

Alessandro De Gloria  
*Editor*

# Applications in Electronics Pervading Industry, Environment and Society

# Lecture Notes in Electrical Engineering

Volume 289

## *Board of Series Editors*

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

For further volumes:

<http://www.springer.com/series/7818>

### *About this Series*

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

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

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

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

Alessandro De Gloria  
Editor

# Applications in Electronics Pervading Industry, Environment and Society

 Springer

*Editor*  
Alessandro De Gloria  
Electronic Engineering  
University of Genova  
Genova  
Italy

ISSN 1876-1100                      ISSN 1876-1119 (electronic)  
ISBN 978-3-319-04369-2            ISBN 978-3-319-04370-8 (eBook)  
DOI 10.1007/978-3-319-04370-8  
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014943115

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

Electronics technology has known a very fast development becoming pervasive of everyday life activities. Nowadays, electronics devices are so common that we seldom pay attention to them considering them as usual objects. Electronics devices are often considered a commodity and the attention is toward the application instead of the devices.

Often the prefix “e” is used to technologically qualify a product or a service (E-mail, E-card, E-commerce, E-banking, E-business, E-book, to cite a few) and to communicate that it is new, modern, advanced. Electronics devices have become a part of our life; they are no more a product used in the industrial environment to improve the features of a product. They have changed our life; you only have to think to a smartphone.

The incursion of electronics devices in life has led to a revision in the electronics engineer’s role. It is not enough to be able to design and implement an efficient device. The design has to consider the context in which the device will be used. Factors like human–machine interaction, usability, scalability, reusability must be included into the specification and drive the design of the device.

These considerations lead to put the attention toward the applications and the development of systems that increasingly simplify human activities.

The APPLEPIES conference aims at bringing together researchers and stakeholders, in order to share the state of the art of research and market in the field of applied electronics. The goal is to discuss the most significant trends, to explore the challenges, issues, and opportunities in the research and to debate on visions about the future of the electronics pervading industry, environment, and society.

The conference also includes an exhibition, where industries can highlight their latest products and technological cornerstones for future applications.

APPLEPIES is an annual conference and it is building a scientific community for shaping the future research in the field. This community represents a significant blend of industrial and academic professionals, mainly at Italian level but with an opening over the international audience, committed to the study, development, and deployment of electronics systems in all the main application fields.

Alessandro De Gloria

# Contents

<b>1</b>	<b>SuperCap-Based Energy Back-up System for Automotive Electronic Control Units</b> . . . . .	<b>1</b>
	Sergio Saponara, Roberto Saletti, Luca Fanucci, Roberto Roncella, Marco Marlia and Corrado Taviani	
<b>2</b>	<b>CH<sub>4</sub> Monitoring with Ultra-Low Power Wireless Sensor Network</b> . . . . .	<b>13</b>
	Davide Brunelli and Maurizio Rossi	
<b>3</b>	<b>Integrated Front-end Electronics for Silicon PhotoMultiplier Readout in Medical Imaging Applications</b> . . . . .	<b>27</b>
	Nahema Marino, Sergio Saponara, Luca Fanucci, Federico Baronti, Roberto Roncella, Francesco Corsi, Cristoforo Marzocca, Gianvito Matarrese, Fabio Ciciriello, Francesco Licciulli, Maria Giuseppina Bisogni and Alberto Del Guerra	
<b>4</b>	<b>Energy Autonomous Low Power Vision System</b> . . . . .	<b>39</b>
	Davide Brunelli, Alberto Tovazzi, Massimo Gottardi, Michele Benetti, Roberto Passerone and Pamela Abshire	
<b>5</b>	<b>A New Space Digital Signal Processor Design</b> . . . . .	<b>51</b>
	Massimiliano Donati, Sergio Saponara, Luca Fanucci, Walter Errico, Annamaria Colonna, Giuseppe Piscopiello, Giovanni Tuccio, Franco Bigongiari, Maximilian Odendahl, Rainer Leupers, Antonio Spada, Vincenzo Pii, Elena Cordiviola, Francesco Nuzzolo and Frederic Reiter	
<b>6</b>	<b>Spatial Sound Rendering for Assisted Living on an Embedded Platform</b> . . . . .	<b>61</b>
	Luca Rizzon and Roberto Passerone	

<b>7</b>	<b>BASIC32: A New ASIC for Silicon Photomultiplier Detectors . . .</b>	<b>75</b>
	Fabio Ciciriello, Francesco Corsi, Francesco Licciulli, Cristoforo Marzocca, Gianvito Matarrese, Alberto Del Guerra and Maria Giuseppina Bisogni	
<b>8</b>	<b>Reconfigurable Implementation of a CNN-UM Platform for Fast Dynamical Systems Simulation. . . . .</b>	<b>85</b>
	Gianluca Borgese, Calogero Pace, Pietro Pantano and Eleonora Bilotta	
<b>9</b>	<b>A Multi Harvester with Hydrogen Fuel Cell for Outdoor Applications . . . . .</b>	<b>103</b>
	Davide Brunelli, Michele Magno, Danilo Porcarelli and Luca Benini	
<b>10</b>	<b>A Dosimetric Device Based on CMOS Image Sensor for Interventional Radiology . . . . .</b>	<b>113</b>
	E. Conti, D. Magalotti, P. Placidi, L. Bissi, M. Paolucci, D. Passeri, A. Scorzoni and L. Servoli	
<b>11</b>	<b>A Novel Wireless Sensor Network for Electric Power Metering . . . . .</b>	<b>121</b>
	Natale Galioto, Francesco Lo Bue, Daniele Rizzo, Leonardo Mistretta and Costantino Giuseppe Giaconia	
<b>12</b>	<b>High Performance Bit-Stream Decompressor for Partial Reconfigurable FPGAs . . . . .</b>	<b>133</b>
	Gian Carlo Cardarilli, Marco Re and Ilir Shuli	
<b>13</b>	<b>A Reconfigurable Functional Unit for Modular Operations . . . . .</b>	<b>141</b>
	Gian Carlo Cardarilli, Luca Di Nunzio, Rocco Fazzolari, Salvatore Pontarelli and Marco Re	
<b>14</b>	<b>Wireless and Ad Hoc Sensor Networks: An Industrial Example Using Delay Tolerant, Low Power Protocols for Security-Critical Applications . . . . .</b>	<b>153</b>
	Claudio S. Malavenda, Francesco Menichelli and Mauro Olivieri	
<b>15</b>	<b>A Social Serious Game Concept for Green, Fluid and Collaborative Driving. . . . .</b>	<b>163</b>
	Francesco Bellotti, Riccardo Berta and Alessandro De Gloria	



# Chapter 1

## SuperCap-Based Energy Back-up System for Automotive Electronic Control Units

Sergio Saponara, Roberto Saletti, Luca Fanucci, Roberto Roncella,  
Marco Marlia and Corrado Taviani

**Abstract** The E-latch is a new automotive mechatronic device that substitutes the door closure mechanical system with electro-actuated parts plus an embedded electronic control unit (ECU) connected to the main vehicle network. Due to severe automotive safety-critical requirements for door closure, an energy back-up system is required. A solution based on supercaps and boost converter is proposed in this work to ensure E-latch operation even in case of main battery failure. An in-depth thermal, electrical and durability characterization of the supercaps proves the reliability of the energy back-up unit for automotive applications. A Components Off the Shelf (COTS) approach has been followed for the E-latch prototype and test phases. A migration towards an Application Specific Integrated Circuit (ASIC) design approach is envisaged for future large volume production.

---

S. Saponara (✉) · R. Saletti · L. Fanucci · R. Roncella  
Department of Information Engineering, University of Pisa, Pisa, Italy  
e-mail: sergio.saponara@iet.unipi.it

R. Saletti  
e-mail: roberto.saletti@iet.unipi.it

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

R. Roncella  
e-mail: roberto.roncella@iet.unipi.it

M. Marlia · C. Taviani  
MAGNA CLOSURES—Motrol Division, Guasticce, Livorno, Italy  
e-mail: marco.marlia@magnaclosures.com

C. Taviani  
e-mail: corrado.taviani@magnaclosures.com

## 1.1 Introduction

A strong tendency in the automotive field is to make the mechanical systems more and more controlled by an Electronic Control Unit (ECU), which properly manages sensors and electro-actuators improves the mechanical system performance and even makes new functions available. Automatic transmission, suspension control, electronic controlled injection in internal combustion engine, brake-by-wire and steer-by-wire systems are just some examples of this consolidated trend [1–6]. As far the door subsystem is concerned window lifter and in some cars also the rear mirror are electronically controlled, while the door open/closure unit is still mechanical, as in [5]. The Advanced Mechatronic Door System (AMDS) project is the framework in which the industry-academic collaboration between Magna Closures and the University of Pisa led to the introduction of a new mechatronic system for door closure called *E-latch*. Several advantages are achieved: reduced weight and size as compared the mechanical-based door closure system; increased flexibility, scalability and re-programmability of the unit to address different vehicle models and vehicle generations; integration of the latch system in the vehicle networks to enable advanced safety features or new comfort functionalities.

The *E-latch* is a new node of the main vehicle network that is connected either through a Local Interconnect Network (LIN) or a Controller Area Network (CAN) bus. It manages all the following functions: reading the car handle and door status by means of Hall sensors or contact sensors; communicating with the car body computer by receiving commands from the users (lock, double lock, child lock, anti theft lock, release) and transmitting the door status or diagnostic info; driving the electric motor actuating the closure/release of the door (operating at 12 V nominal, 8 V minimum, with a current absorption in the order of several amperes); managing the available energy sources, both the main battery and the back-up one (the supercaps and boost converter subsystem proposed in this paper). The widespread adoption of the *E-latch* is strongly challenged by the high level of reliability that is mandatory to achieve, particularly by the energy back-up system. The correct functionality of door release must be guaranteed by the *E-latch* even in case of an accident or a general failure of the main vehicle battery. An energy back-up system with minimum power consumption and weight/size overhead during normal vehicle operation is thus necessary. To overcome this issue a new supercap-based energy back-up system for automotive ECU is proposed in this paper.

Although applied to the E-latch ECU, the proposed energy-back up subsystem is general enough to be applied to any automotive ECU. Hereafter Sect. 1.2 describes the E-latch architecture while Sect. 1.3 deals with the architecture of the energy back-up system. Section 1.4 discusses the thermal, electrical and life-cycle characterization of the new proposed energy back-up system. Conclusions are drawn in Sect. 1.5.

## 1.2 The New E-latch Electronic Control Unit

Figure 1.1a shows the modular architecture of the *E-Latch*, which is divided in two main units (Latch and Cinch). Each unit includes: (1) a micro-controller with LIN connectivity and multiple PWM output channels, (2) a high-voltage protection circuit for the direct connection to a 12 V power supply, (3) an integrated H-bridge power MOS motor driver to drive an electrical motor, (4) electrical motor and Hall sensors to carry on the door lock/release and monitor the door status, respectively. The Latch sub-unit, the detailed architecture of which is shown in Fig. 1.1b, is generally used in all the vehicles, since it manages the basic door locking and release (with special child-lock or double lock or anti-theft lock functionalities), whereas the Cinch is a special function, which automatically and gently closes the door when the door is leaved ajar by the user, to be installed in premium vehicles only.

The Latch sub-unit is connected to the body computer through a LIN port, while the Cinch module, when present, is a slave of the Latch one. The operating temperature of the *E-latch* spans from  $-40$  to  $80^{\circ}\text{C}$ , and thousands of open/lock cycles are expected in its lifetime. The electronics must also withstand temperatures up to  $130^{\circ}\text{C}$  during the repainting process of a vehicle door. The micro-controller and the protection circuitry are realized by a System-on-a-Package (SIP) device with TQFP48 package, the Quest from Freescale [7], which integrates in the same package a digital chip (a 16-b S12 CPU with 20 MHz clock frequency, several kB of FLASH and RAM memory, 16-b timer) and an analog chip sustaining up to 18 V with on-chip temperature sensor, integrated low-drop out 2.5 V/5 V voltage regulator, 10-b ADC, multi-channel PWM module for high/low-side drivers, Hall sensor front-end, GPIO pins.

The integrated motor driver, from STMicroelectronics [8], is provided in a MultiPowerSO-30 package. It contains a dual monolithic high-side driver and two low side switches, with Power MOSFET and intelligent signal/protection circuitry. All it is able to sustain PWM motor control up to 20 kHz with 40 and 30 A voltage and current maximum values, well above the requirements of the Latch or Cinch modules.

The *E-latch* can work in two power modes: full power mode, where all the sub-units are working; power-down mode, where all the devices are off and the ECU is ready to be waken-up by the watchdog timer or an external interrupt. The residual current consumption in this mode useful when the vehicle is parked is a few microampere. The *E-latch* complies with the paradigm of the safety-critical electronic design as dictated by ISO/DIS 26262 [9].

As the E-latch future market volumes are foreseen in millions of pieces, an envisaged evolution of the proposed architecture consists in partitioning the Latch and Cinch units, currently realized via hardware, via software, by adopting a single 32-bit automotive microcontroller, with a 64 pin package at least. Such kind of devices, which represents the next generation of automotive processors [10–12] from different vendors (e.g. TX03 family by Toshiba, SPC56 family by STMicroelectronics, Tricore family by Infineon, Fado and Bolero families by Freescale), are often

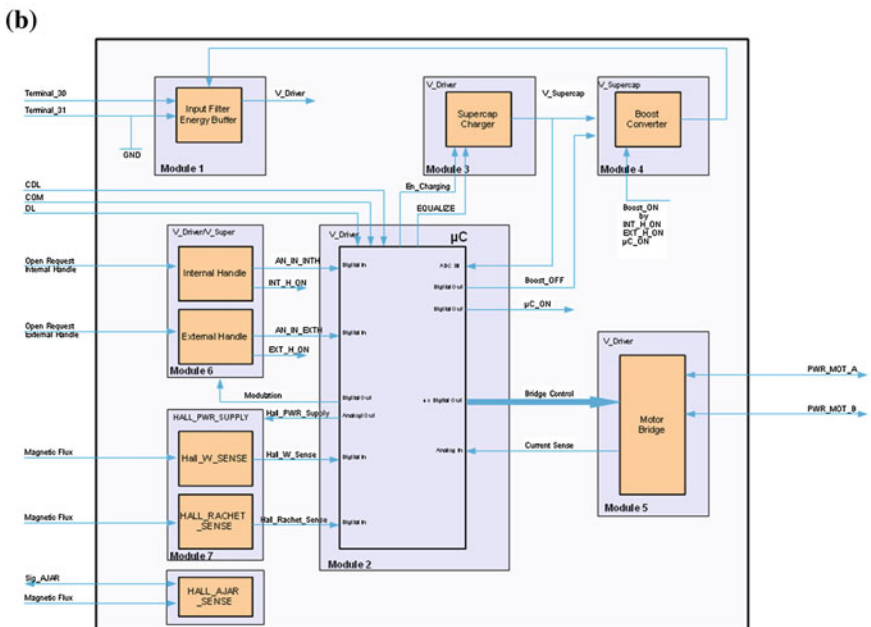
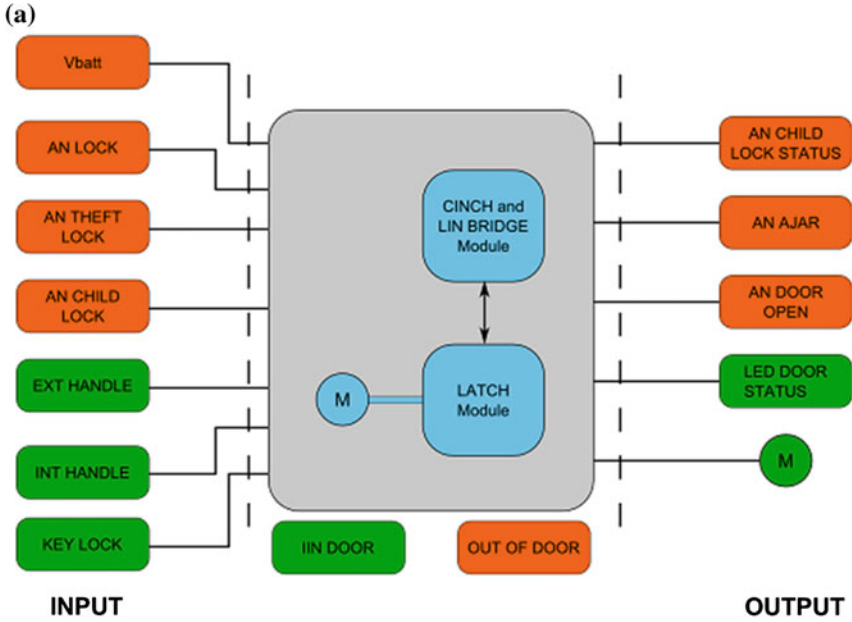


Fig. 1.1 a E-latch block diagram with Latch and Cinch functions. b Schematic of the E-latch unit

equipped with a double core thus increasing redundancy and hence fault-robustness. This way, the Cinch function or other advanced tasks can be added/removed by simply changing the firmware while the hardware of the E-latch ECU remains the same. The microcontroller 12 V protection/power managing circuitry (currently integrated in the single-package Quest device), the integrated motor drivers and the sparse glue logic could be realized single-chip as a custom ASIC, thus reducing the size and assembly cost of the E-latch. This new architecture can be the revolutionary approach to a completely new door system that, beside the E-latch, currently includes other two ECUs, the window lifter (integrating intelligent functionalities as the anti-pinch software) and the mirror control. A single 32-bit powerful automotive microcontroller could manage all the software tasks and the communication with the car body computer, while distributing multiple applications specific ICs for sensor interfacing and motor driving, one for each function (mirror, window lifter, latch/cinch), instead of having 3 different ECUs.

Whichever architecture is adopted, a key issue for door ECUs is guaranteeing the correct behaviour when the main battery fails: a supercapacitor-based energy back-up system has been designed to this aim, and characterized in terms of electrical, thermal and durability performance.

### 1.3 Architecture of the Energy Back-up Unit for Automotive ECU

The energy back-up system of an automotive door systems must operate from  $-40$  to  $80^{\circ}\text{C}$ , and withstand up to  $130^{\circ}\text{C}$  in case of door repainting. The energy back-up unit is kept charged by the main vehicle battery in normal conditions, so that it can provide enough energy (tens of joule in short bursts of about 100 ms, for about 100 W in power, 8–12 V in voltage and 6–10 A in current) to ensure several door releases in case of main battery failure. The energy back-up unit should be close to the ECU, robust to wiring failures, with minimum overhead in terms of cost, size and weight. Supercapacitor based energy storage systems are used in cars, but mainly for higher energy/power levels (tens of kWh/kW) [13–17]. Energy back-up solutions for low-power embedded systems are found in the literature mainly for ICT or consumer applications, not meeting the harsh environment requirements of vehicles. Our choice was exploring the use of super-capacitors as storage devices in the E-latch application, because of the large temperature range and the high power density needed. Lithium batteries, widely adopted [13, 14] for automotive electric or hybrid propulsion, would provide better energy density; however, a burst release of power is needed in the E-latch application when the emergency release is activated (supercapacitors provide better power density [15–21]) and the required temperature range is not covered by Lithium-based rechargeable batteries, typically limited up to  $60^{\circ}\text{C}$ . There are Lithium batteries (3.6 V Li-SOCl<sub>2</sub>) that operate up to  $150^{\circ}\text{C}$ , but they are non rechargeable and with high series resistance. Hence, these batteries

seem more suited as very long-term energy storage devices, useful to keep the energy backup system charged during the winter parking of cabriolet cars or every time the main battery is disconnected for a long time. In conclusion, Electric Double Layer Capacitors (EDLC) with 2.5 V supply and tens of Farads, available from several vendors such as Elna, Nichicon, Cooper-Bussmann, Maxwell, were selected as energy storage devices for the energy back-up system of the E-latch.

Since the electric motor of the Latch or Cinch needs a minimum drive of 8 V, and considering also redundancy issues, the back-up system includes: two EDLC supercapacitors (2.5 V nominal) connected in series plus an on-board x2 boost converter. This solution provides a nominal voltage of 10 V and a minimum of 8 V when the supercapacitors are not completely charged. As an example, two 2.5 V 10 F supercapacitors connected in series provide up to 62.5 J, an energy sufficient for 10 door releases in case of main battery failure. In fact, each release typically requires 10 V and 6 A for 100 ms.

The switching architecture of the boost converter provides a high power efficiency in the voltage doubling. The PWM controller is realized with the TI TL5001A IC and small external RC components in the feedback loop, mounted on the same PCB. The *E-latch* micro-controller properly drives as open the boost converter switch SW1 and the feedback switch SW2 of Fig. 1.2, when the main battery voltage is present.

The boost converter is thus normally off, the supercapacitors maintain their backup energy and the resistors of the converter feedback do not waste power. When the main battery fails, the micro-controller is supplied by the two supercapacitors in series (5 V), and the switches SW1 and SW2 are now turned on, so that the door latch electric motor can be supplied by the supercapacitors.

The switches SW1 and SW2 are realized with low-resistance MOS to maximize power efficiency. The main inductor also has a series resistance of few milliohms. The feedback is realized with a divider of the output voltage realized with two resistors of 30 and 330 k $\Omega$ . Since a complete characterization of supercapacitors of few farads for energy back-up in automotive ECU is missing in the literature, the devices to be used in the *E-latch* have been chosen after a thorough characterization campaign of 2.5 V EDLC supercapacitors in the range 10–25 F, provided from the above cited vendors. Given the available space, not all the experimental data are reported. Instead, the characterization tests are described in Sect. 1.4 and the results obtained for the selected device, the 18 F Nichicon device with  $V_{nom} = 2.5$  V and  $V_{max} = 2.7$  V, are showed.

## 1.4 Thermal, Electrical and Lifecycle Characterization of Supercap for Energy Back-up

Let us define  $C$  as the supercap capacitance,  $V_{nom}$  the nominal voltage,  $V_{max}$  the maximum allowed voltage,  $V_{ref} = 0.9 * V_{max} = 2.43$  V and  $I_{ref} = C * V_{ref}/30 = 1.62$  A. The following tests have been carried out.

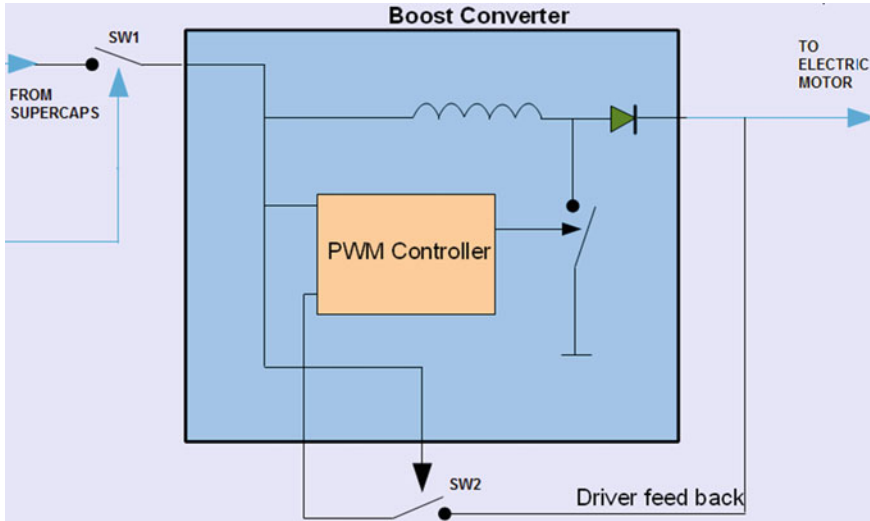


Fig. 1.2 Schematic of the boost-converter used in the back-up energy unit

*Constant-current charge/discharge capacitance test:* the device is charged at 23°C for 3 cycles at a constant current  $I_{test} = I_{ref}/4 = 0.405$  A up to  $V_{ref}$ , then it is kept at this constant voltage for 10 ms and then is completely discharged at constant current  $I_{ref}/4$ ; the 3-cycle test is repeated with current values of  $I_{ref}/2 = 0.81$  A and  $2I_{ref} = 3.24$  A. The supercap capacitance in charge and in discharge modes is calculated as  $C = I_{test} * T_{test} / V_{ref}$ .

*Constant-current ESR test:* the supercap series resistance (ESR) has a visible effect during the above described charge/discharge tests at the start of the discharge phase, where the current step determines a voltage drop. Dividing the voltage drop by the constant discharge current gives the ESR value.

*Leakage test:* the supercap tends to loose charge because of the auto-discharge; this phenomenon is modeled as a parallel time-variant leakage resistance. The supercapacitor is charged from 0 to  $V_{ref}$  at 23°C and is kept at such voltage value for 3 h. The capacitor current  $I_{leak}$  needed to hold the constant voltage value during this time interval is the leakage current. The parallel resistance  $R_p$  is the ratio  $V_{ref}/I_{leak}$  and it is calculated after 30 min, 1, 2 and 3 h. This test is usually repeated at different temperatures and for different durations.

*Technology spreading and thermal tests:* the leakage, ESR and capacitance tests are repeated using different samples of the same device to evaluate the technology spreading. The tests are also repeated on the same super-capacitor at different temperatures to determine the temperature dependence of capacitance, ESR and leakage.

*Durability-temperature test:* after 10 charge/discharge training cycles at 1 A, the supercap is characterized at 23°C using the above described procedures. This is the starting point of a durability test. A loop of 52 cycles is repeated. The loop consists of a first charge from  $V_{max}/2$  to  $0.9 V_{max}$ ; 50 charge/slow discharge cycles between

90 and 80 % of  $V_{max}$  with a charging current of  $I_{ref}/20 = 80$  mA and a discharge current of 10 mA follow; 1 last charge/fast discharge cycle between 90 and 70 % of  $V_{max}$  with a charge current of 80 mA and a discharge of  $I_{ref}/2 = 800$  mA completes the loop. The entire loop is then repeated 60 times. Such tests are repeated at 25,  $-40$  and  $80^\circ\text{C}$  for a total of around 10,000 cycles. The basic ESR-capacitance-leakage characterization at  $23^\circ\text{C}$  is carried out after each temperature value, to analyze degradations caused by the durability test.

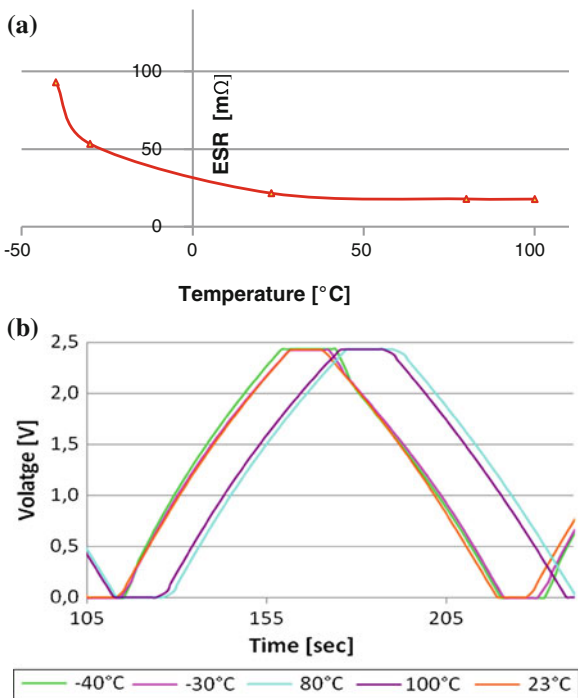
*Repainting test:* It consists of 15 min test at  $130^\circ\text{C}$  followed by 60 min at  $110^\circ\text{C}$ . The supercap is characterized at  $23^\circ\text{C}$  (ESR, leakage, capacitance) before/after the test, to evaluate possible performance degradations due to the repainting cycle. All the thermal tests are carried out in Binder MK53 thermal chamber.

The above described tests have been applied to several EDLC supercapacitors from different vendors, with 2.5 V nominal voltage and capacitance ranging from 10 to 25 F. We report here the main results obtained for the selected device, a 18 F Nichicon EDLC, that demonstrates the suitability of the supercap to solve the energy back-up issue of the *E-latch*. The capacitance and ESR tests at  $23^\circ\text{C}$  shows a measured capacitance and resistance of 17.69 F and 39.42 m $\Omega$ , respectively, at 0.405 A. Values of 18.47 F and 18.92 m $\Omega$  are found at 3.24 A. The measured capacitance differs less than 2.6 % from the nominal value of 18 F; the series resistance is well below 100 m $\Omega$ . The time-variant parallel resistance extracted from the leakage test at  $23^\circ\text{C}$  is in the order of some kilohm after 30 min and rises up to hundreds of kilohm after 3 h. Repeating the tests on different samples of the same super-capacitor gives a spreading of the results limited to few percent, showing a good repeatability of the device characteristics. A limited mismatch of the samples makes negligible the equalization problem that arises when two units are mounted in series, as it happens in the *E-latch* circuit. Instead, a higher dependence of the parameters on the temperature has been found, as expected from theory and from results presented in the literature for much larger size supercapacitors (up to of thousands of Farad) [15–21]. As an example, Fig. 1.3a shows the ESR measured with a  $I_{test} = I_{ref}/2 = 0.81$  A in a temperature range from  $-40$  to  $100^\circ\text{C}$ . The ESR value increases when the temperature decreases, but the series resistance remains always below 100 m $\Omega$ .

Since the capacitance value changes as a function of temperature, the voltage slope of the charge/discharge test changes in its turn, as it is demonstrated in Fig. 1.3b, where 0.81 A constant current tests at different temperatures are reported. While the ESR behavior is monotonic with the temperature and there is a large variation at low temperatures (Fig. 1.3a), the voltage slope value is instead weakly dependent on the temperature between  $-40^\circ\text{C}$  and room temperature (Fig. 1.3b). A difference in the time slope in Fig. 1.3b and hence in the capacitance is noticeable when going from room temperature to  $100^\circ\text{C}$ .

The slope and hence the capacitance increases with the temperature when going from  $-40$  to  $80^\circ\text{C}$ ; instead, the capacitance decreases going from 80 to  $100^\circ\text{C}$ . This behavior agrees with the results published in the literature over larger supercapacitors (thousands of Farad), in which a non-linear behaviour of the capacitance with the voltage is found. In fact, the capacitance is composed of a fixed part  $C_0$ ,





**Fig. 1.3** **a** Thermal dependence of the ESR in the charge-discharge test,  $I_{test} = 0.81$  A. **b** Thermal dependence of the voltage slope in the charge-discharge test,  $I_{test} = 0.81$  A

that increases with the temperature, and a voltage dependent part,  $C_v(V)$ , that instead decreases when temperature increases.

The repainting test up to 130°C does not seem to affect the supercap performance. Indeed, the ESR, leakage current and capacitance values measured after the repainting thermal cycle shows that the capacitance and ESR are 17.61 F and 27.58 mΩ respectively at 0.405 A, and 17.85 F and 20.11 mΩ respectively, at 3.24 A. The leakage resistance varies from 2 to 105 kΩ from 30 min to 3 h. Such values are acceptable for the normal use of a super-capacitor in the *E-latch*. Similar findings are obtained after the durability tests, see Table 1.1. The durability test consists of 10,000 cycles at temperatures from -40 to 80°C. It is found that the ESR only increases of a few percent and the capacitance decrease also is limited to a maximum of 10 %. These values are well acceptable for the application and demonstrate the suitability of the supercaps as energy back-up sources also after thousands of operating cycles. A major effect is instead noticed on the leakage current: the 3 h leakage value increases from 2 to 50 μA after the durability test. It means that after 10,000 cycles the investigated supercapacitor is completely aut discharged in about 10<sup>6</sup> s, i.e. around 11 days, if it is not recharged. This is not a problem when the main battery is working and it is continuously charging the supercapacitors (boost converter off). Should a

**Table 1.1** Performance derating after durability test

	ESR (mΩ )		Capacitance (F)		Leakage current
	0.405 A	1.62 A	0.405 A	1.62 A	
Initial	26.52	19.27	17.52	17.85	2 μA
Max. derating	27.35	19.47	16.04	16	50 μA
Change	3.03 %	1.03 %	8.45 %	10.36 %	–

main battery failure occur, the supercapacitors backup energy source is needed to actuate the door release (e.g. to escape the car after a road accident) and hence 11 days before the autodischarge are still a long time considering the typical *E-latch* application.

## 1.5 Conclusions

A new generation of ECU where the mechanical door closure system is actuated by a motor controlled by an electronic system is showing up on the car market. The *E-latch* brings advantages in system modularity, scalability, cost, size and weight. Due to the severe automotive safety-critical requirements, an energy back-up solution is needed to ensure door release/closure also in case of main battery failure, e.g. after a road accident when the emergency release must be guaranteed. An energy back-up solution based on small-size super-capacitors and a boost converter is proposed to this aim. An in-depth thermal, electrical and durability characterization of the supercaps proves their applicability as energy back-up and the reliability of the energy back-up unit for any automotive small energy back-up applications. The critical point is the right selection of the energy storage device. A thorough test and measurement campaign demonstrates that the selected EDLC supercaps allow for the required energy storage capability, with extended operating temperature range from  $-40^{\circ}\text{C}$  up to (non continuous)  $130^{\circ}\text{C}$ , low series resistance and leakage current, and low performance degradation even after 10,000-cycle durability test.

**Acknowledgments** This work has been supported by Tuscany Region under the project “AMDS: Advanced Mechatronic Door System”.

## References

1. Flamings, B.: Automotive electronics. *IEEE Veh. Technol. Mag.* **1**(1), 40–42 (2006)
2. Baronti, F., Lazzeri, A., Roncella, R., Saletti, R., Saponara, S.: Design and characterization of a Robotized Gearbox System based on Voice Coil Actuators for a Formula SAE race car. *IEEE/ASME Trans. Mechatron.* **18**(1), 53–61 (2013)

3. Costantino, N., et al.: Design and test of an HV-CMOS intelligent power switch with integrated protections and self-diagnostic for harsh automotive applications. *IEEE Trans. Ind. Electron.* **58**(7), 2715–2727 (2011)
4. Baronti, F., Lenzi, F., Roncella, R., Saletti, R., Di Tanna, O.: Electronic Control of a Motorcycle Suspension for Preload Self-Adjustment. *IEEE Trans. Ind. Electron.* **55**(7), 2832–2837 (2008)
5. Zhaoxia, X., Youcheng, L.: IEEE hardware design of automobile door with local interconnect network bus. In: *IEEE Conference on Control, Automation and System Engin (CASE)*, pp. 1–4 (2011)
6. Saponara, S., et al.: A flexible LED driver for automotive lighting applications: IC design and experimental characterization. *IEEE Trans. Power Electron.* **27**(3), 1071–1075 (2012)
7. Quian, H.: MM912F634 (Quest) Workshop. [www.freescale.com.cn/dwf/download/IDCQuestWorkshopPublic.pdf](http://www.freescale.com.cn/dwf/download/IDCQuestWorkshopPublic.pdf) (2010)
8. VNH2SP30: Automotive fully integrated H-bridge motor driver, p. 33 (2008)
9. Hillenbrand, M., et al.: Failure mode and effect analysis based on electric and electronic architectures of vehicles to support the safety lifecycle ISO/DIS 26262. In: *IEEE International Symposium on Rapid System Prototyping (RSP) 2010*, pp. 1–7 (2010)
10. Mayer, A., Hellwig, F.: System performance optimization methodology for Infineon’s 32-bit automotive microcontroller architecture. In: *2008 Design, Automation and Test in Europe (DATE) Conference*, pp. 962–966 (2008)
11. Zhang, G.: Freescale automotive microcontroller roadmap. August 2011, doc. n. FTF-AUT-F0783
12. Saponara, S., et al.: Architectural exploration and design of Time-interleaved SAR arrays for low-power and high speed A/D converters. *IEICE Trans. Electron.* **6**(E92-C), 843–851 (2009)
13. Brandl, M. et al.: Batteries and battery management systems for electric vehicles. In: *IEEE DATE 2012*, pp. 971–976 (2012)
14. Einhorn, M., Conte, F.V., Kral, C., Fleig, J.: Comparison, selection, and parameterization of electrical battery models for automotive applications. *IEEE Trans. Power Electron.* **28**(3), 1429–1437 (2013)
15. Gualous, H., Bouquain, D., Berthon, A., Kauffmann, J.M.: Experimental study of supercapacitor serial resistance and capacitance variations with temperature. *J. Power Sour.* **12**, 86–93 (2003)
16. Zhang, Y.C., Wei, L., Shen X.: Haiquan Liang: Study of supercapacitor in the application of power electronics. *WSEAS Trans.Circ. Syst.* **8**(6), 508–517 (2009)
17. Chang, J.H., Dawson, F.P., Lian, K.K.: A first principles approach to develop a dynamic model of electrochemical capacitors. *IEEE Trans. Power Electr.* **26**(12), 3472–3480 (2011)
18. Rizoug, N., Bartholoméüs, P., Le Moigne, P.: Modeling and characterizing supercapacitors using an online method. *IEEE Trans. Ind. Electron.* **57**(12), 3980–3990 (2010)
19. Gualous, H., Louahlia, H., Gallay, R.: Supercapacitor characterization and thermal modelling with reversible and irreversible heat effect. *IEEE Trans. Power Electr.* **26**(11), 3402–3409 (2011)
20. Kotz, R., Hahn, M., Gallay, R.: Temperature behavior and impedance fundamentals of suercapacitors. *J. Power Sour.* **154**(2), 550–555 (2006)
21. El Brouji, E.L.H. et al.: Aging assessment of supercapacitors during calendar life and power cycling tests. In: *IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 1791–1798 (2009)

# Chapter 2

## CH<sub>4</sub> Monitoring with Ultra-Low Power Wireless Sensor Network

Davide Brunelli and Maurizio Rossi

**Abstract** We propose a novel method to reveal and measure natural gas presence in air, using commercial off-the-self available MOX gas sensors in wireless sensor network applications. This technique reduces the power consumed by the catalytic sensors of a factor 10×, by an analysis on a reduced sampled period and thus extending the autonomy of battery operated systems. The information about the gas concentration is extracted from the sensor transient response through a discrete cosine transform (DCT) analysis and permits to immediately discriminate between clean-air and hazardous situations. The characterization of the sensing device has been conducted using a wide range of humidity and environmental conditions to demonstrate the effectiveness of the approach and a detailed comparison with the standard usage has been performed. Finally, the technique has been implemented in a Wireless Sensor Network designed specifically to measure air-quality in a large area and to share information over the internet.

