFDP entities *Whireshark* [9] was used. Moreover, the latter permits to get the timestamps correspondent to the

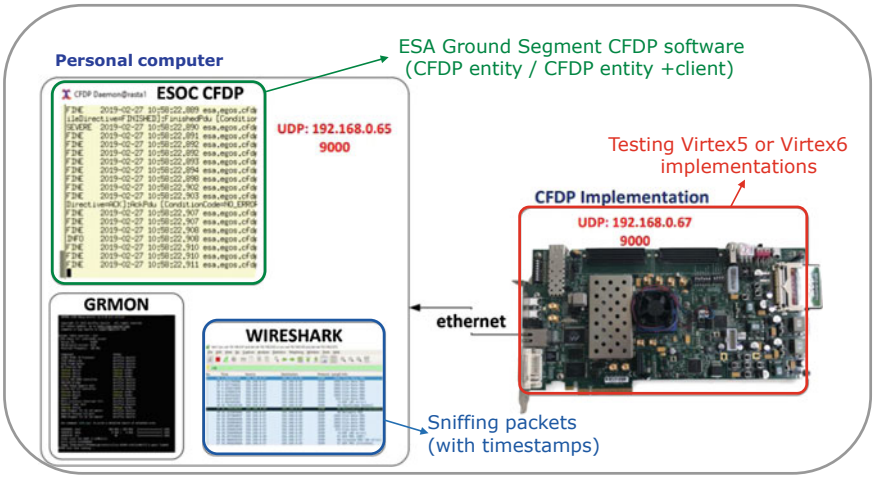


Fig. 60.2 Set-up for performance characterization of prototypes implementing the CFDP IP Core

arrivals of the different PDUs, whose information allows to estimate the transmission data-rate, by measuring the total time necessary for relaying a file of fixed dimension.

60.3.2 Performance Characterization Procedure

The time requested to deliver a file as a function of the system clock frequency in *class 1* and *class 2* modes was measured for both the prototypes. For such aim, multiple syntheses of the CFDP IP Core with different system clock frequencies were carried out for both the FPGA families. This approach permits to measure the dependency of transmission data-rate on the system clock frequency and to identify the value of the clock frequency that leads to the maximum data-rate. Furthermore, the various delivery tests were carried out by transmitting files of different dimensions, such as 5 and 50 kB, to measure eventual dependencies of performances on the file size. As explained in Sect. 60.3.1, the timestamps provided by *Wireshark*, measured in correspondence of the arrival of the PDU packets, were used to estimate the transmission data-rate. In particular, for each test case, which is therefore identified by the following parameters (*FPGA used, file size, system clock frequency, transmission class*), 10 files were transmitted to perform a better estimation of the data-rate. For each file transmitted, the timestamps relatives to the arrival of the first ($T_{F\_F_i}$) and the last ($T_{L\_F_i}$) *FileData PDUs* were acquired. The difference ($T_{L\_F_i} - T_{F\_F_i}$) between these two values corresponds to the time necessary to deliver a file by excluding the service PDUs, i.e., Metadata PDU, End of File, etc. Such information was exploited to estimate the prototype throughput T_{test} for a particular test case, together with the *a priori* knowledge of the file size F_{size}, by using Eq. 60.1:

$$T_{test} = \frac{F_{size}}{\frac{1}{10} \cdot \sum_{i=0}^{9} (T_{L\_F_i} - T_{F\_F_i})} \tag{60.1}$$

60.4 Results

Table 60.1 shows the implementation results of the CFDP IP prototype (including LEON3FT processor) onboard Virtex-5 and Virtex-6 FGPAs in terms of source utilization and maximum system clock frequency $f_{CLK_{MAX}}$. The results of characterization of the transmission data-rate for Virtex6 FPGA are shown in Fig. 60.3. For both the prototypes, in all the cases the transmission data-rate/system clock frequency dependency is well approximated by using a linear trend, described by Eq. 60.2:

$$T_{test}(f_{CLK}) = \theta_0 + \theta_1 \cdot f_{CLK} \tag{60.2}$$

where $T_{test}(f_{CLK})$ is the throughput value in $\frac{Mb}{s}$ and f_{CLK} is the system clock frequency measured in MHz. θ_0, θ_1 parameters were estimated by using a *Least Squares* interpolation, by exploiting the data provided by the different tests. Owing to such linear trend, the maximum transmission throughput is the one measured in correspondence of the maximum clock frequency: Virtex5 = 19.02 Mb/s, Virtex6 = 18.99 Mb/s (PDU data size = 1024 B). Such linear trends demonstrate the CFDP IP Core architecture represents the bottleneck for all the system in the range of system clock frequencies experimented. Indeed, it excludes, for instance, that the UT layer limits the transmission throughput.

In particular, it is possible to notice that T_{test} results higher when the 5 kB files was sent. This fact seems to suggest that performances are dependent on the file size. A

Table 60.1 Implementation results

FPGA	Number of slice LUTs	Number of slice Regs	$f_{CLK_{MAX}}$ (MHz)
Virtex-5 (XUPV5-LX110T)	40.511/69.120 (58%)	27.331/69.120 (39%)	70
Virtex-6 (XC6VLX240T)	91.290/150.720 (60%)	40.511/69.120 (25%)	61.9

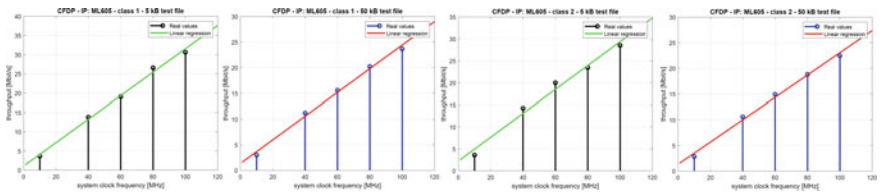


Fig. 60.3 Transmission data-rate/system clock frequency trends for the different test cases on board Virtex6 ML605 board

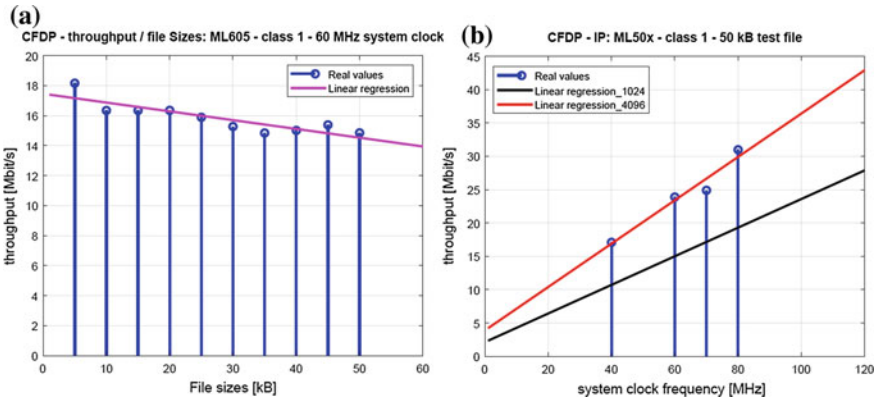


Fig. 60.4 **a** Dependency of throughput performances on file size. **b** Throughput/system clock frequency trends by using different PDU data sizes

possible explanation is that the higher is the file size, the higher is the overhead time due to the numerous accesses and collisions on the AHB bus, owing to the single AHB bus topology and owing to the use of the *Mass Memory* as a temporary storage for those data that require to be processed by different hardware units. Such hypothesis is confirmed by the experiments described in Fig. 60.4. In particular, Fig. 60.4a shows that throughput linearly decreases for growing file sizes by using the same prototype and system clock frequency. In addition, Fig. 60.4b shows that by using two different values of file length transmission, i.e., 1024 and 4096 B, to relay a file of 50 kB, the throughput/system clock trend results lower for all the clock frequencies by choosing the lower packet size. Indeed, a lower PDU data size requires a higher number of bus accesses, by increasing the overhead time. By using a PDU data size of 4096 B, the maximum transmission throughput for Virtex5 resulted of 26.66 Mb/s.

60.5 Conclusions

Preliminary tests confirmed results of the simulations performed by Braunschweig University, which suggested that architecture bottleneck is represented by single AHB bus topology. By transmitting file segments of 1024 and 4096 B, the maximum data-rate for both the implementations was estimated, which resulted to be of 26.66 Mb/s in the best case. More investigations should be performed by using higher file sizes to identify the value of the PDU data size that maximizes the transmission throughput and the measure the correspondent transmission data-rate. Such studies are necessary to provide a full characterization of the IP Core and to figure out the entity of architectural improvements that are necessary to make it usable in modern space platforms.

References

1. Braunschweig University CFDP implementation. https://www.tu-braunschweig.de/c3e/research/cfdp
2. CCSDS file delivery protocol (CFDP) recommended standard (CCSDS 727.0-B-4). https://public.ccsds.org/Pubs/727x0b4.pdf
3. ESA Ground Segment CFDP software. https://www.esa.int/Our_Activities/Operations/gse/CFDP
4. GRMON2. https://www.gaisler.com/index.php/products/debug-tools/grmon2
5. RTEMS. https://www.rtems.org/
6. Valverde A, Taylor C, Magistrati G, Maiorano E, Colombo C, Haddow C (2015) CFDP evolutions and file based operations. In: DASIA 2015—data systems in aerospace, vol 732
7. Virtex-5 ML505/ML506/ML507 evaluation platform user guide. https://www.xilinx.com/support/documentation/boards_and_kits/ug347.pdf
8. Virtex-6 ML605 hardware user guide. https://www.xilinx.com/support/documentation/boards_and_kits/ug534.pdf
9. Wireshark. https://www.wireshark.org/
10. Wu H, Zhang Q, Yang Z, Jiao J, Li Y, Gu S (2016) Double retransmission deferred negative acknowledgement in Consultative Committee for Space Data Systems File Delivery Protocol for space communications. IET Commun 10(3):245–252

Chapter 61
Automatic Detection of the Carotid Artery Position for Blind Echo-Doppler Blood Flow Investigation

Riccardo Matera and Stefano Ricci

Abstract Ultrasound instrumentation is widely employed in everyday clinical practice. Skilled operators, usually trained sonographers, operate the echograph and the ultrasonic probe to acquire B-mode images showing internal tissues/organs morphology, and echo-Doppler images with detailed information about blood flow. Unfortunately, skilled personnel are not always available, in particular in points-of-care distributed in rural areas or developing countries. Echographic systems capable of being operated from non-trained users can be valuable. In this work, an automatic blind procedure for detecting the position of the carotid artery is proposed. The untrained user places the probe transversally on the patient neck and starts the procedure. The machine automatically detects the carotid lumen in the B-mode image and selects its center, which can be used to extract Echo-Doppler data. The method was implemented in the ULA-OP experimental scanner and tested on healthy volunteers.

61.1 Introduction

Thanks to the characteristic of being non-invasive, cheap and reliable, ultrasound techniques are widely used in current clinical practice. Echographic scanners are present not only in every hospital, but they are also common in health centers and point-of-care units.

Ultrasonic tomographic images [1] are made by transmitting ultrasound pulses (typical frequency range 1–10 MHz) into the tissue using a probe composed by hundreds of transducers. The ultrasound pulses, which propagate through the body, are backscattered from the interfaces between tissues with different acoustic properties. Echoes are collected back by the probe and processed to produce a gray-scale image that represents the internal body structures such as organs and vessels (B-mode image). By detecting, in particular, the phase shifts between subsequent ultrasound

R. Matera · S. Ricci (✉)
Department of Information Engineering, University of Florence, Florence, Italy
e-mail: stefano.ricci@unifi.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_61

echoes, it is possible to investigate moving tissue (e.g. heart walls) or flowing particles (e.g. blood cells), and obtain significant information about organs dynamics (echo-Doppler image) [1, 2].