### 2.1 Introduction

The detection of volatile chemicals is an essential to assess the air quality and the safety of indoor environments, because together with surveillance techniques [1], it guarantees to keep the environment safe and secure. Catalytic gas sensors are widely used in environmental monitoring applications because of their low cost, and are available for many kind of chemicals. Moreover, they are more robust with very low maintenance, they exhibit long life time with respect to electrochemical sensors

---

D. Brunelli (✉)  
University of Trento, Via Sommarive 14, 38123 Trento, Italy  
e-mail: [davide.brunelli@unitn.it](mailto:davide.brunelli@unitn.it)

M. Rossi  
University of Trento, Via Sommarive 5, 38123 Trento, Italy  
e-mail: [maurizio.rossi@unitn.it](mailto:maurizio.rossi@unitn.it)

and have fast response time. However the low selectivity and the big energy consumption are challenging problems if the energy availability onboard is constrained. Indeed, nowadays, many environmental monitoring projects are moving toward the use of wireless sensor networks, where any mW of power counts. Usually WSN are designed with low power sensors (e.g. temperature, light, pressure, acceleration...); thus catalytic gas sensors would impact with the highest power consumption with respect to any other component on the sensor node, including radio transceivers. Sensors of this kind are commonly used with continuous power supply leaving it always powered (e.g. in smoke detectors), or, at least, for a time interval sufficient to ensure a reliable response. Furthermore, the influence of air humidity variations has been never investigated in the sensor behavior. In this work, both the energy reduction and the humidity influence are taken into consideration to describe the effectiveness of the proposed method.

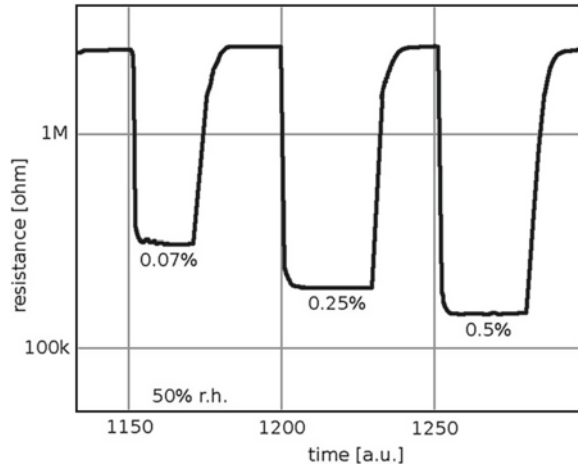
Analyzing features of the transient response such as the DCT, it is possible to determine gas concentration and its dependence on environmental conditions (in particular humidity). The outcome is an estimation of the gas concentration, which is, of course, less accurate than the traditional method, but still reliable and capable to discriminate between clean air and hazardous concentration, saving more than one order of magnitude in terms of energy absorbed by the sensing device. The goal is to outperform the state-of-art gas sensors in terms of energetic efficiency, providing, at the same time, a new method to integrate the traditional time-based characterization [2] for catalytic sensors. A Wireless Sensor Network (WSN) has been developed to characterize the autonomy of the systems when the sensor are used on battery operated boards. To pave the way to future developments, the coordinator mote has been USB-connected with a smartphone to add internet connectivity. With this configuration, the system can upload the data about air quality to the cloud and make them available everywhere. The power budget needed to maintain the network is also mitigated by compressive sampling techniques such as [3].

### ***2.1.1 Related Works***

Chemoresistive sensors are usually targeted at natural gas and combustibles detection, with a focus on performance in terms of ppm/ppb rather than consumption [4, 5]. A great variety of reliable sensors exists, but no one designed to low power applications as it is mandatory for WSNs. However, recently, electrochemical sensors and new catalytic sensors have been presented [6], with low consumption and developed to achieve good performance in environmental monitoring applications.

Unfortunately, an exhaustive characterization of these innovative sensors is not reported, and generally electrochemical devices exhibits a limited lifetime, due to the consumption of the electrochemical reactive elements. Thus, smart and not destructive power management, is still fundamental to achieve ultra-low power consumption with traditional and more robust technology. Some researchers focused their attention on the strategy used to sense the environment. Articles [7–10] propose efficient

**Fig. 2.1** AS-MLK standard response for continuous monitoring



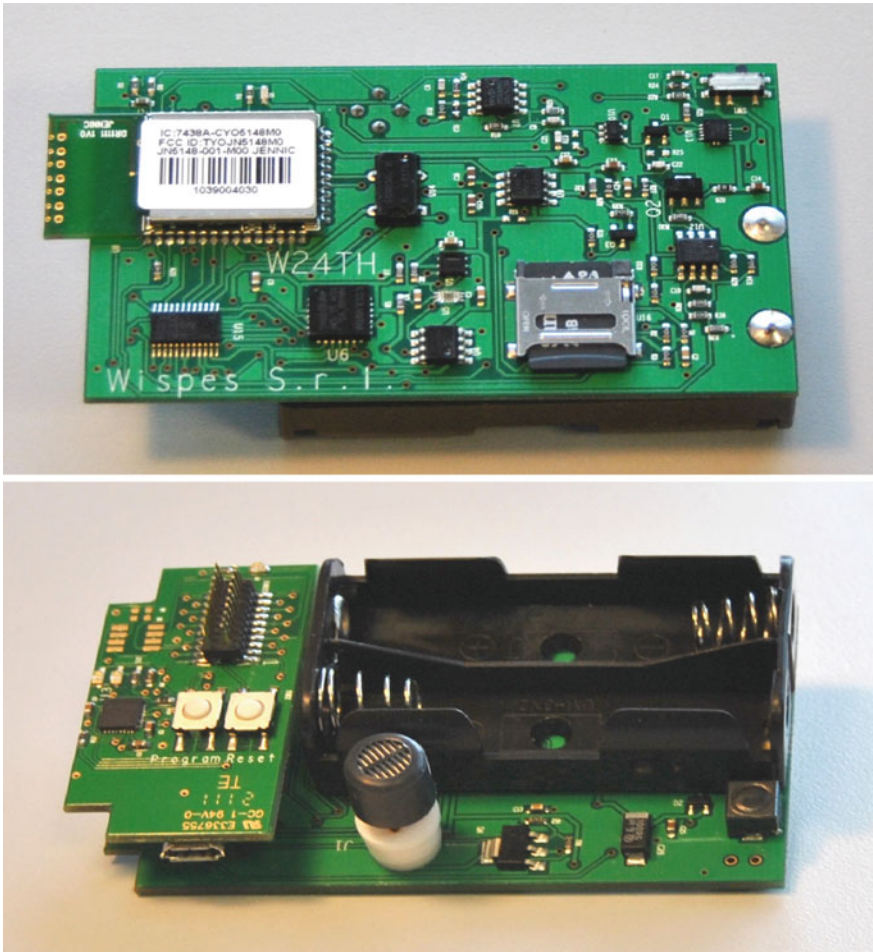
duty-cycle activity of the node, and achieve an extension of the life of a node by  $2\times$  or  $3\times$ , still using the sensor as indicated by the manufacturer. Other kind of optimization were introduced in the hardware of nodes [11, 12], achieving significant reduction of wasted power when the device is in idle or sleeping state.

## 2.2 Gas Measurement Characterization

To validate our approach, we used the AS-MLK natural gas sensor, from the Applied Sensor.<sup>1</sup> This is intended for mass market application which key requirements are long lifetime, low cross sensitivity and long term stability. The AS-MLK is targeted at real-time monitoring applications, this means that it must be always switched on to have a prompt response, as shown in Fig. 2.1, where the characteristic provided by the manufacturer is depicted, and the output resistance (versus time) changes quickly to varying concentrations. This device is able to detect gas level in air in the order of hundreds part per million, well below the explosive threshold (5%), and temperature and humidity slightly influence the measure. Generally, catalytic gas sensors need a constant voltage supply for a reliable measure because the reversible chemical reaction is triggered by heat. The energy consumption are then closely related with the time a sensor needs to reach stability, however, to save power, this time has to be shorten. In environmental monitoring applications, a high frequency of measurement is enough to detect abrupt changes, especially when a dangerous situation is unlikely to happen.

A duty-cycle strategy, in which a measure lasts for less than 6 s and are repeated with intervals of 2 min are shown in Figs. 2.3 and 2.4. The first one was collected with

<sup>1</sup> AS-MLK Datasheet, <http://www.appliedsensor.com>.



**Fig. 2.2** W24TH mote used in our testbed with MOX sensor onboard

30 % of relative humidity in the fluxed mixture (technical air + natural gas), while the latter with 50 %. In both the cases the sensor reaches the stability in the response and it is easy to distinguish the gas level, which traces out the values extracted from the characteristic curve in the datasheet (Fig. 2.1). Unfortunately this approach does not achieve the expected performance, because of the transient response is too long and requires too much power.

### ***2.2.1 Ultra-Low Power Strategy***

The transient responses presented in this paper are illustrated in Figs. 2.3 and 2.4, and are detailed in Figs. 2.5 and 2.6. Generally, the larger the humidity, the slower

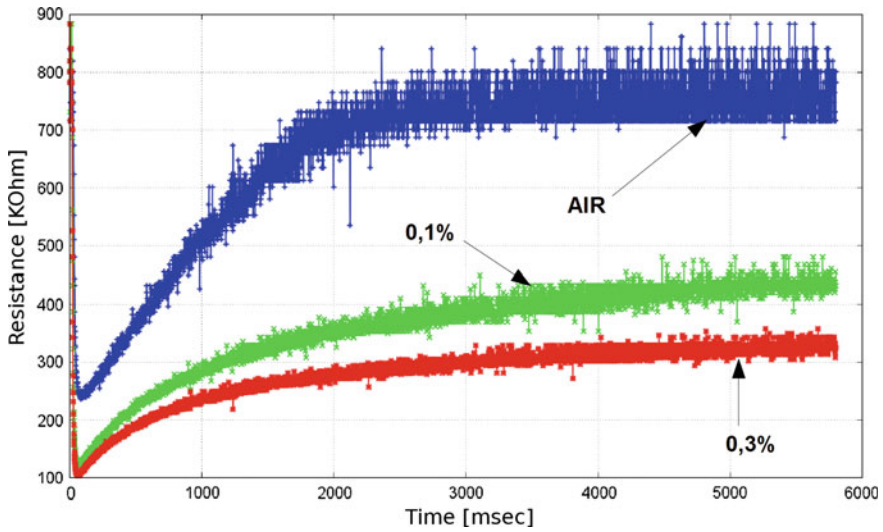


Fig. 2.3 Standard output response 20 °C with 30 % RH, 5 % duty-cycle

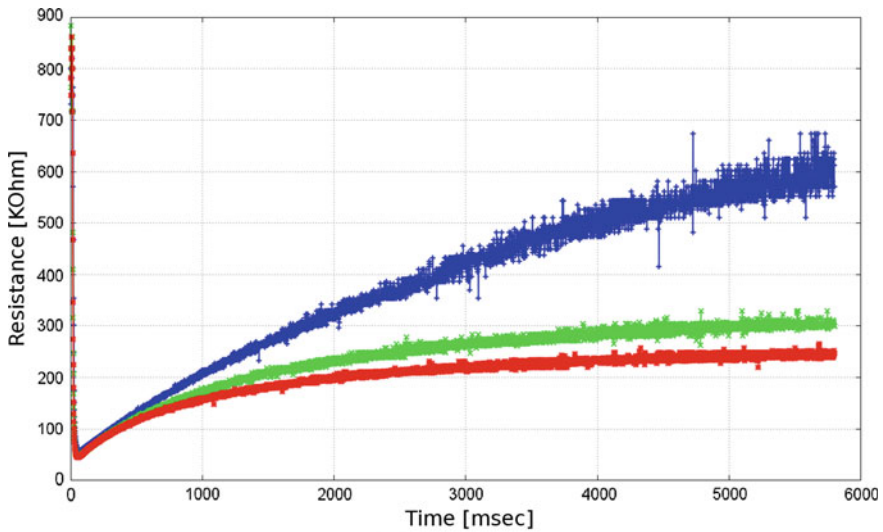


Fig. 2.4 Standard output response 20 °C with 50 % RH, 5 % duty-cycle

the response time. Thus to achieve a good trade-off between energy saving and reliability, reducing the duty-cycle by decreasing the power-on time is not sufficient, because in some environmental conditions it is not possible to discriminate the gas concentration from few samples.



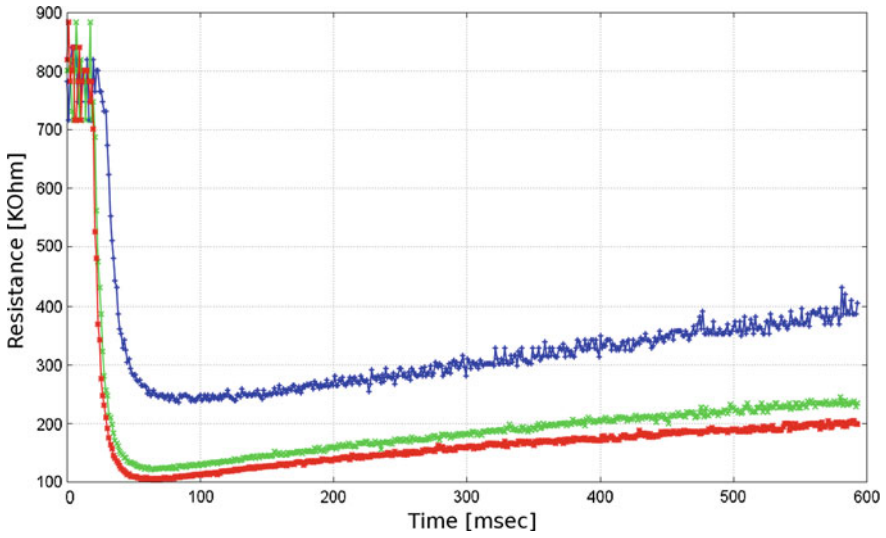


Fig. 2.5 Detail on first 512 samples 20 °C with 30 % RH, 5 % duty-cycle

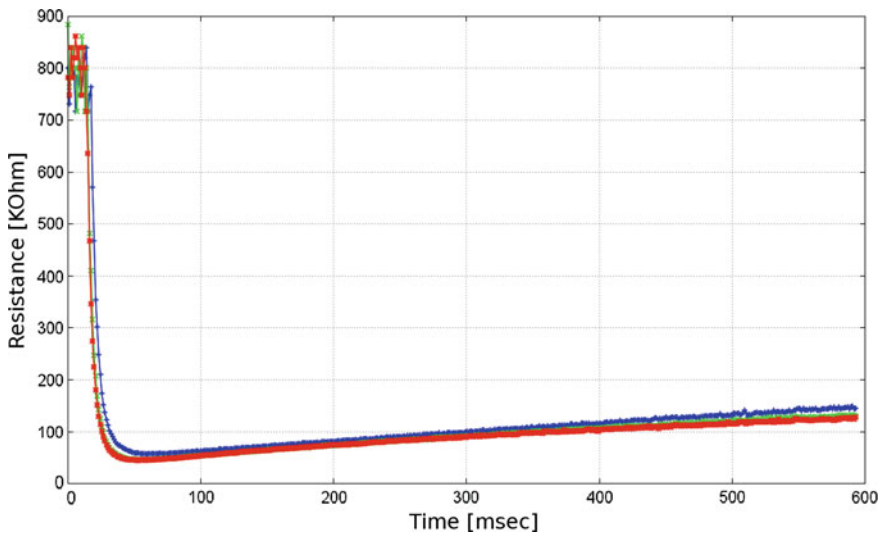


Fig. 2.6 Detail on first 512 samples 20 °C with 50 % RH, 5 % duty-cycle

The output resistance of the catalytic sensor is an aperiodic signal and can be interpolated to extract a continuous spectrum through the Discrete Fourier Transform. From the analysis of the normalized amplitude spectrum, it has been observed that the components around 20 Hz, are pretty proportional to the gas concentration, despite the parameters listed before (i.e. time and humidity). This property has

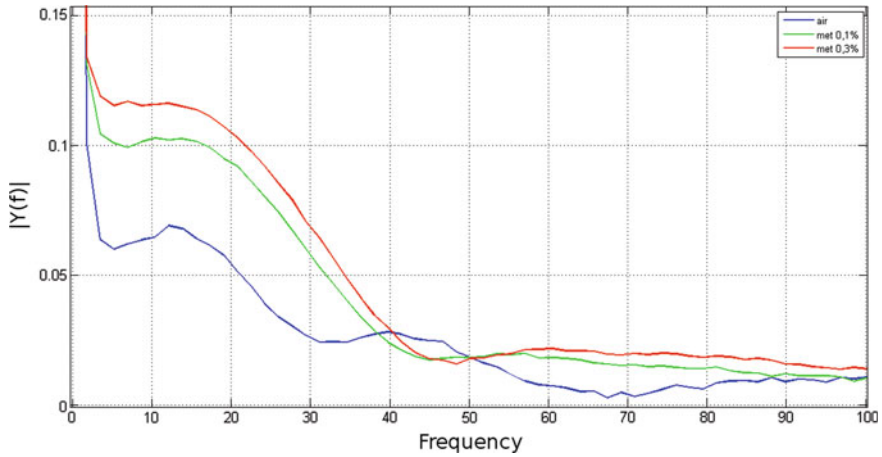


Fig. 2.7 DFT on interpolated standard response 20 °C with 30 % RH, 5 % duty-cycle

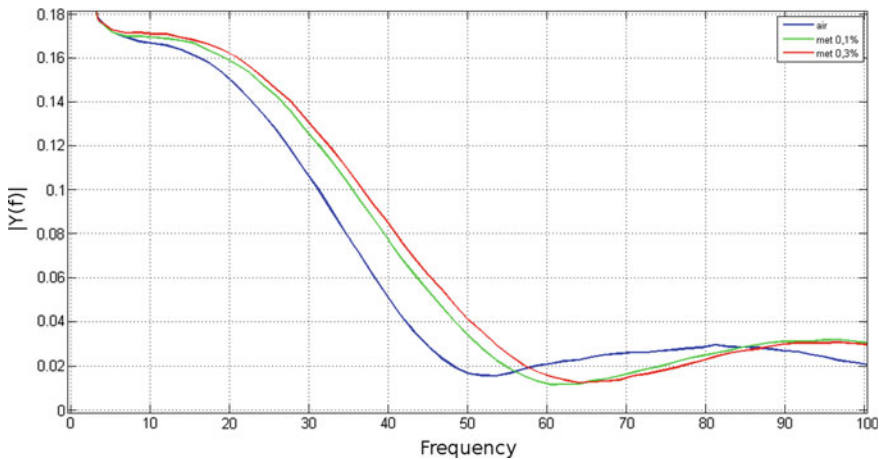
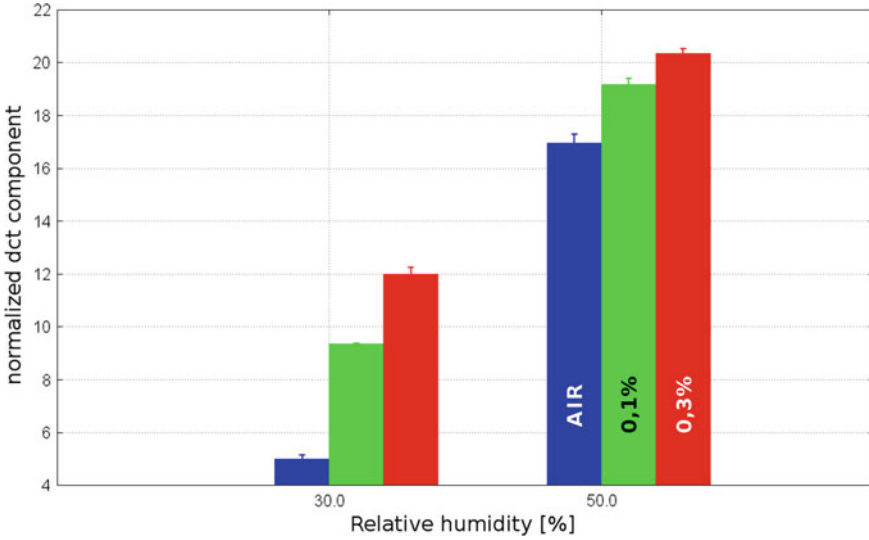


Fig. 2.8 DFT on interpolated standard response 20 °C with 50 % RH, 5 % duty-cycle

been fully characterized and then used as feature to assess the gas concentration, while reducing the energy need by the device. Figures 2.7 and 2.8 show the normalized amplitude spectrum of the first 512 samples extracted from the experiments of Figs. 2.3 and 2.4. A Normalized Discrete Cosine Transform (DCT) has been implemented to concentrate on only one component at a time (20 Hz), as the definition shown in Eq. 2.1.

$$\hat{X}_k = \frac{\sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi \frac{k}{N}n}}{\sum_{n=0}^{N-1} x_n} \quad (2.1)$$



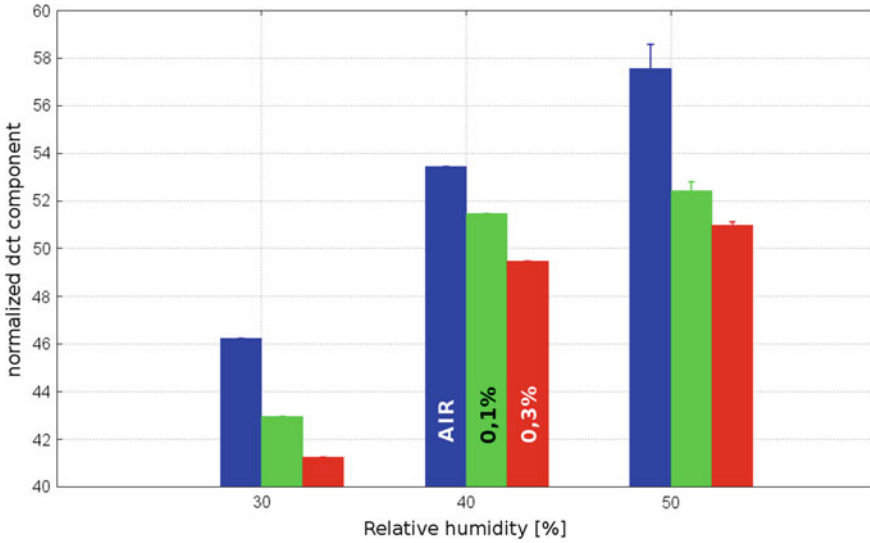
**Fig. 2.9** Characteristic response with DCT analysis 20 °C versus RH, 0.5 % duty-cycle

The main advantage of the DCT is the lack of complex computation, thus reducing the arithmetical operations and the amount of memory required, finally resulting in a fast execution of the task. Thus, the smart characterization of the sensor consists in the analysis of the spectrum obtained by the normalized DCT transform of 512 samples taken every 1  $\mu$ s, but it is strongly related with the strategy employed (i.e. repetition interval between measurements).

### 2.2.2 Characterization

Natural gas is a dangerous volatile substance, thus it is important to guarantee a frequent measurements in the environment, and to assess the features of the proposed technique, we compared it with others implementation characterized by longer repetition interval, namely 2 min: 0.5 % duty-cycle, and 15 min: 0.07 % duty-cycle.

In both the characterization, showed in Figs. 2.9 and 2.10, the results are promising, despite the reduced range of relative humidity condition presented, due to the limits of the gas bench used. The first consideration is related to the very small standard deviation in the measures (the thick lines on top of each bin) which suggests the possibility to reach a smooth characterization and a quantitative determination of this chemical. The other is the behavior of the sensor, strongly related with the sleep time. For short interval, higher the concentration, higher the normalized measure, the opposite in the other case. This underline the importance of defining the sleep interval before each measurements and a careful characterization of the sensor response.



**Fig. 2.10** Characteristic response with DCT analysis 20 °C versus RH, 0.07 % duty-cycle

However, these figures demonstrate that discriminating natural gas presence in air is possible, with a minimal energy. Moreover such techniques are compatible with renewable aware policies such as the scheduling proposed in [13].

### 2.3 Energy Saving in WSN Monitoring Applications

The ultra-low power approach, presented in this paper, permits to reduce of one order of magnitude the energy required to estimate the gas concentration in air, decreasing the response time from 6 s of continuous supply to almost 0,6 s. The whole characterization of the devices and the final example presented in the results were conducted interfacing the sensor to battery operated and resource constraints platform such as a node of a Wireless Sensor Network. Specifications and performance of the nodes can be found in [11]; the most remarkable are the computational architecture (a 32bit microcontroller 32 MHz useful to perform on line processing), with integrated RF module, IEEE 802.15.4 compliant, and integrated antenna, sensors for temperature, relative humidity, light and dock for the catalytic sensors.

The analysis of the consumption is plotted in Fig. 2.11. The trace represents the profile of the power consumption and it is taken using 1  $\Omega$  shunt-resistor. It can be split in two parts: processing and transmission. From the data collected during the tests, we notice that the average duration of the measurement phase (collecting temperature, humidity, battery level, gas sampling, processing and log into SD-card) completes in 2.5 s with an average current consumption of 29 mA, while sleeping

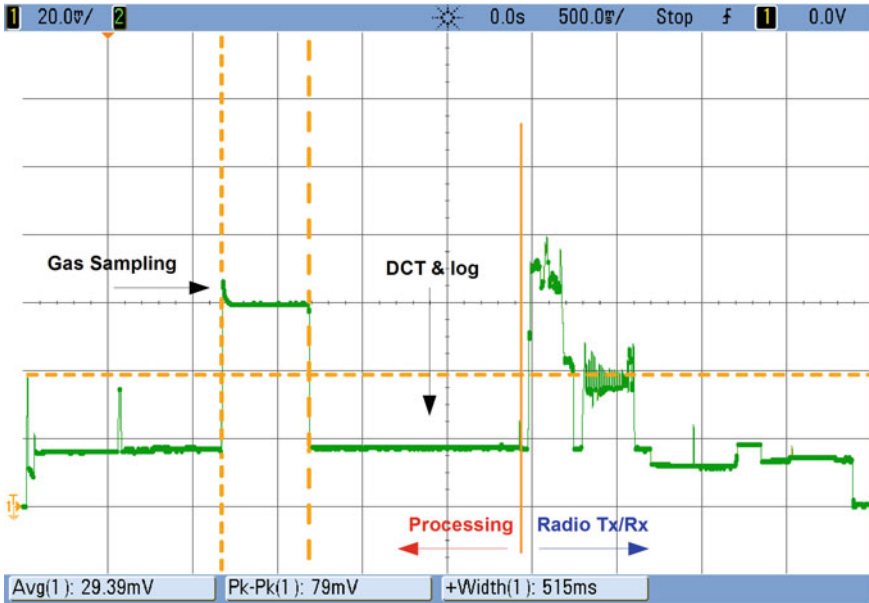


Fig. 2.11 Power consumption profile of the W24TH node

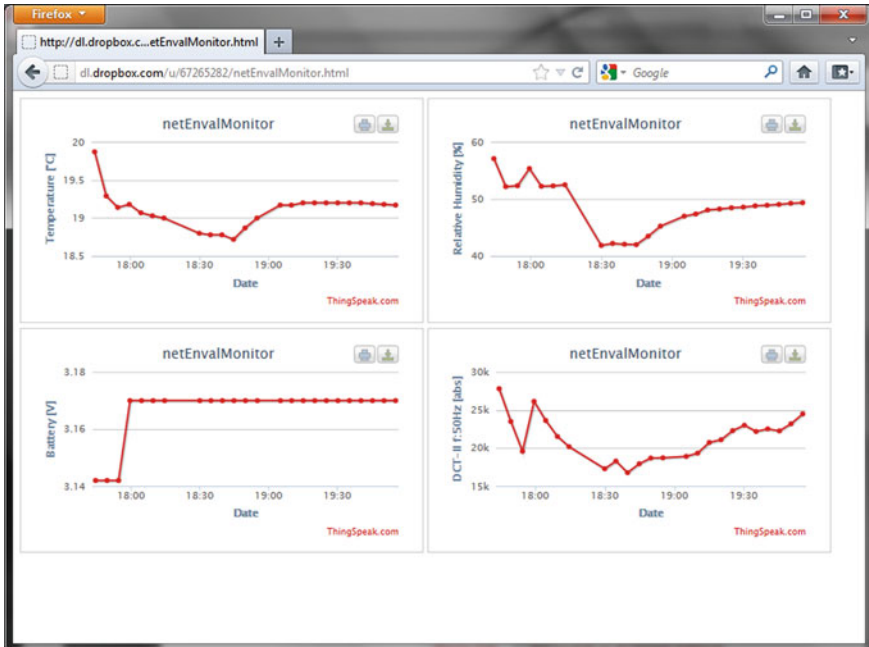


Fig. 2.12 Web based user interface to represent collected data



**Fig. 2.13** Picture of the extended wireless sensor network setup

the motes drains  $8\ \mu\text{A}$  in average (oscillator on during sleep). Using two 2,500 mAh batteries, it is possible to reach nearly 168 weeks of autonomy in the case of 15 min interval, considering that gas sampling is not the only task to execute on the microcontroller. Of course, better autonomy performance can be achieved if the system is equipped with energy harvester devices [14–17] capable to extract and convert energy from the surroundings. Figure 2.13 shows the extended version of the environmental monitoring network, where the coordinator node has been connected to a smartphone, running Android Ice Cream Sandwich (ICS) OS. The data collected by the coordinator are then sent by USB to the smartphone, which uploads the information to the web. Figure 2.12 is a screenshot of the monitoring application’s user interface.

## 2.4 Conclusion

To extend the lifetime and energy autonomy of air monitoring devices, a new strategy of sensing have been investigated for available commercial off-the-shelf gas sensors. The approach is cheaper and faster with respect to developing a new silicon sensor-device. The results presented are straightforward with a reduction of one order of magnitude in energy consumption that has been achieved using the AS-MLK catalytic sensor for natural gas detection, by reducing the sampled interval to  $\approx 500\ \mu\text{s}$  compared to the 5 s, at least, of the standard approach. In the near future we expect to

make a more exhaustive testing and to reduce even more the sampling period of the aerosols, to achieve an aggressive power saving strategy useful for an environmental monitoring application.

**Acknowledgments** The work presented in this paper was supported by the project *GreenDataNet*, funded by the EU 7th Framework Programme (grant n. 609000), and by the Autonomous Province of Trento within *EnerViS* — ‘Energy Autonomous Low Power Vision System’ project.

## References

1. Magno, M., Tombari, F., Brunelli, D., Di Stefano, L., Benini, L.: Multimodal video analysis on self-powered resource-limited wireless smart camera. *IEEE J. Emerg. Sel. Top. Circ. Syst.* **3**(2), 223–235 (2013)
2. Varpula, A., Novikov, S., Haarahiltunen, A., Kuivalainen, P.: Transient characterization techniques for resistive metal-oxide gas sensors. *Sens. Actuators B Chem.* **159**(1), 12–26 (2011)
3. Caione, C., Brunelli, D., Benini, L.: Distributed compressive sampling for lifetime optimization in dense wireless sensor networks. *IEEE Trans. Ind. Inf.* **8**(1), 30–40 (2012)
4. Xu, L., Li, T., Gao, X., Wang, Y.: A high-performance three-dimensional microheater-based catalytic gas sensor. *IEEE Electron Device Lett.* **33**(2), 284–286 (2012)
5. Zhang, P., Vincent, A., Kumar, A., Seal, S., Cho, H.J.: A low-energy room-temperature hydrogen nanosensor: utilizing the schottky barriers at the electrode/sensing-material interfaces. *IEEE Electron Device Lett.* **31**(7), 770–772 (2010)
6. Somov, A., Baranov, A., Savkin, A., Ivanov, M., Calliari, L., Passerone, R., Karpov, E., Suchkov, A.: Energy-aware gas sensing using wireless sensor networks. In: Picco, G., Heinzelman, W. (eds.) *Wireless Sensor Networks*, ser. Lecture Notes in Computer Science, vol. 7158, pp. 245–260. Springer, Berlin (2012)
7. Vito, S.D., Palma, P.D., Ambrosino, C., Massera, E., Burrasca, G., Miglietta, M., Francia, G.D.: Wireless sensor networks for distributed chemical sensing: addressing power consumption limits with on-board intelligence. *IEEE Sens. J.* **11**(4), 947–955 (2011)
8. Rossi, M., Brunelli, D.: Ultra low power wireless gas sensor network for environmental monitoring applications. In: 2012 IEEE Workshop on Environmental Energy and Structural Monitoring Systems (EESMS), pp. 75–81 (2012)
9. Rossi, M., Brunelli, D.: Analyzing the transient response of mox gas sensors to improve the lifetime of distributed sensing systems. In: 2013 5th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), pp. 211–216 (2013)
10. Choi, S., Kim, N., Cha, H., Ha, R.: Micro sensor node for air pollutant monitoring: hardware and software issues. *Sensors* **9**, 7970–7987 (2009)
11. Jelicic, V., Magno, M., Brunelli, D., Paci, G., Benini, L.: A context-adaptive multimodal wireless sensor network for energy-efficient gas monitoring. *IEEE Sens. J.* **13**(1), 328–338 (2013)
12. Bhattacharyya, P., Verma, D., Banerjee, D.: Microcontroller based power efficient signal conditioning unit for detection of a single gas using mems based sensor. *Int. J. Smart Sens. Intell. Syst.* **3**(4), (2010)
13. Moser, C., Brunelli, D., Thiele, L., Benini, L.: Real-time scheduling with regenerative energy. In: 18th Euromicro Conference on Real-Time Systems (ECRTS06), 2006, pp. 261–270. DC, USA, Washington (2006)
14. Dondi, D., Bertacchini, A., Larcher, L., Pavan, P., Brunelli, D., Benini, L.: A solar energy harvesting circuit for low power applications. In: IEEE International Conference on Sustainable Energy Technologies (ICSET 2008), 2008, pp. 945–949 (2008)
15. Magno, M., Marinkovic, S., Brunelli, D., Popovici, E., O’Flynn, B., Benini, L.: Smart power unit with ultra low power radio trigger capabilities for wireless sensor networks. In: Design, Automation Test in Europe Conference Exhibition (DATE), 2012, pp. 75–80 (2012)

16. D. Porcarelli, D. Brunelli, M. Magno, and L. Benini. A multi-harvester architecture with hybrid storage devices and smart capabilities for low power systems. In: 2012 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), pp. 946–951 (2012)
17. Weddell, A.S., Magno, M., Merrett, G.V., Brunelli, D., Al-Hashimi, B.M., Benini, L.: A survey of multi-source energy harvesting systems. In: Design, Automation Test in Europe Conference Exhibition (DATE), 2013, pp. 905–908 (2013)



# Chapter 3

## Integrated Front-end Electronics for Silicon PhotoMultiplier Readout in Medical Imaging Applications

**Nahema Marino, Sergio Saponara, Luca Fanucci, Federico Baronti, Roberto Roncella, Francesco Corsi, Cristoforo Marzocca, Gianvito Matarrese, Fabio Ciciriello, Francesco Licciulli, Maria Giuseppina Bisogni and Alberto Del Guerra**

**Abstract** The 4D-MPET project aims to design a positron emission tomography detection module capable of working inside a magnetic resonant imaging system. The proposed detector will feature a three-dimensional architecture based on two tiles of silicon photomultipliers coupled to a single LYSO scintillator on both its faces. Silicon photomultipliers are magnetic-field compatible photo-detectors with a very small size enabling novel detector geometries that allow the measurement of the depth of interaction. Furthermore they can be fabricated using standard silicon technology, have a large gain in the order of  $10^6$  and are very fast thus allowing evaluating the time of flight. Based on custom integrated circuits, the readout electronics include an innovative current mode front-end coupled to a novel time to digital converter. The former, implemented in AMS  $0.35\ \mu\text{m}$  SiGe-BiCMOS technology, features a very low input impedance ( $17\ \Omega$ ) current buffer and a large bandwidth (1 GHz), which lead to a time resolution of  $\sim 100$  ps FWHM. The time to digital converter exploits the combination of a submicron technology (UMC 65 nm LLLVT) together with a systolic topology so as to work at a high frequency of 2.5 GHz. This yields to a nominal time resolution of 29 ps ( $\sigma$ ) whereas the photon energy is evaluated with a bin size of 400 ps by using a time over threshold technique. Finally, the depth of interaction measurement is performed by an external FPGA with a simulated spatial resolution of 1.3 mm FWHM along the z coordinate.

---

N. Marino (✉) · S. Saponara · L. Fanucci · F. Baronti · R. Roncella  
Dip. di Ing. dell'Informazione-Università di Pisa and INFN Sez. Pisa, Via Caruso 16, 56122  
Pisa, Italy  
e-mail: nahema.marino@iet.unipi.it

F. Corsi · C. Marzocca · G. Matarrese · F. Ciciriello · F. Licciulli  
Dip. di Ing. Elettrica e dell'Informazione-Politecnico di Bari and INFN Sez. Bari, Via Orabona 4,  
70125 Bari, Italy

M. G. Bisogni · A. Del Guerra  
Dip. di Fisica-Università di Pisa and INFN Sez. Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy

### 3.1 Introduction

Positron emission tomography (PET) is a molecular imaging technique that provides images of physiological processes inside the body. After decay of a radiotracer injected into the tissues, a positron is released which annihilates with an atomic electron thus producing two 511 keV photons (event). These pairs of gamma rays are emitted in almost exactly opposite directions within the PET scanner, as depicted in Fig. 3.1. Coincidence detection of the two photons is exploited to provide projection data used for image reconstruction. In clinical applications, PET image quality benefits from the time of flight (TOF) feature which is given by the difference in the arrival time of the two photons on the detectors of the PET ring. This information is used to estimate the position of the annihilation point along the line connecting the centres of the two detectors (LOR) thus enabling the reconstruction algorithm to provide the final image with fewer iterations and less image noise [1]. It has been shown that the quality benefits of TOF increase when increasing the time resolution for large size objects [2]. Thus, high speed electronics are required for the TOF measurement. However, although a single imaging modality can offer some insight into a disease process, in many circumstances the combination of both the morphological and functional information eventually provides a better diagnosis and prognosis. In this scenario, the innovative combination of PET with magnetic resonance imaging (MRI) within a hybrid facility offers better contrast in soft tissues (i.e. human brain) with respect to scanners where PET is coupled to computer tomography (CT) [3]. Furthermore, hybrid systems address most of the problems related to separate multi-modality imaging, where PET and MRI images are software fused, such as different patient positioning on the couch, involuntary movement of internal organs and artefacts due to the fusion techniques [4, 5].

The implementation of TOF PET/MRI systems require innovative detector layouts featuring dedicated front-end designs. In this scenario, the use of SiPM coupled to fast electronic readout is fundamental to achieve good performance in terms of time and spatial resolution thus enabling the implementation of the TOF technique. Indeed, these solid state photo-sensors are MR-compatible and combine the advantages of both Photomultiplier Tubes (PMT), such as high gain and total quantum efficiency (QE), and Avalanche Photodiodes (APD), like the small dimensions which permits extremely compact, light and robust mechanical design.

The 4D-MPET (4 Dimensions Magnetic compatible module for Positron Emission Tomography) INFN project collaboration aims to develop a magnetic-field compatible TOF PET module with good spatial, time and energy resolution. To this end, innovative readout electronics are being developed to be combined with novel scintillator materials and high performing photodetectors.

The proposed TOF PET/MRI module is based on the combination of a single LYSO scintillator crystal coupled to SiPMs which allow a high precision in the determination of the  $(x, y)$  coordinates of the hits on the detector. Both the arrival time and the energy of the events are measured by integrated readout electronics.

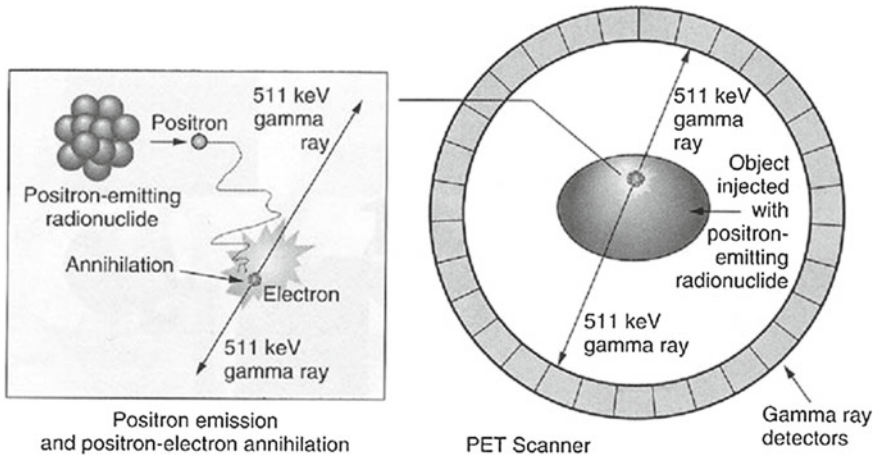
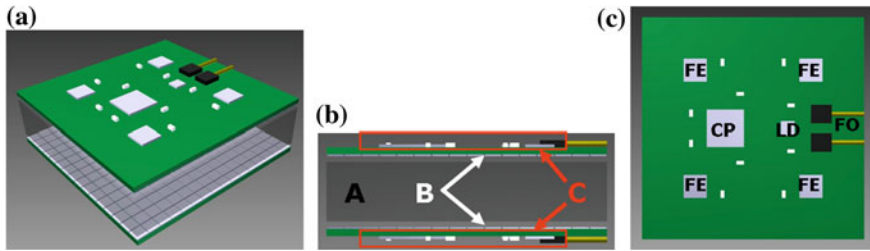


Fig. 3.1 Schematic representation of PET

The detection module will also take advantage of the depth of interaction (DOI) feature related to the  $z$  coordinate of the gamma rays inside the crystal. Indeed, this information reduces the parallax error in the determination of the LOR [6].

## 3.2 Readout of the TOF PET/MRI Module

In Fig. 3.2 the block layout is depicted [7]. The LYSO scintillator slab (A in Fig. 3.2b) has a size of  $48 \times 48 \times 10 \text{ mm}^3$ ; spatial resolution optimization is achieved by painting the side faces of the crystal block. A SiPM layer is laid out both on the top and at the bottom of the scintillator (B in Fig. 3.2b). The detector matrices feature  $16 \times 16$  square pixels with 3 mm pitch and a microcell size of  $50 \mu\text{m}$ , developed by FBK-irst (Trento, Italy [8]). Each pixel is read out by independent and identical multiple-channel electronics (C in Fig. 3.2b). Each board features several electronic blocks as shown in Fig. 3.2c: (1) four front-end (FE) application-specific integrated circuits (ASIC), each featuring 64 channels, for time and energy measurement; (2) a cluster processor (CP) ASIC for data reduction; (3) a laser driver/photodiode receiver/clock reconstruction (LD) ASIC for communication with an external data acquisition system through optical fibres (OF). Communication among ASICs is based on low voltage differential signalling (LVDS) pads for magnetic field compatibility. The electronics have to be mounted and wire-bonded without package; encapsulation for protection and top-side contact cooling will be performed after test. Given the SiPM jitter of 60 ps [9], a time resolution less than  $\sigma = 100 \text{ ps}$  must be reached to implement the TOF technique. Such resolution can be achieved by triggering events at a low threshold ( $TH_{\text{low}}$ ), which corresponds to the first emitted photoelectron so as to measure the interaction time within the crystal with high accuracy. However, the high SiPM dark noise of



**Fig. 3.2** Block detector layout: three dimensional view (a), side view (b) and top view (c)

$\sim 2$  MHz/mm<sup>2</sup> at room temperature (dark count for SiPMs manufactured by FBK-irst at the selected operating  $V_{\text{bias}}$ ) requires the additional use of a validation threshold  $\text{TH}_{\text{high}}$  (corresponding to a programmable number of emitted photoelectrons) to discriminate events from noise as in a double threshold technique.

Simulations have shown that a resolution of  $\sigma = 102$  ps can be achieved [10]. Once a valid event has been detected, the system translates the relevant timestamp into a digital word by a converter and evaluates the associated energy by exploiting a time over threshold (TOT) technique. The best IC implementation of the discussed double threshold technique can be implemented with the aid of two different CMOS technologies. Therefore, the FE ASIC has been split up into two separated chips which communicate with each other: a current-mode (CM) ASIC which converts all SiPM analogue outputs into digital pulses and a time to digital converter (TDC) which provides both TOF and TOT information of valid pulses only. Furthermore, both clustering and DOI evaluation techniques are being examined on an external FPGA before moving to the CP ASIC implementation.

### 3.2.1 Current Mode ASIC

The readout circuitry of the TOF PET module is based on a current-mode ASIC with low input impedance, fast current buffer as very first detector front-end [11]. Such architecture is implemented in AMS 0.35  $\mu\text{m}$  SiGe-BiCMOS technology featuring both MOSFET and fast HBT bipolar transistors. This allows achieving a time resolution of  $\sim 100$  ps FWHM given the very low input impedance ( $17 \Omega$ ) of the current buffer which leads to a short time constant even when combined with the large SiPM capacitive load. Furthermore, a wide bandwidth of about 1 GHz ensures a very short rise time of the response, which limits the jitter due to noise. Power consumption is also controlled at reasonable values, thanks to the large transconductance/current ratio offered by the fast HBT bipolar transistors.

The CM ASIC reads out the SiPM output current and translates it into a digital pulse whose rise edge timestamp represents the arrival time of the signal (TOF data) whereas the fall edge timestamp provides its energy information (TOT data).

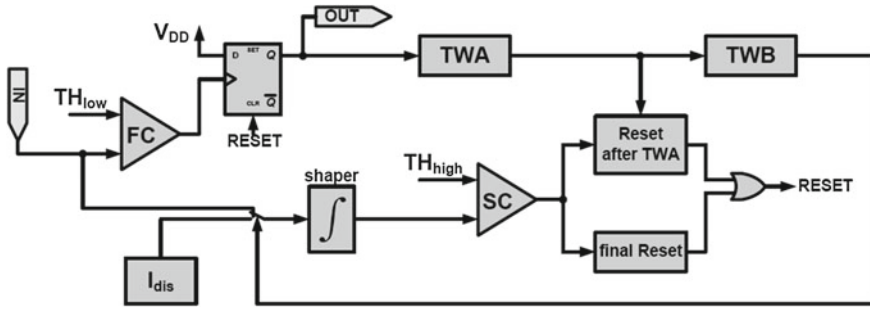


Fig. 3.3 Block diagram of the CM ASIC

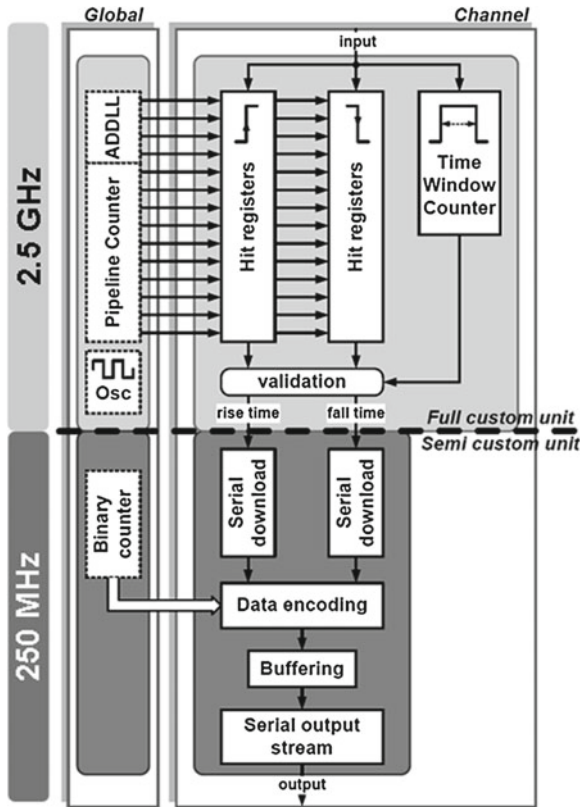
Figure 3.3 shows the block diagram of the CM ASIC. At the input of the circuit, a SiPM output signal is compared with the low threshold  $TH_{low}$  by a fast comparator (FC). As soon as the input goes over  $TH_{low}$ , the FC sets a flip-flop which in turn triggers a counter labelled as Time Window A (TWA).

At the same time a shaper starts integrating the input so that its output, which feeds a slow comparator (SC), increases from zero. When TWA expires, the system evolves towards two different statuses according to the value of the SC output. If it is 0, the SiPM signal is recognized as noise because the shaper output is below the validation threshold ( $TH_{high}$ ); the circuit resets itself and a digital pulse with duration of TWA is produced at the circuit output. If it is 1, a second counter labelled as Time Window B (TWB) is triggered and the shaper keeps on integrating the SiPM output; when TWB expires the shaper is discharged through a constant current thus decreasing its output. As soon as the shaper output goes below  $TH_{high}$  the circuit resets itself. The circuit outputs a digital pulse having duration of TWA plus TWB plus the time required by the shaper to be discharged: this time is proportional to the energy associated to the event and thus it represents the TOT information. It follows that the rising edge of a valid pulse provides the timestamp information (TOF) whereas the trailing edge contains the energy data.

The non-linear, constant current discharge mechanism of the charge integrated during TWA and TWB eliminates any constraint on the stability and uniformity of the SiPM pulse and relaxes precision requirements in the TDC for the TOT measurement, while preserving a good linearity. For calibration issues the counters TWA and TWB will be programmable within the time intervals (6–30) and (50–120) ns, respectively. Thus, in the event of noise, the CM ASIC outputs a short pulse which can be easily distinguished from valid data by the TDC ASIC.

### 3.2.2 Time to Digital Converter ASIC

The digital outputs of the CM ASIC are forwarded to the TDC ASIC which measures and digitizes both the TOF and TOT data of the events and neglects noise pulses. Time



**Fig. 3.4** Block diagram of the time to digital converter ASIC

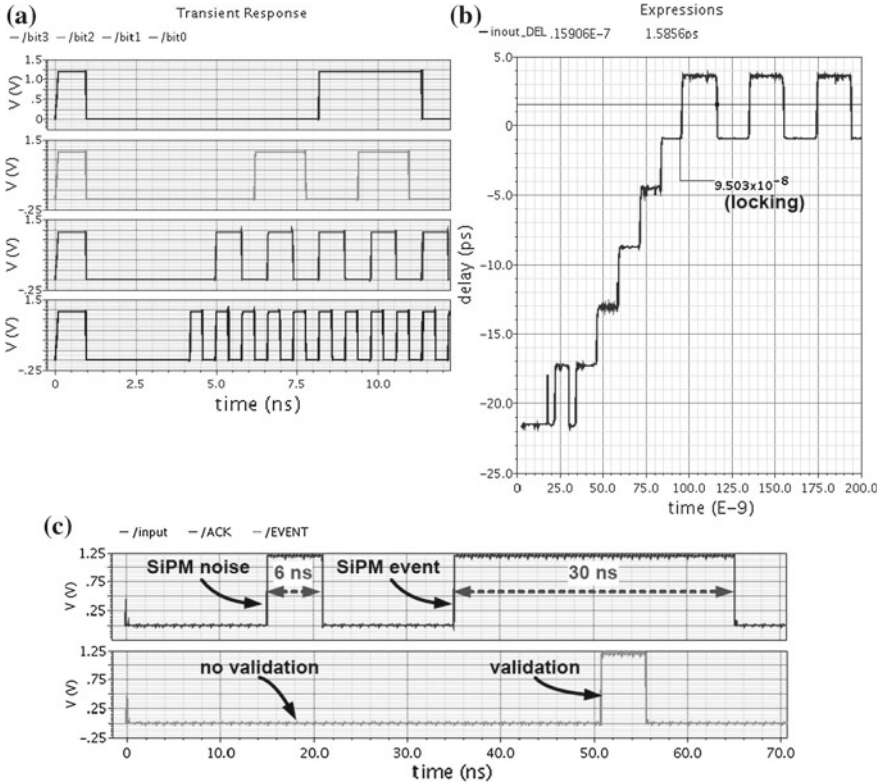
to digital converters are electronic components which quantize small time differences between two signals (referred to as start and stop) and provide a digital representation of this time interval [12]. Technology progress, increased integration level capability and working speed have led to significant improvements in the time measurement performance. Some examples involve the time resolution [13], linearity [14], dynamic range [15], power consumption [16] and low process-voltage-temperature (PVT) variations [17] of the converter. The operation of a TDC is very similar to that of analogue to digital converters (ADC) although the former deals with the time difference rather than voltage or current differences. The measured time is evaluated as the phase difference between the positive edges of the start and stop signals. Thus, the input is defined in the continuous time domain whereas the output is expressed in digital form.

The proposed TDC has been implemented in UMC 65 nm CMOS technology and its block diagram is reported in Fig. 3.4. Here, two sections can be distinguished: a full custom unit which has been designed at transistor level and represents the core of the converter, and a semi custom unit that is based on standard cells for data management.

The TDC core is based on a pipeline architecture where local communication between adjacent cells arranged in a systolic array is exploited to implement any logic function with almost no frequency dependency. In systolic arrays, computations are performed simultaneously in all processing elements (referred to as systolic cells), while data travel from cell to cell, so that parallelism is exploited by partitioning the computation effort over the array elements [18]. In addition, working at high frequencies makes it possible to exploit master-slave dynamic flip flops to implement the systolic cells. Here, parasitic capacitances are used as data storage units and logic functions can be obtained with a smaller number of devices when compared to static logics, thus saving silicon area. These characteristics make this approach suitable to applications where multichannel feasibility is required. The proposed TDC makes use of a 10 bit systolic counter [19, 20] working at 2.5 GHz that provides a timestamp  $T_{\text{coarse}}$  of 400 ps for the TOT measurement. The counter is coupled to a 4 stages delayed locked loop (DLL) which leads to a timestamp  $T_{\text{fine}}$  decreased down to 100 ps to evaluate the TOF data. This corresponds to a theoretical time precision ( $\sigma$ ) of 29 ps which would satisfy the system requirements for the TOF feature. Moreover, the reduced number of delay elements in the DLL is promising in terms of linearity which is fundamental in one shot measurements as for PET. In Fig. 3.5a, the counter bits from b3 to b0 are depicted: it can be noticed that each bit is updated with one clock delay with respect to the previous one because of the pipeline latency. Figure 3.5b shows the behavior of the DLL until it locks to the input clock.

In order to implement a fully synchronous design, the 2.5 GHz system clock is provided by a configurable ring oscillator (RO). The RO is based on 4 tunable inverters where 8 digitally controlled varactors [21] serve as externally configurable load so that the oscillation frequency can be changed by the user with a resolution of 1 % per bit with respect to the central frequency in typical conditions. A fifth inversion is then provided by a NAND gate that receives an external enable signal. The DLL delay units have been designed with a configuration similar to that employed for the RO elements in order to guarantee the achievement of the locking condition between the DLL and the oscillator in any working condition.

The 10 bit systolic counter along with the DLL and the ring oscillator are global blocks inside the TDC and feed 8 readout channels that are accommodated within the same chip. In each channel, two clusters of pipeline hit registers are used: one group is in charge of sampling both the counter and the DLL digits in correspondence to the rise edge of the inputs sent by the CM ASIC. Another cluster samples the counter bits only in order to perform the fall edge measurement with a timestamp of  $T_{\text{coarse}}$ . In addition, a dedicated programmable systolic counter (Time Window Counter in Fig. 3.4) evaluates the width of the channel inputs so as to discard noise pulses without impairing the acquisition capability of the system. Indeed, a validation algorithm is implemented in each channel which allows for a correct data sampling even in the presence of a high noise rate. Figure 3.5c shows an example of both input discard and validation conditions by means of the Time Window Counter with an input width threshold set to 12.8 ns and assuming that the noise pulses sent from the CM ASIC are 6 ns large.



**Fig. 3.5** Systolic counter output bits from b3 to b0 (a); evolution of the DLL behavior from the start to the locking condition (b); example of the input validation performance of the Time Window Counter (c)

The recorded timestamp and energy information are collected at 250 MHz by the semi custom unit with a double hit resolution of 170 ns. Here, an 18 bit binary counter is exploited to extend the TOF dynamic range to 1.048 ms. Each time validation is provided by the Time Window Counter, TOF and TOT bits are serially downloaded and stored in 4 word deep FIFOs after encoding. Finally, a 47 bit word is serially sent out to an FPGA for backend image signal processing.

### 3.2.3 Clustering and Other Features

The FPGA which receives the data sampled by the TDC performs clustering for data reduction. Some algorithms are being investigated with the aim to transfer the design to a custom ASIC once sufficient practical experience has been acquired with the



front-end prototypes. Furthermore, asymmetry in the cluster size on the two crystal faces can be exploited for DOI evaluation.

Several approaches have been proposed in literature in order to optimize the DOI resolution but none of them offers a final solution to the problem given some issues related to costs, complexity or the overall performance of the techniques. Simulations of the method under study have shown that the proposed detection module allows the processing of PET images with a z resolution of 1.3 mm FWHM [9].

In order to minimize the amount of cables inside the magnetic field, the LD ASIC will be designed. Thus, only two optical fibers (one input and one output) will serve the communication between the cluster processor output and the external acquisition board. MRI compatibility packaging issues have to be studied for the laser and photodiode assemblies. It is assumed that they would have limited length pigtailed to avoid connectors within the critical MRI volume.

Other features to be studied include some cooling technique which is fundamental to reduce the SiPM dark noise; shielding to relax MRI compatibility requirements; dual modality (clinical and preclinical) operation. Finally, for the detector connection to the host unit, beside mature technologies (such as Ethernet), also innovative fast and flexible communication links with high data-rate and reliability suitable for nuclear medicine applications are investigated. Particularly communication technologies already proved for high speed space and high energy physics applications will be also considered [22–24].

### 3.3 Conclusions

The design of a hybrid TOF PET/MRI scanner requires fast scintillation materials and light detectors as well as new front-end designs. Within the 4DMPET collaboration, novel ASICs are being developed which have to be employed by an innovative PET block detector in order to read out the output signals of SiPM matrices. The electronics are required to provide the TOF information by measuring the arrival time of the current pulses produced by true events with a resolution less than  $\sigma = 100$  ps. Simulations have shown that such a performance can be accomplished by using a double threshold technique.

Coupling an innovative current mode front-end ASIC with an efficient time to digital converter ASIC leads to a simulated resolution of  $\sigma = 29$  ps for the TOF measurement. In addition, both energy and DOI evaluation must be provided. The former is performed with the employment of a time over threshold technique which is implemented in the CM ASIC. Thereafter, the TDC ASIC digitizes the energy information with a bin size of 400 ps. Finally, an FPGA exploits data clustering to implement an innovative DOI evaluation technique which leads to a simulated spatial resolution of 1.3 mm FWHM.

Prototypes of the ASICs have been submitted to the foundry in early 2013 whereas a CP ASIC will be designed after further investigation of clustering techniques.

## References

1. Conti, M.: Why is TOF PET reconstruction a more robust method in the presence of inconsistent data? *Phys. Med. Biol.* **56**, 55–168 (2011)
2. Karp, J.S., Surti, S., Daube-Witherspoon, M.E., Muehlehner, G.: Benefit of time-of-flight in PET imaging: experimental and clinical results. *J. Nucl. Med.* **49**, 462–470 (2008)
3. Zaidi, H., Del Guerra, A.: An outlook on future design of hybrid PET/MRI systems. *Med. Phys.* **38**, 5667–5689 (2011)
4. Townsend, D.W.: Dual-modality imaging: combining anatomy and function. *J. Nucl. Med.* **49**, 938–955 (2008)
5. Stuart, D.: Fast track finding using radially pointing scintillating fibers. *JINST* **5**, C07006 (2010)
6. James, S.S., Yang, Y., Wu, Y., Farrell, R., Dokhale, P., Shah, K.S., Cherry, S.R.: Experimental characterization and system simulations of depth of interaction PET detectors using 0.5 mm and 0.7 mm LSO arrays. *Phys. Med. Biol.* **54**, 4605–4619 (2009)
7. Marino, N., Ambrosi, G., Baronti, F., Bisogni, M.G., Cerello, P., Corsi, F., Fanucci, L., Ionica, M., Marzocca, C., Pennazio, F., Roncella, R., Santoni, C., Saponara, S., Tarantino, S., Wheadon, R., Del Guerra, A.: An innovative detection module concept for PET. *JINST* **7**, C08003 (2012)
8. Llosa, G., Belcari, N., Bisogni, M.G., Marcatili, S., Collazuol, G., Melchiorri, M., Piemonte, C., Barrillon, P., Bondil-Blin, S., Dinu, N., de La Taille, C., Del Guerra, A.: Monolithic 64-channel silicon photomultiplier matrices for small animal PET. *Phys. Med. Biol.* **55**, 7299–7315 (2010)
9. Collazuol, G., Ambrosi, G., Boscardin, M., Corsi, F., Dalla Betta, G.F., Del Guerra, A., Galimberti, M., Giuliotti, D., Gizzi, L.A., Labate, L., Llosa, G., Marcatili, S., Piemonte, C., Pozza, A., Zorzi, N.: Single photon timing resolution and detection efficiency of the IRST Silicon Photomultipliers. *Nucl. Instr. Met. A* **A581**, 461–464 (2007)
10. Pennazio, F., Barrio, J., Bisogni, M.G., Cerello, P., De Luca, G., Del Guerra, A., Lacasta, C., Llosá, G., Magazzu, G., Moehrs, S., Peroni, C., Wheadon, R.: Simulations of the 4DMPET SiPM based PET module. In: *IEEE Nuclear Science Symposium and Conference (Record MIC)*, pp. 2316–2320 (2011)
11. Corsi, F., Foresta, M., Marzocca, C., Matarrese, G., Del Guerra, A.: ASIC development for SiPM readout. *JINST* **4**, 1–10 (2009)
12. Roberts, G.W., Ali-Bakhshian, M.: A brief introduction to time-to-digital and digital-to-time converters. *IEEE Trans. Circ. Syst. II Express. Briefs* **57**, 153–157 (2010)
13. Hsu, J.C., Su, C.: BIST for measuring clock jitter of charge pump phase-locked loops. *IEEE Trans. Instrum. Meas.* **57**, 276–285 (2008)
14. Baronti, F., Fanucci, L., Lunardini, D., Roncella, R., Saletti, R.: A technique for nonlinearity self-calibration of DLLs. *IEEE Trans. Instrum. Meas.* **52**, 1255–1260 (2003)
15. Guo, J., Sonkusale, S.: A 22-bit 110ps time-interpolated time-to-digital converter. In: *IEEE International Symposium on Circuits and Systems*, pp. 3166–3169 (2012)
16. Wang, K., Liu, Y., Toumazou, C., Georgiou, P.: A TDC based ISFET readout for large-scale chemical sensing systems. *IEEE Biomedical Circuits and Systems Conference*, pp. 176–179 (2012)
17. Hervé, C., Cerrai, J., LeCaër, T.: High resolution time-to-digital converter (TDC) implemented in field programmable gate array (FPGA) with compensated process voltage and temperature (PVT) variations. *Nucl. Instr. Met. A* **682**, 6–25 (2012)
18. Kung, H.T.: Systolic Array. In: Ralston, A., Reilly, E.D., Hemmendinger, D. (eds) *Encyclopedia of Computer Science* 4th ed., pp. 1741–1743. Wiley, Chichester, UK (2000)
19. Vuillemin, J.E.: Constant time arbitrary length synchronous binary counters. In: *Proceedings of 10th IEEE Symposium on Computer Arithmetic*, pp. 180–183 (1991)
20. Stan, M.R.: Systolic counters with unique zero state. In: *Proceedings of the 2004 International Symposium on Circuits and Systems*, vol. 2, pp. 909–912 (2004)
21. Andreani, P., Bigongiari, F., Roncella, R., Saletti, R., Terreni, P.: A digitally controlled shunt-capacitor CMOS delay line. *Analog Integ. Circ. Signal Proc.* **18**, 89–96 (1999)

22. Costantino, N., Borgese, G., Saponara, S., Fanucci, L., Incandela, J., Magazzu, G.: Development, design and characterization of a novel protocol and interfaces for the control and readout of front-end electronics in high energy physics experiments. *IEEE Trans. Nucl. Sci.* **60**, 352–364 (2013)
23. Saponara, S., Fanucci, L., Tonarelli, M., Petri, E.: Radiation tolerant spacewire router for satellite on-board networking. *IEEE Aerosp. Electron. Syst. Mag.* **22**, 3–12 (2007)
24. Magazzù, G., Borgese, G., Costantino, N., Fanucci, L., Incandela, J., Saponara, S.: Design exploration and verification platform, based on high-level modelling and FPGA prototyping, for fast and flexible digital communication in physics experiments. *JINST* **8**, P02021 (2013)

# Chapter 4

## Energy Autonomous Low Power Vision System

**Davide Brunelli, Alberto Tovazzi, Massimo Gottardi, Michele Benetti, Roberto Passerone and Pamela Abshire**

**Abstract** This paper presents the design and the development of a novel vision system, capable of sensing and describing the visual world it observes under physical constraints that include ultra-low power consumption, easy deployment, low maintenance cost, and a small unobtrusive form-factor. Energy aware vision processing algorithms have been developed based on the custom hardware. Simulation and design of an energy harvester using solar cells has been addressed to become the power supply unit of the proposed vision system. We describe the hardware-software architecture of the video sensor node and provide a characterization in terms of power consumption and power generation and energy efficiency of the harvester. Different strategies of energy harvesting, based on low energy DC–DC converter, and different types of storage device are analyzed, focusing on different battery technologies and comparing the different characteristic curves (charge and discharge curves). Specific attention will be reserved to different types of solar cells (amorphous and monolithic) in indoor environment.

---

D. Brunelli (✉) · A. Tovazzi · R. Passerone  
University of Trento, via Sommarive 5, 38123 Trento, Italy  
e-mail: davide.brunelli@unitn.it

A. Tovazzi  
e-mail: alberto.tovazzi@unitn.it

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

M. Gottardi · M. Benetti  
Fondazione Bruno Kessler, via Sommarive 18, 38123 Trento, Italy  
e-mail: massimo.gottardi@fbk.it

M. Benetti  
e-mail: michele.benetti@fbk.it

P. Abshire  
University of Maryland, College Park, MD 20742, USA  
e-mail: pabshire@umd.edu

## 4.1 Introduction

Low-cost video surveillance systems based on wireless embedded electronics have already entered the marketplace with the promises of flexibility, quick deployment, and accurate real-time visual data. However, many technical problems must still be overcome for widespread diffusion of such a technology. For instance, although research continues to develop higher energy-density batteries, the amount of available energy on board still severely limits the lifespan of distributed battery-operated embedded systems. This leads to the fact that a major challenge of battery-operated vision sensor systems is to maximize the lifetime of the network. One of the best solutions to achieve this goal is implementing alternative power sources which increase the autonomy of the systems considerably.

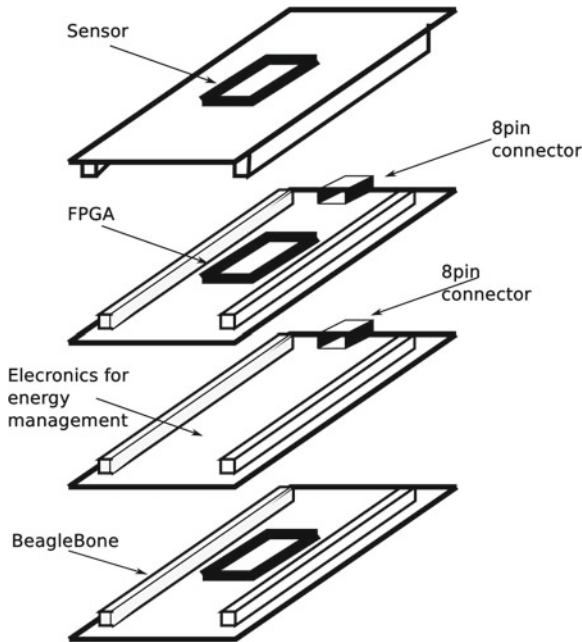
In this paper, we present an energy autonomous vision system, featuring an ultra low power camera sensor, interfaced with an energy-aware processor and a multiple source energy harvesting unit for supplying the whole system. The energy harvesting unit has been added to take a significant step forward with respect to current embedded implementations. For the vision sensor, the design considerations to optimize the autonomous system are twofold:

1. ultra low-power consumption has been a primary objective both from the hardware and the software design;
2. adaptive power management is an important design objective because power cannot be continuously available from the energy harvesting device as it happens in the case of standard battery supplied systems.

In the literature, few energy-autonomous vision and safety systems have been reported so far [1–7]. Although several recent works have been presented on ultra-low power imagers [8–10], many of them refer to imagers rather than vision sensors, delivering data all the time, and continuously loading, in turn, the processor. The vision sensor presented in this paper, by extracting spatio-temporal contrast from the scene, is one of the sensors with the best performance on power consumption among all the available vision chips.

The proposed system, as depicted in Fig. 4.1, consists of four stacked PCB boards:

- a sensor board that contains the imager;
- an acquisition board that contains an FPGA to perform image acquisition, processing and management;
- an energy supply and energy harvesting board that collects and supplies the energy required by the system;
- a control board that manages the system, stores the images in local memory and communicate with other systems.

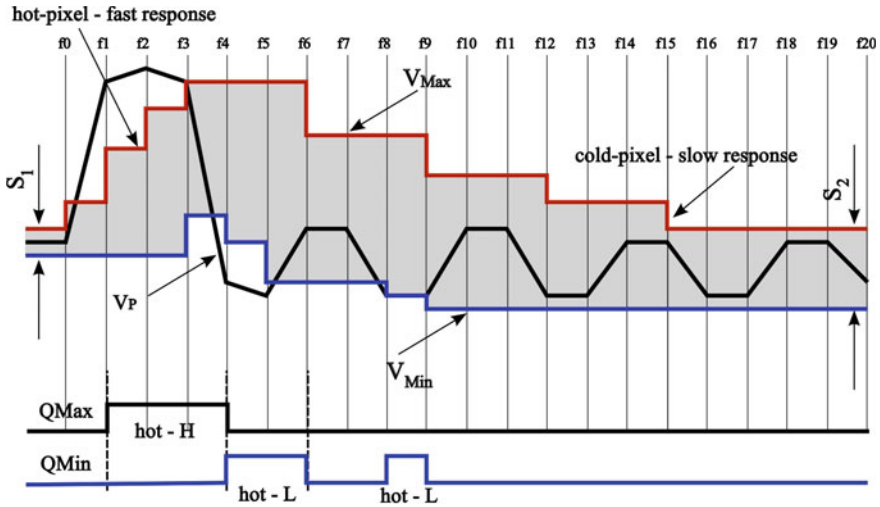


**Fig. 4.1** Stacked boards which build the system

## 4.2 Imager Architecture

Custom digital processors have been proposed for early visual processing [11–13], thanks to their high parallelism. They are mainly based on Single-Instruction Multiple-Data (SIMD) architectures, offering massive parallel processing capabilities. However, their performance are limited because they need to be fed by the video signal, which is slow. Embedding some intelligence at the sensor level, making it able to recognize and deliver only those data related to image features in the scene which are of interest, drastically increases the energy efficiency of the entire system, without losing performance.

Our designed ultra-low power vision sensor integrates a proprietary image processing algorithm for unusual event detection. The sensor embedded algorithm is based on a *Hot Pixel Learning and Detection* (HPLD) algorithm for scene analysis and interpretation. It has been designed using primary arithmetical operators such as *increment* and *decrement*, and *compare*, thus it is suitable to be implemented in CMOS, at pixel-level, turning into an efficient Single Instruction Multiple Data sensor architecture.



**Fig. 4.2** Dynamic background-subtraction, implemented at pixel-level.  $V_P$  is the current signal;  $V_{Max}$ ,  $V_{Min}$  are the boundaries defining the *COLD-pixels*

### 4.2.1 Imager Architecture

The basic operating principle of the algorithm, applied to a single pixel, is represented in Fig. 4.2. From frame to frame, the pixel light intensity ( $V_P$ ) is compared with the two dynamic thresholds ( $V_{Max}$ ,  $V_{Min}$ ), which take into account the pixel past behavior. As long as the current signal  $V_P$  changes within the two thresholds, no unusual events are detected (*COLD-pixel*). Trespassing one of the two thresholds, the pixel is recognized as anomalous and one of the two bits, either  $Q_{Max}$  or  $Q_{Min}$ , is asserted (*HOT-pixel* =  $Q_{Max} + Q_{Min}$ ). After the pixel readout, the two thresholds are updated according with the result of the detection phase.

The pixel schematic is shown in Fig. 4.3. The photodiode works in storage mode, buffered by a source follower. This last is turned on by  $V_{P\_clk}$  only when necessary, reducing the pixel dc power consumption. It also embeds two analog memories ( $Max$ ,  $Min$ ), keeping trace of the photodiode activity along time. The output of the sensor is binary, with two-bits/pixel. This binary image is ready to be processed outside the chip by the higher-layer algorithm, implementing complex vision tasks.

Figure 4.4 shows how the HPLD algorithm works on real image sequences acquired with a VGA camera. The algorithm delivers binary images, as shown in Fig. 4.5, where the black pixels are the *HOT-pixels* detecting anomalous events.

Figure 4.6 shows the block diagram of the sensor architecture for an array of  $64 \times 64$ -pixel prototype. The imager is an addressable array of pixels, with a 64-cell ROW DECODER and a 64-cell COLUMN DECODER. The vision sensor was designed using a  $0.35 \mu\text{m}$  CMOS technology. A prototype of  $64 \times 64$  pixel sensor was realized and fully tested (Fig. 4.7).

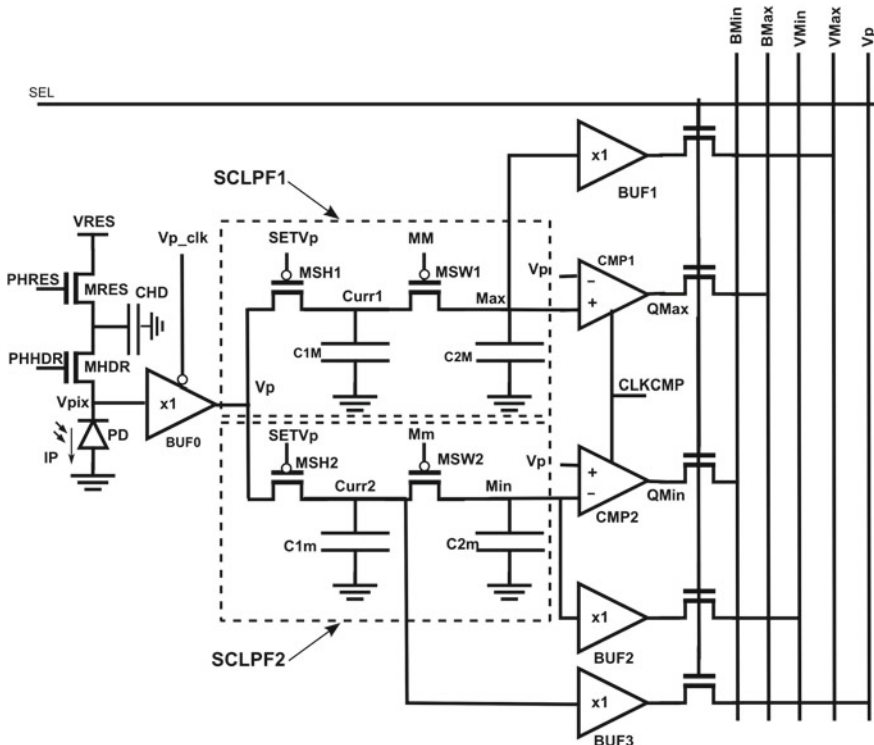


Fig. 4.3 Pixel schematic implementing dynamic background subtraction

### 4.3 Energy Harvesting System

Typical energy harvesting sources include solar radiation [14], vibrations [15], thermal gradient [16], kinetic energy [17], wind [18, 19], and electromagnetic energy [20]. The type of energy source to harvest depends on the type and amount of energy available in the considered environment. In this case the system is supposed to operate in indoor environments, such as offices, meeting rooms, or the entry hall of commercial buildings. Solar energy harvesters convert the sun’s radiation into electric energy. In a typical indoor environment, there is not much solar radiation unless the device is placed near a window on a sunny day. Nonetheless, the usage of indoor solar cells permits the scavenging of energy from very low levels of illumination (50–100 lux) even in case of artificial light. In the following sections, we will compare different types of indoor solar cells in order to determine the surface area needed to supply the system under typical indoor conditions.