Blood flow investigation of the carotid artery is a common ultrasound exam suitable to investigate general hemodynamic conditions and to monitor the presence of dangerous atherosclerotic plaques. It can represent a life-saving exam [3]. An expert sonographer searches for the correct probe position (e.g. transverse probe position) on the patient neck by checking the B-mode image. In this condition the carotid looks like a dark almost-circular structure. Then the sonographer selects the region of interest (ROI) in the middle of the carotid lumen and switches the echograph in Echo-Doppler modality to save the flow data.

Recently, several very economic and simplified small scanners have been introduced. Some of these are simple smartphone add-ons [4]. Despite the evident limitations of such instrumentation with respect to hospital scanners, these tools can be precious in developing countries, points-of-care located in remote areas or emergency situations. Unfortunately, in addition to the echograph, the presence of an expert sonographer is necessary for exams like the carotid blood flow investigation.

In this paper we present a method for the automatic detection of the carotid artery lumen in the B-mode live image. This method can be used for the automated setting of the ROI in the carotid lumen for acquisition of flow data. The user (not necessarily an expert professional) is requested only to place the probe on the patient neck so that it crosses transversally the artery and to start the automatic procedure.

61.2 Materials and Methods

The method is based on the automatic detection of the position of the carotid artery in a transverse B-mode image. The image is obtained by positioning the ultrasound linear probe on the patient's neck, about at half of its extension. The probe is rotated roughly at 90° with respect to the neck axis. In this way the image plane cuts transversally the common carotid artery (CCA), which, being the blood almost transparent to ultrasound, is represented in the image by an anechoic (dark) region and presents a roughly circular shape. The surrounding tissue has variable echogenicity and, in general, it appears of variable clearer gray tone (see Fig. 61.1 for an example of carotid B-mode image). Other dark portions are present in the image, in particular in the deeper region where the echoes are weaker, but their contours and dimensions are quite discernable from the artery.

From this premise, an image-processing algorithm has good chances to autonomously detect the artery. In the proposed method we employed the Hough transform, a well-established method for detecting curves and shapes in images [5]. In particular, the Circular Hough Transform (CHT) algorithm is adapted for finding circular shapes. The method potentiality was first investigated in Matlab (The Mathworks, Natick, MA, USA). A sequence of 25 B-Mode frames is used for each detection. The image sequence is pre-processed to reduce noise and adjust contrast

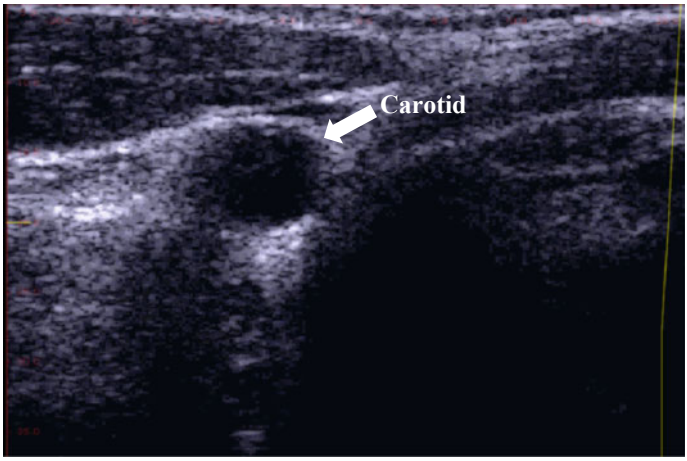


Fig. 61.1 B-mode image of the common carotid artery in a healthy volunteer. The image plane cuts the vessel transversally. The vessel is represented by a dark region of circular shape

and brightness. Figure 61.2 shows the processing steps. For each frame, CHT detects all the dark circles whose radii are within a physiological range. Their center coordinates (x_i, y_i) and radii (r_i) are collected in the so-called Dark Circles Matrix (DCM). For each of the dark circles that have been detected, the brightness values of their pixels are collected in the RGB Matrix (RGBM). The circle showing the darkest value, i.e. the minimum brightness value, is selected as "candidate circle", and its center (x_{ci}, y_{ci}) coordinates with radius (r_{ci}) are saved in the Candidates Matrix (CM). Once the 25 B-Mode frames sequence has been processed, the most frequent (x_c, y_c, r_c) triad occurring in the CM is selected as the final choice.

61.3 Experiments

The method was tested with the experimental scanner ULA-OP [6] managed in real-time by Matlab®, and connected to a LA533 linear probe (Esaote s.p.a). A script running in Matlab® configured the scanner (see Table 61.1) and started the acquisition of the B-Mode frame sequence. The sequence was immediately available in Matlab®. The carotid was detected by the described procedure and the lumen center was used to set-back to ULA-OP the sample volume [1] position suitable for a possible Doppler investigation.

The non-trained user placed the probe on the volunteer's neck roughly in the position where the CCA is expected to be traced, without the help of the B-mode display. The automatic procedure started and about 4 s later it was concluded. As reference, Matlab® saved the B-mode and the estimated CCA position. Figure 61.3 reports an example of automatic segmentation of the carotid. The image shows the

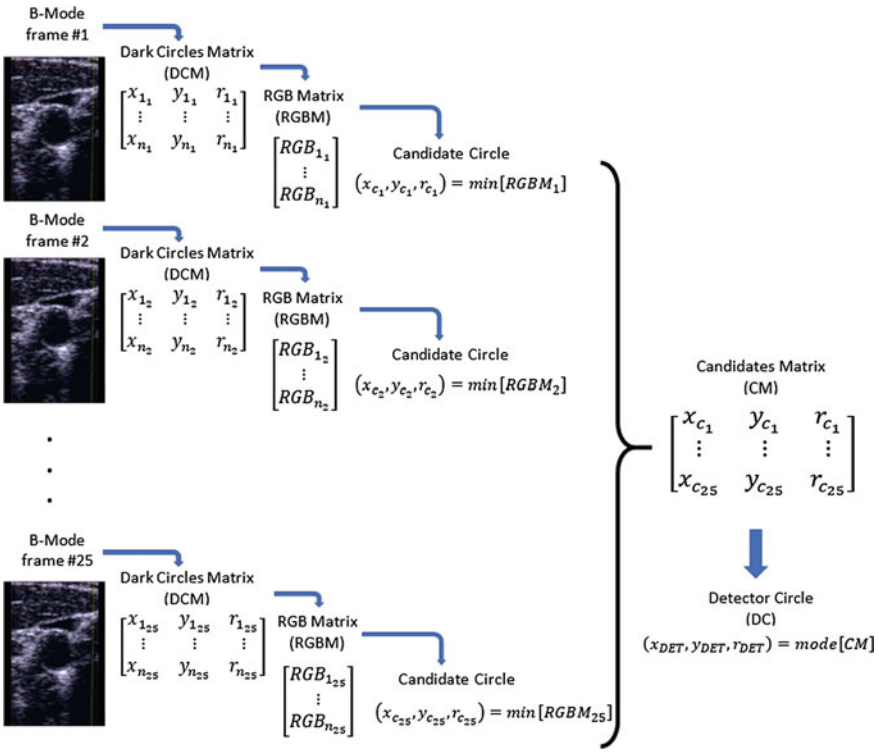


Fig. 61.2 Procedure for locating the carotid position. Candidate circles are located on a sequence of 25 B-mode images. The most frequent one is then selected as output decision

Table 61.1 Transmission/reception parameters used for tests

Parameter	Value
Probe type	Linear
Element pitch	245 μm
Bandwidth	6–11 MHz
Number of elements	192
TX/RX aperture	16 elements (3.9 mm)
TX azimuthal focus	20 mm
Transmission	3 sinusoidal cycles @ 9 MHz
RX focus	Dynamic focusing $F_{\#} = 2$
Image creation	Standard

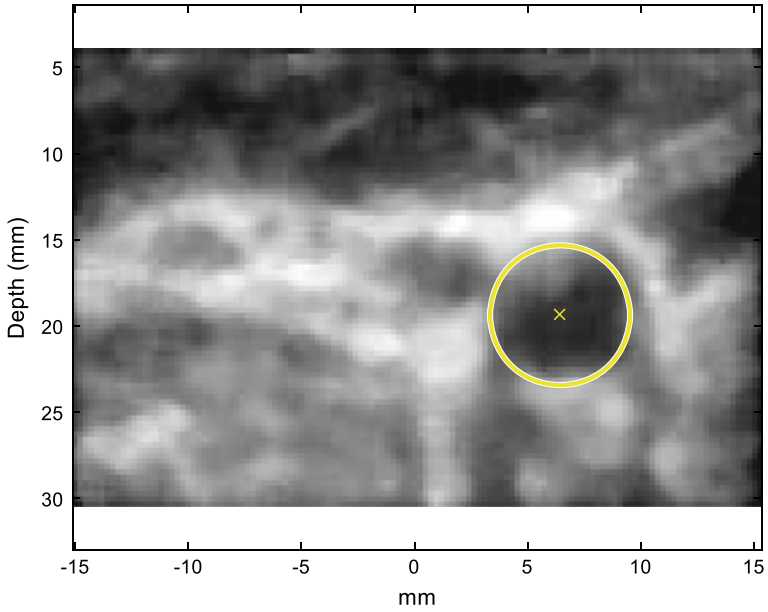


Fig. 61.3 Carotid position (yellow circle) detected in the pre-processed image sequence. 'X' represents the circle center, i.e. the region where the flow analysis should be carried out

result of the pre-processing, and although it appears more confused with respect to the original B-mode image (see Fig. 61.1 as an example), it is more effective when processed by the detection algorithm. The yellow circle represents the position of the carotid automatically located. The yellow 'X' represents the center of the carotid lumen (located as the center of the circle) and is passed back to the scanner as the point where to focus the flow investigation. In the experiments the carotid position was located correctly in about 90% of the tests. When non-located, it was sufficient for the user to slightly move the probe and retry to obtain a correct result.

61.4 Discussion and Conclusion

Compared to other medical imaging methods, ultrasound techniques have several advantages. They provide real-time images, are substantially less expensive and do not use dangerous ionizing radiation. Moreover, very compact and economic systems, tailored for point-of-care or emergency use, are currently available. In this paper we have discussed a simple method for the automatic detection of the position of the carotid artery in a B-mode image sequence. The linear array produces a B-mode image of about 3 cm wide (see Fig. 61.3). Thus, it is quite simple to intercept the carotid in the field-of-view of the probe also for non-expert operators. A bit more attention should be paid to avoid positioning the probe too high in the neck, where

the carotid bifurcates in 2 branches and generates a more complex morphology. In some cases, the jugular vein can be present in the field-of-view, and theoretically it can disturb the localization process. However, according to our tests, this was not a big issue, especially because this vein, when compressed by the probe, is hardly circular.

Once the center of the carotid is located, the echograph can proceed for flow data investigation through echo-Doppler. Even though Doppler investigation at transverse beam-to-flow angle is feasible [7] operating at angles different from 90° is more convenient [1]. This can be achieved also by the non-expert operator by tilting the probe slightly. The carotid section from circular becomes elliptical, but the eccentricity is so low that does not impact the performance of the proposed algorithm.

We are currently working on a special probe, composed by 2 parallel linear arrays with a 30° angle between scanning planes [8]. With this probe, the proposed algorithm can locate the carotid lumen in the B-mode images from both the arrays and proceed to an automatic vector flow investigation [9]. The proposed method can be valuable also for maintaining the correct ultrasound beam position in small "operator-free" ultrasound systems to be used for long-time monitoring of carotid flow in patients [10].

Acknowledgements Authors thank Prof. Piero Tortoli for the valuable advices that contributed to improve the paper. This work is part of the AMICO project funded from the National Programs (PON) of the Italian Ministry of Education, Universities and Research (MIUR): code ARS01_00900 (Decree n. 1989, 26 July 2018).