The energy harvesting circuit (see Fig. 4.8) is designed to optimize the energy collected by the solar cell and to charge an energy accumulator. For a more flexible design, the circuit provides an adjustable output voltage for charging the battery and



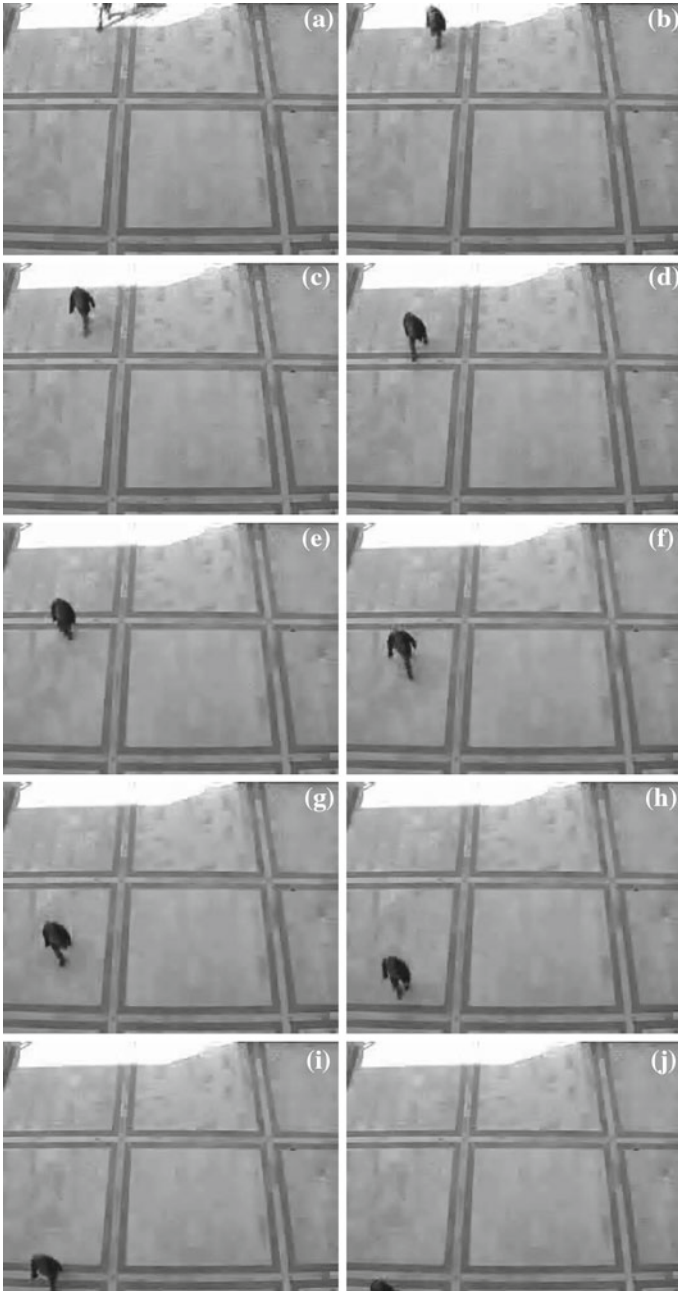


Fig. 4.4 Examples of outdoor dataset

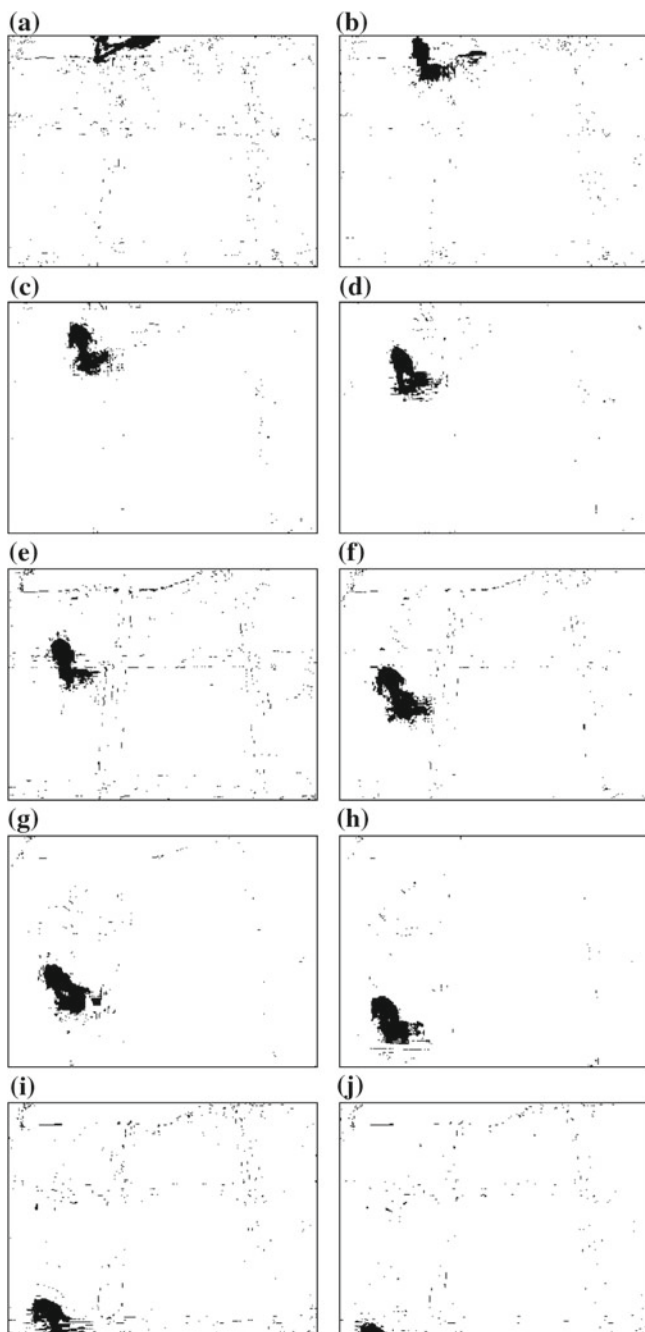


Fig. 4.5 Binary images after dynamic background subtraction and binarization

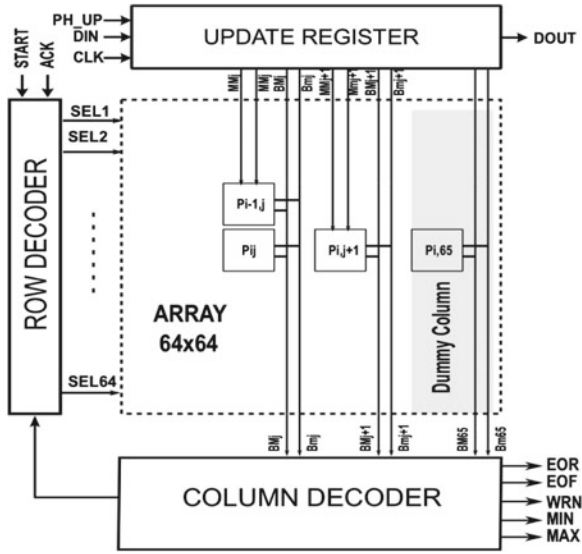


Fig. 4.6 Block diagram of the vision sensor architecture

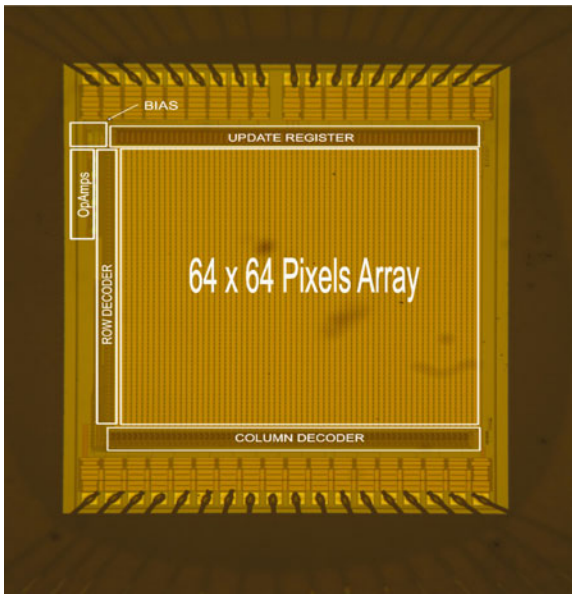
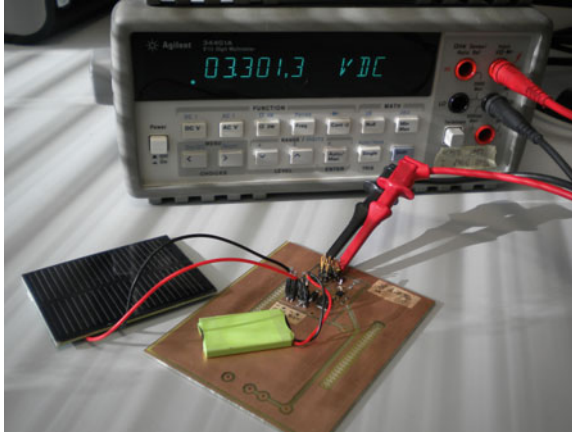


Fig. 4.7 Chip layout microphotograph



**Fig. 4.8** The harvesting board under test

two fixed output voltages at 3.3 and 1.8 V. To reach this specification we use two different DC–DC converters: the LTC3105 and the TPS61201 which operates with an input voltage range from 0.3 to 5 V. So even under worst case conditions of no photovoltaic energy, the harvester module can provide the required supply from the accumulator.

### ***4.3.1 Characterization of the Solar Cell***

Nowadays there are different types of outdoor and indoor solar cells, each of which has different dimensions and exhibits a different current-voltage curve. Moreover the curve and the maximum power point change with the incident radiation level. Under outdoor conditions, the direct solar radiation is variable within a range of 32–100 K lux, while indoors the radiation is typically much less (100–500 lux). This huge difference highlights the fact that a good outdoor solar cell has poor performance under indoor conditions and a PV cell tailored for indoor irradiance underperforms with direct sun light in comparison to outdoor device.

The table below reports a comparison between different indoor cells considering the possible number of cells that the available space permits. In this table the operative voltage and current data are referred to a 200 lux illumination. The best configuration is represented by AM1820 because it guarantees a maximum estimated power over a wide range of load conditions.

### ***4.3.2 Storage Device Analysis***

In this work we used an innovative type of battery, the *BatteryCloth*, provided by *FlexEl Inc.* [21] which is a flexible thin film battery. The thickness and the

Manufacture part number	AM-1819CA	AM-1820	AM-1454CA	AM-1801CA	ECS300
Voper (V)	3	3	1.5	3	3
Loperat (uA)	5.5–6.9 mA	13.3	21.0–31.0	12.0–18.5	4.5
Voc (V)	4.9	3	2.4	4.9	4
Isc (uA)	7.5 uA	13.3 uA	35	20	6.5
Pn (uW)	18	39.9	39	45	13.5
Size	31 × 24	43 × 26	41.6 × 26.3	53 × 25	35 × 12.8
Area (mm <sup>2</sup> )	744	1118	1094.08	1325	448
N° cells on PCB	4	4	4	3	8
Estimated current uA	24	53.2	104	45	36
Estimated power uW	72	159.6	156	135	108

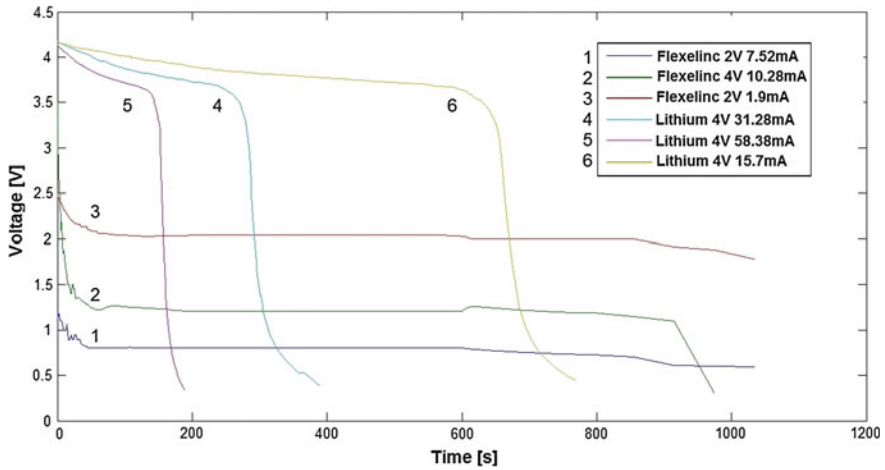


Fig. 4.9 Discharge curves of the batteries considered for the vision system

flexibility of the *BatteryCloth* are suitable for wearable devices, in fact the thickness can range between 0.2 and 1 mm, and the battery can be rolled to a radius of less than 0.15 mm. Moreover FlexEl's secondary cells are particularly useful in energy harvesting because it recharges at lower voltages compared with other technologies. The flexibility is not the only advantage in using these type of batteries. Indeed, the shape of the battery is an important characteristic which can be defined at design time allowing to fit any available space on board. In our case, the possibility to design the battery of the same size of the board and to insert it between two stacks without wasting space on board is a precious feature. Since size and capacity are design parameters, to assess the performance of the *BatteryCloth*, we compared the discharge curves with a comparable lithium battery. Results are plotted in the Fig. 4.9. The lithium batteries support higher discharge currents but exhibit lower capacity.

## 4.4 Conclusion

An integrated self sustainable vision sensor system has been presented. The solar harvester used to supply the node leads to several benefits such as the possibility to extend the autonomy when battery operated. Future work aims at integrating smart video analysis with an autonomous power supply and adaptive configuration, smart scheduling of the tasks onboard [22], and compression techniques [23] on the imagers.

**Acknowledgments** This work was supported by the Autonomous Province of Trento within *EnerViS—Energy Autonomous Low Power Vision System* project.

## References

1. Magno, M., Tombari, F., Brunelli, D., Di Stefano, L., Benini, L.: Multimodal video analysis on self-powered resource-limited wireless smart camera. *IEEE J. Emerg. Sel. Top. Circ. Syst.* **3**(2), 223–235 (2013)
2. Rossi, M., Brunelli, D.: Ultra low power wireless gas sensor network for environmental monitoring applications. In: 2012 IEEE Workshop on Environmental Energy and Structural Monitoring Systems (EESMS), pp. 75–81 (2012)
3. Rossi, M., Brunelli, D.: Analyzing the transient response of *mox* gas sensors to improve the lifetime of distributed sensing systems. In: 2013 5th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), pp. 211–216 (2013)
4. Jelcic, V., Magno, M., Brunelli, D., Paci, G., Benini, L.: A context-adaptive multimodal wireless sensor network for energy-efficient gas monitoring. *IEEE Sens. J.* **13**(1), 328–338 (2013)
5. Somov, A., Baranov, A., Spirjakin, D., Spirjakin, A., Sleptsov, V., Passerone, R.: Deployment and evaluation of a wireless sensor network for methane leak detection. *Sens. Actuators A Phys.* **202**, 217–225 (2013)
6. Somov, A., Baranov, A., Savkin, A., Spirjakin, D., Spirjakin, A., Passerone, R.: Development of wireless sensor network for combustible gas monitoring. *Sens. Actuators A Phys.* **171**(2), 398–405 (2011)
7. Somov, A., Spirjakin, D., Ivanov, M., Khromushin, I., Passerone, R., Baranov, A., Savkin, A.: Combustible gases and early fire detection: an autonomous system for wireless sensor networks. In: Proceedings of the First International Conference on Energy-Efficient Computing and Networking, Passau, Germany, 13–15 Apr 2010
8. Cottini, N., Gottardi, M., Massari, N., Passerone, R., Smilansky, Z.: A 33uW 42 GOPS/W 64 × 64 pixel vision sensor with dynamic background subtraction for scene interpretation. In: Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design, ISLPED '12, pp. 315–320. ACM, New York, NY, USA (2012)
9. Cottini, N., Gottardi, M., Massari, N., Passerone, R.: A bio-inspired APS for selective visual attention. *IEEE Sens. J.* **13**(9), 3341–3342 (2013)
10. Cottini, N., Gottardi, M., Massari, N., Passerone, R., Smilansky, Z.: A 33μW 64×64 pixel vision sensor embedding robust dynamic background subtraction for event detection and scene interpretation. *IEEE J. Solid-State Circuits* **48**(3), 850–863 (2013)
11. Broggi, A., Conte, G., Gregoretti, F., Passerone, R., Reyneri, L.M., Sansoé, C.: Design and implementation of the PAPRICA parallel architecture. *J. VLSI Sig. Process. Syst. Sig. Image Video Technol.* **19**(1), 5–18 (1998)
12. Komuro, T., Ishii, I., Ishikawa, M., Yoshida, A.: A digital vision chip specialized for high-speed target tracking. *IEEE Trans. Electron Devices* **50**(1), 191–199 (2003)

13. Komuro, T., Kagami, S., Ishikawa, M.: A dynamically reconfigurable SIMD processor for a vision chip. *IEEE J. Solid-State Circ.* **39**(1), 265–268 (2004)
14. Dondi, D., Bertacchini, A., Larcher, L., Pavan, P., Brunelli, D., Benini, L.: A solar energy harvesting circuit for low power applications. In: ICSET 2008, IEEE International Conference on Sustainable Energy Technologies, pp. 945–949 (2008)
15. Olivo, J., Brunelli, D., Benini, L.: A kinetic energy harvester with fast start-up for wearable body-monitoring sensors. In: 2010 4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 1–7 (2010)
16. Rizzon, L., Rossi, M., Passerone, R., Brunelli, D.: Wireless sensor networks for environmental monitoring powered by microprocessor heat dissipation. In: 1st International Workshop on Energy Neutral Sensing Systems, ENSSys13, p. 6. ACM, The Association for Computing Machinery, 2 Penn Plaza, Suite 701 New York, New York, 10121–0701, Nov 2013
17. Weimer, M.A., Paing, T.S., Zane, R.A.: Remote area wind energy harvesting for low-power autonomous sensors. In: Proceedings of 37th IEEE power, electronics 1–5, Jun 18–22 2006
18. Porcarelli, D., Brunelli, D., Magno, M., Benini, L.: A multi-harvester architecture with hybrid storage devices and smart capabilities for low power systems. In: International symposium on power electronics, electrical drives, automation and motion (SPEEDAM) 946–951, 2012
19. D. Carli, D. Brunelli, D. Bertozzi and L. Benini. A high-efficiency wind-flow energy harvester using micro turbine. In Power Electronics Electrical Drives Automation and Motion (SPEEDAM), 2010 International Symposium, pages 778–783, Jun 2010
20. D. Porcarelli, D. Balsamo, D. Brunelli, and G. Paci. Perpetual and low-cost power meter for monitoring residential and industrial appliances. In Design, Automation Test in Europe Conference Exhibition (DATE), 2013, pages 1155–1160, 2013
21. FlexEl's BatteryCloth website. <http://www.flexelinc.com>
22. Moser, C., Brunelli, D., Thiele, L., Benini, L.: Real-time scheduling with regenerative energy. *Real-Time Systems*, 2006. 18th Euromicro Conference on, ECRTS '06, pp. 261–270. DC, USA, Washington (2006)
23. Caione, C., Brunelli, D., Benini, L.: Distributed compressive sampling for lifetime optimization in dense wireless sensor networks. *Industrial Informatics, IEEE Transactions on* **8**(1), 30–40 (2012)

# Chapter 5

## A New Space Digital Signal Processor Design

**Massimiliano Donati, Sergio Saponara, Luca Fanucci, Walter Errico, Annamaria Colonna, Giuseppe Piscopiello, Giovanni Tuccio, Franco Bigongiari, Maximilian Odendahl, Rainer Leupers, Antonio Spada, Vincenzo Pii, Elena Cordiviola, Francesco Nuzzolo and Frederic Reiter**

**Abstract** The increasing demand of on-board real-time processing represents one of the critical issues in forthcoming scientific and commercial European space missions. Faster and faster signal and image processing algorithms are required to accomplish planetary observation, surveillance, Synthetic Aperture Radar imaging and telecommunications, especially due to the importance of elaborate the sensing data before sending them to the Earth, in order to exploit effectively the bandwidth to the ground station. The only available space-qualified Digital Signal Processor (DSP) free of International Traffic in Arms Regulations restrictions (ATMEL TSC21020) faces a poor performance of 60 MFLOPs peak, and it is becoming inadequate to fulfill the computation demand of the space missions. For this reason, the development

---

M. Donati (✉) · S. Saponara · L. Fanucci  
Department of Information Engineering, University of Pisa, via Caruso 16,  
56120 Pisa, Italy  
e-mail: massimiliano.donati@for.unipi.it

S. Saponara · L. Fanucci  
Consorzio Pisa Ricerche SCARL, Corso Italia 116, 56120 Pisa, Italy  
e-mail: sergio.saponara@iet.unipi.it

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

W. Errico · A. Colonna · G. Piscopiello · G. Tuccio · F. Bigongiari  
SITAEI spa, via Livornese 1019, 56122 Pisa, Italy  
e-mail: walter.errico@sitael.com

A. Colonna  
e-mail: annamaria.colonna@sitael.com

G. Piscopiello  
e-mail: giuseppe.piscopiello@sitael.com

G. Tuccio  
e-mail: giovanni.tuccio@sitael.com

F. Bigongiari  
e-mail: franco.bigongiari@sitael.com



of a new generation of space-qualified DSP is well known in the European space community. The space-qualified DSP architecture proposed in this work fills the gap between the computational requirements and the available devices. Additionally, it has been implemented using technologies available in Europe without any restriction. The DSP processor leverages a pipelined and massively parallel core based on the Very Long Instruction Word paradigm, with 64 registers and 8 operational units. The rest of the System-on-Chip architecture consists in the instruction and the data cache memories, the memory controllers and two SpaceWire interfaces. The processor, implemented in CMOS 65 nm technology, reaches an operational frequency of 120 MHz and area occupation of around 350 Kgates. The correlated Software Development Environment (SDE), with compiler, assembler, linker, debugger and instruction-level simulator, allows for an easy programming of the device in C language.

## 5.1 Introduction

In recent years, the data rates and the data volumes produced by both scientific and commercial space missions and the space-oriented applications have dramatically grown as consequence of some technological advancements, especially in the fields of sensing devices and spacecrafts fabrication. To accomplish planetary observation, surveillance, Synthetic Aperture Radar (SAR) imaging and telecommunication, it is becoming mandatory to have a high on-board computational power to execute signal and image processing algorithms such as Fast Fourier Transform (FFT), Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) filters, data compression, images compression, matrix calculations, sample decimation and cryptography.

---

M. Odendahl · R. Leupers

Institute for Communication Technologies and Embedded Systems,  
RWTH Aachen University, Sommerfeldstraße 24, 52074 Aachen, Germany  
e-mail: maximilian.odendahl@iss.rwth-aachen.de

R. Leupers

e-mail: leupers@iss.rwth-aachen.de

A. Spada · V. Pii · E. Cordiviola

INTECS spa, Polo Attività Montacchiello via Forti Trav. A5, 56121 Pisa, Italy  
e-mail: antonio.spada@intecs.it

E. Cordiviola

e-mail: elena.cordiviola@intecs.it

F. Nuzzolo · F. Reiter

Space Applications Services NV, Leuvensesteenweg 325, 1932 Zaventem, Belgium  
e-mail: francesco.nuzzolo@spaceapplications.com

F. Reiter

e-mail: frederic.reiter@spaceapplications.com

As stressed by the European Space Agency (ESA), there is an urgent need for more performing data processing platforms [1] to elaborate on-board the sensing data and to refine gathered information before sending them to the ground station. Reducing the amount of data transferred to the Earth is a crucial task to exploit the available downlink bandwidth effectively, that in such systems represents the bottleneck. For example in naval surveillance using SAR, 2D image transformations allow to extract from the sensing data the position and heading of the ships in the observed area providing data that are many order of magnitude smaller than the complete image.

Unfortunately the space-qualified Digital Signal Processor (DSP) available in Europe without incurring in ITAR restrictions, i.e. the rad-hard TSC21020F by ATMEL [2] (40 MFLOPS performance, 60 MFLOPS peak), are becoming obsolete under these conditions. For example, the ExoMars vehicle [3] and the planetary observation missions EUCLID [4] and PLATO [5] require 100 MFLOPS while the Meteosat Third Generation Infrared Sounder [6] expects up to 10 GFLOPS. Indeed the US-made alternative products are ITAR-restricted and do not meet the need to reduce the dependence for critical technologies outside the Europe.

The need of a novel general-purpose high performance DSP entirely implemented with European space technologies has been underlined by ESA since the ADCSS workshop in October 2007 [7]. The main set of requirements envisaged for the new generation of space DSP are: processing power  $\geq 1000$  MIPS, Radiation Hardness (TID  $\geq 100$  Krad), EDAC-protected memories, support for space standard I/O interfaces (i.e. SpaceWire) along with a stable and high quality C language toolchain.

The goal of the DSPACE European research project [8] was the development of an innovative and radiation-hardened by design DSP architecture. The developed System-on-Chip (SoC) meets the requirements established by ESA, in order to fill the gap between the actual processing demand and the available devices. The processing core in the SoC reaches around 1 GOPS peak performance operating at the clock frequency of 120 MHz, while the rest of the building blocks are the instruction and data caches, the DMA and memory controllers, the SpaceWire interfaces and the on-chip interconnection busses. Moreover, the GCC-based Software Development Environment (SDE) enables the software architects to easily implements programs for this architecture and to program the device. Whereas the final target technology is indubitably a European space-qualified CMOS ASIC process, one of the key objectives of the project was to emulate the system into a commercial FPGA to validate its performance, and make this prototype available to the space community. Hereafter the paper is organized as follows. Section 5.2 gives an overview of the complete DSP SoC architecture. Section 5.3 details processing core and Sect. 5.4 describes the programming environment. Section 5.5 reports the main achievements. Finally, Sect. 5.6 concludes the paper.

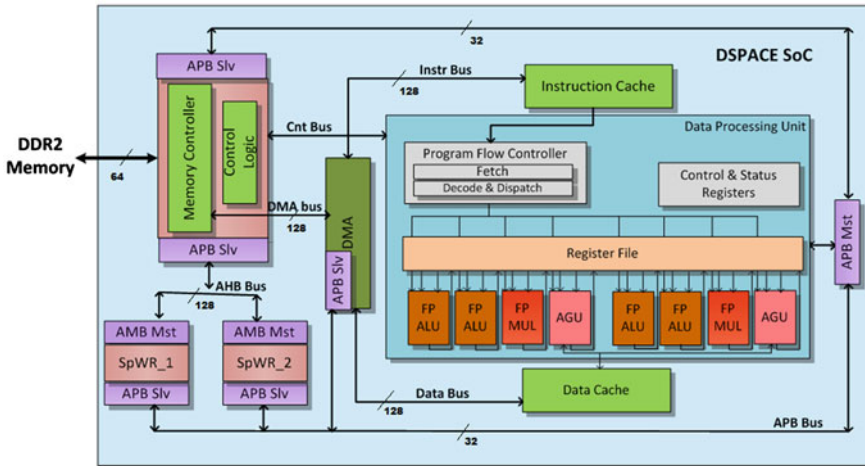


Fig. 5.1 DSPACE system-on-chip schematic diagram

## 5.2 System-on-Chip Architecture

The DSPACE System-on-Chip (SoC) architecture is shown in Fig. 5.1. It consists of a minimal and complete set of building blocks arranged in order to supply the Data Processing Unit (DPU) with the program instructions and the data to be elaborated, allowing also the communication with the hosting environment. In this design, the DPU represents the computational element of the chip. Moreover, there are some on-chip peripherals able to provide I/O capabilities for operative and control purposes. AMBA busses are involved within the chip to connect the building elements.

The system works with cache memories for both instructions and data, aiming to improve its performance. The D-Cache is a direct-mapped memory (64 KB, 32 byte of block size) with copy-back strategy for the cache miss condition. The size allows accommodating datasets typical of image/video processing algorithms. It provides to the AGU units two read ports and two write ports working in parallel for Read-Read, Read-Write, Write-Read, Write-Write operations. The actual size of the I-Cache is 32 KB and the block size is 256 byte. It is implemented in direct-mapped fashion and contains up to 8 kilo-instructions of the program.

All the communications with the external main memory (i.e. in case of cache miss) are driven by the multi-channels Direct Memory Access (DMA) controller through the generic memory controller able to deal with Double Data Rate (DDR2) RAM memories. Moving the control of this datapath outside the caches gives the programmer the possibility to manage directly the pre-fetch/save according to a cache-less paradigm.

The two SpaceWire interfaces, allow for the information exchange between the DSPACE chip and the hosting system (i.e. read the status of a sensor, actuate a control or output a result).

While the implementation of the DPU followed the LISA flow, the rest of the SoC building blocks were developed following the traditional hardware design flow (i.e. VHDL).

### 5.3 Digital Signal Processing Core

The processing core represents the computational element of the DSPACE SoC architecture. It was designed starting from a deep survey of the ground DSP, because of these devices generally have interesting performance compared to the emerging space requirements. To achieve a significant performance increase versus the previous space DSP generation, without relying on high clock frequency due to space technology limitations, the core is based on a pipelined and massively parallel architecture. Triple Modular Redundancy (TMR) with voting logic to protect internal registers and EDAC protection for memories to correct single error and to detect double error events (SEC/DED) allow maintaining the functionality in radiation environments.

The Digital Processing Unit (DPU) is based on a Very Long Instruction Word (VLIW) architecture featuring 8 computational units organized in parallel. It allows performing of up to 8 RISC instructions at every clock cycle. Assuming an operational frequency of 120 MHz, it ensures near 1000 MIPS peak performance as expected. Figure 5.1 shows a block diagram of the structure of the DPU collocated in the SoC environment. The role of this component is to fetch, decode and execute the program instructions coming from the Instruction cache (I-Cache) elaborating the data provided by the Data cache (D-Cache) or by the SpaceWire ports.

The DPU datapath (see Fig. 5.2) consists of:

- a register file with 64 general purpose 32-bit registers
- four instances of a arithmetical-logical unit (FP\_ALU) capable of all the logic, arithmetic, with exclusion of multiply, conversion and I/O operations
- two multiplier units (FP\_MUL) dedicated to multiply operations
- two address generation units (AGU) for accessing the memory
- two 64-bit ports (LD) to load data from the memory to the register file
- two 32-bit ports (ST) to store data contained in the register file to the memory.

The eight functional units accept immediate operands or values contained in the register file. Supported formats are the 32-bit 2's Complement Fixed-Point and the IEEE 754 single precision Floating-Point.

The instruction set offers general-purpose instructions, to be compliant with the wide range of applications typical of space missions, but at the same time, it provides a lot of specific and optimized instructions to speed up the recurrent operations in signal and image/video processing. FP\_ALU, FP\_MUL and AGU have assigned 92, 24 and 34 instructions respectively. All the instructions can be executed conditionally depending on the content of some registers for a reduction of the pipeline hazard.

The 7-stages pipelined architecture ensures a single-cycle throughput. The Fetch phases contain 3 stages of the pipeline. It is in charge of retrieve instructions from the

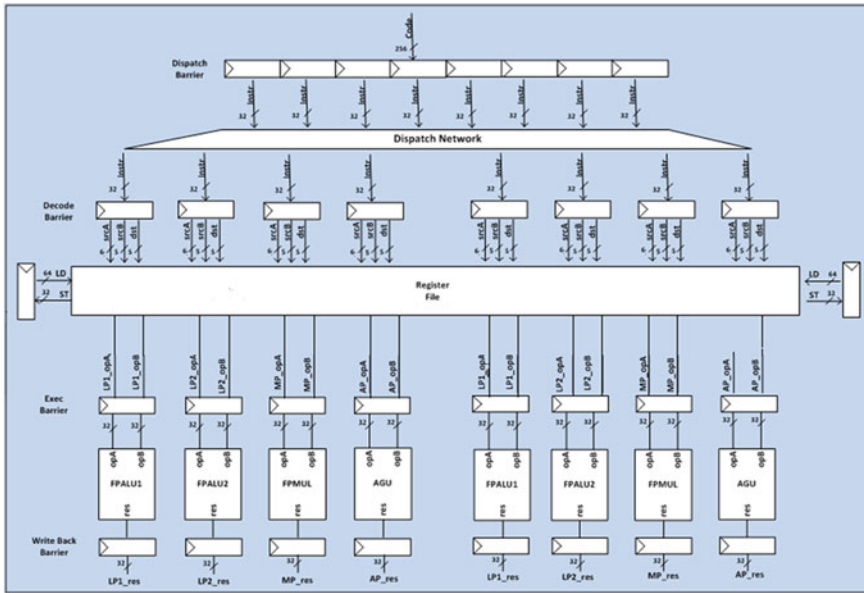


Fig. 5.2 Internal structure of the DSPACE data processing unit

I-cache according to the program flow. The following 2 stages of the pipeline (Decode phase) provide the source operands decoding and the dispatching of the instructions to the right execution unit. During the Execution stage the units perform the assigned instructions and then provide the results of the computation in the register file in the final write-back stage. The pipeline is stalled in case of cache misses until the missing block has been copied from the external memory.

The DPU fetches instruction-packets (IP) composed by 8 instructions. The instructions within each IP that can be executed in parallel are grouped at compilation time exploiting a specific bit in the opcodes. In this way, each IP may generate from one to eight sets of instructions executed in parallel, called execution-packets (EP). Every cycle, if the previous IP has been completely dispatched a new one is addressed to enter in the pipeline, otherwise the fetching is stalled. The Decode phase automatically dispatches NOP to the idle computational units during the current EP, reducing the memory occupation of the program. The assignment among instructions and execution units is made explicit into opcode at compile time, thus dispatcher hardware is minimal while the compiler is rather complex.

The DPU presented in this work supports three kinds of interrupts, served with different levels of priority: reset, non-maskerable and 7 maskerable interrupts. The source of interrupts may be on-chip or off-chip. The I/O space of the DPU consists of 32 virtual registers that map the peripherals in the SoC. FP\_ALU can access them for reading or writing these registers causing respectively input or output operations involving the mapped peripheral. Finally, the DPU includes 20 control and status

registers grouped in a dedicated register file to control the status of the computation of arithmetical, logical and field operations (i.e. carry, saturation, overflow, etc.), the interrupts management (i.e. masking flag, address of handlers table, etc.) and the addressing mode and other DPU settings. FP\_ALU can access these registers.

The design of the DPU module went through a methodology of concurrent design of the hardware platform and the related software tools [9]. The DPU was modeled using LISA (Language for Instruction Set Architectures) and the associated Processor Designer by Synopsys, allowing obtaining both the synthesizable VHDL description of the processor and the software development tools (assembler, linker, simulator, debugger) from this high level model. To complete the SDE and to avoid the cost of the C compiler development, a large reuse of an already existing and stable SDE was adopted as described in Sect. 5.4.

## 5.4 Software Development Environment

Since the definition and design of the entire SDE (Software Development Environment) from scratch is a very costly task, the approach taken is based on the reuse of the C compiler component of an open-source GCC-based software collection for a ground DSP. All the rest of the development tools come from a LISA concurrent hardware/software design approach.

In particular, the final SDE architecture consists of the four consecutive steps depicted in Fig. 5.3:

- the well-known GCC-based compiler produces the assembly for the reference architecture of the reused toolchain. This code is optimized for the target platform and it needs to be optimized to leverage the parallel architecture of DSPACE processing core;
- the Glue Software module translates the GCC assembly code into DSPACE Linear Assembly. It removes all optimizations and registers allocation in order to be hardware-independent;
- the Code Optimizer optimizes the DSPACE linear assembly to exploit the parallelism of the DSPACE hardware. Registers allocation, instructions scheduling, etc. are performed in this level;
- the Assembler and Linker generated from the LISA model produce the final binary file to be loaded in the program memory.

In order to simplify the production and the deployment of the executable files for the DSPACE platform, all the previous transformation steps have been integrated into a single procedure called DSPACE C Compiler (DCC). Programmers can use DCC both with the command line interface or directly from the Eclipse programming tool installing the plug-in developed. In particular, in the latter case the users have to use the classical and familiar Eclipse software development also for the DSPACE platform.

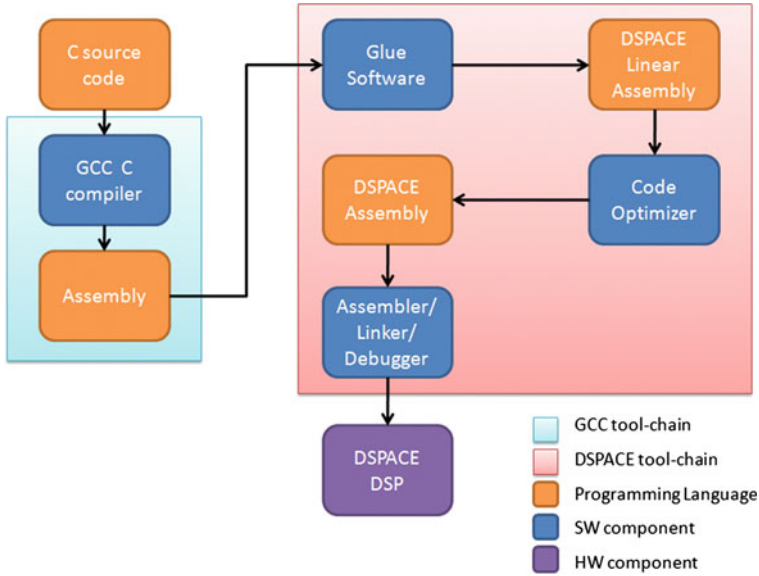


Fig. 5.3 DSPACE software development architecture

### 5.5 Results

In order to validate the performance of the system a demonstrator board based on the ground Xilinx Kintex7-XC7K325T FPGA was designed. The complete SoC occupies 30 % of the resource budget on this device, operating at 25 MHz of clock frequency. This means that such board represents the performance of the final system scaled down by a factor 5.

The board has Compact-PCI 3U size and includes a DDR2 SORDIMM connector with 1 Gbyte EDAC memory, two SpaceWire connectors, one USB interface, one header connector for RS232 communication links (for remote control and program upload/download) and a JTAG interface for the FPGA programming. Furthermore, the DSPACE Demo Board is provided with expansion connectors giving the possibility to attach a mezzanine board to improve its communication features. The mezzanine board contains a FPGA that acts as bridge between the DSPACE Demo Board and the some communication interfaces. Optionally these interfaces can be two MIL-STD-1553B buses, interfaced via Twinax BJ77 bulkhead connectors, two Controller Area Network (CAN) buses, interfaced via RJ-45 connectors and two I2C links, both interfaced via front-panel header connectors. The DSPACE Demo Board was designed to be housed in a Compact PCI crate within a computer or alternatively in a stand-alone desktop version. An example of the desktop version, with 2 RS232 expansion ports is shown in Fig. 5.4.

Preliminary synthesis using CMOS standard cell 65 nm technology show an area of around 350 Kgates and a peak performance of near 950 MOPS (720 MFLOPS) at



**Fig. 5.4** DSPACE Demo Board desktop version

120 MHz. The ATMEL ATC18RHA (standard cell 180 nm) is the target European space technology, but also the ST DSM65 65 nm technology will be considered for the final implementation. Moreover, the entire design is compatible with the implementation in the radiation-hardened Xilinx Virtex5-QV XQR5VFX130 FPGA device.

The level of performance reached by the DSPACE DSP has been assessed using the benchmarks established in 2008 by ESA [10]. The main kernel functions identified in the benchmark applications are: FIR filtering, FFT functions, and data compression algorithm (CCSDS lossless data compression). In addition, some auxiliary functions also have been developed (i.e. routines for data copy, format conversions, etc.). These functions have been developed either in C or in DSPACE assembly language, and executed both on the simulator and on the DSPACE Demo Board. All functions passed their tests when executed on the simulator and on the DSPACE Demo Board, and they are used to set up the benchmarking scenarios B1, B2 and B4 described in ESA [10]. FIR filter benchmark results confirmed the peak performance of 2 single-precision floating-point MACs per cycle.

## 5.6 Conclusions

This work presents an innovative DSP design for space applications. The high performance of the general-purpose cache-aided architecture and its wide instruction set allow to deal with a large variety of applications envisaged in the future European missions (i.e. images compression and elaboration, data fusion).

The project meets all the requirements underlined by ESA, resulting in a valid candidate for the next generation space DSP. It runs at 120 MHz on European space-qualified technology and provides a peak performance of around 1GOPS, outperforming the outdated TSC21020 device 17 times. With respect to the TSC21020F



processor, the DSPACE DSP has a much extended instruction set, including both general purpose and optimized instructions to cope with the wide range of space applications expected in the near future.

The DSPACE Demo Board has been thoroughly tested and the ESA benchmark have been used to assess the real performance of the DSPACE DSP.

Finally, the DSPACE System-on-Chip will be made available on “Design and Reuse” website ([www.design-reuse.com](http://www.design-reuse.com)). The package will include the architectural model (LISA Core and VHDL Code), the C-Compiler and binary tools, the instruction set simulator, the SDE and the set of benchmarks. On the website also the possibility to order the Demo Board for demonstrator-based refinement activities on DSPACE will be offered.

**Acknowledgments** The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under grant agreement n°262798.

## References

1. Trautner, R., Vitulli R.: Ongoing developments of future payload data processing platform at ESA (2010)
2. ATMEL TSC21020F, available: <http://www.atmel.com/devices/tsc21020f.aspx>
3. Hult, T., et al.: The ExoMars rover vehicle OBC. In: Proceedings of DASIA Conference (2010)
4. EUCLID M-class mission. Available at <http://sci.esa.int/euclid>
5. PLATO M-class mission. Available at <http://sci.esa.int/plato>
6. Meteosat Third Generation Infrared Sounder Mission. Available at <http://www.eumetsat.int>
7. European Space Next Generation Processor for On Board Payload Data Processing Application. ESA Round Table Synthesis (2007)
8. Errico, W., Cordiviola, E., Fanucci, L., Saponara, S., Donati, M., Nuzzolo, F., Leupers, R., Odendahl, M., et al.: DSPACE: a new space DSP development. In: Proceedings of DASIA Conference **701**, 1–6 (2012)
9. Odendahl, M., Yakoushkin, S., Leupers, R., Errico, W., Donati, M., Fanucci, L.: A next generation digital signal processor for European space missions. In: Proceedings of ESTEL 2012 Conference (2012)
10. ESA Next Generation Space Digital Signal Processor Software Benchmark. TECEDP / 2008.18/RT (2008)

# Chapter 6

## Spatial Sound Rendering for Assisted Living on an Embedded Platform

Luca Rizzon and Roberto Passerone

**Abstract** 3-D sound can be used to synthesize audio stimuli able to describe spatial information. This can be used as a sensorial substitution of sight for the visually impaired to help them in the task of autonomous orientation and mobility. However, commonly used techniques are computational demanding, therefore not optimal for being implemented in embedded systems. Moreover, the sound localization is specific to each individual and complex to measure or customize. We chose to develop a bottom-up physical model to synthesize a simplified transfer function and playback audio signals over headphones. The model permits the computational requirements to be reduced at the cost of lower accuracy of representation. Still the proposed system can meet the goal of describing spatial information to the listener. Moreover, it can be a promising solution for on-the-go individualization. In this paper we describe the algorithm, the implementation on an embedded platform and present the comparison between HRIR-based synthesis and the proposed simplified physical approach.

### 6.1 Introduction

Humans use five senses to analyze the world around them, make decision based on the stimuli they collect and react accordingly. In particular two of those senses are involved in the task of orientation, localization and movement: sight and hearing. Normal people commonly use at most the sight to orientate when walking, even if this means capturing information lying only inside the field of view. Sound information is only partially used for orientation, for example when a sound representing a dangerous situation is perceived as coming from out of sight. In this case, audio

---

L. Rizzon (✉) · R. Passerone  
University of Trento, via Sommarive 5, 38123 Povo, TN, Italy  
e-mail: luca.rizzon@unitn.it

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

information becomes an interesting cue for orientation. We can say that sight and hearing work in strong conjunction helping each other for the safety of individuals. At the same time, visually impaired people are more trained to use audio signals to orientate themselves. Unfortunately, for those people who suffer from visually impairments, hearing is not sufficient to guarantee autonomous orientation since it cannot efficiently represent obstacles.

A navigational aid is able to tell the user the action and direction to take to reach a certain target. Some aids interface with the user in the same manner as the navigation system for automotive: by telling instructions via speech commands [1]. Other systems communicate to the user using different media simultaneously and giving perceptual information rather than instructions [2]. This solution is particularly interesting because it exploits a key feature of the human sensing system: the ability of the brain to integrate information from different senses and compensate one another [3]. Moreover, using this approach stimulates the user at a higher level of cognition, by exciting senses and independent thinking [4].

In this work we develop a method for representing spatial information to a visually impaired users via sound signals over headphones, in a way such that it would be possible for those people to autonomously orientate in an unknown environment. In our application, spatial information is described to the user exploiting positional audio, stimulating higher level of cognition with respect to standard descriptive audio aids. Sound signals have to be processed so that they give the illusion to the listener of coming from a particular point in space even if they are played-back through standard headphones. To this end, the physical and perceptual mechanisms that are involved in sound localization have to be replicated using DSP techniques to create a synthesized virtual soundscape. We refer to this technology as 3-D Audio [5].

Our work fits in the context of the DALi Project (Devices for Assisted Living) [6]. The scope of DALi is the development of a smart walker for the elderly. The walker complements the standard auxiliary navigational aid for the physical impaired and can be used by people who suffer from frail cognitive disabilities or visual impairments. The walker will be equipped with electronics able to scan the surrounding environment in order to create a map, record and elaborate the attitude of the user and present to the assisted person information useful for autonomous navigation using audio as a sensory substitution in order to overcome the user's cognitive impairments.

## 6.2 Related Work

The ability of humans to recognize the point in space where a sound signal originates comes from the displacement and shape of the ears. The geometry of the head and pinnae introduces some modification in the audio signals impinging the ear. These spectral changes depend on the path the sound waves traverse when moving from the sound source to the listener eardrum. The human brain can detect these phenomena, correlate the signals at the two ears and recognize with a given accuracy the position of the auditory event.

There exist many algorithms able to reproduce these phenomena over headphones. The most frequently used method to render 3-D sound is based on the Head Related Transfer Function (HRTF) or their Fourier transform (HRIR) [7]. HRTFs represent the response of the human ears and body for a given direction of the incoming sound and individual anthropometry. Since those responses may vary from person to person, there is the need to measure them. However, measuring HRTF is a time-demanding method and requires sophisticated equipment. Fortunately, some HRTFs databases are freely available for research activity [8]. This kind of implementation requires large memory to store the filter coefficients and the resulting filtering process is computationally demanding. Another disadvantage of HRTF-based sound rendering is that it lacks the sensation of distance and movement; therefore these effects have to be separately implemented by means of proper algorithms (e.g., Doppler effect, reverberation) [9].

This translates into having a large and complex amount of computation to be performed. For the purpose of mobility aids it is not required to achieve high accuracy of representation, it should be sufficient to evocate the sensation of spaciousness and displacement to the user. Our work is motivated by the intent of lowering computational needs, in order to permit the implementation in an embedded system. The goal is to create a model with a reasonable good spatial resolution, a low complexity and a limited number of parameters that can be used to customize the model to the listener needs. Moreover, we would like to overcome the need of measurements of the ears response.

Our proposal is to use simplified synthesized HRTFs instead of measured HRTFs. Synthesis is done by using simplified physical models, the obtained response is then translated into a set of filters. Each set of filters represents some physical phenomena the sound waves encounter while traveling in space and interacting with the listener body. Our analysis is focused on sound sources placed on the horizontal plane without considering the elevation angle. For this reason we introduce in the model only basic binaural quantities.

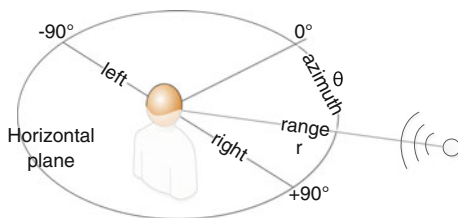
### 6.2.1 Binaural Cues for Sound Localization

One of the major cues for sound localization is the Interaural Time Difference (ITD) that is the transient time difference between the two ears due to the differential distance between ears and the sound source. A simple model for capturing this effect is the one proposed by Brown and Duda [10]:

$$\Delta T = \frac{a}{c}(\sin \theta + \theta), \quad (6.1)$$

where  $a$  represents the head radius,  $c$  the speed of sound in normal condition (343 m/s) and  $\theta$  is the azimuthal angle. We adopt the vertical-polar coordinate system, so  $\theta$  assumes value 0 in front of the listener, negative values on the left hand side and positive value on the right (Fig. 6.1).

**Fig. 6.1** The coordinate system



For an average value of the head radius we have a maximum value for  $\Delta T$  of 660 ms corresponding to a delay of about 32 samples at 44.1 kHz. This time difference is the primary cue for localization of sounds below 1.5 kHz.

Another cue for the localization of sounds lying on the horizontal plane is the Interaural Level Difference (ILD) or Intensity Difference. In fact, sounds impinging the ear coming from different paths are attenuated according to the different length of the path and due to the presence of the head of the listener. The effect of the path is easily inserted in the model considering the attenuation in amplitude is equivalent to the inverse of the path length. The shadowing effect introduced by the head is frequency dependent. However, as pointed out in the literature [10], this effect can be modeled as a single pole/single zero filter:

$$H(\omega, \theta) = \frac{1 + j \frac{\alpha \omega}{2\omega_0}}{1 + j \frac{\omega}{2\omega_0}}. \quad (6.2)$$

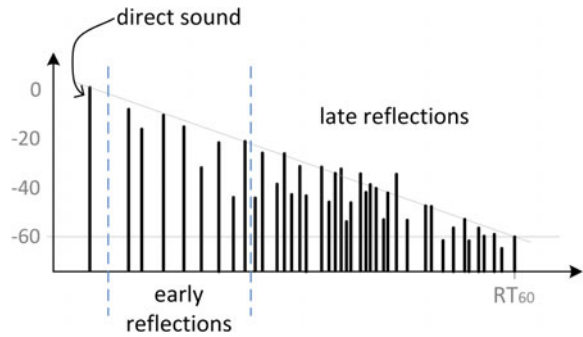
The coefficient  $\alpha$ , which depends on the angle of incidence  $\theta$ , controls the location of the zero. With this approximation the head shadowing effect is modeled as a tunable low-pass filter for the contralateral (far) ear. Using MATLAB we estimated a FIR filter with 35-taps is required to get a good sound quality.

For sounds with an elevation angle different from zero, many points in space will cause an identical value for both ITD and ILD composing the so-called cone of confusion. Humans can distinguish the elevation of a sound source thanks to other cues mainly coming from the geometry of the external ear. The presence of concha and helix introduces small echoes on the impinging sound signals whose resonant frequency depends upon the elevation angle. At the moment we do not include in our application the sensation of elevation.

## 6.2.2 Rendering the Sensation of Distance

In order to give the listener the illusion of range and spaciousness, synthesized sounds are shaped accordingly to the distance. From a physical viewpoint, in a simplified approximation, the change in range corresponds to a variation in intensity level of the sound, which is the primary cue for distance judgment. Precisely, sound intensity falls

**Fig. 6.2** Reverberation decay in time

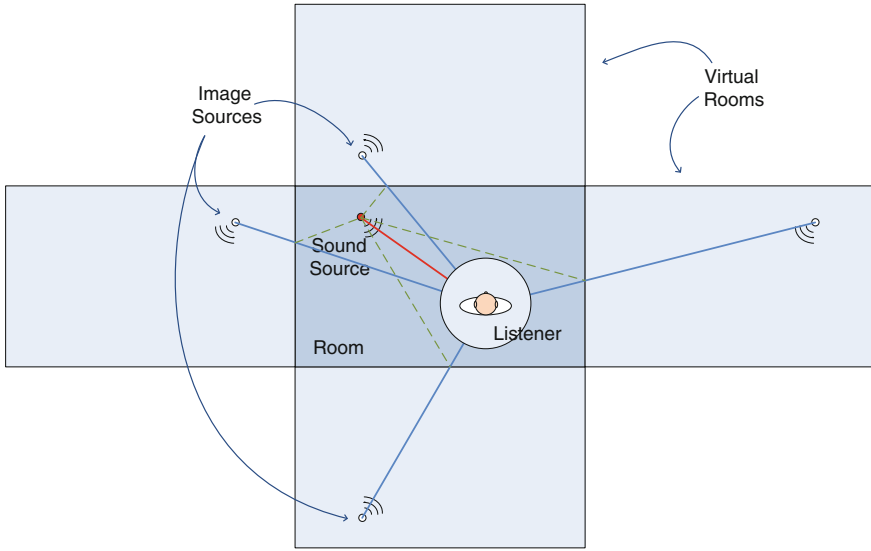


inversely proportional to the square of the distance ( $1/r^2$ ) from the sound source, so that in our algorithm the amplitude of each sound is reduced by half when doubling the distance. However, in a real world scenario, humans take advantage of more than one cue for judging the distance of the auditory event, also because the intensity of sound cues depends on the familiarity of listeners with the sounds. For example, the reverberation introduced by reflective surfaces inside a room is integrated by listeners for estimating the spaciousness of the room and also the distance from the sound source.

Generally, we refer to the reverberation as the succession in time of attenuated replicas of the direct sound, as in Fig. 6.2. The first vertical line represents the direct sound that reaches the ear. It is followed by first-order reflections (early or dry), which are reflections of that direct sound that have bounced off the walls once. Higher order reflections (wet or late) have bounced off more than one surface before impinging the ears.

Reverberation has the effect of decorrelating the signal at the two ears. As a consequence, it increases the sensation of immersion and spaciousness and reduces the “inside the head” phenomena, that causes the sensation of perceiving the sound as coming from inside the listener’s head instead of coming from the surrounding space. Listeners estimates distance during the attack portion of sounds, where ITD is most effective. So, reverberation that falls in a small time gap are associated to the same sound event and interpreted as a single phenomenon. On the contrary, late reverberation may be distinguished and the listener’s may wrongly interprets them as different auditory events, as what happens with echoes. In particular, first order reflection are perceived from humans as a indication of the distance of the sound source. Higher order reflections not only are useless for the location task but they may corrupt the generated signal causing a decrement of the informative audio content. For example, the reverberation in a cathedral has a large wet contribution, that causes the sensation of being surrounded by sound sources and makes the localization task almost impossible because the observer receives acoustic sound waves from all directions, interfering each other and confusing him.

So, a good degree of spaciousness can be implemented by means of a reverberant algorithm. There exist many reverberation techniques [11]. Many of them are



**Fig. 6.3** Graphical representation of the image source method

very complex and are able to produce at the output an effect very similar to the reverberation of the real room. Unfortunately, this translates into a time demanding and memory hungry implementation. Moreover, these reverberation algorithms are parameterized on room geometry, size and materials but do not take into account the position of sound source and observer inside the room.

Our choice for reducing the computational complexity of the system and in the meantime taking care of relative positions of sound elements in the room is to adopt the so-called Image Source Method (ISM). In the ISM each wall is equivalent to a reflective surface. A sound wave reflecting off a wall is equivalent to having a virtual source on the mirror image of the original, behind the wall [12].

In our work a simple ISM technique is implemented, considering only four contributions inside a virtual rectangular room whose dimensions are proportional to the distance of the farther objects to be rendered. Those four contributions are first order reflections. In the virtual room, walls are the borders of our region of interest. These contributions are spatialized according to the displacement, delayed consistently with the distance and summed up with the direct path sound ray (Fig. 6.3). The obtained signal sounds more pleasant and the sensation of spaciousness of the surrounding scene is not detrimental for the task of localization.

Denoting the spatial coordinates of the listener in 2D space as  $(x_l, y_l)$  and the sound source coordinates as  $(x_s, y_s)$  we can compute the reverberation parameters for the ISM in a rectangular room of dimension  $(x_r, y_r)$  as follows:

$$(x_1, y_1) = (2 \cdot x_r - x_s - x_l, y_s - y_l)$$

$$(x_2, y_2) = (x_s - x_l, 2 \cdot y_r - y_s - y_l)$$

$$(x_3, y_3) = (-x_s - x_l, y_s - y_l)$$

$$(x_4, y_4) = (x_s - x_l, -y_s - y_l)$$

In formulas, index denotes the virtual source starting from right proceeding in counter-clock wise direction. From coordinates, the algorithm compute the azimuthal angle  $\theta_i$ , the delay and attenuation for each of the four image contributions ( $i = 1, \dots, 4$ ).

$$\theta_i = a \tan(x_i/y_i).$$

Formally, negative values of  $y$  imply the sound source is behind the listener, therefore  $\phi = 180^\circ$ , otherwise the value for the elevation angle is always zero. The distance is computed as

$$d_i = \sqrt{x_i^2 + y_i^2}.$$

The distance influences the attenuation, and the delay computed as

$$\delta t_i = d/c$$

where  $c$  represents the sound speed (343 m/s).

### 6.3 Implementation

The idea proposed in this paper is to use synthesized spatial sound to orientate visually impaired user in an unknown environment. To this end we associate a sound to each interesting event and spatialize it according to the point in space in which the event is placed. Our intention is to associate to a point in space a virtual sound source. The point we want to render can be the safety path to traverse for the user (i.e., the line to follow in order to avoid obstacles) or, alternatively, a sound can be associated to an obstacle to avoid [13].

In this paper we focus on the synthesis of audio signals. However, a complete navigational aid is composed by acquisition sensors, actuators, and multiple user interfaces. All components are coordinated by the control unit which shares its limited hardware resources and resolves conflicting needs in terms of computational needs, memory usage and interoperability. Still, the system has to react efficiently to multiple inputs, controls multiple actuators, and must prove predictability and robustness features in order to guarantee the safety of the individuals. When integrating the proposed auditive interface into a complete navigational system, designers have to adopt a proper design methodologies [14] and tools [15] specifically tailored for the design of Cyber-Physical Systems. Given the binding of our system with the timing



features of the perceptual mechanism of the hearing sense of humans, at this stage of development, we focus on the timing requirements of the audio algorithm. The technical challenges a designer can hit while working at the complete system design can be managed by adopting a platform based design approach, in order to satisfy specific timing requirements of a complex systems made of concurrent components, and exploring different architectural implementation.

We have evaluated different approaches using a model written in MATLAB:

- HRTF-based 100-tap filtering, with Doppler Effect and reverberation or ISM;
- Inter-Positional Transfer Function (resulting from HRTF interpolation, used to avoid the needs for synthesizing Doppler);
- Proposed algorithm, with 35 taps and ISM.

The general pseudocode of the implementation is:

```

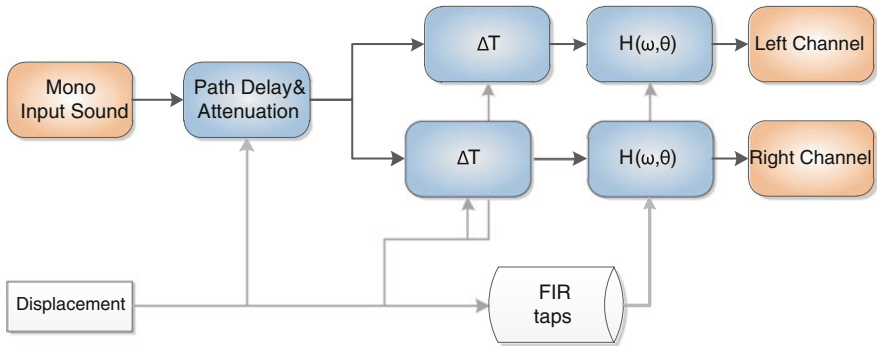
initialize sound card;
get FIR coefficient set;
open input sound;
while 1 do
    get spatial coordinates;
    compute coordinates for ISM;
    compute ITD;
    get FIR coefficients;
    spatialize each voice;
    compute delay between voices;
    mix all voices;
    playback;
end

```

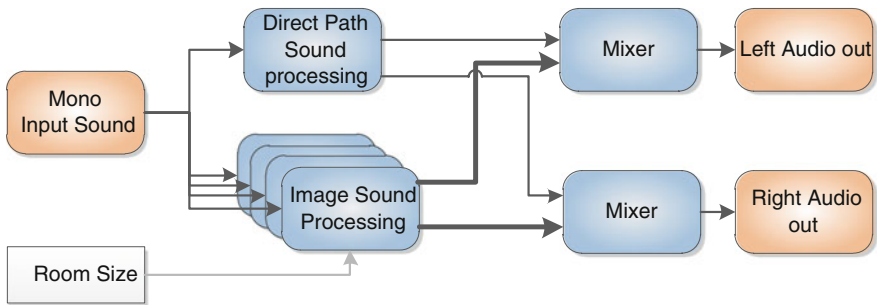
**Algorithm 1:** Pseudocode of the application

Our proposed algorithm was then implemented in the C language and was run on a laptop (Intel i5 2.5 GHz, 4 Gb RAM) and on an embedded device (BeagleBoard-xM; ARM Cortex A8 1 GHz with 512 Mb of RAM [16]), both running Linux (kernel: 3.1), to evaluate its performance. For the aforementioned implementation we use CD quality, 44.1 kHz sample rate, 2-channels, interleaved PCM 16-bit synthesized samples. The software we have implemented is based on Advanced Linux Sound Architecture libraries (ALSA), which provide a high level programming interface for sound cards without directly dealing with kernel drivers.

The processing of audio has been done with both of the discussed techniques, HRIR-based and binaural cues only. Figure 6.4 represents the application block-diagram used for all the implementations. To give the sensation of left and right displacement to the sound, the signal on the farthest channel has to be delayed and attenuated according to the angle of arrival. The delay depends on the head radius, while the FIR filter models the shadowing effect introduced by the presence of the head. In the case of measured HRIR, FIR filter coefficients contains the contribution of many folds of the pinnae and already captures the interaural time difference contribution.



**Fig. 6.4** Block diagram of the spatial algorithm, single voice spatialization is performed using proper delay and FIR-filtering



**Fig. 6.5** The system process sound with ISM by mixing together the contribution of five spatialized voices

The block diagram in Fig. 6.5 shows how the Image Source Method is implemented in the sound engine. Whatever is the spatialization method, the sensation of immersion and surrounding space dimension is realized by means of ISM reverberation. This is achieved by superimposing five spatialized voices, one that models the direct path between the virtual sound source and the listener. Other four contributions model the first order reflections on the four walls of the rectangular virtual room. If one wants to increase the order of the reverberation, the number of voices to mix up increases as well.

In order to estimate the possibility of future development of the proposed algorithm on embedded electronic systems, we compared the CPU execution time needed to process one second of sound. We chose to use ALSA without any other library in order to keep the software latency as low as possible. Notice that the implementation does not exploit DSP or RT capability of the BeagleBoard. Table 6.1 show that the time required to process a sound with the HRTF+IMS algorithm on the BeagleBoard is higher than the playback time; therefore continuous audio reproduction is not possible. On the contrary our implementation is able to process sound within the playback time, thus is suitable for real-time applications. Currently, the C application

**Table 6.1** Computation time required to compute one second of spatial sound

Algorithm	Laptop	Embedded device
Measured HRIRs	20	360
Measured HRIRs and ISM	80	1940
Synthesized HRIRs	5	130
Synthesized HRIRs and ISM	38	790

Time is expressed in milliseconds

plays back sound with standard write operations; however the application will be implemented using the MMAP DMA-like writing mode to perform audio processing while the sound is played at the output.

## 6.4 Testing

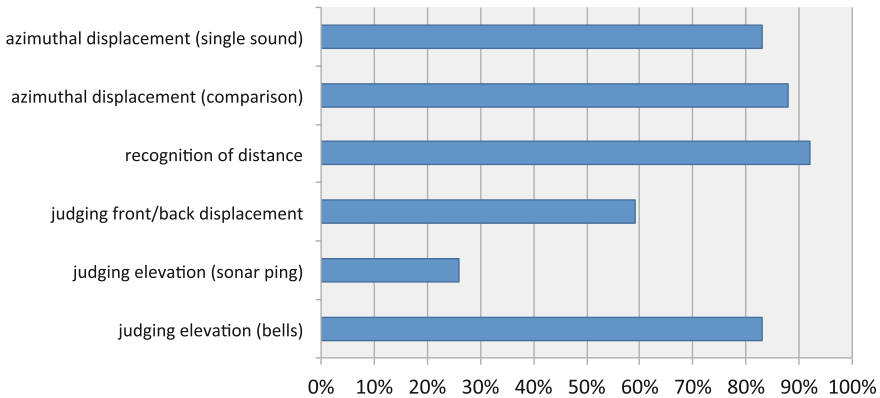
The effectiveness of the proposed sound enhancement algorithm was verified by asking individuals to provide feedback through a listening test and a survey. The scope of this work is to get an indication of how much the system is effective for a user, and eventually how to enhance it. The testing procedure is described in this section, along with some consideration on the collected responses. The survey includes different type of questions: multiple choice, check boxes and open answers. The survey is composed of a total of 57 questions relative to 36 different sounds scenes. It is organize in order to investigate the following aspects:

- the perception of source displacement and movements;
- the sensation of left and right displacement along with elevation;
- the perception of changes in the distance from the sound source;
- how the perception of the azimuth angle changes when comparing two sounds instead of hearing only one.

Each part of the questionnaire is subdivided in two phases: the first is a training part, while the second phase consists of the actual test from which we extract useful information. The main scope of the training part is to let the respondents familiarize with the problem and check if the listener correctly wears the earphones or has issue with his/her sound system. During this phase the listener hears a sound, and reads a description of what the sound is intended to mean. In the survey, the listener must agree or disagree with the description, or indicate if the description is incomplete or partially wrong. The familiarization part is composed of a total of 11 sounds, whose answers are not considered in the result section.

In the survey, different types of input sounds have been used, in particular:

- a single sinusoidal tone centered at A4, corresponding to 440 Hz,
- an Italian female voice pronouncing numbers from one to five,
- sampled stimuli of a bell, a sonar beep and a doorbell.



**Fig. 6.6** Percentage of right answers collected with the survey

In all cases, sounds were quantized with 16 bit at a sampling rate of 44.1 kHz.

The questionnaire is subdivided into four sections. The first part of the survey is focused on how effective binaural techniques are to determine the movements of the sound around the listener. In the second part, the listener must identify the path that the virtual sound source traverses. The input sound consists of several repetitions of the same sound, while the values of azimuthal angle and range vary, standing on the horizontal plane. The third part involves both perception of distance and of movements along the horizontal plane. The listener must say if he/she feels the sound is moving from left to right (or vice versa) and if the voice is approaching, departing or if the distance is perceived to be constant. In the last section, we investigate the recognition of elevation displacement of different types of sounds. This part is the only one exploiting variations of the elevation angles ranging from  $-45^\circ$  to  $+45^\circ$ .

We collected responses from 34 different individuals whose age ranges from 18 to 63. Each of the respondents uses a personal headphone. More than half of the respondents use earbuds, 30 % use in-ears and only six of them use full size headphones to answer the survey. Figure 6.6 depicts a summary of the analysis of the responses.

The listeners perform well in the judgment of the position of sound sources lying on the horizontal plane ( $\phi = 0$ ). In particular, recognition of a succession of sounds is slightly better (88 %) with respect of the recognition of sound position for a single sound (83 %). Recognizing sound source distance by comparison is generally an easy task, in fact listeners correctly answer to 92 % of the questions.

A problematic task is the judgment of front/back displacement of the sound. In this case, we consider sound stimuli computed at  $\phi = 0$ , or  $\phi = 180^\circ$ . In fact, almost half of the answers were wrong (59 % of correctness). This problem is related to the so called cone of confusion, or reversal error. In fact, there are many points in space that produces stimuli with exactly the same interaural differences and are therefore indistinguishable by the listeners.

In this test, we also investigate the perception of elevation. For this experiments, we synthesize sounds using measured HRTF only, since the discussed approach does not consider elevation cues. Interestingly, the percentage of correct responses varies consistently according to the type of input sound. Sonar pings are difficult to interpret (only 26 % of correct answers) compared to the sound of bells (83 %). This result can be explained because the bell sound have a spectral content spread over a wider range of frequency. Consequently, the output signal carries more spectral cues that the listener uses for the elevation judgment.

## 6.5 Conclusion and Future Work

In this work we presented navigational aid for visually impaired users. We proposed a simplified algorithm that exhibits lower computational demands with respect to the state of the art regarding 3D audio enhancement. The proposed solution is feasible, since it achieves real time rendering in portable embedded devices. Moreover, we described and commented the result of a survey. Results show the efficacy of the system and suggest to investigate further the reversal error and the choice of input sounds that can be integrated in the final application. Our method is based on a small number of parameters, and is therefore a promising solution for the individualization of sound spatialization filter sets, with the advantage of not requiring complex and time-demanding measurement sessions. In the next stage of development we plan to integrate the sensation of elevation of the sound source. In the future, we would also like to integrate the proposed system with an inertial platform placed over the headphone to capture the listener's head orientation with respect to the virtual scenario. This way, the software can align the displacement of the sound source with the environment. By doing this coherently with the integration time of perception of sounds the problem of front/confusion will be solved. Our testing approach will also be extended to include a personalization algorithm.

**Acknowledgments** This work was supported by the EU project DALi, grant number ICT-2011-288917.

## References

1. Zöllner, M., Huber, S., Jetter, H.C., Reiterer, H.: NAVI—a proof-of-concept of a mobile navigational aid for visually impaired based on the microsoft kinect. In: Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT'11), Lisbon, Portugal (2011)
2. Dunai, L., Fajarnes, G.P., Praderas, V.S., Garcia, B.D., Lengua, I.L.: Real-time assistance prototype—a new navigation aid for blind people. In: IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, pp. 1173–1178 (2010)
3. Murch, G.M.: Visual and Auditory Perception. Bobbs-Merrill, Indianapolis (1973)

4. Begault, D., Wenzel, E.M., Godfroy, M., Miller, J.D., Anderson, M.R.: Applying spatial audio to human interfaces: 25 years of NASA experience. In: Proceedings of AES 40th International Conference, Tokyo, Japan (2010)
5. Begault, Durand R.: 3-D Sound for Virtual Reality and Multimedia. Academic Press Professional, Cambridge (1994)
6. Devices for Assisted Living, DALi FP7 project, website (2013) <http://www.ict-dali.eu/dali>
7. Blauert, J.: Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, Cambridge (1983)
8. Algazi, V.R., Duda, R.O., Thompson, D.M., Avendano, C.: The CIPIC HRTF database. In: IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, pp. 99–102. Mohonk Mountain House, New Paltz (2001)
9. Begault, D., Wenzel, E., Lee, A., Anderson, M.: Direct comparison of the impact of head tracking, reverberation, and individualized HRTFs on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.* **49**(10), 904–916 (2001)
10. Brown, C.P., Duda, R.O.: A structural model for binaural sound synthesis. *IEEE Trans. Speech Audio Process.* **6**(5), 476–488 (1998)
11. Jot, J.M.: Efficient models for reverberation and distance rendering in computer music and virtual audio reality. In: Proceedings of International Computer Music Conference, ICMC (1997)
12. Borish, J.: Extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Am.* **75**(6), 1827–1836 (1984)
13. Rizzon, L., Passerone, R.: Embedded soundscape rendering for the visually impaired. In: Proceedings of the 8th IEEE International Symposium on Industrial Embedded Systems (SIES13), Porto, Portugal (2013)
14. Pinto, A., Bonivento, A., Sangiovanni-Vincentelli, A.L., Passerone, R., Sgroi, M.: System level design paradigms: Platform-based design and communication synthesis. In: *ACM Transactions on Design Automation of Electronic Systems*, vol. 11, no. 3. ACM Press, New York, USA (2006)
15. Davare, A., Densmore, D., Guo, L., Passerone, R., Sangiovanni-Vincentelli, A.L., Simalatsar, A., Zhu, Q.: metro II: a design environment for cyber-physical systems. In: *ACM Transactions on Embedded Computing Systems*, vol. 12, no. 1s. ACM Press, New York, USA (2013)
16. Beagle Board website <http://beagleboard.org>

# Chapter 7

## BASIC32: A New ASIC for Silicon Photomultiplier Detectors

Fabio Ciciriello, Francesco Corsi, Francesco Licciulli, Cristoforo Marzocca,  
Gianvito Matarrese, Alberto Del Guerra and Maria Giuseppina Bisogni

**Abstract** Silicon Photomultipliers (SiPM) have shown to be excellent substitutes for more traditional and bulky Photomultiplier Tubes (PMT) in many photon detection applications, thanks essentially to their high quantum efficiency, low bias voltage and immunity to magnetic fields. However, they pose some challenging design constraints on the design of the electronic front-end (FE) due to their intrinsic high gain and speed. In particular, when changing from low to high light levels of exposition, they produce output signals whose amplitude may span over about three order of magnitude with rise times of less than 1 ns. This is of particular concern when developing integrated multichannel electronics in deep submicron technology. We report here on the realization of a 32-channel ASIC in CMOS technology, based on an innovative current-mode architecture. Besides presenting the experimental results of the ASIC characterization, some perspective indications are given concerning the work currently in progress.

### 7.1 Introduction

Among the various types of solid state devices presently employed in the detection of low energy photons, SiPMs have gained growing diffusion over different fields of applications. Aside the field of high energy physics experiments where they first

---

F. Ciciriello · F. Corsi (✉) · F. Licciulli · C. Marzocca (✉) · G. Matarrese  
Dip. di Ing. Elettrica e dell'Informazione-Politecnico di Bari and INFN Sez. Bari,  
Via Orabona 4, 70125 Bari, Italy  
e-mail: francesco.corsi@poliba.it

C. Marzocca  
e-mail: cristoforo.marzocca@poliba.it

A. Del Guerra · M. G. Bisogni  
Dip. di Fisica-Università di Pisa and INFN Sez. Pisa, Largo Bruno Pontecorvo 3,  
56127 Pisa, Italy

have found application, they are presently employed in many new generation medical imaging and astroparticle detection. This is essentially due to their excellent single photon detection capability, high quantum efficiency and relatively rugged physical structure. In particular, they share the same basic structure and operation principle of other common devices, i.e. the avalanche photo diode (APD) and the single photon avalanche device (SPAD) [1] but, in contrast with these latter, they are constituted by an array of elementary  $p-n$  junctions rather than a single one. These elementary devices, called micro-cells, are reverse biased slightly beyond their breakdown voltage so as to operate in Geiger mode [2], rather than in proportional mode as in APDs. Compared to the more traditional PMTs, largely employed until recent years, SiPMs offer the advantages of a much lower bias voltage (30–70 V), higher quantum efficiency, insensitivity to magnetic fields, smaller size and cost, thus explaining the enormous diffusion of these devices in all those applications where time and spatial resolution are at a premium such as, for example, in positron emission tomography (PET) scanners and in PET\_MRI (magnetic resonance imaging) systems.

However, the peculiar structure and operating mode of these devices poses some very demanding constraints on the FE electronics which has to be employed to fully exploit their intrinsic excellent gain and timing capabilities.

As a first point, due to the high gain associated to the avalanche process ( $10^5 - 10^6$ ) and to the number of elementary micro-cells, the amplitude of the current pulse produced in response to a  $\gamma$ -photon may range from  $10^{-6}$  to  $10^{-3}$  A, thus calling for a wide dynamic range of the preamplifier employed to read the SiPM signal.

Another issue is related to the fast rise time of the current pulse (typically less than 1 ns) which has to be read by the front-end without sensible degradation. This is a rather difficult task since the SiPM is characterized by a high intrinsic capacitance produced essentially by the parallel combination of all the reverse biased elementary  $p-n$  junctions constituting the micro-cells and by the metal grid used to distribute the bias within the device. The device capacitance may vary from a few tens (for SiPM with 400–600 micro-cells) to a few hundreds (for SiPM with 3000–4000 micro-cells) of picofarads. Thus, preserving the signal bandwidth imposes to realize a low input resistance for the preamplifier.

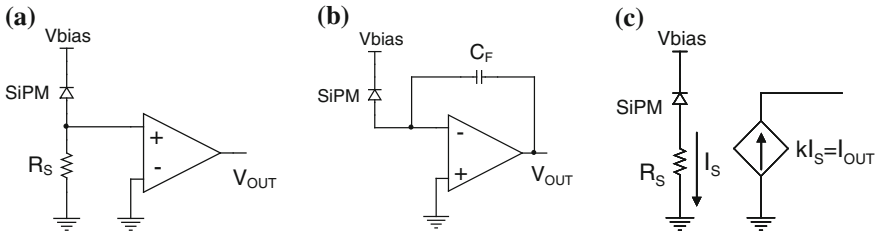
The main source of noise in these devices is related to the Geiger elementary discharges caused by the thermally generated carriers within the sensitive volume (dark pulses).

Designing suitable FE electronics for these devices requires accounting for the above characteristics and for the inevitable parasitics associated to the interconnection between the SiPM and the FE.

## 7.2 Current-Mode Front-End Architecture

A comprehensive electrical model of the SiPM coupled to the FE has been proposed [3] and has proved to be of great help in evaluating the dependence of the system performance on some crucial parameters of the FE. One of the most widely used



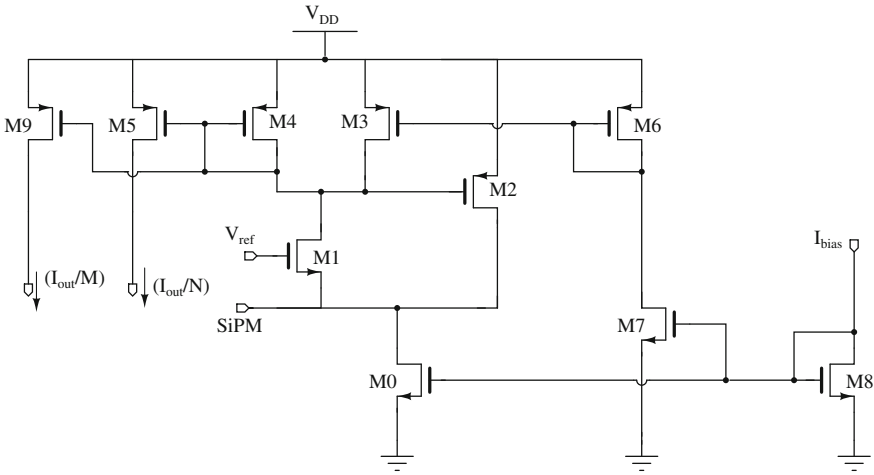


**Fig. 7.1** SiPM read-out by means of: **a** a resistor  $R_S$  and a voltage amplifier; **b** a charge-sensitive amplifier; **c** a current buffer

approaches involves a current-to-voltage conversion of the SiPM signal through a resistor followed by a fast voltage amplifier to achieve the desired signal level, as shown in Fig. 7.1a.

This scheme is often used for the characterization of SiPMs and for timing measurements performed on a limited number of detectors, as it can be easily implemented in a discrete environment and thus can be made fast with a proper choice of the components. Since the amplifier input is not a virtual ground, the value of the conversion resistor must be small enough to avoid fluctuations of the SiPM bias voltage under signal; on the other hand, a small value of  $R_S$  is needed also to reduce the time constant at the amplifier input which could slow down the response of the FE, depending on the SiPM input capacitance. With this approach timing can be made fast, but when the application requires to extract the charge released by the detector (and thus the energy associated with an event), the output voltage must be integrated and this involves a further voltage-to-current conversion and an integration stage [4].

The most straightforward solution appears to be a charge sensitive amplifier (CSA), in which the charge delivered by an event is collected on a feedback capacitance, as shown in Fig. 7.1b. This configuration is able to guarantee the best noise performance and has represented the standard scheme adopted for the read-out of radiation detectors. However, in case of SiPM detectors, a number of issues arises, especially when several FE channels must be integrated on the same chip, to read-out an array of SiPMs. Since the charge delivered by each SiPM in response to an event is very large, the CSA dynamic range can represent a serious limitation. For instance, considering a typical gain of  $10^6$  for the SiPM, and a standard  $0.35\ \mu\text{m}$ ,  $3.3\text{V}$  CMOS process, in which the maximum allowed supply swing is  $\Delta V \cong 3\text{V}$ , the feedback capacitance required to collect a charge  $Q = 50\text{ pC}$  (corresponding to about 300 micro-cells hit) should at least be  $C_f = Q/\Delta V \cong 16.7\text{ pF}$ . This value is impractical in a FE multi-channel ASIC, due to the excessive area occupancy. The situation worsen when a deeper sub-micron technology is used and/or when the number of micro-cells of the detector increases. Moreover, the output stage of the amplifier used to implement the CSA must be able to drive the large capacitive load given by the series of the feedback and the SiPM capacitances. To avoid stability issues and to achieve the speed constraints required in many applications, the output stage must be biased with a large current, thus increasing also the power consumption.