References

1. Evans DH, McDicken WN (2000) Doppler ultrasound physics, instrumentation and signal processing. Wiley, Chichester. ISBN: 978-0471970019
2. Urban G, Vergani P, Ghidini A, Tortoli P, Ricci S, Patrizio P, Paidas MJ (2007) State of the art: non-invasive ultrasound assessment of the uteroplacental circulation. Semin Perinatol 31(4):232–239. https://doi.org/10.1053/j.semperi.2007.06.002
3. Grant E, Benson C, Moneta G et al (2003) Carotid artery stenosis: gray-scale and doppler US diagnosis—society of radiologists in ultrasound consensus conference. Radiology 229(2):340–346. https://doi.org/10.1148/radiol.2292030516
4. Huang CC, Lee PY, Chen PY, Liu TY (2012) Design and implementation of a smartphone-based portable ultrasound pulsed-wave doppler device for blood flow measurement. IEEE Trans Ultrason Ferroelectr Freq Control 59(1):182–188. https://doi.org/10.1109/tuffc.2012.2171
5. Duda RO, Hart PE (1972) Use of the Hough transformation to detect lines and curves in pictures. Commun ACM 15(1):11–15. https://doi.org/10.1145/361237.361242
6. Tortoli P, Bassi L, Boni E, Dallai A, Guidi F, Ricci S (2009) ULA-OP: an advanced open platform for ultrasound research. IEEE Trans Ultrason Ferroelectr Freq Control 56(10):2207–2216, https://doi.org/10.1109/tuffc.2009.1303
7. Tortoli P, Guidi G, Pignoli P (1993) Transverse Doppler spectral analysis for a correct interpretation of flow sonograms. Ultrasound Med Biol 19(2):115–121. https://doi.org/10.1016/0301-5629(93)90003-7

8. Di Palma V et al (2019) "Medical assistance in contextual awareness" (AMICO): a project for a better cardiopathic patients quality of care. In: Proceedings of IEEE international workshop on advances in sensors and interfaces (IWASI), Otranto, June 2019. https://doi.org/10.1109/iwasi.2019.8791308
9. Ricci S, Ramalli A, Bassi L, Boni E, Tortoli P (2018) Real-time blood velocity vector measurement over a 2D region. IEEE Trans Ultrason Ferroelectr Freq Control 65(2):201–209. https://doi.org/10.1109/tuffc.2017.2781715
10. Song I, Yoon J, Kang J, Kim M, Jang WS, Shin NH, Yoo Y (2019) Design and implementation of a new wireless carotid neckband Doppler system with wearable ultrasound sensors: preliminary results. Appl Sci 9:2202. https://doi.org/10.3390/app9112202

Chapter 62
Efficient Mathematical Accelerator Design Coupled with an Interleaved Multi-threading RISC-V Microprocessor

Abdallah Cheikh, Stefano Sordillo, Antonio Mastrandrea, Francesco Menichelli and Mauro Olivieri

Abstract Interleaved multi-threaded architectures (IMT) have proven to be an advantageous solution to maximize the pipeline utilization, when it comes to executing parallel applications, as different threads operate different instruction processing phases in the same cycle. In this study, we expand the target applications of an IMT microarchitecture by introducing an efficient yet handy special-purpose mathematics engine, operating on local scratchpad memories that give low latency and wide data-bus access.

62.1 Introduction

The RISC-V open instruction set architecture paved the way for computer architects to design innovative capable cores able to execute complex instruction extensions [1]. The instruction set was designed to support 32/64/128 bit architectures in bare metal, supervisor, or user modes [2], and has available encoding space that allows processor designers to augment their own custom instruction set, for educational, research or industrial application purposes.

The family of RISC-V processor cores Klessydra has been designed to support domain-specific features, while fitting in the Pulpino microcontroller platform

A. Cheikh (✉) · S. Sordillo (✉) · A. Mastrandrea · F. Menichelli · M. Olivieri
Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, Rome, Italy
e-mail: abdallah.cheikh@uniroma1.it

S. Sordillo
e-mail: stefano.sordillo@uniroma1.it

A. Mastrandrea
e-mail: antonio.mastrandrea@uniroma1.it

F. Menichelli
e-mail: francesco.menichelli@uniroma1.it

M. Olivieri
e-mail: mauro.olivieri@uniroma1.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_62

[3]. Here we present the Klessydra T13 architecture, which is an extension of the Klessydra T03 version [4, 5, 8] in the domain of computation intensive embedded microcontrollers.

This study presents the features of an efficient accelerator named Special Purpose Mathematical Unit (SPMU) facilitates vector execution on an instruction level basis. Then it shows how this SPMU can be scaled to run fast convolutions on embedded systems, and identify what is the most convenient data level parallelism setting that brings out the most acceleration while still maintaining a relatively small area occupation. Section 62.2 explains the architecture of the accelerator. Section 62.3 shows the synthesis results on the FPGA, and the maximum speed of each layout generated. Section 62.4 shows the performance on the instruction level, and the acceleration contributed by the different implementations in the SPMU, and then again is shown how the SPMU faired when executing a set of convolutions in which it was configured for different data level parallelism settings. The last section determines which configuration gives a good performance boost while still maintaining a relatively small area occupation.

62.2 SPMU Architecture

62.2.1 Klessydra Processor General Architecture Features

Klessydra processor cores support multiple thread execution by means of is interleaved multithreading (IMT) [6]. The T03 core is a four pipeline stage in-order processor that interleaves three or more hardware threads, and supports the RV32IA instruction set in bare metal execution. The T13 core maintains T03 pipeline organization and IMT support, however it decouples the execution stage into three parts to allow a partial superscalar execution. Figure 62.1 shows the block organization of the T13 core. T13 expands the instruction set of T03 with two extensions; the first being the "M" (multiply/divide) extension which is handled in the IE block, and the second is the "K" custom instruction set extension, specifically designed to facilitate vector calculations, that is managed by the SPMU. So, the ISA supported by the T13 core is RV32IMAK.

62.2.2 Special Purpose Mathematical Unit Architecture

The architecture of the SPMU is divided into two main sub-systems. The Special Purpose Engines (SPE), and the Scratchpad Memory Interface (SPI), as shown in the block diagram in Fig. 62.2. The SPMU can be configured with many different parameters.

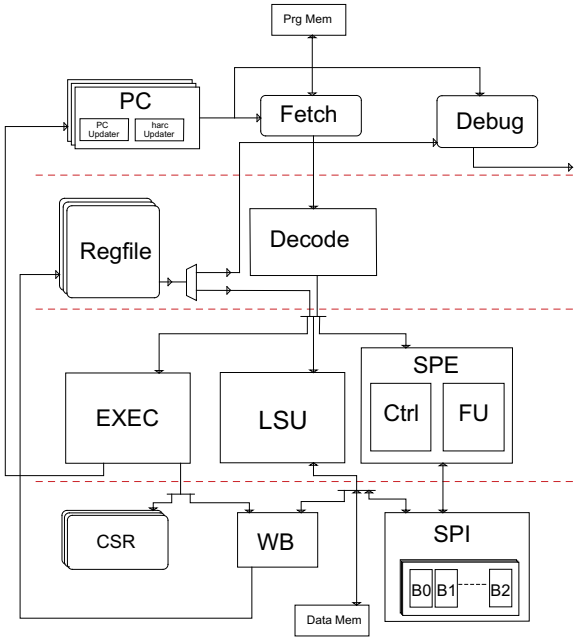


Fig. 62.1 T13 organization

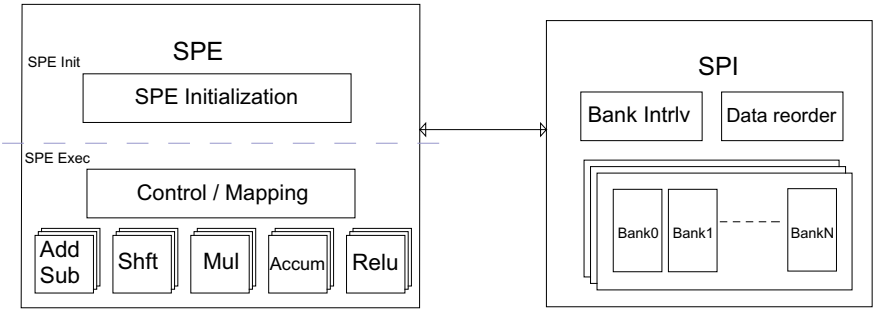


Fig. 62.2 SPMU block diagram

The T13x core comes with multiple configuration parameters to generate a set of different designs:

- The first sets the number of scratchpad memories (SPM); this cannot be set to lower than two, since a two-source operand instruction requires that we read from two different SPMs simultaneously.
- The second sets the SPM size: It can be set to any number which is a power of two. The total size of the SPM will be divided on the number of banks in the SPM.

- The third parameter changes the SIMD capability of the engine, for example; if this value was set to 4, all the functional u-nits become multiplied by 4, and each SPM will have four banks of 32-bit words.

Now this study is NOT going to explore the best number of SPMs per core, since this setting is used per application basis, and every user might utilize his SPM space differently. However, note that the more the user increases the number of SPMs, the more complex the crossbar connecting the SPMs to the SPI will become.

62.2.3 Special Purpose Execution Unit

The SPE is the engine that executes the special purpose instructions with the "K" extension. It is divided into multiple subsystems.

The exception handler is the controller which is a part of the initialization phase that checks and predicts for any exceptions at the very first cycle of the execution of a custom instruction from the "K" extension. The reason behind checking in the first cycle is that in the case of encountering an exception, the core can recover the state of the processor precisely to the time before the exception occurred. After the first cycle, another instruction might be issued to execute in parallel with the accelerator, and the program counter will change its value.

the following is a list of what might trigger an exception:

1. Out of bound SPM access; in this case, one of the pointers to a data element is pointing to an address not belonging to any of the SPM memories.
2. Dual SPM read access; a SPM has one read port, and if the two operands point to the same SPM, we encounter an exception.
3. Overflow data read and write; this happens when the SPM pointer plus the vector size will overflow the address of the SPM being indexed. This overflow exception only traps when the operand being indexed is used as a vector, and not scalar.
4. Misaligned access; SPMs are 32-bit word aligned and any misaligned access will trigger this exception.

The SPE initialization block configures the functional units correctly in order to execute the current instruction in flight. An example of some configurations might be; Setting the *FU* controls to execute the data type to be computed on, such as; *chars*, *shorts* or *ints*. Other configurations might also be to transform the input operands into their two's complement or they might be to configure outputs to either become sign extended or zero extended.

In the execute state of the SPE, the **hardware loops** start incrementing the vector addresses to fetch the next element, and decrementing the vector length. When the vector size becomes zero, the hw-loops stop, and the instruction is considered done. A masking vector is created depending on the number of elements left, such that if the number of elements is less than the number of bytes processed in one cycle, the mask will disable the upper bytes of the fetched elements. This is essential when

elements fetched get accumulated. In this case, we need to avoid accumulating data not belonging to the instruction in order to get a correct accumulation result.

The fetched input operands go into a **mapping unit**, which maps the inputs to their corresponding functional units. The operands can be scalar or vector, and they can be fetched from the SPM or the register file. The outputs of the functional units will connect again to the mapping unit, in which they will be written back to the SPMs.

A control unit in the SPE, controls the fetching of the inputs, and the writing of the results, as well the halting the vector processor in case the source SPMs are being accessed by the load-store unit simultaneously. When the SPE gets a halt signal, all the data in the pipes will maintain their state, and the hardware loops will stop counting until the halt signal returns back to zero, and then the SPE recovers its previous state.

The SPE has five different **functional units (FUs)**. All the units work with different data width (8-bit, 16-bits, 32-bit). Three of the FUs work in partial mode; the adder, shifter, and multiplier. The partial FUs increase the parallelism for smaller data width elements while maintaining a small area occupation. Table 62.1 shows how many operations we do in one cycle in every FU and for each data type when the SIMD parameter is configured to be 1. Bigger SIMD configurations will double the number of parallelisms on all the functional units.

The partial adder as seen in Fig. 62.3 is a set of four 8-bit adders cascaded together. To produce 8-bit sums, no carries are propagated from the partial sums. For 16-bit additions, only the first and the third adders are allowed to propagate their carries, while for the 32-bit sum all carries are allowed to be propagated. However, the adders are split into two pipe stages, so the carry from the lower 16 bits, goes to the upper sixteen bits through a register and not a wire.