**Fig. 7.2** Basic structure of the input current buffer

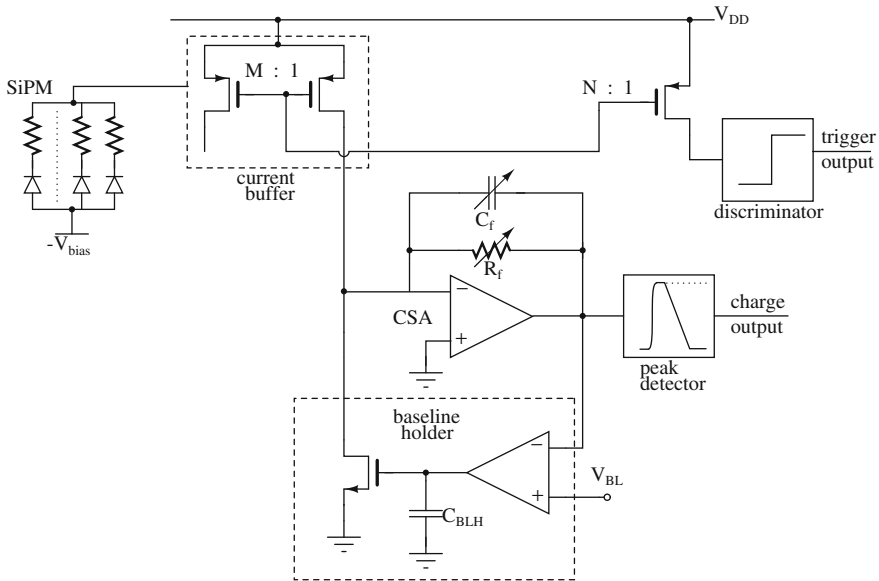
On the basis of these considerations and to avoid multiple conversions between current and voltage signals, we adopted a FE architecture based on a current-mode approach. The current signal from the detector can be directly read by means of a current buffer with low input impedance. This approach offers great flexibility since the output current of the buffer can be a scaled replica of the detector signal at high impedance, according to the principle schematic in Fig. 7.1c. This output current can be easily replicated, by means of current mirrors, and sent to a simple integration stage, to extract the charge information, and/or to a current discriminator, to obtain a fast timing signal. The circuit is inherently fast and the current mode of operation enhances the dynamic range, since it does not suffer from possible voltage limitations posed by a deep-submicron implementation.

By using the accurate electrical model of the SiPM, proposed in [3], it has been shown [4] that the peak value of the current delivered by the SiPM increases with decreasing values of the input impedance of the buffer; moreover, the rise time and the peak value of the current buffer output are strongly affected by the bandwidth of the circuit.

For these reasons, current feedback techniques have been used to improve the performance of the buffer. The final configuration is shown in Fig. 7.2.

Basically, the circuit consists of a current follower, the common gate  $M1$ , closed in a feedback loop, built around the common source  $M2$  and the diode-connected MOSFET  $M4$ . The negative feedback decreases the input impedance of the buffer and enhances its frequency response by a factor which depends on the loop gain  $T$ , equal to  $g_{m2}/g_{m4}$ .

The resulting FE structure, which has been implemented in a  $0.35\ \mu\text{m}$  CMOS process, is shown in Fig. 7.3; it exhibits two signal paths: a fast path, used to extract the timing of the detected event, which includes a current discriminator, and a slow path,

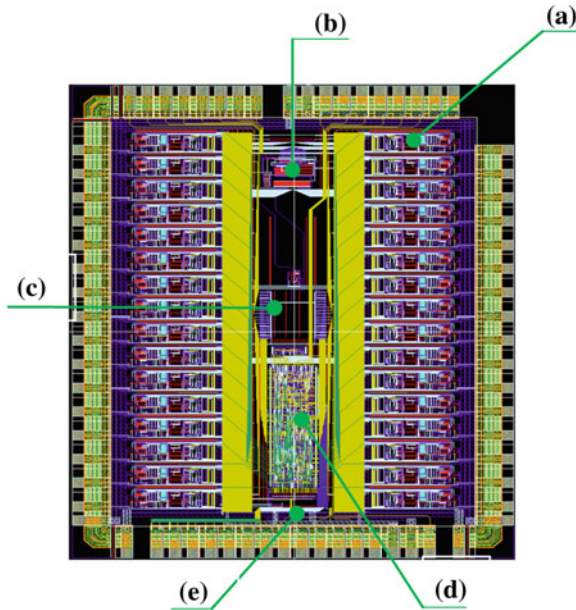


**Fig. 7.3** Schematic architecture of the analog FE channel

which integrates the charge by means of a CSA and provides the energy information. The input current of the slow signal path is a replica of the output of the current buffer, scaled down by a factor  $M = 10$ , in order to allow the use of a smaller integration capacitance, while keeping the CSA output in the range of  $0.3 \div 3$  V. The damping time constant of the integrator  $R_f C_f$  has been fixed at 200 ns. The CSA features a gain which can be selected among three values (1; 0.5; 0.33 V/pC) by means of two bits ( $a_0, a_1$ ). A baseline holder circuit (BLH) is used to stabilize the DC baseline voltage at the CSA output. The BLH circuit, enclosed in a feedback loop with the CSA, compares the CSA output voltage to a reference one,  $V_{BL}$  and, if the loop gain of the feedback system is sufficiently high, injects at the CSA input a current needed to keep the virtual short-circuit between the BLH inputs, thus allowing to set the baseline at the reference voltage  $V_{BL}$  (300 mV).

The discriminator threshold can be set by means of a DAC between 0 and  $40 \mu\text{A}$ . The rise time of the output voltage is about 300 ps with a 4 pF load.

Finally, the information related to the energy of incoming events is associated with the peak voltage pulse value at the output of the CSA. A peak detector (PD) circuit is employed to hold the peak value during the analog-to-digital conversion.



**Fig. 7.4** BASIC32 layout: **a** analog channels; **b** bias circuit; **c** analog multiplexer; **d** digital logic block; **e** fast-OR

### 7.3 BASIC32: A 32-Channel ASIC

Based on the analog channel described in the previous section, a 32-channel ASIC (BASIC32) has been designed, including a digital logic block needed for managing the configuration and read-out phases [5]. In Fig. 7.4 the layout of BASIC32 is shown, with the 32 analog channels placed in two columns on the opposite sides of the chip (Fig. 7.4a), whereas in the central region it is possible to identify a bias circuit (Fig. 7.4b), an analog multiplexer (Fig. 7.4c), the read-out logic (Fig. 7.4d) and a fast-OR circuit (Fig. 7.4e). Each channel, which provides the trigger signal (fast-path) and the analog signal proportional to the charge released by the SiPM, is equipped with digital inputs that allow to program and control the mode of operation (gain, time constant, threshold and track-and-hold phases of the peak detector). The bias circuit generates the needed voltage and current references by exploiting the output of a bandgap, whereas the analog mux allows to transfer on an analog output pad the charge signals of the channels, one at a time. All the signals generated by the fast-path of each channel are collected and processed by the fast-OR circuit, so that a trigger signal is fired as soon as at least one of the channels is found over threshold and the read-out procedure is started consequently.

The digital logic block manages the SPI interface used for setting both the register containing the channel configuration bits (all the channels share the same configuration) and the bits which control the read-out acquisition mode (internal/external or

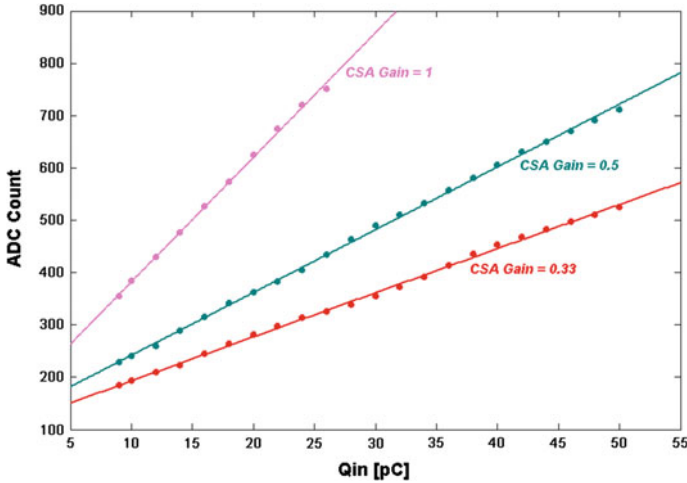


Fig. 7.5 PD output as a function of the injected charge

sparse/serial). The track-and-hold mode of operations of the peak detectors and the multiplexing of the channels on the output analog pad are automatically operated by the internal logic if the internal read-out mode is selected. In this acquisition mode, the read-out procedure is started by the output of the fast-OR, which fires as soon as one of the channels goes above threshold. Two time windows are then opened: the channels found above threshold within the first window (TWA), of very short duration, are marked in a trigger register. During the second time window (TWB), of longer duration, the presence of an external “coincidence” signal is checked. If the coincidence signal is not activated within TWB, the event is discarded and the read-out logic and the PDs are reset, so that the chip is again ready to receive a further trigger from the fast-OR. On the contrary, if the coincidence signal goes high within TWB, the outputs of the PDs are multiplexed according to the selected acquisition mode (in sparse mode, only the outputs of the PDs associated to the channels found above threshold within TWA, whereas in serial mode all the channels are read out). The coincidence signal can be very useful when the application requires the acquisition of an event only if certain conditions are fulfilled (for example, the coincidence of two events in a PET scanner). The read-out logic also provides a digital output which is activated only during the read-out procedure and marks the time when the last channel has been multiplexed in the sparse read-out mode.

In the following some results from the characterization tests are reported. Figure 7.5 shows the PD output (ADC counts) as a function of the injected charge, for three different values of the CSA gain. The linearity error is within 1 % up to 70 pC if the smallest possible gain value is selected.

A SiPM equipped with a small LYSO scintillator has been coupled to a channel of BASIC32 and exposed to different radiation sources, such as  $^{176}\text{Lu}$  (203 and

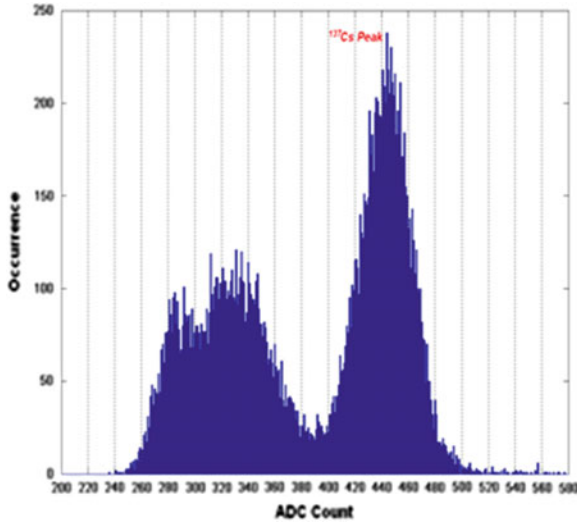


Fig. 7.6 Spectrum of  $^{137}\text{Cs}$ , acquired with a SiPM coupled to a LYSO crystal

307 keV),  $^{22}\text{Na}$  (511 keV),  $^{137}\text{Cs}$  (662 keV),  $^{57}\text{Co}$  (122 keV). In Fig. 7.6, the emission spectrum for the Cesium is shown, which exhibits a 12 % FWHM energy resolution.

## 7.4 Future Developments

Latest studies concerning PET diagnostics aim at improving the signal-to-noise ratio (SNR) of the imaging system through the introduction of a time measurement. Modern PETs (ToF-PET: Time of Flight PET) measure, beside the impact position of the  $\gamma$ -ray on the scanner, the time of the photon. In order to achieve real improvements of the SNR, it is mandatory to identify the arrival time of the very first photon generated by the scintillation crystal with an accuracy of the order of hundreds of picoseconds [6]. This kind of measurements require the development of front-end electronics with two essential features: wide bandwidth and low noise. The first is needed to preserve the very fast slope of the input pulse, whereas low noise allows the reduction of the threshold that is used to detect the single photon and decreases the arrival time jitter. In fact, the standard deviation  $\sigma_t$  of the time jitter is given by:

$$\sigma_t = \sigma_n / \left( \frac{dV_o}{dt} \right)$$

where  $\sigma_n$  is the standard deviation of the noise and  $dV_o/dt$  is the slope of the signal [7].

Recent researches aimed to PET scanner design use SiPMs with large dimensions in order to maximize the total detection efficiency. The downside of this solution is the increase of the parasitic capacitance and the dark rate of the detector. FE circuits based on a feedback structure, like the one used in BASIC32, exhibit strong limitations when coupled to such large capacitance detectors, since the first pole, that defines the bandwidth of the circuit, is related to the input node and so to the capacitance of the detector. Thus the heavy capacitive load leads to a reduced bandwidth. This results in a slow rising edge which prevents from obtaining a very low time jitter. Solutions that adopt open loop architectures are thus preferred in these cases. The most suitable circuit is the common gate/base configuration, where the source/emitter terminal is the input, while the drain/collector is the output. This solution has two independent poles, one related to input node and the other to the output. For instance, a common base input stage realized by a SiGe HBT allows the reduction (compared to a MOST with the same bias current) of both the input resistance and output capacitance. In other words an HBT allows remarkable power consumption and size reductions respect to a MOST.

Further improvements in signal conditioning concern new ways to measure the charge related to the input pulses. Basically, the choice is between direct integration of the input current pulse and time over threshold (TOT) techniques, based on the association of the charge contained in the scintillation pulse to the time during which the signal stays above a given threshold. The latter offers the advantage that only a Time to Digital Converter (TDC) is needed for both time and charge measurements. In both cases a suitable mechanism should be established to get rid of the dark pulses: a double threshold approach can be used, with a low threshold used for the timing measurements and a higher threshold used to reject all the triggers coming from small signals, classified as dark pulses.

Direct integration provides better energy accuracy than TOT techniques, since the amplitude of the signal is strongly dependent from the statistic behavior of the scintillator. A mixed approach can guarantee very good results in both timing and energy accuracy: an integrator can be used to validate a trigger when the charge associated to the pulse reaches a low threshold in a given short time window. In case of valid signal, the integration proceeds over a suitable longer time window and then is stopped. After that, the integration capacitance is discharged at constant current, so that the discharge time is proportional to the total integrated charge. In this way the TDC is able to measure with the required accuracy both the arrival time of the event and the discharge time, thus the charge associated to the event.

A 4-channel ASIC, based on this approach, is currently under development. The very first FE is an open loop current buffer based on a SiGe HBT, whereas an implementation of the mixed integration-TOT technique is exploited for the charge measurements.

## References

1. Cova, S., Ghioni, M., Lacaita, A., Samori, C., Zappa, F.: Avalanche photodiodes and quenching circuits for single-photon detection. *Appl. Opt.* **35**(12), 1959–1976 (1996)
2. Buzhan, P., Dolgoshein, B., Filatov, L., et al.: Silicon photomultiplier and its possible applications. *Nucl. Instrum. Methods Phys. Res. A* **A504**, 48–52 (2003)
3. Corsi, F., Dragone, A., Marzocca, C., et al.: Modelling a silicon photomultiplier (SiPM) as a signal source for optimum front-end design. *Nucl. Instrum. Methods Phys. Res. A* **A572**, 416–418 (2007)
4. Corsi, F., Foresta, M., Marzocca, C., Matarrese, G., et al.: Preliminary results from a current-mode CMOS front-end circuit for silicon photomultiplier detectors. In: *IEEE Nuclear Science Symposium (NSS-MIC'07) Conference Record*, pp. 360–365. Honolulu, USA (2007)
5. Corsi, F., Argentieri, A.G., Foresta, M., Marzocca, C., Matarrese, G., et al.: Front-end Electronics for silicon photo-multipliers coupled to fast scintillators. In: *IEEE Nuclear Science Symposium (NSS-MIC'10) Conference Record*, pp. 1332–1339. Knoxville, USA (2010)
6. Spanoudaki, VCh., Levin, C.S.: Photo-detectors for time of flight positron emission tomography (ToF-PET). *Sensors* **2010**(10), 10484–10505 (2010)
7. Spieler, H.: *Semiconductor Detector System*, pp. 179–188. Oxford Science Publications, Oxford (2005)



# Chapter 8

## Reconfigurable Implementation of a CNN-UM Platform for Fast Dynamical Systems Simulation

Gianluca Borgese, Calogero Pace, Pietro Pantano and Eleonora Bilotta

**Abstract** In this work we present a distributed computing system, called DCMARK, aimed at solving partial differential equations at the basis of many investigation fields such as Solid State Physics, Nuclear Physics and Plasma Physics. This distributed architecture is based on the Cellular Neural Network (CNN) paradigm which allows to divide the differential equation system solving into many parallel integration operations to be executed by a custom multiprocessor system. We pushed the number of processors to the limit of one processor for each equation. In order to test the present idea, we chose to implement DCMARK on a single FPGA, designing the single processor in order to minimize its hardware requirements and to obtain a large number of easily interconnected processors. This approach is particularly suited to study the properties of one-, two- and three-dimensional locally interconnected dynamical systems. In order to test the computing platform, we implemented a 200 cells, Korteweg de Vries (KdV) equation solver and performed a comparison between simulations conducted on high performance PC and on our system. Since our distributed architecture takes a constant computing time to solve the equation system, independently of the number of dynamical elements (cells) of the CNN array, it allows to reduce the elaboration time more than other similar systems in the literature. To ensure a high level of reconfigurability, we designed a compact System on Programmable Chip (SoPC) managed by a softcore processor which controls the

---

G. Borgese (✉) · C. Pace  
Department of Informatics, Modeling, Electronics and System Engineering,  
University of Calabria, 87036 Arcavacata di Rende, Cosenza, Italy  
e-mail: gianluca.borgese@dimes.unical.it

C. Pace  
e-mail: cpace@unical.it

P. Pantano · E. Bilotta  
Department of Physics, University of Calabria, 87036 Arcavacata di Rende, Cosenza, Italy  
e-mail: piepa@unical.it

E. Bilotta  
e-mail: bilotta@unical.it

fast data/control communication between our system and a PC Host. An intuitively Graphical User Interface (GUI) allows to change the calculation parameters and plot the results.

## 8.1 Introduction

Nowadays in many fields of physics and engineering there is the need to analyze and tackle various complex problems such as numerical differential equation solving, high-definition image processing, target recognition and tracking, etc., in which the elaboration time is an important and critical factor. The problem can be technically addressed using higher and higher performance computers, such as multi-core processor grid. The problems of this approach are the complexity in programming parallel code, the managing and controlling of a large grid and the necessity to have access to such facilities. Moreover, these systems are usually cumbersome, expensive and power-hungry, thus making their use in embedded applications such as industrial controllers, missile guidance systems, video surveillance, etc. To solve the problem, the generality of the hardware architecture can be sacrificed re-designing the system in order to obtain very significant advantages in specific applications. Several hardware technologies such as digital signal processing (DSP) devices [1], reconfigurable digital devices (e.g.: FPGAs) [2–4], symmetric multi-processing (SMP) machines [5] as well as neural network systems [6] and cellular neural network (CNN) paradigms [7, 8] are employed to achieve the purpose. The CNN architecture takes advantage over the neural network one thanks to its lower implementation complexity. It is convenient to investigate locally interconnected dynamical structures thanks to the CNN distributed computing approach. The evolution of this idea is the cellular neural network universal machine (CNN-UM) which is a new CNN computing structure formed by an array of  $N \times M$  dedicated processors [9]. The CNN paradigm can be implemented in many kind of technologies such as analog devices [10, 11], hybrid digital devices, embedded architectures, digital signal processing systems and reconfigurable platforms as FPGAs [12, 13]. Clearly, there are pros and cons in the choice of each technology. By using the analog approach there is the advantage to have higher computing performances but a lower accuracy due to non-linearity and dispersion of analog component parameters. With a reconfigurable digital approach we have lower design costs, higher accuracy but lower elaboration performances. In the recent years, the possibility, for the common user, to write specific code for graphic processing unit (GPU) systems, thanks to programming environments such as the Compute Unified Device Architecture (CUDA), has led to significant improvements in elaboration speed due to the highly parallelized hardware architecture typical of graphic processors [14]. Speedup ratios ranging from tens to hundreds, w.r.t. standard CPU approach, have been demonstrated in several application fields such as image processing [15] and fluid dynamic simulations [16]. GPU are relatively low cost and natively integrated in common PC platforms but their architecture is, as expected, optimized for the algorithms to be executed by

the PC graphic board. So the implementation of the CNN UM always requires the reorganization of the algorithm in order to efficiently use the GPU resources in terms of number of parallel processes and internal data transfer rate [17]. In this work we propose an FPGA based distributed computing microarchitecture (DCMARK) based on the CNN UM paradigm. This approach was already used in many research fields [18–21] in confirmation of its efficiency. Our guiding idea was to push the number of digital computing units, working in parallel, to the limit of the CNN cell’s number, replicating the CNN architecture, taking advantage, at the same time, of the efficiency of the hardware implementation of the local interconnection.

The choice of an FPGA platform for the development phase was obvious, but the possibility to maintain a high degree of reconfigurability in the system convinced us to select it as the final implementation technology. To this aim a calculation system (DCMARK Calculator) was developed, of which DCMARK is the computing part. This system [22] is aimed to solve partial differential equation systems, in particular, as a well known benchmark, we chose the one dimensional Korteweg de Vries equation system [23].

Using this type of architecture it is possible to reduce the elaboration time, increase the CNN array size (number of computing cells), increase equation solution accuracy and obtain a run-time fast calculator.

## 8.2 General System Architecture

The complete block diagram of the whole calculation system is shown in Fig. 8.1. It is divided into three main parts:

- a Terasic DE4-230 FPGA Development Board;
- the System on Programmable Chip (SoPC) which contains our Distributed Computing Microarchitecture block implemented on the FPGA;
- a host PC with a Graphic User Interface (GUI) for calculation management.

The Development board and the host PC communicate using an Ethernet protocol connection. SoPC is a complete system controlled by an Altera NIOS II IP softcore processor. In the system there are: a Triple Speed Ethernet (TSE) module for communications, our Distributed Computing System (DCSYS), and Scatter-Gather (SG) DMA modules for fast transmission operations between devices. NIOS II processor manages all main SoPC operations such as communications between TSE module and the host PC, devices interrupt handling, MicroC/OS II Operative System (OS) supervising, etc. NIOS II is programmed by the user using the Eclipse software, based on C language. The SoPC allows to interconnect system devices thanks to an Altera communication bus facility called Avalon. There are several kinds of Avalon buses for every need such as Memory-Mapped (MM) bus with a Master (M) and Slave (S) structure for device command and control operations and Streaming (ST) bus with a Source (SRC) and Sink (SNK) structure for continuous data transmissions.

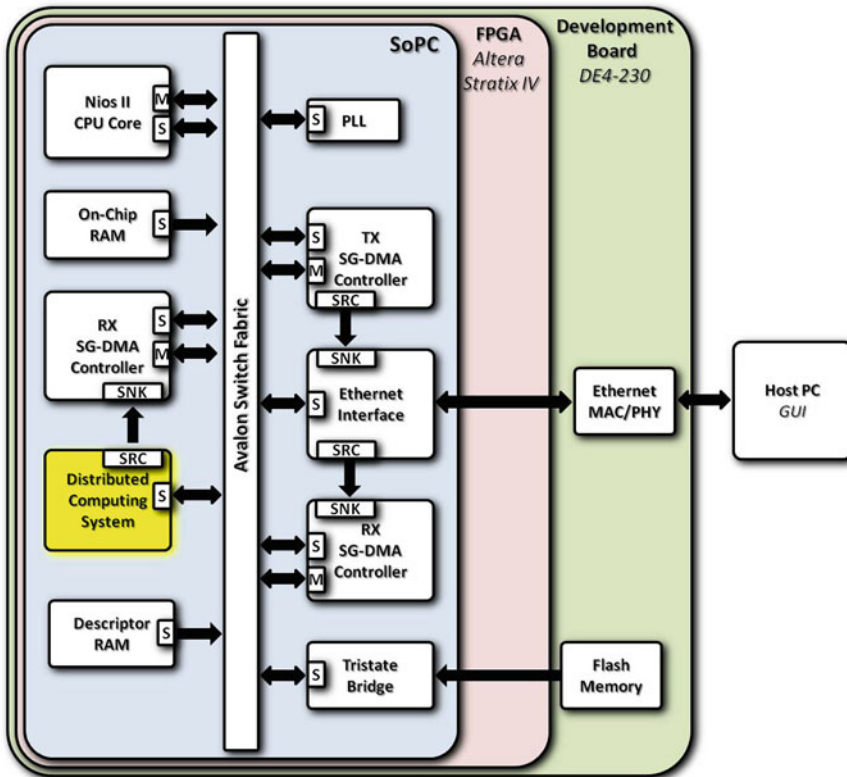


Fig. 8.1 DCMARK calculator block diagram

### 8.3 Distributed Computing Microarchitecture

The DCMARK, included into the Distributed Computing System (DCSYS) block of SoPC, is based on the CNN-UM approach in which several custom processors execute a group of sequential operations at the same time in order to elaborate particular information. At the first computing step each processor acquires status data from the neighboring processors (with a limited sphere of influence) and after a determined number of clock cycles (elaboration time) they will give their local results. These results can be analyzed versus time and/or space. With a time analysis we study just one single processor (a fixed spatial point) results versus time while with a space analysis we study all processors (all spatial points) results at a fixed time step. In this paper, by way of example, a 1-D locally interconnected dynamical system is investigated. This system is based on a discretized partial differential equation where every point of the spatial array is a dynamical element. Each processor of DCMARK is dedicated to a dynamical element of the spatial array. In our case every dynamical element has a neighborhood formed by four dynamical elements (two on its right and two on its left).

### 8.3.1 Single Cell Block

Each calculation unit of DCMARK is called a Cell and is based on a Von Neumann elaboration architecture, thus its RAM memory stores both Data and Micro-Code (MCode). The MCode approach was chosen in order to easily modify the physical phenomenon investigation just changing the implemented equation (as long as it is expressible with sums and multiplications). The Cell (Fig. 8.4) has a 40 bit Data Bus and an 8 bit Address Bus. Its Arithmetic Logic Unit (ALU) allows to execute Floating Point (FP) additions and multiplications. In order to demonstrate the DCMARK idea, this first implementation of the Cell uses ALTERA blocks for the adder and multiplier, leading to a maximum amount of about 200 Cells to be integrated in the FPGA device adopted. This number is expected to be substantially increased by working on the customization of the ALU block and thanks to the continuing growth of the available device size. The Cell, that can be clocked up to 180 MHz, contains:

- a Control Module, implemented as a Finite State Machine (FSM) which controls the micro-code execution and enables the control signals;
- a 40 bit  $\times$  256 RAM memory;
- a 32 bit Floating Point (FP) Adder;
- a 32 bit Floating Point (FP) Multiplier;
- a 8 bit Program Counter;
- a 40 bit Instruction Register;
- three 32 bit Operation Registers (A, B and C) for arithmetical operations;
- five 32 bit input/output (I/O) Registers (I, M2, M1, P1, P2) for acquiring current status data from its neighborhood;
- a 6  $\times$  1 40 bit Data Multiplexer;
- a 2  $\times$  1 8 bit Multiplexer;
- a 2  $\times$  1 32 bit Result Multiplexer.

The Micro-Code is written using a group of 12 custom micro-instructions: loading data in register A, B and I (LDA, LDB, LDI), storing I/O register data and operation result on RAM (STM1, STM2, STP1, STP2, ST), adding (ADD), subtracting (SUB), multiplying (MUL) and jumping (JUMP). Each instruction stored in RAM has the following format (Fig. 8.2).

### 8.3.2 Parallel Cell Configuration Module

In order to quickly program the RAMs of all the Cells, we designed a Parallel Cell Configuration Module (PCCM) (Fig. 8.3).

The PCCM is formed by:

- a Configuration Block: it is a Finite State Machine (FSM) which reads a configuration File from Configuration ROM and programs all the Cells;

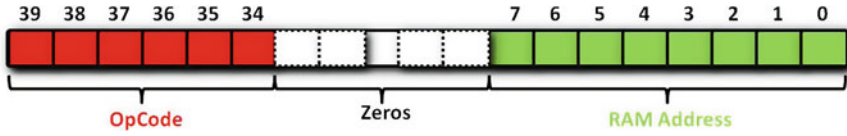


Fig. 8.2 RAM data word structure

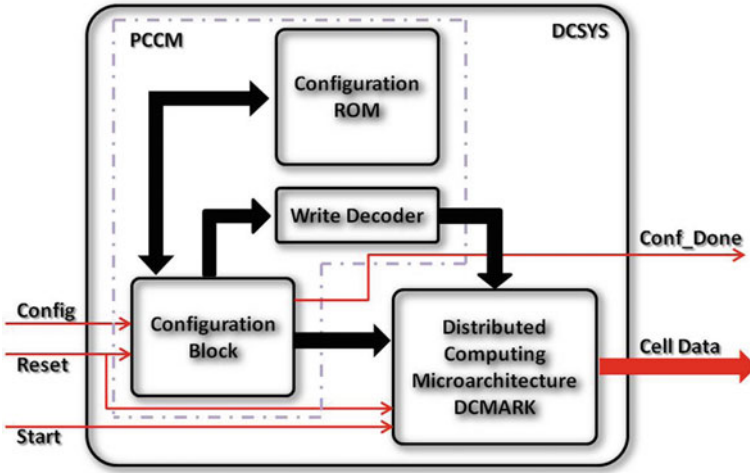
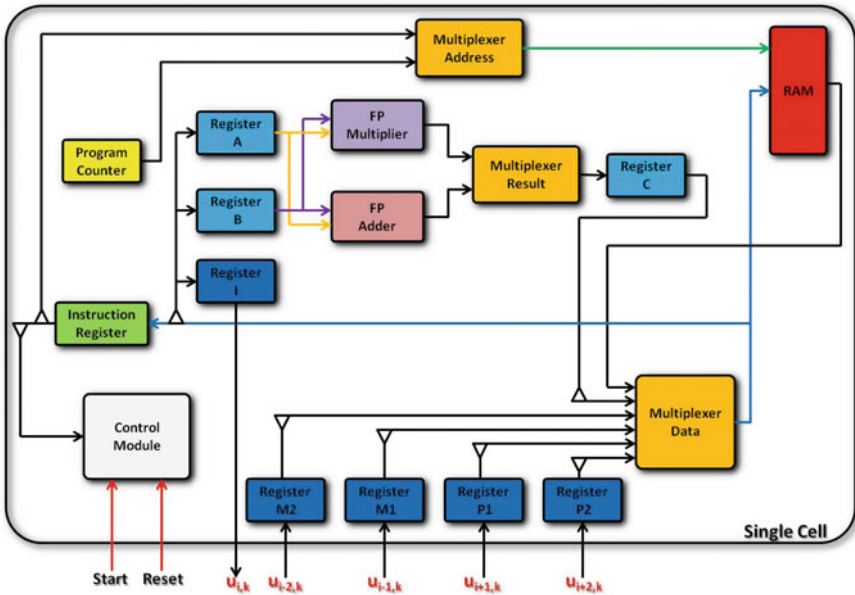


Fig. 8.3 DCSYS block diagram

- a Configuration ROM: it stores a Configuration File containing MCode, initial status variables and constants;
- a Write Decoder: it allows to address every Cell for one-to-one and parallel programming. One-to-one programming allows to store a different initial status variable on each cell RAM while parallel programming allows to store MCode and constants on all cell RAM at the same time.

### 8.4 Complex Physical Dynamics Investigation

One of the most important topic of contemporary science focuses on the study of continuous [24, 25] and discrete [26, 27] dynamical systems, analyzing their organization as non linear evolving structures [28]. Chaos is the most striking feature of their behavior. The concept of dynamical system is connected to a mathematical model which describes its time evolution and it is often characterized by differential equations [29]. Differential equation solving allows to define and forecast the future evolution of the system in time and space. To allow a more and more detailed analysis of dynamic systems it is absolutely necessary to perform long and heavy



**Fig. 8.4** DCMARK single cell block diagram (working registers are in sky-blue and I/O registers are in dark blue; the blue and green lines represent the 40 bit data bus and the 8 bit address bus respectively)

numerical simulations which would require powerful, fast and expensive elaborators (sometime multi-core grid). In order to verify the quality of our DCMARK approach we began to investigate a no complex problem characterized by a one-dimension partial differential equation.

### 8.4.1 A Case Study: 1-D Korteweg de Vries Equation

The Korteweg de Vries takes its name from Diederik Korteweg and Gustav de Vries who, in 1895, proposed a mathematical model which allowed to predict the waves behaviour on shallow water surfaces [23]. The solutions of this equation were self-reinforcing solitary waves named Solitons and had several interesting properties. Mainly, these solutions have permanent shape and are localized within a region and when they interact with other solitons they don't change their speed or shape (neither a signal amplification or signal fading) but they just have a phase shift [30, 31]. There are many research topics explained by the KdV equation, such as the already mentioned shallow-water waves [32], the ion-acoustic waves in plasma [33, 34], the wave propagation in nonlinear lattice [35], the non-linear transmission networks [36, 37] and the Fermi-Pasta-Ulam recurrence problem [38]. The main idea is modulating solitons and transmitting them on communication lines such as optical fibers.

The one-dimension (1-D) Korteweg de Vries differential equation is the following [35]:

$$\frac{\partial u(t, x)}{\partial t} = -6u(t, x)\frac{\partial u(t, x)}{\partial x} - \frac{\partial^3 u(t, x)}{\partial x^3} \quad (8.1)$$

where  $u(t, x)$  is the solitonic propagating wave. The progressive wave (called soliton) of the KdV equation has the following expression:

$$u(t, x) = -\frac{\nu}{2} \cdot \operatorname{sech}^2 \left[ \frac{\sqrt{\nu}}{2} (x - \nu t - x_0) \right] \quad (8.2)$$

where  $\nu$  is the wave velocity and  $x_0$  is the initial spatial constant. Furthermore, the KdV equation can be analytically solved by the inverse scattering transform [39, 40].

### 8.4.2 Discretization of KdV Equation

In order to implement the equation on DCMARK we had to discretize (8.1). Considering the second term in right-end side of (8.1) we can lay down:

$$u(t, x)\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} \frac{\partial [u(t, x)]^2}{\partial x} \quad (8.3)$$

hence, as in [32], the (8.1) becomes:

$$\frac{\partial u(t, x)}{\partial t} = -3 \frac{\partial [u(t, x)]^2}{\partial x} - \frac{\partial^3 u(t, x)}{\partial x^3} \quad (8.4)$$

We used for the numerical discretization of spatial derivative terms of (8.4), a space-centered finite difference method [41] and we divided the KdV equation in  $N$  single equations [42]:

$$\frac{\partial u_i}{\partial t} = \frac{1}{2\Delta x^3} [(u_{i-2} - u_{i+2}) + 2(u_{i+1} - u_{i-1})] + \frac{3}{2\Delta x} [u_{i-1}^2 - u_{i+1}^2] \quad (8.5)$$

where  $i = 0, \dots, N$  is the space iteration index and  $\Delta x$  is the space step of the discrete grid. For the time derivative term of (8.5), just for the first iteration, we used a forward time finite difference method (8.6) as in [31, 33] because there is no preceding value at the first step of numerical integration process. Hence, for the other iterations, we used a centered-time finite difference method (8.7). We set  $K_{i1} = 1/2\Delta x^3$ ,  $K_{i2} = 3/2\Delta x$  and  $K_1 = 1/\Delta x^3$ ,  $K_2 = 3/\Delta x$



$$u_i^{k+1} = u_i^k + \Delta t \left\{ K_{i1} \left[ \left( u_{i-2}^k - u_{i+2}^k \right) + 2 \left( u_{i+1}^k - u_{i-1}^k \right) \right] + K_{i2} \left( u_{i-1}^k{}^2 - u_{i+1}^k{}^2 \right) \right\} \quad (8.6)$$

$$u_i^{k+1} = u_i^{k-1} + \Delta t \left\{ K_1 \left[ \left( u_{i-2}^k - u_{i+2}^k \right) + 2 \left( u_{i+1}^k - u_{i-1}^k \right) \right] + K_2 \left[ u_{i-1}^k{}^2 - u_{i+1}^k{}^2 + u_i^k \left( u_{i-1}^k - u_{i-1}^k \right) \right] \right\} \quad (8.7)$$

where  $k = 0, \dots, M$  is the time iteration index,  $i = 0, \dots, N$  is the space iteration index and  $\Delta t$  is the integration time. Using this combined approach we have a stable loop propagation of a soliton through all cells for all time cycles. This kind of discretization is less accurate than other types, such as Runge-Kutta methods, but it is also the best technique in terms of implementation easiness and resources saving on embedded systems. The linchpin of the calculator idea is to consider every single  $u_i$  with  $i = 0, \dots, N$  a single solitonic state cell which calculates its future state value on the basis of state values of its first and second neighbors that is  $u_{(i\mp a)}$  with  $a = 1, 2$  as in [11, 43].

## 8.5 KdV Implementation on DCMARK

The implementation of KdV equation on DCMARK consists of two main implementation steps: a MCode step and a Cells Network step.