For the 32 bit **multiplier** the partial multiplication structure is based on four 16-bit multipliers, according to the following implementation:

$$\boldsymbol{A_{31-0} * B_{31-0} = \left[(A_{31-16} \ll 16) + A_{15-0}\right] * \left[(B_{31-16} \ll 16) + B_{15-0}\right]}$$

This method can generate two 8-bit, or two 16-bit MULs per cycle, or one 32-bit MUL per cycle. The reason this was not divided to do 8-bit partial multiplications instead, was because one DSP slice is utilized in the FPGA whether an 8-bit or16-bit

Table 62.1 Number of operations per FU

FU/data type	8-bit	16-bit	32-bit
Adder	4	2	1
Multiplier	2	2	1
Shifter	4	2	1
Accumulator	2	2	1
ReLu	4	2	1

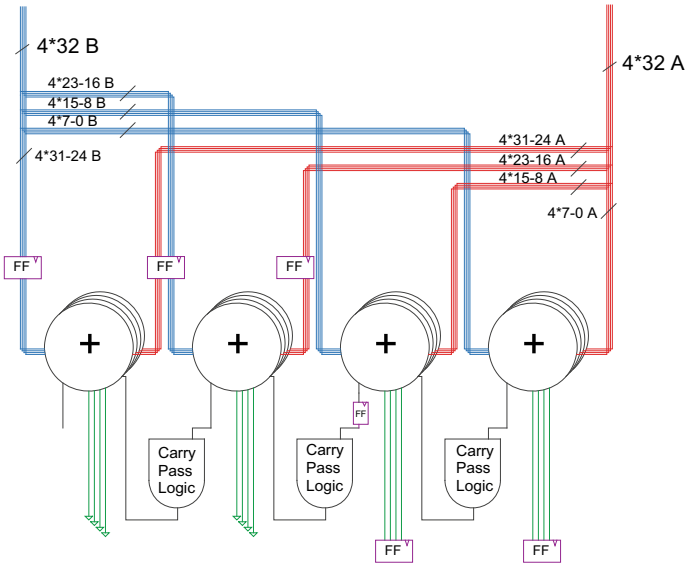


Fig. 62.3 Partial adder circuit in SIMD = 4

multiplication is done. So for vector operations using multipliers, we will only get twice the speed-up for 8-bits of data and not four times as in the case of the partial adders.

The **partial shifter** in the SPE works in the opposite manner (Fig. 62.4). One 32-bit right logic shifter slides the input operands and computes one 32-bit shifted output. If the data width was 16-bits, it will execute as follows: The two 16 bits data

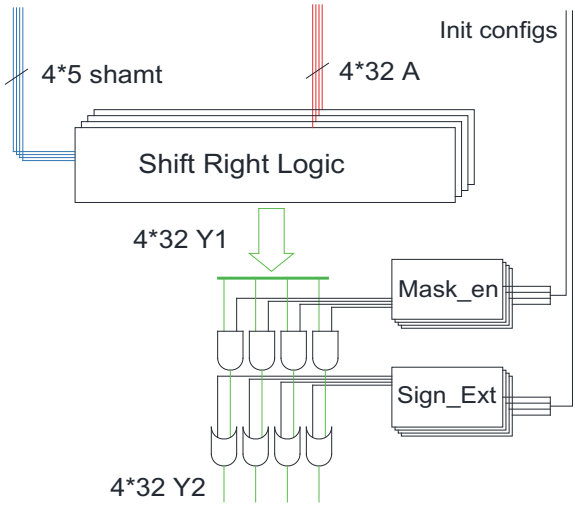


Fig. 62.4 Partial shifter circuit in SIMD = 4

will go into the right shifter, the output of the shifter will be sent to the next stage where the lower bits of the **upper** 16-bit input that were slided into the upper bits of the **lower** 16-bit input will be masked with a zero if the shift was logical sign extended if the shift was arithmetic. A similar approach is applied for 8-bit data types.

The remaining two functional units are a **2-stage accumulator**, which accumulates an input vector source into a scalar output, and a **ReLu unit** that rectifies all negative vector elements to zero.

62.2.4 Scratchpad Memory Interface Unit

The engine is interfaced with a set of SPMs. Each SPM has a read and write port, and every SPM line has a set of banks that hold a 32-bit word. An SPM read or write access will fetch or write an entire line in one cycle. If the fetch pointer was not pointing to the beginning of the line, the data fetched will be from the current line being indexed, and the next line, therefore maintaining the fetching of one complete line per cycle.

Misaligned fetches go into a read-rotator circuit to make it appear as if the fetching is from the beginning of the line. In this manner operand_a[i] will always be aligned with operand_b[i] and go the same functional unit. Without rotation, misaligned accesses might send operand_a[i] and operand_b[i+2] to go to the same functional unit, and that produces erroneous outputs. During the result write, the result to be written will be rotated back to point to the correct bank indicated in the write address.

The SPI has a serialized access grant unit, in which the instruction that comes first in program order will either lock the read or write access of a certain scratchpad. And since T13 is an in-order processor, there will never be data hazards with the serialized access grant.

62.3 Synthesis Results

The T13 core was synthesized with different configuration parameters of the SIMD variable. Synthesis results were generated by Vivado 2018.2. A clock constraint of 1 ns was placed in order to compel Vivado to generate the fastest netlist possible from each configuration. The Genesys 2 was our target board for implementation [7]. Table 62.2 shows the element utilization for each SIMD configuration, while Fig. 62.5 shows the maximum frequency of each element. You can see that the LUT utilization almost doubles when going from SIMD 1 to SIMD 8, and the number DSP and BRAM slices are four times or more. Looking at the maximum operating frequency of each generated layout, we see that the maximum clock frequency of each configuration lies in the same range going from 140 to 150 MHz, we also note that the overhead of the extra SIMD did not affect the net-delay by much.

Table 62.2 Element Utilization on FPGA

		FPGA utilization			
		LUT	FF	BRAM	DSP
FU	1	12681	6673	3	8
	2	14647	7032	3	12
	4	17236	7703	6	20
	8	23068	9059	12	36

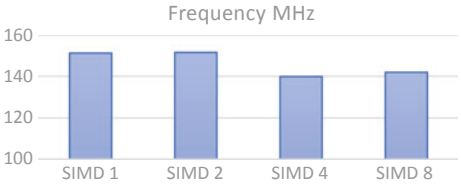


Fig. 62.5 Maximum frequency with different SIMD configurations

62.4 Acceleration Speed Results

62.4.1 Instruction Level Testing

In order to benchmark the SPMU, a set of simple arithmetic tests were run to see what approaches bring forth the biggest performance boosts. Figure 62.6 shows the number of cycles taken to run an arithmetic operation without using the SPMU, that is used as a reference for comparisons. All data types take the same time to execute when not using the accelerator. Figure 62.7 shows the speed when using the SPMU without any SIMD capabilities, and also the fetching of the next element is done by software loops instead of hardware. In this configuration, the speed-up only comes from having burst loads and stores to the SPM, and the low latency in the SPM, so the superscalar execution can present significant boosts for big vectors, while for very small vectors they can be slower than the non-accelerated approach.

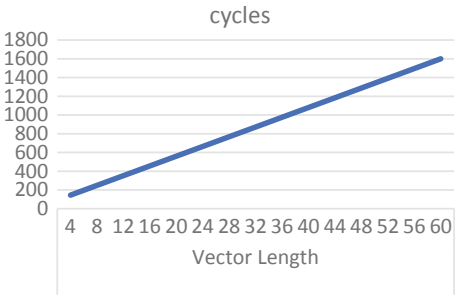


Fig. 62.6 Number of cycles taken to perform an arithmetic vector operation without the SPMU

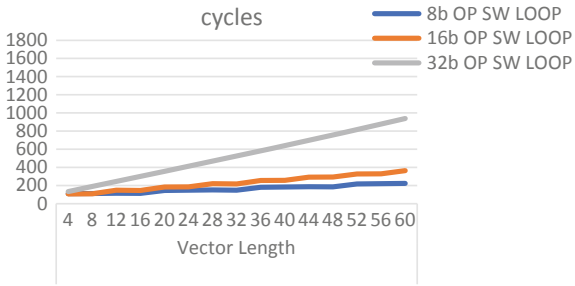


Fig. 62.7 Cycle time using the SPMU with SIMD = 1 and hardware loops disabled

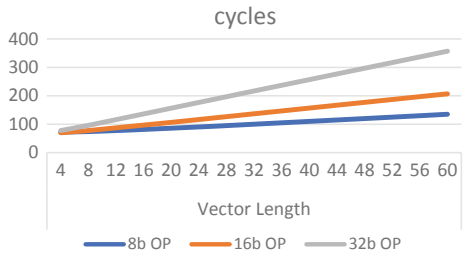


Fig. 62.8 Cycle time using the SPMU with SIMD = 1 and hardware loops enabled

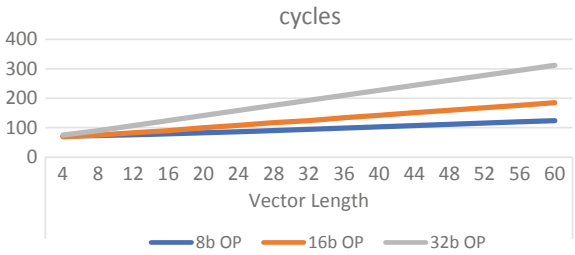


Fig. 62.9 Cycle time using the SPMU with SIMD = 4 and hardware loops enabled

Figure 62.8 shows hardware loop impact, with a speed boost gain over 200% for all vector sizes. Going from SIMD 1 to SIMD 4 as shown in Fig. 62.9 boosts the speed by a small margin for big vectors, and by a barely detectable margin for small vectors. The SIMD boost can be better seen when running more complex tests.

62.4.2 Routine Level Testing

We ran a set of matrix convolutions as shown in Table 62.3 ranging from matrix sizes of 4×4 to 32×32. It appears that SPMU no matter what configuration it is set to, cannot generate good performance when the matrix is 4×4. That is probably is due to a few reasons; the first being that every custom instruction takes from six to ten cycles of latency to process one scratchpad data line, and every other line fetched

Table 62.3 Speed of running a convolution test

SIMD		1	2	4	8	NO ACCEL
Conv_2D	4 × 4	3498	3320	3257	3233	2519
	8 × 8	6030	5653	5521	5453	7038
	16 × 16	12,238	9647	8793	8353	20,411
	32 × 32	37,563	25,021	20,419	18,314	79,474

from the SPM the vector augments one extra cycle of latency. While on the other hand in the normal execute stage, T13 takes from one to three cycles to execute an instruction. Second of all the SPMU instruction operands indirectly reference their data values, while the normal RISC-V instructions do a direct referencing of their data. This indirect referencing requires one to two cycles to create the pointer, plus the interleaving of two threads in every pointer creating instruction we have. So, for creating the three pointers of the source and destination operands, the SPMU adds a significant time overhead if the vectors are small.

It is apparent that as the matrix size gets bigger, the acceleration becomes more apparent, such that for 32 × 32 convolutions we get more than double the speed boost, and that is just by using hardware loops, and the low latency SPMs. While for data-level parallelism, higher-order SIMD configurations got a speed boost of more than four times.

62.5 Conclusions

Our tests show that data level parallelism was the smallest contributor to the speed boosts in the accelerator in the basic tests, however its significance was more apparent for the relatively big applications. The Hardware loops in fact showed the greatest speed improvements while they only contributed to a very small area utilization. The number of cycles in the convolution test show that SIMD 8 configuration saturates in the performance boost, but gives a steady linear growth in the area utilization. So, for our current design, if the matrix sizes do not exceed 32 × 32, we recommend the configuration of the core to be set to SIMD 2 or SIMD 4, in order to get a good boost and maintain a small layout. Finally, it is suggested that small vector calculations less than four elements are preferred to be executed without using the SPMU, while vectors bigger than four elements should be processed by the SPMU.

References

1. Waterman A, Asanovic K (eds) (2017) The RISC-V instruction set manual—volume I: User-Level ISA—Document Version 2.2, May 2017. https://riscv.org/specifications/
2. Waterman A, Asanovic K (eds) (2017) The RISC-V instruction set manual—volume II: Privileged ISA—Document Version 1.10, May 2017. https://riscv.org/specifications/
3. Traber A, Zaruba F, Stucki S, Pullini A, Haugou G, Flamand E, Gurkaynak FK, Benini L (2016) PULPino: a small single-core RISC-V SoC. In: 3rd RISCV workshop
4. Olivieri M, Cheikh A, Cerutti G, Mastrandrea A, Menichelli F (2017) Investigation on the optimal pipeline organization in RISC-V multi-threaded soft processor cores. In: Proceedings of 2017 new generation of CAS (NGCAS). IEEE, pp 45–48
5. Cheikh A, Cerutti G, Mastrandrea A, Menichelli F, Olivieri M (2017) The microarchitecture of a multi-threaded RISC-V compliant processing core family for IoT end-nodes. In: International conference on applications in electronics pervading industry, environment and society. Springer, Cham, pp 89–97
6. Bechara C, Berhault A, Ventroux N, Chevobbe S, Lhuillier Y, David R, Etiemble D (2011) A small footprint interleaved multithreaded processor for embedded systems. In: 2011 18th IEEE international conference on electronics, circuits, and systems. IEEE, pp 685–690
7. Genesys 2 Reference Manual by Digilent. https://reference.digilentinc.com/reference/programmable-logic/genesys-2/reference-manual
8. Blasi L, Vigli F, Cheikh A, Mastrandrea A, Menichelli F, Olivieri M (2019) A RISC-V Fault-Tolerant microcontroller core architecture based on a hardware thread full-weak protection and a thread-controlled watch-dog timer. In: Applications in electronics pervading industry, environment and society. ApplePies

Chapter 63
AXI4LV: Design and Implementation of a Full-Speed AMBA AXI4-Burst DMA Interface for LabVIEW FPGA

Luca Dello Sterpaio, Antonino Marino, Pietro Nannipieri, Gianmarco Dinelli and Luca Fanucci

Abstract The Advanced Microcontroller Bus Architecture Advanced eXtensible Interface is a memory mapped protocol intended for internal System on Chip communications. However, there is no mean to directly exchange data between AXI devices and LabVIEW applications. This work proposes a novel and streamlined bridge solution to transfer data directly and effectively from/to an AXI target memory and any LabVIEW FPGA application, essential to integrate AXI-based architectures into PXI FPGA peripheral modules. The block diagram of the proposed IP solution and synthesis results are presented for target programmable devices of interest.

Keywords Hardware design · System on chip · SoC · LabVIEW · FPGA · PXI

63.1 Introduction

63.1.1 Objectives and Challenges

This work aims to design an IP core bridge module to interface directly any *LabVIEW FPGA* (LVFPGA) *Virtual Instrument* (VI) with an AMBA 4 AXI interconnect fabric. This will allow hardware designers to include complex architecture based on this kind of bus into LVFPGA projects or, viceversa, LVFPGA VIs as part of a

L. Dello Sterpaio (✉) · A. Marino (✉) · P. Nannipieri (✉) · G. Dinelli (✉) · L. Fanucci (✉)
Department of Information Engineering, University of Pisa, Pisa, Italy
e-mail: luca.dellosterpaio@ing.unipi.it

A. Marino
e-mail: antonino.marino@ing.unipi.it

P. Nannipieri
e-mail: pietro.nannipieri@ing.unipi.it

G. Dinelli
e-mail: gianmarco.dinelli@ing.unipi.it

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

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_63

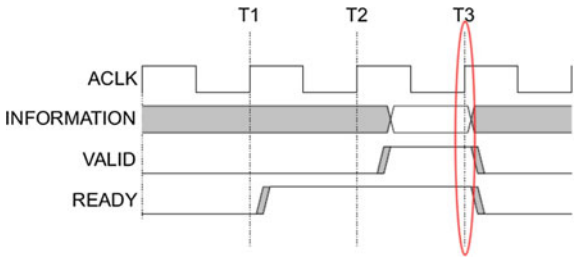


Fig. 63.1 The valid-ready paradigm on which AMBA AXI protocol is based on

larger FPGA architecture. Greatest challenge in proposed work is to obtain maximum performance, i.e. without affecting negatively the LVFPGA/AXI exchange data-rate.

63.1.2 AMBA AXI Protocol

The *Advanced Microcontroller Bus Architecture* (AMBA) is an open and a de facto standard protocol for *System on Chip* (SoC) interconnections [1] that collects several interface specifications. Published in 2010, the fourth version of the AMBA specifications also improved the *Advanced eXtensible Interface* (AXI). AXI is based on a *valid-ready* paradigm, as shown in Fig. 63.1: data is considered transferred whenever both valid and ready signals are asserted at rising clock edge. The last version of the AXI interface, AXI4, improves the previous version (AXI3) including a wider bus support and increasing to 256 the number of elements that is possible to include in a single burst. AXI4 is backward compatible with legacy AXI3 devices, if the burst length is equal or lower than 16 elements.

AXI is a memory-mapped protocol; there are simplified versions for point-to-point streaming (AXI-Stream) or for reduced complexity (AXI-Lite). *AXI-Burst* is used to refer to the standard full AXI protocol.

AXI-Stream, is stripped no-more-needed address channels (AW and AR); AXI-Lite uses less signals per channels and can transfer only a single word per transaction (1-length burst).

63.1.3 LabVIEW FPGA

Laboratory Virtual Instrument Engineering Workbench (LabVIEW) is a widely adopted development environment with a data-flow block-diagram graphical approach. LVFPGA, one of the many available LabVIEW module extensions, proposes for hardware description and FPGA programming the very same peculiar

graphical approach. Hardware/software partitioning is carried out by the environment, so that users can benefit from FPGA implementations without any Hardware Description Language (HDL) specific knowledge.

LabVIEW is largely adopted in the industry. There are many different hardware modules available that allow final users to create their own custom hardware and software instruments. Many natively supported target FPGAs are available for the LabVIEW environment as peripheral modules based on the *Peripheral Component Interconnect (PCI) eXtension for Instrumentation* (PXI) standard.

63.1.4 Typical Use-Case Scenario

Complex architectures can be implemented on PXI FPGA targets exploiting LVFPGA Socketed Component-Level IP (Sck-CLIP) macro [2].

The typical use-case scenario of proposed IP is indeed the implementation of an AXI-based system on PXI FPGA targets, to be controlled and operated underneath LabVIEW ecosystem, including any other LabVIEW-derived test environment like applications or *Hardware In the Loop* (HIL) solutions.

63.2 Architecture

AXI4 for LabVIEW (AXI4LV) IP is a configurable bridge module that operates translations of AMBA AXI4-Burst transactions into LVFPGA simplified 3-Wire handshake transfers [3, 4], and vice versa. The main challenge is represented by the translation of memory-mapped requests into direct stream transfers, or by the translation of stream into memory-mapped transactions.

A single AXI4LV instance bounds two *LVFPGA DMA-FIFO* structures to a single AMBA AXI4. *Host-to-Target* (H2T) DMA-FIFO is linked to write channels (AW, W and B); *Target-to-Host* (T2H) DMA-FIFO is linked to read channels (AR and R).

AXI4LV module configuration and control during operation is accomplished by setting its register file values. Registers can be accessed for write and read operations through a dedicated AMBA AXI4-Lite Slave interface.

Thanks to its AMBA AXI4-Burst Master interface, an instance of an AXI4LV module is capable of operating read/write Direct Memory Access (DMA) from/to any slave target mapped into its address space: without CPU mediation, it can issue AXI4 transaction to access address locations toward any slave device connected to the bus.

From a hierarchical point of view, the AXI4LV entity is internally organized into two independent sub-modules dedicated to issue and execute H2T/Write (H2T/W) and T2H/Read (T2H/R) transactions respectively, as shown in Fig. 63.2. Indepen-

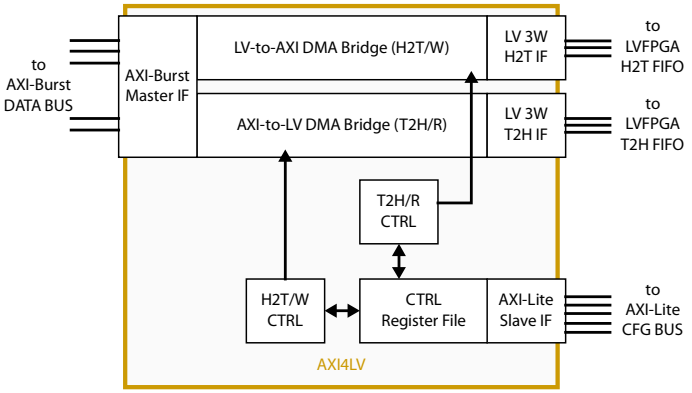


Fig. 63.2 A detailed block diagram of the AXI4LV IP design, illustrating internal hierarchy and functional blocks

dence of these two modules is nevertheless inescapable in order to achieve full-duplex capability, as per AXI4 protocol. Register file locations, which control each submodule, are split into two non-contiguous segments within the assigned address space.

Functionalities are well-separated in the internal organization as well. Control register file, control FSMs and DMA bridges are implemented as distinct units.

63.2.1 AXI4LV Controller Module

Transactions in both H2T/Write and T2H/Read directions are handled by independent FSM control units. Control units are instance of the same entity for both directions.

In order to operate correctly, control units shall be configured by system central control unit (i.e. the program of a CPU) through the AMBA AXI4-Lite interface of the register file submodule.

Figure 63.3 shows how the Control FSM module operates and its state transitions. When reset signal is removed, controller module is in CTRL_IDLE state and register file is editable (i.e. it is possible to update values of all writable locations); when enable register is set to a non-zero value and there are enough elements available to carry out the request completely, state evolves toward CTRL_WORK. In CTRL_WORK state, register values cannot be modified anymore, and control signals are issued to the DMA Bridge module of reference. Control FSM stays in CTRL_WORK state until a request is acknowledged as fulfilled by a feedback signal from the DMA module. State is restored to CTRL_IDLE automatically after recover time in CTRL_DONE state.

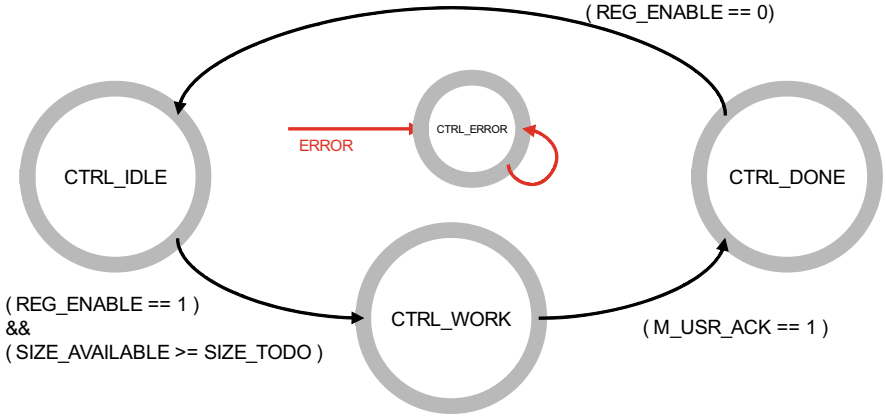


Fig. 63.3 State diagram of the control finite state machine for both H2T/W and T2H/R transactions

63.2.2 AXI4LV Register File Module

This submodule implements an AMBA AXI4-Lite slave interface to read and write 32-bit values of the control register file. Table 63.1 shows the structure of the register file.

Address space is 32 bit long, according to the hexadecimal pattern of 0xYYYYzRRR. The four most significant bits are the base address of the AXI4LV instance. H2T and T2H register subsets are separated with a 0x00001000 offset. The least significant bits address the register of interest. Considering AXI4 is byte addressed and data word are on 32 bits, each register is therefore located with an offset of 4 from others.