### 8.5.1 MCode Implementation Step

The MCode step consists of dividing (8.6) and (8.7) in single micro-instructions to be stored on RAM. We defined 14 arithmetical operations for (8.6), where ROp $x$  is the operation result which is stored on RAM:

- Opi1 :  $(u_{i-2} - u_{i+2}) \Rightarrow$  ROp1
- Opi2 :  $(u_{i+1} - u_{i-1}) \Rightarrow$  ROp2
- Opi3 :  $(ROp2 + ROp2) \Rightarrow$  ROp3
- Opi4 :  $(ROp1 + ROp3) \Rightarrow$  ROp4
- Opi5 :  $(K_1 * ROp4) \Rightarrow$  ROp5
- Opi6 :  $(u_{i-1} * u_{i-1}) \Rightarrow$  ROp6
- Opi7 :  $(u_{i+1} * u_{i+1}) \Rightarrow$  ROp7
- Opi8 :  $(ROp6 - ROp7) \Rightarrow$  ROp8
- Opi9 :  $(K_2 * ROp8) \Rightarrow$  ROp9
- Opi10 :  $(ROp5 + ROp9) \Rightarrow$  ROp10

- Opi11 :  $(\Delta t * \text{ROp10}) \Rightarrow \text{ROp11}$
- Opi12 :  $(u_i + \text{ROp11}) \Rightarrow \text{ROp12}$
- Opi13 :  $(u_i + \text{ZERO}) \Rightarrow (\text{updating } u_i^{k-1})$
- Opi14 :  $(\text{ROp12} + \text{ZERO}) \Rightarrow (\text{updating } u_i)$

For (8.7) we defined 17 arithmetical operations, but the first eight are the same as those for (8.6):

- Op9 :  $(u_{i-1} - u_{i+1}) \Rightarrow \text{ROp9}$
- Op10 :  $(u_i * \text{ROp9}) \Rightarrow \text{ROp10}$
- Op11 :  $(\text{ROp8} + \text{ROp10}) \Rightarrow \text{ROp11}$
- Op12 :  $(K_2 * \text{ROp11}) \Rightarrow \text{ROp12}$
- Op13 :  $(\text{ROp5} + \text{ROp12}) \Rightarrow \text{ROp13}$
- Op14 :  $(\Delta t * \text{ROp13}) \Rightarrow \text{ROp14}$
- Op15 :  $(u_i^{k-1} + \text{ROp14}) \Rightarrow \text{ROp15}$
- Op16 :  $(u_i + \text{ZERO}) \Rightarrow (\text{updating } u_i^{k-1})$
- Op17 :  $(\text{ROp15} + \text{ZERO}) \Rightarrow (\text{updating } u_i)$

The Opi13, Opi14 and Op16, Op17 have the task to update the Cell Status variables at the end of every iteration, that is the old value of  $u_i$  becomes  $u_i^{k-1}$  and the new value of  $u_i$  is updated. After the arithmetic operations definition we started to write all MCode copies according to a well-defined process. This process starts loading the Cell Status variable  $u_i^k$ , from RAM and storing it on I/O Register I to be available for other neighbor Cells and then storing on RAM the four neighbor Cell Status variables  $u_{i-2}^k, u_{i-1}^k, u_{i+1}^k, u_{i+2}^k$  stored in I/O Registers M2, M1, P1 and P2, respectively. After every iteration step, this process is re-executed. This loading/storing operations are conducted using the micro-instructions LDI, STM2, STM1, STP1 and STP2. In the RAM structure, shown in Fig. 8.5, we find four main parts: MicroCode part to store the 137 micro-instructions, Status Variables part to store the Cell Status variables, Constants part to store the constants defined in (8.6) and (8.7) and Operation Variables to store the partial operation results. We have also four free locations for possible modifications.

The MCode is formed by 137 micro-instructions but after the system start-up (after the first iteration) only 74 micro-instructions are executed in the computing loop. The content of Configuration File stored on Configuration ROM has the structure as shown in Fig. 8.5.

### 8.5.2 Cells Network Implementation Step

Cells Network step lies in connecting properly every Cell with its first and second neighbors according to Cell relationship shown in (8.6) and (8.7). In particular, using the approach in [11], we connected the Cells building a Ring network as in Fig. 8.6.

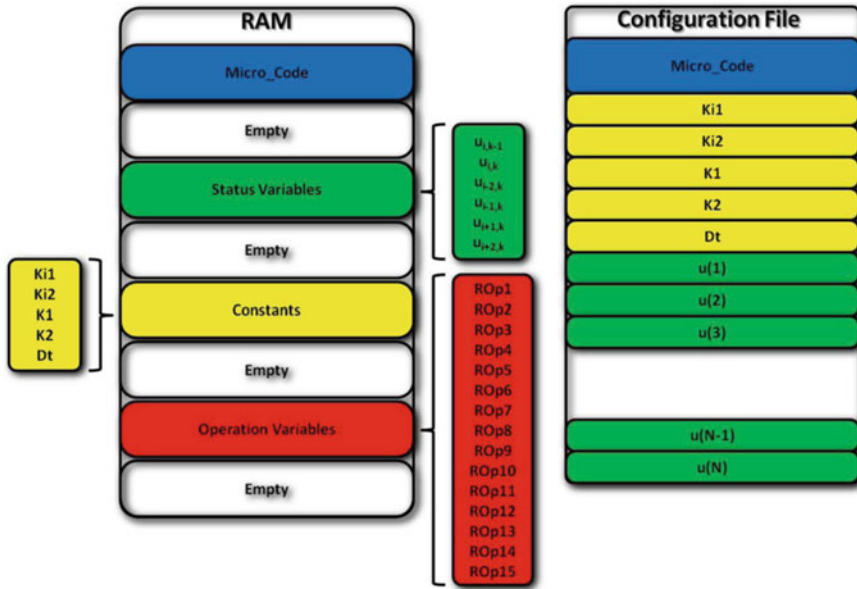


Fig. 8.5 RAM structure and ROM configuration file structure

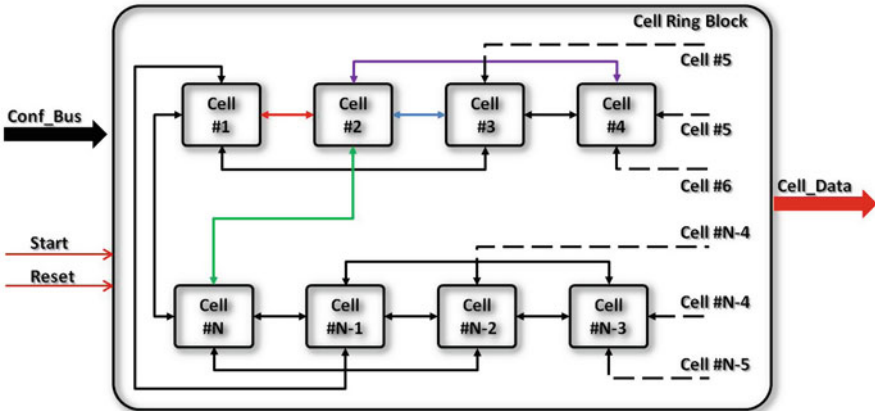


Fig. 8.6 Cell ring block diagram (for Cell #2 case: red line shows the link to first left neighbor, blue line the link to first right neighbor, green line the link to second left neighbor and violet line the link to second right neighbor)

### 8.5.3 DCMARK Performances and Used Resources

The preliminary version of a Single Cell processor and of DCMARK is implemented on FPGA Stratix IV GX. In Table 8.1 we find the resources used, without any kind of

**Table 8.1** Single cell used resources

FPGA resources	Used resources
ALMs	561
Combinatorial ALUTs	860
Total registers	800
Total block memory bits	10,240
DSP block 18-bit elements	4
DSP $36 \times 36$	1

design optimization. As regards the Single Cell computing time, every Cell executes a KdV iteration (integration step) producing a 32 bit output value in about  $3.77 \mu\text{s}$  with a 100 MHz system clock hence with a throughput of about 8.5 Mbit/s. But considering a Cell Ring of N Single Cells this throughput has to be multiplied by N. Using the FPGA Altera Stratix IV GX, we can implement up to 200 Single Cells on our DCMARK.

## 8.6 Analysis Settings and Results

We executed two kinds of analysis:

- a high-level test by means of a MatLab software simulator in order to verify the quality of equation discretization and to study the variation of parameters such as number of cells,  $\Delta x$ ,  $\Delta t$ , Initial Cell Status and number of iterations;
- a calculation test using DCMARK Calculator to verify the correspondence with the former results and to measure the new solution elaboration speed.

### 8.6.1 KdV Simulation Tests

The first parameters to tune are  $\Delta x$  and  $\Delta t$ . According to [31] these two parameters have to be related on the basis of (8.8), called Courant-Friedrichs-Lewy (CFL) condition, to have convergence.

$$v \cdot \Delta t / \Delta x \geq C \quad (8.8)$$

where  $v$  is the wave velocity by which the wave goes from  $x_i$  to  $x_{i+1}$  and  $C$  is a constant which depends on the equation. In a nutshell,  $\Delta t$  has to be smaller than  $\Delta x / v$ . Then we chose the number of Cells closed in the Ring Network and the Initial Cell Status. A hyperbolic secant squared function is chosen:

$$u_i = K \cdot \operatorname{sech}^2 x_i \text{ with } \min < x_i < \max, 0 < i < (\max - \min) / \Delta x,$$

$$\Delta x = x_{i+1} - x_i.$$

This function avoids divergence integration problems, thanks to its zero-tangent envelope for  $x \rightarrow \pm\infty$ . We conducted three types of simulations (10,000 iterations): time, space and time/space simulations, considering a ring-like network of 100 cells and  $\Delta x = 0.5$  mm,  $\Delta t = 0.01$  s. The results confirm both the stability of KdV equation numerical solving after many integration steps and the physical phenomenon emergence of soliton propagation.

### 8.6.2 Calculation Results

In our tests we deployed up to the maximum number of Single Cells implementable on DCMARK, that is 200 Single Cells. The Single Cell is still a prototype core and so it is not optimized for saving FPGA resources. Many Altera library IP cores, such as floating point adders and multipliers, RAMs, counters, etc, with several unused features, are instantiated on the Single Cell. Our idea for future developments is to design our own cores in order to significantly decrease the Single Cell FPGA requirements.

As previously said, using the Analysis GUI we monitored the analysis evolution. In the Configuration File, stored on the Configuration ROM, we set as Initial Cell Status the hyperbolic secant squared function and the calculation parameter values for a typical KdV analysis as in simulation tests. Figure 8.7 shows the LabWindows GUI image plotting a KdV calculation result obtained using the DCMARK Calculator according to the same parameters settings of the simulation test. Therefore, we noticed that the same result of the Matlab simulation is obtained. This result is also confirmed by a numerical comparison between Matlab and DCMARK data.

## 8.7 Performance Comparison

After the Test Phase (Simulation/Calculation) we underlined the differences between MatLab KdV simulations on PC and KdV calculation on FPGA using our DCMARK approach. As elements of comparison we chose two PCs with the following processors:

- Intel Core-i7 2630QM, 2 GHz clock speed, 4 Cores, 8 Threads, 64 bit Instruction Set, 6 MB Intel Smart Cache.
- Intel Pentium M 760, 2 GHz clock speed, 1 Core, 1 Thread, 32 bit Instruction Set, 2 MB L2 Cache.

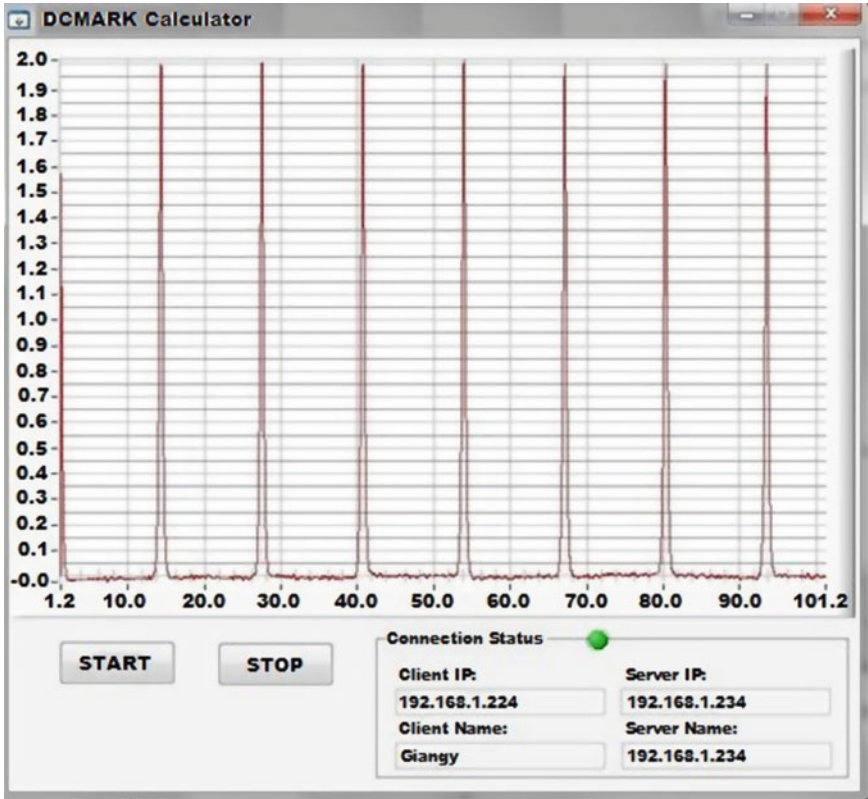
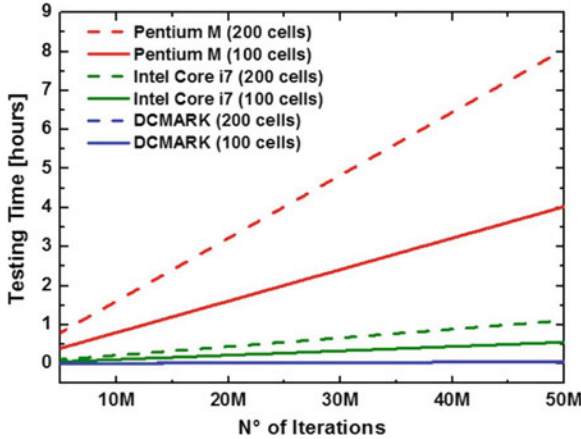


Fig. 8.7 LabWindows<sup>TM</sup> GUI with results of a KdV time calculation using a 100 cells DCMARK (10th cell output) with a  $2 * \text{sech}^2$  as initial status function

The parameters settings ( $\Delta x$ ,  $\Delta t$  and Initial Cell Status) are the same of simulation test. From calculation results, the elaboration time for DCMARK system is about 10 times shorter than Intel Core i7 PC and about 70 times shorter than Intel Pentium M PC already for the 100 cells problem.

Doubling the cell number, as expected, the performance gap increases. DCMARK performances are unrelated from the number of cells and so, as long as the FPGA resources are saturated, from the complexity of the investigated problem. In Fig. 8.8, we find linear fitted curves about testing time variation with respect to number of iterations for the two study cases: 100 and 200 cells system. It is again underlined the independence of DCMARK performances from the number of cells.



**Fig. 8.8** Comparison between PCs and DCMARK elaboration time increasing the number of iterations and the number of Cells

## 8.8 Conclusions and Future Developments

In this work we introduced an innovative kind of distributed computing architecture, called DCMARK, for investigating complex physical dynamical problems. DCMARK is the union of a FPGA-based extremely parallelized computing platform and a PC based user interface for setting and analyzing the results of calculations. The main features of this system are the total system reconfigurability for analysing different types of cell-based phenomena and an elaboration time independent from the complexity (in terms of number of cells) of the studied problem. This hardware calculation approach allows to exploit many concurrent processes executed at the same time, decreasing the elaboration time. Besides, using an FPGA device we exploited its intrinsic reconfigurability and flexibility. The results are promising since, for example, a 100 Cells DCMARK allows to execute KdV equation integration steps 10 times faster than a 4-core processor. The future development steps to increase the performances will be: the optimization of Single Cell in terms of used resources in order to tackle more and more difficult problems and the improvement of the GUI usability. Taking advantage of the reconfigurability, the DCMARK Calculator can be used in order to implement innovative learning techniques, as in [44] or in analogy to [45].

## References

1. Valasoulis, K., Fotiadis, D.I., Lagaris, I.E., Likas, A.: Solving differential equations with neural networks implementations on a DSP platform. In: Proceedings of 14th International Conference on Digital Signal Processing, Santorini, Greece, July 2002
2. Piuél, L., Martin, I., Tirado, F.: A special-purpose parallel computer for solving partial differential equations. In: Proceedings of 16th Euromicro Workshop on Parallel and Distributed Processing, (PDP'98), pp. 21–23. Madrid, Spain, 21–23 January 1998
3. Osana, Y. et al.: ReCSiP: an FPGA-based general-purpose biochemical simulator. *Electron. Commun. Jpn. Part 2* **90**(7), 1–10 (2007)
4. Huang, C., Vahid, F., Givargis, T.: A custom FPGA processor for physical model ordinary differential equation solving. *IEEE Embed. Syst. Lett.* **3**(4), 113–116 (2011)
5. He, K., Jiang, Y., Dong, S.: A hybrid parallel framework for cellular potts model simulations. In: Proceedings of 15th International Conference on Parallel and Distributed Systems, Shenzhen, Guangdong, China, 11 December 2009
6. Hertz, J., Palmer, R.G., Krogh, A.S.: *Introduction to the Theory of Neural Computation*. Perseus Books, Reading (1990). (ISBN 0-201-51560-1)
7. Chua, L.O., Yang, L.: Cellular neural networks: theory. *IEEE Trans. Circuits Syst.* **35**, 1257–1272 (1998)
8. Chua, L.O., Yang, L.: Cellular neural networks: applications. *IEEE Trans. Circuits Syst.* **35**, 1273–1290 (1998)
9. Roska, T., Chua, L.O.: The CNN universal machine: an analogic array computer. *IEEE Trans. Circuits Syst. II* **40**(3), 163–173 (1993)
10. Arena, P., Fortuna, L., Rizzo, A., Xibilia, M.G.: Extending the CNN paradigm to approximate chaotic systems with multivariable nonlinearities. In: ISCAS 2000, IEEE International Symposium on Circuits and Systems, Geneva, Switzerland, 28–31 May 2000
11. Fortuna, L., Rizzo, A., Xibilia, M.G.: Modeling complex dynamics via extended PWL-based CNNs. *Int. J. Bifurcat. Chaos* **13**(11), 3273–3286 (2003)
12. Cheung, O.Y.H., Leong, P.H.W., Tsang, E.K.C., Shi, B.E.: A scalable FPGA implementation of cellular neural networks for gabor-type filtering. In: Proceedings of International Joint Conference on Neural networks, Vancouver, BC, Canada, 16–21 July 2006
13. Magazzu, G., Borgese, G., Costantino, N., Fanucci, L., Incandela, J., Saponara, S.: Design exploration and verification platform, based on high-level modeling and FPGA prototyping, for fast and flexible digital communication in physics experiments. *J. Instrum.* **8**(2), P02021 (2013)
14. Soos, B.G., Rak, A., Veres, J., Cserey, G.: GPU powered CNN simulator (SIMCNN) with graphical flow based programmability. In: Proceedings of 11th International Workshop on Cellular Neural Networks and Their Applications (CNNA'08), pp. 14–16. Santiago de Compostela, Spain, 14–16 July 2008
15. Dolan, R., DeSouza, G.: GPU-based simulation of cellular neural networks for image processing. In: Proceedings of International Joint Conference on Neural Networks (IJCNN'09), Atlanta, Georgia, USA, 14–19 June 2009
16. Griebel, M., Zaspel, P.: A multi-GPU accelerated solver for the three-dimensional two-phase incompressible Navier-Stokes equations. *Comput. Sci.-Res. Dev.* **25**(1–2), 65–73 (2010)
17. Ho, T.Y., Lam, P.M., Leung, C.S.: Parallelization of cellular neural networks on GPU. *Pattern Recognit.* **41**(18), 2684–2692 (2008)
18. Nagy, Z., Szolgay, P.: Configurable multilayer CNN-UM emulator on FPGA. *IEEE Trans. Circuits Syst. I: Fundam. Theor. Appl.* **50**(6), 774–778 (2003)
19. Nagy, Z., Vrshazi, Z., Szolgay, P.: Emulated digital CNN-UM solution of partial differential equations. *Int. J. Circuit Theor. Appl.* **34**, 445–470 (2006)
20. Vrshazi, Z., Kiss, A., Nagy, Z., Szolgay, P.: FPGA based emulated-digital CNN-UM implementation with GAPU. In: Proceedings of 11th International Workshop on Cellular Neural Networks and their Applications, pp. 14–16. Santiago de Compostela, Spain, July 2008



21. Kocsardi, S., Nagy, Z., Csik, A., Szolgay, P.: Two-dimension compressible flow simulation on emulated digital CNN-UM. In: Proceedings of 11th International Workshop on Cellular Neural Networks and their Applications (CNNA'08), pp. 14–16. Santiago de Compostela, Spain, July 2008
22. Borgese, G., Pace, C., Pantano, P., Bilotta, E.: FPGA-based distributed computing microarchitecture for complex dynamical physics investigation. *Trans. Neural Netw. Learn. Syst. (TNNLS)* **24**(9), 1390–1399 (2013)
23. Korteweg, D.J., de Vries, G.: On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. *Philos. Mag.* **39**, 422–443 (1895)
24. Bilotta, E., Stranges, F., Pantano, P.: A gallery of Chua attractors: part III. *Int. J. Bifurcat. Chaos* **17**(3), 657–734 (2007)
25. Bilotta, E., Di Blasi, G., Stranges, F., Pantano, P.: A gallery of Chua attractors. Part VI. *Int. J. Bifurcat. Chaos* **17**(6), 1801–1910 (2007)
26. Bilotta, E., Pantano, P.: Emergent patterning phenomena in 2D cellular automata. *Artif. Life* **11**(3), 339–362 (2005)
27. Bilotta, E., Pantano, P.: Structural and functional growth in self-reproducing cellular automata. *Complexity* **11**(6), 12–29 (2006)
28. Bilotta, E., Pantano, P.: The language of chaos. *Int. J. Bifurcat. Chaos* **16**(3), 523–557 (2006)
29. Hirsch, M.W., Smale, S., Devaney, R.: *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press, New York (2003). (ISBN 0-12-349703-5)
30. Drazin, P.G., Johnson, R.S.: *Solitons: An Introduction*, 2nd edn. Cambridge University Press, Cambridge (1989). (ISBN 0-521-33655-4)
31. Zabusky, N.J., Kruskal, M.D.: Interaction of solitons in a collisionless plasma and the recurrence of initial states. *Phys. Rev. Lett.* **15**(6), 240–242 (1965)
32. Hereman, W.: Shallow water waves and solitary waves. In: *Encyclopedia of Complexity and Systems Science*, pp. 8112–25. Springer, New York (2009)
33. Washimi, H., Taniuti, T.: Propagation of ion acoustic solitary waves of small amplitude. *Phys. Rev. Lett.* **17**, 996–998 (1966)
34. Giamb, S., Pantano, P.: Three-dimensional ion-acoustic waves in a collisionless plasma. *Lett. al Nuovo Cimento* **34**, 380–384 (1982)
35. Wadati, M.: Wave propagation in non linear lattice. *J. Phys. Soc. Jpn.* **38**, 673–680 (1975)
36. Fukushima, K., Wadati, M., Kotera, T., Sawada, K., Narahara, Y.: Experimental and theoretical study of the recurrence phenomena in nonlinear transmission line. *J. Phys. Soc. Jpn.* **48**, 1029–1035 (1980)
37. Pantano, P.: Inhomogeneous dispersive and dissipative nonlinear transmission lines and solitons. *Lett. al Nuovo Cimento* **8**, 209–214 (1983)
38. Gallavotti, G. (ed.) *Fermi-Pasta-Ulam problem: A status report*, A Status Report, Lecture Notes in Physics, vol. 728 (2008)
39. Gardner, C.S., Greene, J.M., Kruskal, M.D., Miura, R.M.: Method for solving the Korteweg de Vries equation. *Phys. Rev. Lett.* **19**, 1095–1097 (1967)
40. Gardner, C.S., Greene, J.M., Kruskal, M.D., Miura, R.M.: The Korteweg de Vries equation and generalizations VI. Methods for exact solution. *Commun. Pure Appl. Math.* **27**, 97–133 (1974)
41. Vliegthart, A.C.: On finite-difference methods for the Korteweg-de Vries equation. *J. Eng. Math.* **5**(2), 137–155 (1971)
42. Fortuna, L., Frasca, M., Rizzo, A.: Generating solitons in lattices of nonlinear circuits. In: *ISCAS 2001, The 2001 IEEE International Symposium on Circuits and Systems*, pp. 680–683. Sydney, NSW, Australia, May 2 2001
43. Remoissenet, M.: *Waves Called Solitons*. Springer, Berlin (1996)
44. Luitel, B., Venayagamoorthy, G.K.: Decentralized asynchronous learning in cellular neural networks. *IEEE Trans. Neural Netw. Learn. Syst. (TNNLS)* **23**(11), 1755–1766 (2012)
45. Papadonikolakis, M., Bouganis, C.: Novel cascade FPGA accelerator for support vector machines classification. *IEEE Trans. Neural Netw. Learn. Syst. (TNNLS)* **23**(7), 1040–1052 (2012)

# Chapter 9

## A Multi Harvester with Hydrogen Fuel Cell for Outdoor Applications

**Davide Brunelli, Michele Magno, Danilo Porcarelli  
and Luca Benini**

**Abstract** Energy availability and long term operation are key challenges for wireless sensor networks and for all the applications where the devices are battery operated. For this reason energy harvesting is becoming very important for powering ubiquitously deployed sensor networks and mobile electronics. One of most important goal for the next generation of power supply units for standalone embedded systems is to power nearly perpetually the devices when the scavenger is exposed to reasonable environmental energy conditions. However, due to the unpredictable nature of the environmental sources, prolonged lacks of energy intake usually happen. The last frontiers of perpetual operating systems is combining different energy harvesters in a single unit and using green energy supply with high energy density as micro hydrogen fuel cells. In this paper we introduce a Smart Power Unit (SPU) for embedded system which incorporates energy harvesters from sun and wind and uses hydrogen fuel cell as alternative energy storage. The power unit can work as a long-term battery or providing serial communication to exchange power information and to perform power management. In fact the core of the SPU is an ultra low power micro controller which is in charge to do the power activities such as Maximum Power Point Tracking for the harvesters, fuel cell activation, energy prediction, adaptive power management on board, battery monitoring and communications with powered systems. Experimental results and simulations shows the high efficiency (up to 90 %)

---

D. Brunelli (✉)

University of Trento, via Sommarive 14, 38123 Trento, Italy  
e-mail: [davide.brunelli@unitn.it](mailto:davide.brunelli@unitn.it)

M. Magno · D. Porcarelli · L. Benini

University of Bologna, viale Risorgimento 2, 40136 Bologna, Italy  
e-mail: [michele.magno@unibo.it](mailto:michele.magno@unibo.it)

D. Porcarelli

e-mail: [danilo.porcarelli@unibo.it](mailto:danilo.porcarelli@unibo.it)

L. Benini

e-mail: [luca.benini@unibo.it](mailto:luca.benini@unibo.it)

of the power conversion subsystem. Finally a real deployment in a structural health monitoring site in Switzerland shows as the energy neutral condition is achieved on field.

## 9.1 Introduction

Wireless sensor networks (WSNs) are used today in a wide variety of application areas, from environmental monitoring [1–3] to security management [4], and from medical applications to smart homes [5]. WSNs consist of wireless sensor nodes which are supplied by batteries, which have usually small form factor and limited life time in contrast with the major requirements of most of the applications which aim at perpetual operation. The need of energy availability in wireless sensor networks (WSN) is the most important issue to address to confirm the effectiveness of this technology in the widest range of applications. Energy harvesting technology makes possible batteries recharging using small amount of energy collected from environmental sources, saving the maintenance and battery replacement costs. In principle, all energy sources should be exploited, see [6–9]; among the others, the solar one is generally the most effective in outdoor applications for the high power density provided and exploitable through solar cells. In last few years also micro fuel cells are gaining interest in the WSN research community, to increase the energy availability due to high power density of fuel cell technology [10]. Therefore architectures based on energy harvesters which convert and collect energy from the environment combined with fuel cell hybrid systems are very attractive solutions, because they guarantee a superior level of reliability also in long absence of energy intake and improve both power and energy density.

Concerning energy harvesting, the major challenges are mainly focused on tracking the maximum power point with the minimum cost in terms of power consumption and achieving a positive balance between the energy harvested and energy consumed by the load. While in micro fuel cells the challenge is designing the control of the fuel flow to activate or deactivate the fuel combustion only when is needed. Moreover the small size solution for power supply is a challenging constraint since WSN applications often require non-invasive devices.

In this paper we present a Smart Power Unit (SPU) which is a circuit designed to provide power supply from multiple sources, with the capability of the  $H_2$  fuel cells. The SPU is designed to efficiently scavenge energy from two different environmental outdoor sources: airflow from wind and sunlight from the sun. To allow the power unit to collect efficiently energy from both sources, each harvester firstly stores energy in a dedicated local super capacitor and then recharges a battery (i.e. Li-Ion Battery) as shown in Figs. 9.1 and 9.2.

The advantage of this approach is twofold:

- to avoid the possibility to waste important energy when both sources are collecting energy and to increase the overall available energy of the power unit;

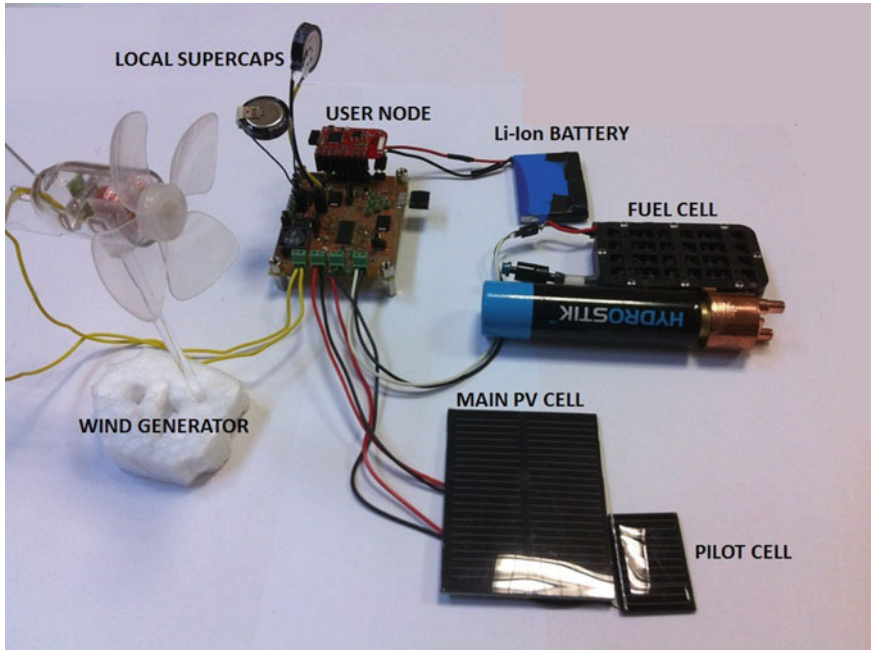


Fig. 9.1 Smart power unit hardware setup

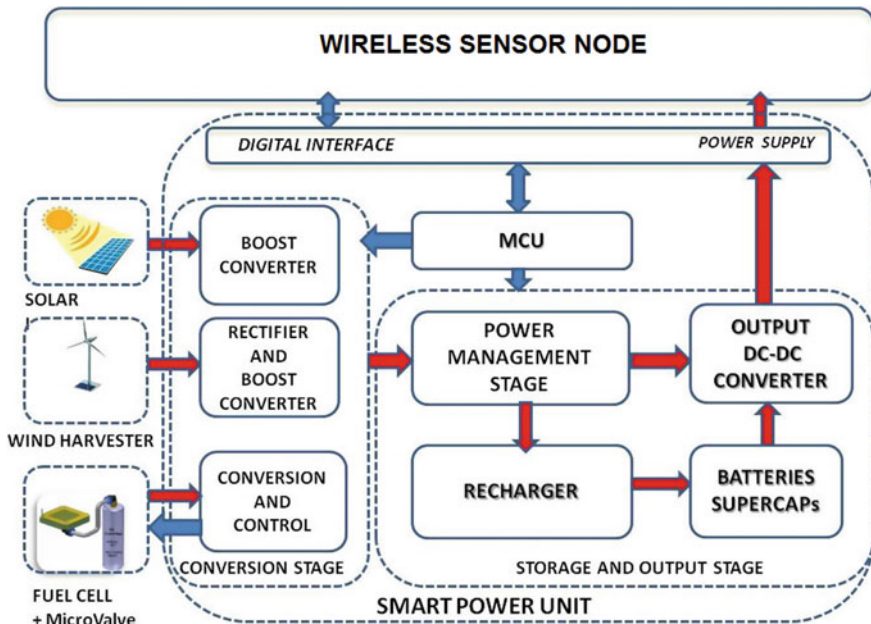


Fig. 9.2 Smart power unit architecture

- Furthermore the power unit incorporates and controls a hydrogen micro fuel cell to be prompt in urgent events or critical situation when the renewable sources are not enough and the battery is running out of energy.

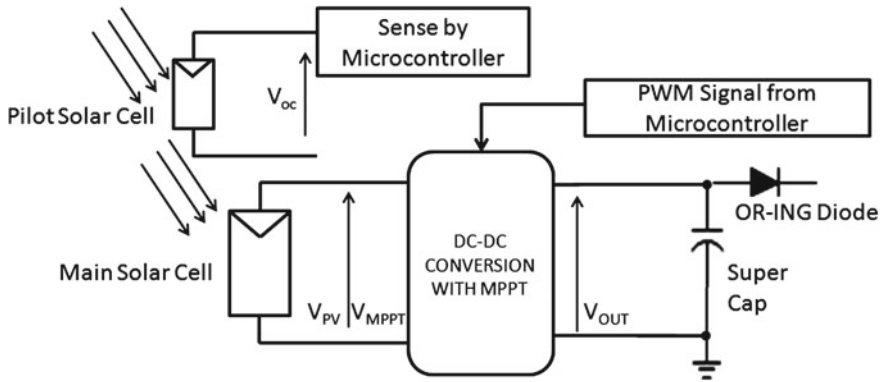
The SPU has an on board microcontroller to provide a dedicated ultra low-power management which extracts efficiently the energy from two environmental sources and activates or deactivates the electrovalve to control the hydrogen flow. Finally the power unit can be configured and used a smart battery.

## 9.2 Energy Source and System Design

The architecture and the developed prototype of the proposed multi-source energy harvester SPU is shown in Fig. 9.2. The power unit is *smart* because it permits to monitor runtime the status of the harvesters, batteries, fuel cell and to manage parameters. In fact due to presence of a MSP430 from Texas Instruments, every power decision (activation of resources, maximum power point tracking, fuel cell hydrogen flow control, communication, duty cycling, etc.) is triggered directly on-board. Moreover one of most important feature of the power unit is the interface to the powered system that provides the energy supply, GPIOs, SPI and I2C electrical interfaces. As a result, in addition to being supplied with the required energy, the primary microcontroller of the node can communicate with the power unit in order to increase the battery life performing power management policies such as changing the duty cycle parameters or sleep/wake up techniques or selectively choosing the optimal harvesting unit.

It is possible to change the operating frequency used by internal DC-DC converters and chargers in order to be adaptable to different conditions and to achieve the maximum conversion efficiency. In fact it has been designed to provide advanced features in terms of power management and energy efficiency directly on board in a flexible way. In particular it is possible to monitor the current state of the harvesters, batteries and micro fuel-cells and change the power polices run time. Furthermore, it is possible to change the operating frequency used by internal DC/DC converters and chargers and microcontroller to reduce the power consumption.

Many kinds of energy sources are available to the designers and the incoming power depends on the category of the ambient source. Focusing on the energy transducers we can distinguish between  $\mu\text{W}$  generators, such as piezoelectric or thermal and milliwatt generators which include air-flow and solar. So the topology of the transducers is fundamental to determine the class of the harvester, its efficiency and the design methodology.



**Fig. 9.3** Solar conversion stage based on boost converter topology. The pilot cell is used to provide open circuit voltage in FOC algorithm

### 9.2.1 Solar Path

The solar harvester subsystem consists of a solar energy harvesting unit, a boost converter driven by pulse-width modulation (PWM) signal controlled by on board microcontroller, and a supercapacitor (Fig. 9.3). The PWM is used to achieve the maximum power point tracking. In fact the power unit hosts both the main solar panel, which will give the energy to recharge the battery, and a smaller PV cell, made with the same technology of the main PV, and it is used to sense the open circuit voltage and to perform the MPPT algorithm as explained in [11, 12]. The driven boost circuit forces the main cell to operate near the maximum power point by means of the tracking algorithm. In the proposed solution, the SPU use the Fractional Open Circuit algorithm (FOC) which exploits the nearly linear proportionality between the MPPT and the open circuit voltage. The solar cell was tested under different light intensity conditions and the results, plotted in Fig. 9.4, are compared with curves obtained from our simulation model. In addition, the plot highlights how the maximum power point varies with the light intensity. The shift from the MPP results in a significant variation in solar cell power output and it justifies the implementation of a MPPT routine to maximize the energy transfer.

### 9.2.2 Air-Flow Path

The wind harvester path is depicted in Fig. 9.5, which consists of a hybrid full wave ac/dc converter, a COTS buck-boost converter operating in Fixed Frequency-Discontinuous Current Mode (FF-DCM), supercap for local storage and the OR-ring diode for the Li-Ion battery recharger. It is an enhancement of the wind energy harvester presented in [13, 14].

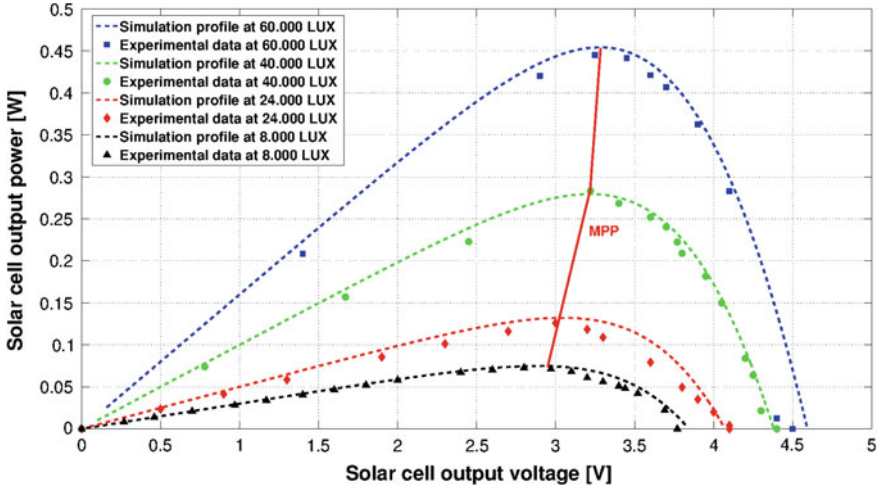


Fig. 9.4 Comparison between power curves obtained from SPICE model (dashed curves) and measured output power of the main solar cell. The solid curve also shows the MPP voltage curve

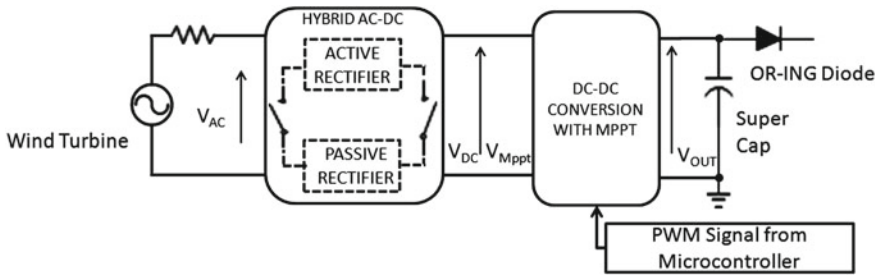


Fig. 9.5 Air-flow conversion stage based on hybrid rectifier and boost converter

The curves plotted in Fig. 9.6 represent the power delivered to the output of the passive rectifier. We notice that the maximum power point is located in a impedance range of 500–700  $\Omega$  at all the three air-flow speed used in the tests. Within that range, on the other hand, the shift from the MPP results in negligible variations of the output power. Remarking this, an impedance matching circuit is more appropriate than a maximum power point tracker.