Table 63.1 Register file organization of an AXI4LV instance

Register offset	Register description	Read or write
000	Transfer enable register	RW
004	Slave base address	RW
008	Size in bytes	RW
00C	Interrupt enable register	RW
010	Status register	R
014	Slave high address bound	RW
018	Requested size in bytes	R
01C	Effective size in bytes	R
020	Available elements in DMA FIFO	R

Structure is replicated for both H2T/W channel and T2H/R channel

63.2.3 DMA Bridge Modules

DMA modules are blocks that carry out data transactions and operate stream flow control. H2T/Write DMA can issue requests independently on AW, W and B channels; T2H/Read DMA can issue requests on AR and R channels.

When enabled by the control FSM, this module internally stores configuration register values of interest and then orders AXI request(s) accordingly. Once operations end, a feedback signal is provided to its control FSM unit and, if enabled by user, an external interrupt pulse is generated.

AMBA AXI4-Burst transactions are carried out by requesting read or write access to the target slave on the address channel, starting from provided base address and of requested length. Burst length may differ: it is indeed the least long among (1) requested amount of data, (2) maximum allowed burst length or (3) distance from next *4k-boundary* address [5]. If needed, AXI4LV IP is capable of automatically split the whole requested amount of data into multiple AXI transactions, taking into account all of the above, up to the complete fulfillment of requested data packet. Moreover, the module supports overlapping (optional AXI4 feature) for maximum parallelization and pipeline capabilities (i.e. AXI4LV instances can immediately order another transaction on the address channel before the previous request is completed on the data channel).

Data is simply forwarded between the two interfaces and flow control can be operated either on the AXI bus signals or onto the LabVIEW 3-Wire protocol, in order to stabilize the streams across and, thus, to not lose any data to be transferred. On AXI data channels (W and R), flow control is operated driving the *ready* or *valid* signals high or low, exactly as for *dready* and *dvalid* signals on the LV 3-Wire IF.

63.3 Synthesis Results

In this section, synth results are presented. AXI4LV sources are tech-independent code, yet results refer to Xilinx 6-Series and 7-Series devices on which National Instruments PXI FPGA modules are based on.

Frequency analysis returns 156 MHz as the highest possible clock. At 100 MHz clock already, no data-rate degradation is exhibited in transfers. Introduced latency is just 6 clock ticks between request and actual start of the AXI-Burst transaction: considering this bridge will be used to move large amount of data, it is negligible. Table 63.2 reports IP synthesis results, carried out by ISE (for 6-Series devices) and Vivado (for 7-Series devices), IDE of reference tools.

Table 63.2 Synthesis results on target technologies of interest

Device family	AXI parameters			Resource utilization							
	Addr width	Data width	Max burst	Slices	LUTs						Registers
					TOT	LUT2	LUT3	LUT4	LUT5	LUT6	
Virtex-6	32	32	16	522 0.44%	1540 0.32%	282	175	382	267	434	725 0.08%
Kintex-7	32	32	16	497 0.67%	1295 0.43%	237	146	319	229	364	737 0.12%

Percent values refers to largest device model in the product family (i.e., XC6VLX760 and XC7480T chips)

63.4 Conclusions

Presented novel IP efficiently exchange large amount of data between any LabVIEW applications on a host PXI controller PC and any memory mapped AXI target implemented on a PXI FPGA peripheral module. This IP is essential for AXI-based architecture implementation onto PXI FPGAs. It supports the most advanced features introduced in the latest version of the AMBA AXI4 standard, such as overlapping transactions, wider bus width and 256-elements burst. Provided synthesis results targets the main reference technologies and highlight the overall very small footprint in terms of absolute FPGA resource cost.

References

1. AMBA AXI and ACE Protocol Specifications. Issue D, ARM Holdings, Cambridge, Oct 2011, IHI 0022D
2. National Instruments, "Importing External IP Into LabVIEW FPGA," LabVIEW FPGA 2014 Support. Accessed on 05/2019. Available at www.ni.com/tutorial/7444/en/
3. National Instruments, "Using FPGA FIFOs (FPGA Module)," LabVIEW FPGA 2010 Support. Accessed 05/2019. Available at http://zone.ni.com/reference/en-XX/help/371599F-01/lvfpgahelp/creating_fpga_fifos/
4. National Instruments, "How DMA Transfers Work (FPGA Module)," LabVIEW FPGA 2012 Support. Accessed on 05/2019. Available at http://zone.ni.com/reference/en-XX/help/371599H-01/lvfpgaconcepts/fpga_dma_how_it_works/
5. Dello Sterpaio L et al (2019) Exploiting LabVIEW FPGA socketed CLIP to design and implement soft-core based complex digital architectures on PXI FPGA target boards. In: 2019 International conference on synthesis, modeling, analysis and simulation methods and applications to circuit design (SMACD), Lausanne (CH), July 2019

Chapter 64
3D-HEVC Neighboring Block Based Disparity Vector (NBDV) Derivation Architecture: Complexity and Implementation Analysis

Waqar Ahmad, Naveed Khan Baloch, Fawad Hussain, Muhammad Asif Khan and Maurizio Martina

Abstract HEVC (High Efficiency Video Coding), the state-of-the-art video coding standard has 3D extension known as 3D-HEVC, which is established by JCT-3V. In current design of 3D-HEVC, to exploit the redundancies of the 3D video signal, various tools are integrated. In 3D-HEVC, the neighboring block disparity vector (NBDV) mode is used to replace the original predicted depth map (PDM) for inter-view motion prediction. A new estimated disparity vector depth oriented neighboring block disparity vector (DoNBDV) is used to enhance the accuracy of the NBDV by utilizing the coded depth map. In this paper, the complexity and implementation analysis of the NBDV and DoNBDV architectures are analyzed in terms of performance, complexity, and other design considerations. It is hence concluded that NBDV and DoNBDV for 3D-HEVC video signals provide attractive coding gains with comparable complexity as traditional motion/disparity compensation.

W. Ahmad (✉) · N. K. Baloch · F. Hussain · M. A. Khan
University of Engineering and Technology Taxila, Taxila, Pakistan
e-mail: waqar.ahmad@uettaxila.edu.pk

N. K. Baloch
e-mail: naveed.khan@uettaxila.edu.pk

F. Hussain
e-mail: fawad.hussain@uettaxila.edu.pk

M. A. Khan
e-mail: masif.khan@uettaxila.edu.pk

M. Martina
Politecnico di Torino, Turin, Italy
e-mail: maurizio.martina@polito.it

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4_64

64.1 Introduction

The substantial progress and growing acceptance in 3D technologies and 3D contents together with the displays, acquisition and rendering, numerous 3D products and applications are increasing quickly and becoming nearer to realism in current years, including 3D gaming and IMAX cinemas, Free Viewpoint Video [1], 3D Televisions [2]. Glasses are required to empower depth perception for stereoscopic displays, much of the work has been done emphasizing these displays. Autostereoscopic displays, a new generation of displays, depending on the position of the observer's eye, emit different pictures and there is no need of glasses for viewing, is starting to arise and commercially become accessible [3, 4]. Depth image-based rendering are employed in autostereoscopic displays to produce a thick set of pictures to the scene [5]. The high-quality depth maps are desirable, which need to be coded and represented sideways with the texture, to render the intermediate views/pictures with adequate quality. Stereo correspondence techniques [6] can be used to obtain depth maps from multicamera or stereo setup. Depth maps can also be obtained by a dedicated depth camera. This specific topic has seen prominent developments in current ages with design based on time-of-flight-based imaging [7] and structured light [8]. Added use case of depth maps is the stereo depth perception adaptation for heterogeneous display strategies. Lastly, depth map is an essential fragment of computer-generated images, which is prevalent in numerous movies makings.

Moving Picture Experts Group (MPEG) issued a call for proposals (Cfp) [9] on 3D video coding technologies for cutting-edge use cases as stated above [10–15].

For the statistical redundancy compression, existing between different views, 3D and multiview compression techniques and tools are employed. For both 3DV coding and multiview coding, disparity vector (DV) has a major role in the identification of inter-view redundancy. The legacy devices mainly designed for stereo displays which are not able to produce synthesized views, may request the 3DV bitstream. In such a case, to circumvent the redundant bandwidth increase, to communicate the depth, the calculations gets doubled for decoding the corresponding depth maps, which is essential to have the disparity vectors resulting from corresponding textures, as MPEG CfP requirement [16], this is named as the multiview/stereo compatibility. For the texture image decoding in such a multiview compatibility requirement, depth maps are needed. Thus, derivation of disparity vector method must be intended in an integrated manner.

The fundamental problem in 3D and multiview coding is the design of the disparity derivation process. For the coding complexity of multiview/3D and coding efficiency, the accurateness of this method is vital. In video codecs, the disparity vector addition is often carried out at slice or block level. This paper presents the complexity analysis and high-level hardware design of the derivation of disparity vector, which is integrated in the 3D-HEVC and 3D-AVC video standards with name of Neighboring Based Disparity Vector (NBDV) method for disparity vector derivation, it generates the disparity vector at block-level by examining neighboring blocks.

The rest of this paper is organized as follows. Standard contributions and previous academic research associated with disparity vector generation and estimation are presented In Sect. 64.2. NBDV has been integrated as a necessary technique in 3D-AVC and 3D-HEVC standards. The technical specifics of the NBDV and a more refined form of the NBDV i.e. Depth-oriented NBDV (DoNBDV) are presented in detail. The Complexity of the NBDV is presented in Sect. 64.3. By extensive discussions and experimental results, the complexity is analyzed. The high-level hardware design of the NBDV is presented in Sect. 64.4 to confirm the benefits of the NBDV method in real-world codecs. Section 64.5 presents the conclusion of this article.

64.2 NBDV Derivation

The solution presented in [17] becomes the basis of the development of NBDV derivation method. The NBDV is integrated as an important method in 3D-HEVC and 3D-AVC standards, after the validation of its effectiveness in 3D platforms. Already coded motion fields are used for the generation of the NBDV based disparity vectors with no additional signaling. There is no conversion between depth and disparity during this disparity generation process. The NBDV does the same process at both the decoder and encoder with reduced complexity with respect to any main component of the video codec. Inspiration of the NBDV is presented in this section with a broad picture of the NBDV process. Further, the NBDV optimizations i.e. DoNBDV is discussed.

The actual NBDV method was presented in [18, 19] for 3D-HEVC. The basic concept of this method was to use the temporal and spatial neighboring blocks as shown in Fig. 64.1. An effective NBDV method depends on the subsequent features: (i) The likelihood of discovering a disparity-based motion vector, therefore a disparity motion vector could be obtained; (ii) If obtainable, the rate distortion optimization correctness based on disparity vector; (iii) Memory access increment based on reference frames; (iv) Number of additional blocks and additional calculations required to be checked; and (v) Extra impermanent memory vital to complete NBDV [20].

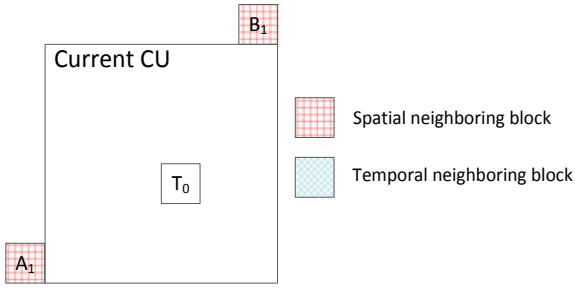


Fig. 64.1 Location of spatial and temporal neighbour blocks [21]

An important feature of the NBDV method is, to decode a current block. If the required blocks are already used by the HEVC decoding, then supplementary retrieving these blocks is not measured as an extra load particularly as of memory bandwidth viewpoint.

Major milestone in the NBDV optimizations methods propose the reduction of extra memory accesses. The decoded depth of the main view (also called as base view) already exist, while coding the dependent view texture. The availability of the base view depth map might enhance the dependent view texture coding disparity derivation method. A better disparity vector can be obtained by the following steps:

1. NBDV based disparity vector derivation.
2. Then this NBDV vector is used to trace the matching block in the reference view's coded depth. This reference view must have the same view order index as the NBDV based disparity vector has its view order index. If the traced matching depth block locate on the boundary or outside the depth picture, then the pixels located outside the depth picture are clipped to boundary of the picture. The samples located within the depth picture are reserved untouched.
3. The matching block's depth in the base view is supposed to be the "virtual depth block" of the dependent view's current block.
4. The four edge pixels of the virtual depth block are used for the retrieval of maximum value of the depth.
5. The disparity is achieved by converting the maximum depth value obtained in step 4.

D0 denotes the coded depth map of the view 0 as shown in Fig. 64.2. T1 is the texture to be coded. Using the disparity vector estimated by NBDV, depth block from the coded depth D0 for the current block (CB) is derived. In NBDV the estimated disparity vector is obtained as stated in step 1, this method is already integrated in the Test Model of HEVC [22]. The candidates for disparity vectors can be from temporal/spatial motion compensated predicted (DV-MCP) neighbouring blocks, temporal/spatial disparity compensation prediction (DCP) neighbouring blocks. The

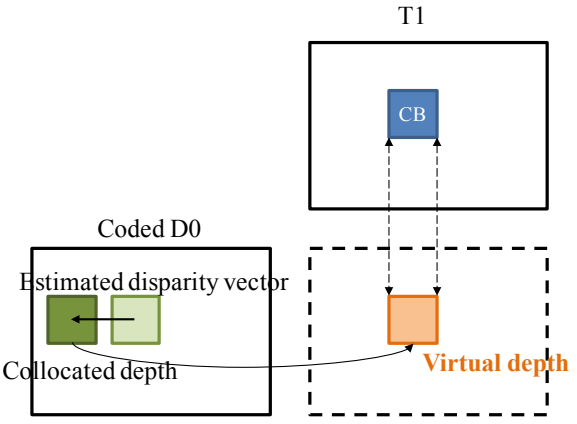


Fig. 64.2 The virtual depth retrieval method [21]

main purpose of NBDV optimization is to utilize the extracted virtual depth to retrieve a more precise disparity vector for prediction. In the existing implementation, the maximum disparity of the virtual depth is converted into the new disparity vector. The camera parameters and view position are used for the conversion of depth to depth values. The new improved disparity vector is termed as "depth oriented neighbouring block disparity vector" (DoNBDV). The DoNBDV has overhead of only accessing the reference depth buffer. The other coding tools can use the virtual depth obtained during the estimation of DoNBDV. The merge mode and Advanced Motion Vector Prediction (AMVP) makes the utilization of DoNBDV in obtaining the inter-view motion prediction.

64.3 Complexity Analysis of Neighboring Block Based Disparity Vector Derivation

According to the HTM reference software and the algorithm description, the steps listed in Sect. 64.2 are used for the disparity vector derivation. For each 4×4 block, in the depth map, one depth value is maintained. During this process, the holes may occur for the depth map and hole-filling process may be carried out. The hole-filling process fills the depth hole by means of the available foremost depth value of the same line.

64.3.1 Increase in Memory Bandwidth

One of the most important issue in the hardware design is the worst-case memory bandwidth requirement. As, the system on chip (SoC) has fixed total data transfer rate and hardware design of video must share it with further applications. In 3D-HEVC, all HEVC tools are supported, the extra memory bandwidth required for the tools of 3D-HEVC should be controlled well to reduce memory bandwidth requirement.

Due to the motion prediction of HEVC design, no extra memory is needed for the spatial neighboring blocks of NBDV because those blocks are already accessible.

Since, half of the temporal neighboring blocks belong to the Temporal Motion Vector Prediction (TMVP) co-located picture, hence, extra memory access is needed for two blocks only of the extra candidate frame. In general, we can say, for the NBDV process the additional memory bandwidth required is the identical as required for the HEVC TMVP process. The subsequent candidate image is accessed the identical way as per TMVP.

Overall memory access analysis is presented by evaluating the number of samples to be retrieved for minutest (i.e. 8×8) coding unit decoding in the scenario of the worst-case. The worst-case occurs in case of merge mode and Prediction Unit (PU)

size of 8 × 8 with bi-prediction. If we assume 4 bytes for motion vector and 1 byte for reference picture index representation, though it is implementation dependent.

The analysis of the complexity of the dependent view decoding in 3D-HEVC can be performed by comparing its complexity with single-view HEVC coding, the complex modes such as Advanced Residual Prediction (ARP), inter-view motion prediction are excluded while analyzing the complexity. Hence, only major modules such as merge list construction and motion compensation are considered as anchor for memory bandwidth.

Motion Compensation. In HEVC, 8-tap interpolation filters [23, 24] are used to interpolate one PU of size 8 × 8. For an 8 × 8 block bi-predicted, 450 (i.e., (8 + 7) × (8 + 7) × 2) pixels need to be accessed. This figure can be even more for the motion compensation depending on the memory access pattern. Thus, even reduced percent of total memory bandwidth may be required for the NBDV process.

Construction process of Merge list. For the construction of merge list, motion information together with two motion vectors and indices of two reference pictures, of up to 2 temporal neighboring block and 5 spatial neighboring blocks are retrieved in the co-located picture. For creating the two combine lists of the said block, the total bytes needed to retrieve are: (2 + 8) × 2 × (5 + 2) = 140.

NBDV. Like, we described earlier, the motion data, including the indices of the reference linked with the two more blocks of extra candidate picture and two motion vectors required to be accessed for a block of size 8 × 8, totaling up to (2 + 8) × 2 = 20 bytes. Because of the inherent characteristic of NBDV, blocks of the temporal positions can be used early to examine the four reference indices of the two blocks for motion vectors retrieval, this happens only in case of identification of the reference image of the inter-view prediction situation. The additional bytes required to be retrieve in this situation are 4 + 4 = 8 bytes.

Table 64.1 shows, for the NBDV the increase in memory access is about one (1) %. For 3D-HEVC, additional tools such as ARP and motion prediction of inter-view motion may need additional accesses of memory as compared to NBDV. Reason for this extra memory access is because of the pixels from extra blocks and inter-view reference picture motion vectors. The major tools of the 3D-HEVC are based on NBDV, this estimated 1% memory bandwidth increment is acceptable.

Table 64.1 NBDV memory accessed and other methods in 3D-HEVC

CU size 8 × 8	Required bytes	Percent (%)
Merge mode HEVC	140	23.7
Motion compensation HEVC	450	76.3
Extra memory access for NBDV 3D-HEVC	8	1.4

64.3.2 Increase in Memory Storage

In NBDV, the motion data of the two aspirant images in Decoded Picture Buffer (DPB) is used, each image in the DPB holds motion vectors, thus, no extra information storage in DPB is required to rise the memory storage. Even though, the Derived Disparity Vector (DDV) may need to store one vector per slice, but, for the current picture decoding the temporary memory required is negligible.

64.3.3 Computational Complexity

NBDV process required to check the Disparity Motion Vector (DMV). Thus, in comparison with the motion compensation the complexity is negligible. The extra reference indices access by the NBDV is up-to nine blocks to both lists of reference pictures. Hence, only 18 extra conditions per CU are inserted. In each access, the reference index must be checked, thus, this checking is approximately as expensive (in terms of hardware) as an addition process.

For a bi-predicted 8×8 CU block, the interpolation of 64×2 pixels for luma part may be needed. Eight multiplications and seven additions are required for the interpolation of each pixel of luma component. If we suppose the complexity of the operation of one multiplication is about 5 times the complexity of the addition operation. Then the complexity for the process of motion compensation of the luma component will be roughly about $[(5 \times 8 + 7) \times 128]$ 6016 operations of additions. Thus, the complexity of the luma component of motion compensation is about 334 (6016/18) times the NBDV process complexity. Therefore, in comparison with the overall decoder's computation, the NBDV process has negligible computational complexity as shown in Table 64.2.

Table 64.2 NBDV complexity and complexity of bi-prediction of luma block of motion compensation

CU Size 8×8	Required approximate hardware	Complexity
NBDV	18 additions	Negligible as compared to bi-prediction
Bi-predicted luma component motion compensation HEVC	6016 additions	334 xNBDV

64.4 High-Level Hardware Representation of Implementation

The high-level hardware representation of the NBDV and DoNBDV is presented in Figs. 64.3 and 64.4, respectively. In Fig. 64.3, block level representation of the NBDV is shown. It can be observed that the NBDV process can be carried-out depending on

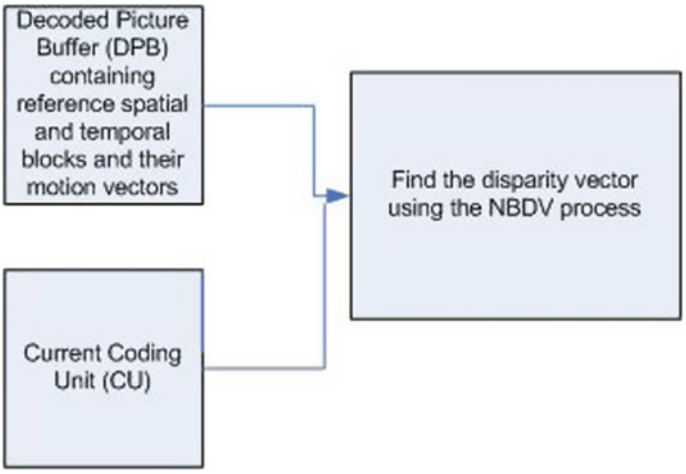


Fig. 64.3 NBDV block level hardware representation

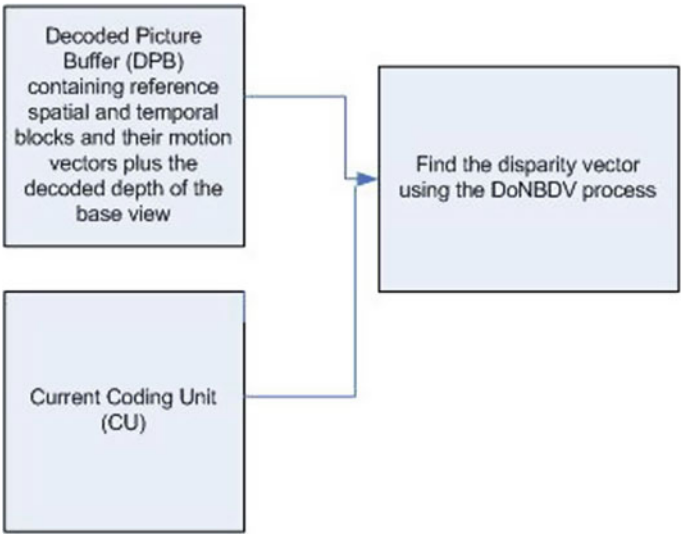


Fig. 64.4 DoNBDV block level hardware representation

the availability of the basic requirements of the NBDV process i.e. spatial/temporal neighboring blocks of the CU.

In Fig. 64.4, the DoNBDV block level hardware implementation is presented it involve the inclusion of already coded depth map of base view for the more accurate derivation of disparity vector of the current coding block.

64.5 Conclusion

In this paper, the complexity and implementation analysis of the NBDV and DoNBDV designs are examined in terms of complexity, performance and other design deliberations. It is determined that NBDV and DoNBDV for 3D-HEVC video signals offers striking coding gains with analogous complexity as traditional motion/disparity compensation. NBDV and DoNBDV are the important tools of the 3D extensions of HEVC and H.264/AVC. As discussed above, these tools have appropriate coding gains and the hardware implementation of these tools can further help to improve the coding gain and performance of video coding standards.

References

1. Tanimoto M (2006) Overview of free viewpoint television. Signal Process Image Commun 21(6):454–461
2. Vetro A, Matusik W, Pfister H, Xin J (2016) Coding approaches for end-to-end 3D TV systems. In: Proceedings of the 23rd picture coding symposium (PCS'04), San Francisco, California, USA, Dec 2004, pp 319–324
3. Urey H, Chellappan KV, Erden E, Surman P (2011) State of the art in stereoscopic and autostereoscopic displays. Proc IEEE 99(4). https://doi.org/10.1109/JPROC.2010.2098351
4. Dodgson NA (2005) Autostereoscopic 3D displays. IEEE Comput 38(8):31–36
5. Fehn C (2004) Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In: Proceedings of SPIE, stereoscopic displays and virtual reality systems XI, vol 5291, May 2004, p 93
6. Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int J Comput Vis 47(1):7–42
7. Foix S, Alenya G, Torras C (2011) Lock-in time-of-flight (ToF) cameras: a survey. IEEE Sensors J 11(9):1917–1926
8. Salvi J, Pagès J, Batlle J (2004) Pattern codification strategies in structured light systems. Pattern Recognit 37(4):827–849
9. Call for proposals on 3D video coding technology, N12036, MPEG of ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Mar 2011
10. Kauff P, Atzpadin N, Fehn C, Müller M, Schreer O, Smolic A, Tanger R (2007) Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability. Signal Processing: Image Communication. Special Issue on 3DTV, Feb 2007
11. Advanced video coding for generic audiovisual services, Rec. ITU-T H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2012
12. High efficiency video coding, Rec. ITU-T H.265 and ISO/IEC 23008-2, Jan 2013
13. Sullivan GJ, Ohm J-R, Han W-J, Wiegand T (2012) Overview of the high efficiency video coding (HEVC) standard. IEEE Trans Circuits Syst Video Technol 22(12):1649–1668

14. Hannuksela MM, Chen Y, Suzuki T, Ohm J-R, Sullivan G (2013) 3D-AVC draft text 8. Presented at the 6th meeting joint collaborative team on 3D video coding extension development, Geneva, Switzerland, 25 Oct–1 Nov, 2013, Doc. JCT3V-F1002
15. Tech G, Wegner K, Chen Y, Yea S (2013) 3D-HEVC Draft Text 2. Presented at the 6th meeting joint collaborative team on 3D video coding extension development, Geneva, Switzerland, 25 Oct–1 Nov 2013, Doc. JCT3V-F1001
16. Applications and requirements on 3D video coding, N12035, MPEG of ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Mar 2011
17. Schwarz H, Wiegand T (2012) Inter-view prediction of motion data in multiview video coding. In: Proceedings of picture coding symposium, May 2012, pp 101–104
18. Zhang L, Chen Y, Karczewicz M (2012) CE5.h: disparity vector generation results. Presented at the 1st meeting joint collaborative team on 3D video coding extension development, Stockholm, Sweden, 16–20 July 2012, Doc. JCT3V-A0097
19. Zhang L, Chen Y, Karczewicz M (2013) Disparity vector based advanced inter-view prediction in 3D-HEVC. In: Proceedings of IEEE international symposium circuits system, May 2013, pp 1632–1635
20. Tech G, Müller K, Ohm J-R, Vetro A, Overview of the multiview and 3D extensions of high efficiency video coding. IEEE Trans Circuits Syst Video Technol 26(1)
21. Chen Y et al (2014) Test model 10 of 3D-HEVC and MV-HEVC. Joint collaborative team on 3D video coding extensions of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Document JCT3V-J1003. 10th meeting, Strasbourg
22. Gerhard Tech et al (2012) 3D-HEVC Test Model 1. Joint collaborative team on 3D video coding extension development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-A1005, July 2012, Stockholm
23. Ahmad W et al (2016) High level synthesis based FPGA implementation of H. 264/AVC sub-pixel luma interpolation filters. In: Modelling symposium (EMS), 2016, European. IEEE
24. Ahmad W, Martina M, Masera G (2015) Complexity and implementation analysis of synthesized view distortion estimation architecture in 3D High Efficiency Video Coding. In: 2015 International conference on 3D imaging (IC3D). IEEE

Author Index

A
Abbatessa, Emanuele, 413
Ahmad, Waqar, 549
Akkad, Ghattas, 145
Albanese, Andrea, 65
Alessandrini, Adriano, 293
Ambrosini, Dario, 425
Andrighetti, M., 173
Apicella, Tommaso, 103
Asciolla, Dario, 91
Aurioso, Daniele, 103
Ayoubi, Rafic, 145

B
Baldanzi, Luca, 117, 381
Baloch, Naveed Khan, 549
Barbieri, Riccardo, 293
Baronti, Federico, 285, 397
Bassoli, Marco, 163
Battistini, Daniele, 125
Belli, Jacopo, 117
Bellitti, P., 81
Bellitto, Jessica, 349
Bellotti, Francesco, 103, 349, 469
Benini, Luca, 125
Bertacchini, Alessandro, 251
Berta, Riccardo, 103, 349, 469
Berti, L., 25
Berzi, Lorenzo, 293
Bianchi, Valentina, 163
Blasi, Luigi, 505
Boni, Andrea, 443
Boni, Enrico, 371, 389
Bortoli De, Luca, 341
Breglio, Giovanni, 259
Brunelli, Davide, 65, 125, 355, 477
Bruschi, Paolo, 225
Bruzzone, Andrea, 301

C
Cammarata, S., 11
Capra, Maurizio, 137
Cardarilli, Gian Carlo, 325
Carloni, Andrea, 397
Carrato, Sergio, 341, 489
Caselli, Michele, 443
Cheikh, Abdallah, 505, 529
Ciarpi, G., 11, 25
Cicalini, Mattia, 225
Cignini, Fabio, 293
Cirillo, Gennaro, 73
Cococcioni, Marco, 213
Conti, Massimo, 51, 333
Crocetti, Luca, 117
Croci, Tommaso, 3
Cucchi, Francesca, 463

D
d'Acunto, Donato, 65
D'Addona, D. M., 267
Dallai, Alessandro, 371
Danese, Luisa, 103
D'Arco, M., 183
De Bortoli, Luca, 489
De Caro, Davide, 155
Della Corte, F. G., 267
Dello Sterpaio, Luca, 483, 499, 541
Demarchi, Danilo, 207
De Marchi, Luca, 355, 363

S. Saponara and A. De Gloria (eds.), *Applications in Electronics Pervading Industry, Environment and Society*, Lecture Notes in Electrical Engineering 627,
https://doi.org/10.1007/978-3-030-37277-4

De Munari, Ilaria, 163
Dente, Davide, 413
Dilillo, Luigi, 91
Di Meo, Gennaro, 155
Dinelli, Gianmarco, 483, 499, 541
Di Nunzio, Luca, 325
Di Pascoli, Stefano, 463
Di Pasquale, Fabrizio, 233
Di Rienzo, Roberto, 285, 397
D'Ortenzio, Alessandro, 191

E
ElHassan, Bachar, 145

F
Falaschi, Francesco, 117
Falbo, Vincenzo, 103
Fanucci, Luca, 117, 381, 483, 499, 513, 541
Faralli, Stefano, 11, 233
Fazzolari, Rocco, 325
Fernandes Carvalho, D., 81
Ferrari, P., 81
Ferri, Giuseppe, 425
Feyen, D. A. M., 243
Fienga, Francesco, 259
Filosa, Mariangela, 381
Fiore, Gaia, 433

G
Gagliardi, Alessio, 405
Gaiduk, Maksym, 333
Gaioni, L., 19
Gallina, Paolo, 489
Gastaldo, Paolo, 301
Genovese, Antonino, 293
Giammatteo, Paolo, 191
Gianoglio, Christian, 301
Giardino, Daniele, 325
Girolami, Alberto, 355
Girolami, Marco, 33
Giuffrida, Gianluca, 381
Gloria De, Alessandro, 103, 349, 469
Graziano, M., 173
Guidi, Francesco, 371
Guzzi, Francesco, 341, 489

H
Hussain, Fawad, 549

I
Iannaccone, Giuseppe, 463
Iero, D., 267
Irace, Andrea, 259

K
Khan, Muhammad Asif, 549
Kobeissi, Ahmad, 469

L
Lasagni, Marco, 251
Laurendi, D., 267
Lopomo, N. F., 81

M
Magazzù, G., 11
Magistrati, Giorgio, 513
Malatesta, Michelangelo Maria, 363
Manghisoni, M., 19
Mangraviti, G., 25
Mansour, Ali, 145
Maresca, Luca, 259
Marino, Antonino, 483, 499, 541
Marrazzo, Vincenzo Romano, 259
Marsi, Stefano, 341, 489
Martina, Maurizio, 137, 549
Martínez Madrid, Natividad, 333
Marzani, Alessandro, 355, 363
Masera, Guido, 137
Massari, Luca, 381
Mastrandrea, Antonio, 109, 505, 529
Matera, Riccardo, 521
Matta, Marco, 325
Maya López, Armando, 43
Meacci, Valentino, 371
Melo, Douglas, 91
Menichelli, Francesco, 109, 505, 529
Menicucci, Alessandra, 91
Meoni, Gabriele, 499, 513
Mercola, M., 243
Merenda, M., 267
Messina, E., 243
Mestice, Marco, 451
Mihet-Popa, Lucian, 433
Monda, D., 25
Montagni, Marco, 389
Morozzi, Arianna, 3
Motto Ros, Paolo, 207
Muanenda, Yonas, 233
Muttillo, Mirco, 425

N
Nannini, Andrea, 225
Nannipieri, Pietro, 483, 499, 541
Napoli, Ettore, 155, 183
Neri, Bruno, 451

O
Oddo, Calogero Maria, 381
Olivieri, Mauro, 109, 505, 529
Orcioni, Simone, 51, 333
Orsini, Andrea, 33
Ortenzi, Fernando, 293
Oton, Claudio J., 233
Ottavi, Marco, 91

P
Palla, F., 11
Palma, F., 243
Palmisano, Giuseppe, 277
Panicacci, Silvia, 381
Parisi, Alessandro, 277
Passeri, Daniele, 3
Passerone, Roberto, 73
Peloso, Riccardo, 137
Petra, Nicola, 155
Pettinato, Sara, 33
Pezzoli, M., 19
Piedimonte, P., 243
Pierini, Marco, 293
Pierpaolo, Dini, 309
Piotto, Massimo, 225
Placidi, Pisana, 3
Polonelli, Tommaso, 125
Posenato, Antonio, 73
Pugi, Luca, 293, 389

R
Ragonese, Egidio, 277
Ragusa, Edoardo, 301
Ramponi, Giovanni, 341, 489
Ratti, L., 19
Re, Marco, 325
Renzi, M., 243
Re, V., 19
Ria, Andrea, 225
Riccio, Michele, 259
Ricci, Stefano, 199, 521
Riceputi, E., 19
Rizzon, Luca, 73
Roncella, Roberto, 285, 397
Rosales, Ricardo Maximiliano, 207
Rossi, Fabio, 207
Rossi, Federico, 213
Rossi, Maria Cristina, 33
Rubeis de, Tullio, 425
Ruffaldi, Emanuele, 213
Ruo Roch, Massimo, 137, 173
Rusci, Manuele, 125
Russo, Dario, 199

S
Saletti, Roberto, 285, 397
Salvatori, Stefano, 33
Santoro, G., 173
Santos, Douglas, 91
Saponara, Sergio, 11, 25, 117, 213, 309, 405, 413, 433, 451
Scavongelli, Cristiano, 51
Schenone, Valentina, 349
Seepold, Ralf, 333
Sereni, Gabriele, 251
Serpelloni, M., 81
Simonte, Gianluca, 285
Sisinni, E., 81
Solanas, Agusti, 43
Sole, Luigi, 137
Sonzogni, M., 19
Sordillo, Stefano, 529
Spanò, Sergio, 325
Spina, Nunzio, 277
Stazi, Giulia, 109
Strollo, Antonio G. M., 155

T
Tagnani, Diego, 33
Terruso, Giuseppe, 381
Testoni, Nicola, 355, 363
Tonietti, Giovanni, 381
Tortoli, Piero, 371
Traversi, G., 19
Turvani, G., 173

V
Vacca, M., 173
Valente, Giacomo, 191
Valverde, Alberto, 513
Velha, P., 11
Vigli, Francesco, 505

W
Wöhrl, Hugo, 477

Z
Zacharelos, E., 183
Zalteri, Martina, 381
Zamboni, M., 173
Zonzini, Federica, 355