### 9.2.3 Fuel Cell Path

The Smart Power Unit provides a fuel cell section as a reservoir energy source. When the energy storage elements are going to deeply discharge and the ambient energy intake is scarce, the power manager activates the fuel cell interface to rapidly recharge

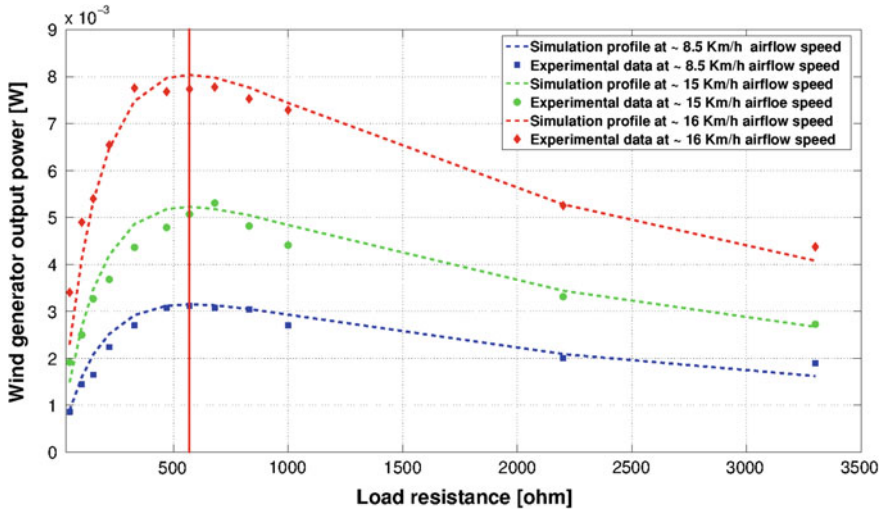


Fig. 9.6 Power delivered by the wind generator to the rectifier output. Markers shows power curves obtained from experimental measurement. Dashed line represent power profiles obtained from PSPICE simulations. Vertical line shows the selected operating point

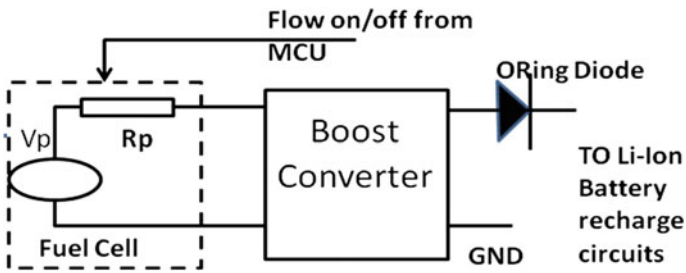


Fig. 9.7 Fuel cell path subsystem

the lithium ion battery avoiding the system shutdown. As Fig. 9.7 illustrates, the fuel cell subsystem consists of a DC–DC boost converter with over voltage protection, in addition to an ORing diode, to guarantee the minimum voltage (4.5 V) needed for the recharger. In fact a single cell under normal operation typically produces between 0.9 and 1.4 V, but several cells can be connected in series, arranging a stack that can supply 5 V or more. However to recharge a Li-Ion battery typically used for WSN is enough that the voltage of FC is stable at 4.3 V. In the proposed approach, a boost converter has been inserted to permit the recharge of the battery and to overcome some stability issues of the FC. In fact on one hand, adding a DC–DC introduces a conversion power loss; on the other hand, it gives a stable output needed to recharge the battery and can reduce the size of whole system using only a single cell FC with 1.1 V output.



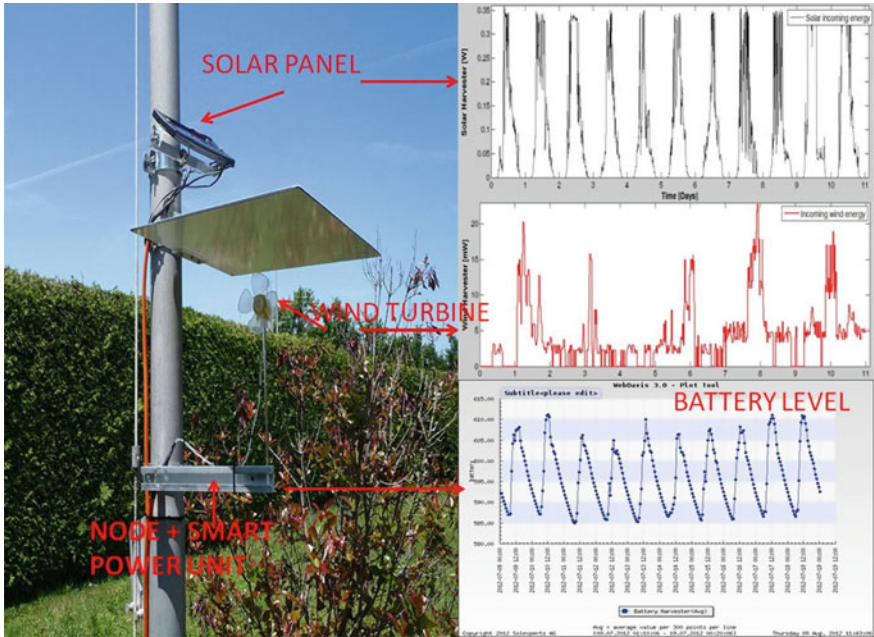


Fig. 9.8 Test deployment of the smart power unit in a wireless sensors monitoring

### 9.3 Experimental Results

The prototyped power unit has been used in a real deployment in Zurich (Switzerland) on the test site of *Solexperts AG* [15] which supplies real nodes for structural health monitoring. The power unit has supplied the company’s node for last 6 months, it is still working and the perpetual operation is achieved at the time the paper is written. Figure 9.8 shows the real deployment with the environmental scavenger. The power unit is inside of the grey box together with the wireless node which include a temperature, pressure, humidity sensors and a GSM transmitter to send the data to the base station. On the top there is a 112 cm<sup>2</sup> PV module which provides max 450 mW, and in the middle, the wind turbine is used for air-flow energy harvester with a max 10 mW output. It is four-bladed horizontal-axis plastic turbine, with a diameter of 6.3 cm. Efficiencies was measured, with MPPT active and recharging both a super capacitor and Li-Ion batteries. A smaller PV panel, shown in Fig. 9.1 close to the wind generator, acts as a pilot cell, used by the power unit to enable execution of the MPPT algorithm. Figure 9.1 shows also the cylindrical tank used in the laboratory which can contain 10 l of Hydrogen and 12 Wh energy capability.

## 9.4 Conclusion

The design, implementation and characterization of a Smart Power Unit prototype with hybrid energy harvesting capability and electrochemical fuel cell integration have been presented. The SPU with its sources, coupled with intelligence and interoperability, represents a significant improvement over the current state-of-the-art. The system was designed for ultra-low power and high efficiency energy conversion, with less than 1 mW in sleep mode to achieve continuous operation using only one 800 mAh battery, one fuel cell, solar and wind harvesters. As a real deployment has demonstrated, when it is interfaced with the appropriate wireless sensor network hardware infrastructure, it is suitable for long-term perpetual wireless structural health monitoring. Experimental results demonstrate that, even with an extra MCU to provide additional novel features, overall efficiency is still comparable with state-of-the-art of harvesting solutions, giving very high energy conversion efficiency up to 86 %, and a low quiescent current of only 5  $\mu$ A. Finally, future work will investigate existing power management policies [16] and compression techniques [17] to enhance and improve the performance.

**Acknowledgments** The authors would like to thank the FP7 GENESI project (Green sEnsor NEtworks for Structural monitoring) funded grant number 257916.

## References

1. Rossi, M., Brunelli, D.: Ultra low power wireless gas sensor network for environmental monitoring applications. In: 2012 IEEE Workshop on Environmental Energy and Structural Monitoring Systems (EESMS), pp. 75–81 (2012)
2. Rossi, M., Brunelli, D.: Analyzing the transient response of *mox* gas sensors to improve the lifetime of distributed sensing systems. In: 2013 5th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), pp. 211–216 (2013)
3. Jelicic, V., Magno, M., Brunelli, D., Paci, G., Benini, L.: A context-adaptive multimodal wireless sensor network for energy-efficient gas monitoring. *IEEE Sens. J.* **13**(1), 328–338 (2013)
4. Magno, M., Tombari, F., Brunelli, D., Di Stefano, L., Benini, L.: Multimodal video analysis on self-powered resource-limited wireless smart camera. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **3**(2), 223–235 (2013)
5. Porcarelli, D., Balsamo, D., Brunelli, D., Paci, G.: Perpetual and low-cost power meter for monitoring residential and industrial appliances. In: Design, Automation Test in Europe Conference Exhibition (DATE), 2013, pp. 1155–1160 (2013)
6. Weimer, M.A., Paing, T.S., Zane, R.A.: Remote area wind energy harvesting for low-power autonomous sensors. In: Proceedings of 37th IEEE Power, Electronics, pp. 1–5, 18–22 Jun 2006
7. Park, C., Chou, P.H.: Power utility maximization for multiple-supply systems by a load matching switch. In: Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED), pp. 168–173. Newport Beach, CA, 9–11 August 2004
8. Jiang, X., Polastre, J., Culler, D.E.: Perpetual environmentally powered sensor networks. In: Proceedings of 4th International Symposium on IPSN, pp. 463–468, 25–27 April 2005

9. Raghunathan, V., Kansal, A., Hsu, J., Friedman, J., Srivastava, M.B.: Design considerations for solar energy harvesting wireless embedded systems. In: Proceedings of IPSN, pp. 457–462, 25–27 April 2005
10. Magno, M., Porcarelli, D., Benini, L., Brunelli, D.: A power-aware multi harvester power unit with hydrogen fuel cell for embedded systems in outdoor applications. In: 2013 International Green Computing Conference (IGCC), pp. 1–6 (2013)
11. Dondi, D., Bertacchini, A., Larcher, L., Pavan, P., Brunelli, D., Benini, L.: A solar energy harvesting circuit for low power applications. In: IEEE International Conference on Sustainable Energy Technologies, 2008, ICSET 2008, pp. 945–949 (2008)
12. Weddell, A.S., Magno, M., Merrett, G.V., Brunelli, D., Al-Hashimi, B.M., Benini, L.: A survey of multi-source energy harvesting systems. In: Design, Automation Test in Europe Conference Exhibition (DATE), 2013, pp. 905–908 (2013)
13. Porcarelli, D., Brunelli, D., Magno, M., Benini, L.: A multi-harvester architecture with hybrid storage devices and smart capabilities for low power systems. In: 2012 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), pp. 946–951 (2012)
14. Magno, M., Marinkovic, S., Brunelli, D., Popovici, E., O’Flynn, B., Benini, L.: Smart power unit with ultra low power radio trigger capabilities for wireless sensor networks. In: Design, Automation Test in Europe Conference Exhibition (DATE), 2012, pp. 75–80 (2012)
15. Solexpert AG. Swiss precision Geomonitoring <http://www.solexperts.com>
16. Moser, C., Brunelli, D., Thiele, L., Benini, L.: Real-time scheduling with regenerative energy. In: 18th Euromicro Conference on Real-Time Systems, 2006, ECRTS ’06, pp. 261–270. Washington, DC, USA (2006)
17. Caione, C., Brunelli, D., Benini, L.: Distributed compressive sampling for lifetime optimization in dense wireless sensor networks. IEEE Trans. Industr. Inf. **8**(1), 30–40 (2012)

# Chapter 10

## A Dosimetric Device Based on CMOS Image Sensor for Interventional Radiology

E. Conti, D. Magalotti, P. Placidi, L. Bissi, M. Paolucci, D. Passeri,  
A. Scorzoni and L. Servoli

**Abstract** Interventional radiologists and staff members, during all their professional activities, are frequently exposed to protracted and fractionated low doses of ionizing radiation. The authors present a novel approach to perform on line monitoring of the staff during their interventions by using a device based on an Active Pixel Sensor (APS). The performance of the sensor as an X-ray radiation detector has been evaluated with a proper experimental set-up: the number of photons and the generated charge have been assessed as dosimetric observables. The correlation of these observables with the dose measured by the passive dosimeters has been analyzed: a good linearity has been demonstrated and the response difference between pulsed and continuous operational modes is reduced to less than 10 %, marking a distinct improvement with respect to commercial Active Personal Dosimeters.

### 10.1 Introduction

Interventional Radiology (IRad) techniques embrace minimally invasive diagnostic and therapeutic procedures, with the guidance provided by radiological devices. There are numerous radiation protection issues in such a field, both for patients and operators, because of the exposure to ionizing radiation [14] which can cause detrimental effects such as hands depilation or cancer in the worst case [15].

---

E. Conti (✉) · P. Placidi · D. Passeri · A. Scorzoni  
University of Perugia, Perugia, Italy  
e-mail: elia.conti@studenti.unipg.it

D. Magalotti  
University of Modena and Reggio Emilia, Modena, Italy

E. Conti · P. Placidi · D. Passeri · A. Scorzoni · D. Magalotti · L. Bissi · L. Servoli  
Istituto Nazionale di Fisica Nucleare (INFN), Rome, Italy

M. Paolucci  
Servizio di Fisica Sanitaria, AUSL Umbria 2, Perugia, Italy

**Table 10.1** Features of commercial active personal dosimeters

Device name	Energy range	Dose rate range
Unfors EDD-30 [13]	14–120 keV ( $\pm 10\%$ )	0.03 mSv/h–2 Sv/h
Thermoscientific EPD Mk2+ [7]	15–10,000 keV ( $\pm 20\%$ )	0 $\mu$ Sv/h–4 Sv/h
Dosilab EDM III [4]	20–6,000 keV	0.5 $\mu$ Sv/h–1 Sv/h
RaySafe i2 [11]	33–101 keV	40 $\mu$ Sv/h–300 mSv/h

We focus our attention on the radiation protection of operators, who absorb ionizing dose from scattered radiation during procedures: international guidelines have defined limits for equivalent dose and equivalent effective dose [2] and this restricts the number of procedures that the radiological staff can undertake. The interest in improving radiation protection is thus strong, taking also into account the aim of developing more complex procedures that can involve higher doses and exposure times [6].

During IRad procedures, dose monitoring of operators is carried out using certified passive dosimeters (e.g. ThermoLuminescence Dosimeters, TLDs) but their major drawback is their inability to provide a real time dose measurement; for this reason a wide range of semiconductor-based active personal dosimeters is commercially available. Their response to the X-ray fields commonly used in IRad (low energies and pulsed fields) is however not satisfactory: it has been reported that the response of most Active Personal Dosimeters decreases as the equivalent dose rate increases ( $>2$  Sv/h) and decreases from 10 to 40 % when pulse frequency increases from 1 to 20 pulse per second (pps) [3]. It is also to be noted that the energy range of most active personal dosimeters has a lower bound which is greater than that of passive dosimeters (Table 10.1).

We are currently assessing the capability of exploiting CMOS Active Pixel Sensors (APS) for dosimetry of operators in IRad. The aim is to develop a portable device capable of measuring in real time the accumulated dose, with the following requirements: (i) on line monitoring of staff operation, with the production of an alarm when the dose exceeds a warning level; (ii) off line storage of dose measurements in order to correlate them with the specific activities of the staff during procedures. In this way it is possible to plan and subsequently optimize the number and type of procedures that interventionists can undertake.

A portable dosimetric device should meet a series of requirements, both technical and related to the operators' workplace: first of all the dose measurement should have an accuracy of few percent in the energy range of interest; then the device should be battery powered and feature wireless data transmission, because of practical issues connected to the use of wired systems [12]. Furthermore, if it is needed to monitor the dose on hands and eye lens, for example, the dosimeter should be worn on wristbands or headbands and thus it should be lightweight, with a small form factor.

In this work, set in the framework of the Italian *RAPID* INFN project, we focus on a commercial device we have investigated: its performance as a radiation sensor has been evaluated with respect to TLD measurements, calibrated in terms of personal

equivalent dose, by using a dedicated experimental setup. The paper is organized as follows: Sect. 10.2 describes the main features of the sensor which is the heart of the system we are currently designing; Sect. 10.3 presents and discusses the experimental results of the sensor evaluation. Finally, Sect. 10.4 presents the conclusions.

## 10.2 The sensor

The sensor under evaluation is a standard VGA imager with a pixel size of  $5.6 \times 5.6 \mu\text{m}^2$ , optimized for an output frame rate of 30 fps. Pixel gain and integration time are programmable (from 1 to 16 and from 41  $\mu\text{s}$  to 267 ms, respectively). The energy range of the readout chain for each pixel goes from 2 to 150 keV deposited energy, even with a variable efficiency as a function of photon energy. The package form factor is  $11.43 \times 11.43 \text{ mm}^2$ , suitable for future insertion in a wearable dosimeter we aim to develop. This device will feature the following building blocks (Fig. 10.1): (i) the image sensor; (ii) a digital signal processing unit to retrieve dose information from sensor output data; (iii) a control unit; (iv) a wireless interface to transmit data to a remote workspace; (v) a Graphical User Interface (GUI) to manage received data.

The sensor is the heart of the system we are planning to design and therefore its evaluation is mandatory [10]. In this work we focus on the performance characterization of the commercial sensor above described as a radiation sensor. To this purpose we have used the experimental set-up described in Sect. 10.3. It should be underlined that so far we collected all the acquired data for remote post processing analysis. Because each pixel of the matrix is represented with 10 bits, a VGA frame carries more than 300 kbytes of information. Therefore in the real time and low power wearable system we aim to design, a data reduction strategy will be mandatory; a relevant data reduction to be performed by using low power circuits will be suitable to meet low power wireless protocol requirements [8].

## 10.3 Results

We have investigated the sensor response to the X-ray radiation scattered by a phantom during an IRad procedure. When a single X-ray photon interacts within the sensitive volume of the sensor, it generates a cloud of electron-hole pairs that are collected by more than one pixel (Fig. 10.2); the amount of the charge collection depends on the charge collection efficiency profile of the sensor [9]. Hence we have used two system observables, the number of photons detected and the sum of the reconstructed photon signals in a frame, to study their capability to serve as

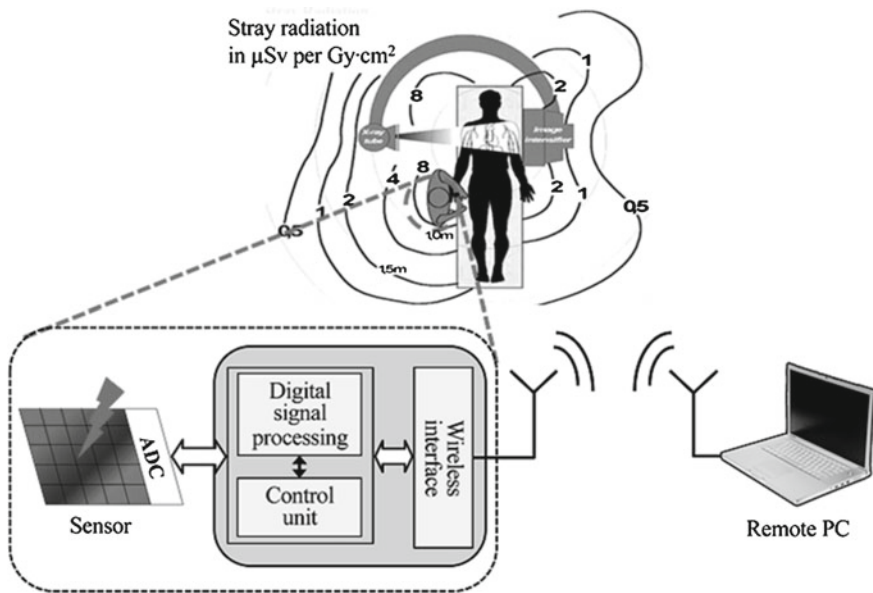


Fig. 10.1 Block diagram of the a wearable portable dosimeter architecture

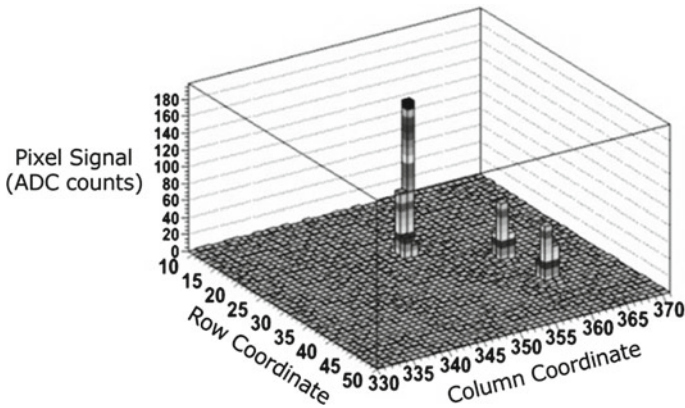
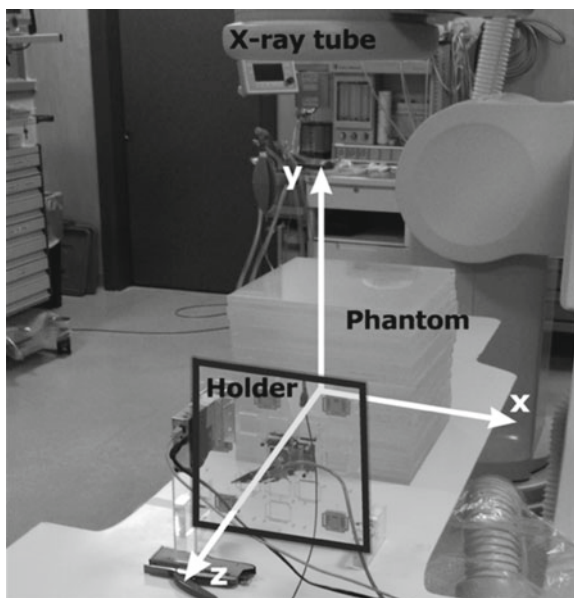


Fig. 10.2 Display of photons hitting the sensor matrix

dosimetric quantities. Each observable has been carried out using a devoted clustering algorithm [5]; data acquisitions have been performed in order to investigate the uncertainty on the measurement of each observable and the uncertainty in the calibration relation of each observable with the dose measured by a certified dosimeter.



**Fig. 10.3** Detail of the interventional angiography system showing the coordinate space adopted for the experimental set-up

### ***10.3.1 Experimental Set-up***

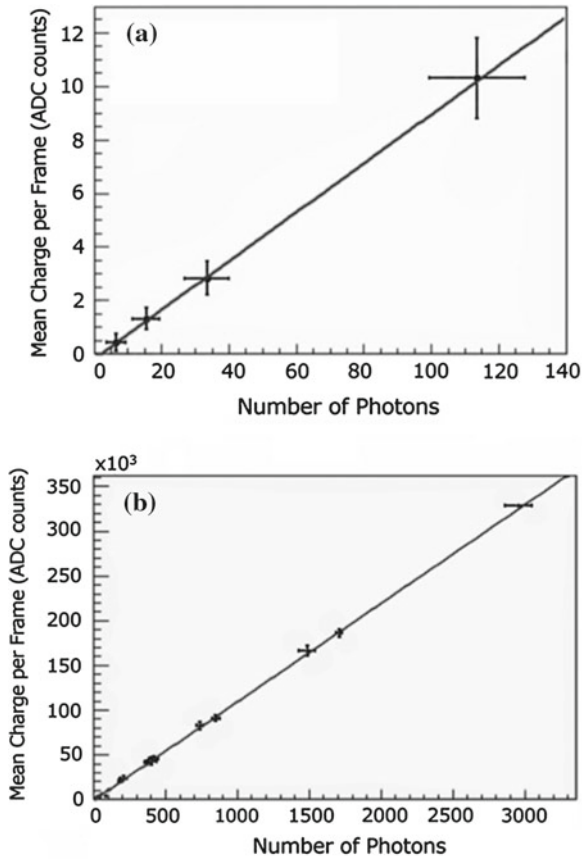
The interventional angiography system used to produce X-ray radiation is a Toshiba Infinix VC-I, available at the San Giovanni Battista Hospital in Foligno (Perugia, Italy). The radiation is scattered by a phantom made of  $20 \times 20 \times 3 \text{ cm}^3$  PMMA slabs.

The X-ray tube parameters during Interventional Radiology may vary, due to both the protocol and the patient-specific case. Two common configuration settings for the X-ray tube have been chosen: (a) tube voltage and current equal to 80 kV and 4 mA respectively (continuous mode); (b) tube voltage equal to 80 kV, tube current equal to 50 mA, pulse rate equal to 15 pps and pulse width equal to 1.9 ms (pulsed mode).

The spectrum region of interest for these operating modes has been determined with the Amptek X-123 precision spectrometer, used as a reference detector [1] and ranges from 10 keV up to few tens of keV. This energy range features a lower bound that is smaller than the one of the commercial Active Personal Dosimeters.

With respect to the coordinate space highlighted in Fig. 10.3, the X-ray tube radiates in the  $y$  direction. The spectrometer device and the sensor, surrounded by five TLDs, were mounted in a plastic holder, which was moved along the  $z$  axis at a variable distance of 0–100 cm from the phantom: that is a typical range between the medical staff and a patient during IRad treatment. TLDs have been used for evaluating the dose at the sensor position for each irradiation session [5].





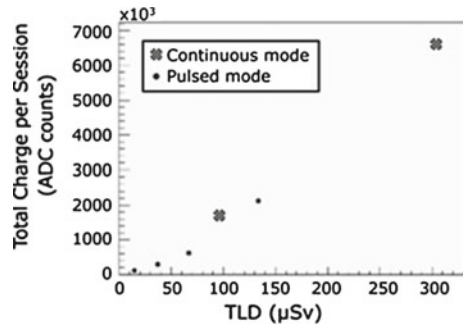
**Fig. 10.4** Correlation between number of detected photons and sum of reconstructed photon signals in **a** continuous and **b** pulsed mode

### 10.3.2 Study of Dosimetric Observables

A large number of data acquisitions was carried out by using the experimental set-up previously described with a fixed sensor gain value of 1. The goal of these measurements is to retrieve information about statistical distribution and relative uncertainty of each observable. The first observable, number of detected photons in a frame, features a statistical distribution compliant with a Gaussian function; the correlation between the observable and its relative uncertainty instead agrees with a Poissonian distribution.

The same results are obtained for the second observable, the sum of reconstructed photon signals in a frame. Nevertheless, it should be pointed out that the relative uncertainty of this observable includes both the statistical uncertainty contribution (Poissonian) and the sensor response fluctuation to a photon at a given energy.

**Fig. 10.5** Correlation between photon signal and dose measured using TLD



For this reason the relative uncertainty is bigger than that of the first observable, but its agreement with a Poissonian distribution suggests that the greatest contribution to the uncertainty comes from the photon counting observable.

The two dosimetric observables are linearly correlated, as shown in Fig. 10.4, where the average number of photons and the average sum of photon signals in a frame are plotted, for both continuous and pulsed operational modes. In order to compare the observables with TLD measurements, which are accumulated over an entire session, the total number of detected photons in a session has been calculated. Because the photon flux changes according to the position in the plastic holder, a compensation has been performed by using the method described in [5]. The correlation of the photon signal and the dose measured using passive dosimeters for both pulsed and continuous mode is shown in Fig. 10.5.

The linear correlation holds for both pulsed and continuous operational modes: this is a step forward with respect to the response reported for other dosimetric systems, where the difference among the two modes reached levels of 20–40 % [3].

## 10.4 Conclusion

The use of a CMOS pixel sensor to monitor the X-ray field diffused by a PMMA phantom during a standard IRad procedure has been tested in two different operational modes, continuous and pulsed. Two dosimetric quantities have been evaluated: the number of detected photons and the sum of reconstructed photon signals.

A correlation between the two quantities has been demonstrated and both could reach an uncertainty in the measurement below 10 %. A good linearity with dose measurements using passive dosimeters (TLD) has also been demonstrated, and the response difference between pulsed and continuous operational modes is reduced to less than 10 %, marking a distinct improvement with respect to other Active Personal Dosimeters.

**Acknowledgments** We would like to thank M. Biasini, A. Calandra, B. Checcucci, S. Chiocchini, R. Cicioni, R. Di Lorenzo, A. C. Dipilato, A. Esposito, A. Pentiricci for helpful discussions and for their technical support. This work was supported by *Fondazione Cassa di Risparmio di Perugia*, Italy (project reference: 2010.011.0421) and by the Italian RAPID INFN project.

## References

1. Amptek Complete X-Ray Spectrometer (2011): Available via AMPTEK. [www.amptek.com/pdf/x123.pdf](http://www.amptek.com/pdf/x123.pdf). Accessed Jan 2011
2. Annals of ICRP: Avoidance of Radiation Injuries from Medical Interventional Procedures (ICRP Publication 85), vol. 30(2). Elsevier, Amsterdam (2000)
3. Clairand, I., et al.: Use of active personal dosimeters in interventional radiology and cardiology: tests in laboratory conditions and recommendations. In: Radiation Measurements (International Workshop on Optimization of Radiation Protection of Medical Staff, ORAMED 2011), vol. 46(11), pp. 1252–1257 (2011)
4. Clairand, I., et al.: Use of active personal dosimeters in interventional radiology and cardiology: tests in laboratory conditions and recommendations. Available via: <http://www.oramed-fp7.eu/~media/Files/ORAMED/Presentations/25ICLAIRAND.pdf>. Accessed Jul 2011
5. Conti, E., et al.: Use of a CMOS image sensor for an active personal dosimeter in interventional radiology. *IEEE Trans. Instr. Meas.* **62**(5), 1065–1072 (2013). doi:10.1109/TIM.2012.2223331
6. Corbett, R.H., et al.: IRPA-10: international radiation protection association. *Br. J. Radiol.* **74**, 883–885 (2001)
7. EPD Mk2<sup>TM</sup>: Available via: [http://www.thermo.com/eThermo/CMA/PDFs/Product/productPDF\\_52873.pdf](http://www.thermo.com/eThermo/CMA/PDFs/Product/productPDF_52873.pdf). Accessed Jan 2011
8. Lee, J-S., et al.: A comparative study of wireless protocols: bluetooth, UWB, ZigBee and Wi-Fi. In: Proceedings of 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON), 5–8 Nov 2007, Taipei, Taiwan (2007)
9. Meroli, S., et al.: A grazing angle technique to measure the charge collection efficiency for CMOS active pixel sensors. *Nucl. Instr. Meth. A* **650**, 230–234 (2011)
10. Passeri, D., et al.: RAPS: an innovative active pixel for particle detection integrated in CMOS technology. *Nucl. Instr. Meth. A* **518**, 482–485 (2004)
11. RaySafe i2: Available via RaySafe. <http://www.raysafe.com/en/Products/Staff/RaySafe%20i2#Downloads> (2013). Accessed Jan 2011
12. Sornjarod, O., et al.: Radiation dose to medical staff in interventional radiology. *J. Med. Assoc. Thai.* **90**(4), 823–828 (2007)
13. Unfors EDD-30: Available via UNFORS. <http://www2.unfors.se/products.php?prodkey=55&catid=9> (2010). Accessed Jan 2011
14. UNSCEAR: Sources and Effects of Ionizing Radiation. General Assembly with Scientific Annexes, United Nations (2000)
15. Venneri, L., et al.: Cancer risk from professional exposure in staff working in cardiac catheterization laboratory: national research biological effects of ionizing radiation VII report. *Am. Heart J.* **157**(1), 118–124 (2009)

# Chapter 11

## A Novel Wireless Sensor Network for Electric Power Metering

Natale Galioto, Francesco Lo Bue, Daniele Rizzo,  
Leonardo Mistretta and Costantino Giuseppe Giaconia

**Abstract** Wireless sensor network with lightweight and efficient communication infrastructures are today a great interest, both at research and industrial level, as they are the basis for reliable monitoring services. Mesh networks are ideal candidates in this scenario, as they can be very fault-tolerant. After introducing wireless network systems requirements, we show the design of a mesh network routing protocol based on AODV schemas. Its implementation runs on top of low-cost off-the-shelf components, and allowed us to build a custom power meter wireless sensor node.

### 11.1 Introduction

Wireless networks (WSN) are today experiencing a tremendous development due to the potential benefit they deliver to a large number of application scenarios where the freedom from wires is a key advantage. E-health and Smart Grid are only two examples of research fields where WSN play a crucial role, and these application frameworks imposes quite a number of constraints on the characteristics of suitable network. Among these requirements it is valuable to mention:

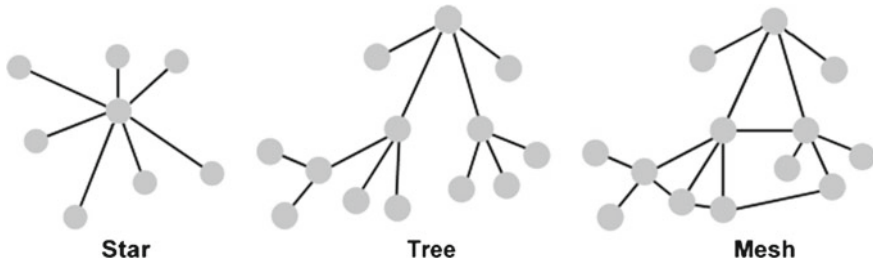
---

N. Galioto (✉) · F. Lo Bue · L. Mistretta · C. G. Giaconia  
DEIM, University of Palermo, Palermo, Italy  
e-mail: natale.galioto@unipa.it

F. Lo Bue  
e-mail: francesco.lobue@unipa.it

L. Mistretta  
e-mail: leonardo.mistretta@gmail.com

D. Rizzo  
Maxim Integrated Design Center, Milano, Italy  
e-mail: daniele.rizzo@maximintegrated.com



**Fig. 11.1** Various networks topologies: *star networks* have one main a central node that interconnects all other nodes. *Tree networks* are organized in parent–child relationship, where any node can have only one parent and multiple children, and one node being the root of the tree. *Mesh networks* have no special nodes at all, and every node can directly communicate with each other

- Resilience to nodes failures: network nodes in fact can fail, due to simple reasons as going outside the radiofrequency range or because they experience hostile environments (deserts, freezers etc.). Good WSN must possess the ability to reconfigure themselves and to perform unaffected by the failure of single nodes;
- Scalability: nodes count belonging to wireless networks is variable and depends on the specific application. In a very dense network each node could interfere with others nodes and this can potentially cause inefficiency or even network problems;
- Node Cost: since a lot of application needs many nodes, they should keep a very low cost both in commercial and energy request terms.

Nodes are typically embedded systems with low or very low resources, energy saving requirements, and low transmitting power. In order to fulfill these requirements, mesh networks show much better robustness when compared with typical point-to-multipoint solution (as in the wifi or Bluetooth case). Maximum attainable range can easily be extended by allowing data to hop from node to node, and reliability is increased by creating alternative paths when one node fails and/or a connection is lost. Various mesh networking protocols are already on the market, ZigBee [10] being the most popular one, which is specifically designed for low-data rate and low-power applications.

The ZigBee PRO version defines three types of nodes: Coordinators, Routers and End Device. While all nodes can send and receive data, there are differences in the specific roles they play.

Coordinators are the most capable of the three node types. Every ZigBee network relies on a single coordinator and it is the device responsible of originating and managing the established network. It is able to store information about the network, including security keys. This can potentially be considered as a major drawback, since a coordinator failure will disrupt the entire network till a new node with coordination capabilities, if any, will take over to form a new network. Routers act as intermediate nodes, relaying data from one device to another. End Devices instead are the simplest devices and they are able to monitor their own sensing and actuator and store or transmit their collected data. They have sufficient functionality to talk to their parents

(either the coordinator or a router) but they cannot relay data from other devices. This reduced functionality cuts down their cost and energy budget, thus suggesting low-power/battery-powered solutions.

The ZigBee PRO has ultimately standardized a new optional specification by introducing novel devices type, named Green Power nodes and enabling proxy and sink functionalities. In short, these have been specifically introduced for energy harvesting solutions, by assigning to a proxy node the task to store network information regarding sink nodes even when their energy runs out [11].

## 11.2 Proposed Solution

After a development of single chip solution implementing a ZigBee PRO network [5, 6] the authors observed a few important drawbacks that could potentially limit the node sensing and/or actuating capabilities. In particular most of the proposed ZigBee based solutions are implemented by using a microcontroller, with or without an embedded transceiver, and an Operating System acting as a Hardware Abstraction Layer. The most used OS is TinyOS [9] among other proposed solutions. These OSES are responsible both to manage the network, as a high priority task, and to collect data, trapped by the microcontroller I/Os or the AD converters, usually with the lowest priority among all the needed processes. This in turn leads to an uncompressible delay between successive measurements, and it hardly goes below a few tenths of milliseconds. In some scenarios this could be unacceptable if a faster refresh rate of measured data is necessary.

In order to overcome the above described limits the authors propose a custom network layer for mesh networks designed to run on low-power and low-end devices. Specifically, our implementation consists of an Atmel ATmega1284P microcontroller [3] running at 8 MHz and an IEEE 802.15.4 RF transceiver, MRF24J40MA from Microchip [7]. Even if integrated single chip solutions are available on the market we would keep separate the two devices in order to have a better control for testing and debugging purposes. All the firmware needed to generate the mesh network and the test application described in the next paragraph were written in C++, and any underlying Operating System or abstraction layer was avoided by hypothesis. This solution allows us to get rid of all the OS related overhead.

### 11.2.1 Custom Network Layer Architecture

#### 11.2.1.1 Overview

Developing a mesh network stack from scratch is not a trivial task. Network communication stack is generally a complex machine, and physical and datalink layers are often available and already manageable through off-the-shelf components. Those

**Table 11.1** Network packets layout

Field	Length	Definition
Message type	1 byte	
Source address	2 bytes	Network header
Destination address	2 bytes	
Payload	Variable	Payload

components are an invaluable aid for rapid prototyping and the chosen transceiver module manages physical and datalink layer of our mesh network.

While the Microchip MRF24J40MA is doing its work, the microcontroller must manage the Network Layer operations. The routing protocol we designed is *flat-based*, and then every node within the network has the same functionalities of all other nodes. Moreover in order to minimize network traffic, a *reactive* protocol was chosen. The *minimum-hop* metric was used to build routes between two nodes, which is relatively easy to implement while minimizing the activation of middle-route nodes.

In literature, two protocols satisfying these criteria are reported: the Dynamic Source Routing protocol (DSR) [8] and the Ad-hoc On-demand Distance Vector (AODV) [1]. Because of the high memory overhead associated to DSR implementations and memory constraints we imposed in order to target low complexity microcontrollers by design, the AODV was chosen.

### 11.2.1.2 Network Packet Format

Table 11.1 summarized network packets layout. A single byte was destined to message type, two bytes to source address and destination address respectively, and a variable space (up to the 802.15.4 maximum packet size, minus MAC header and Network Header) to the effective data. The payload length is then extracted from the underlying MAC frames. Our network layer supports the following four message types: Route Request (RREQ), Route Replay (RREP), Route Error (RERR) and Data (DATA).

### 11.2.1.3 Route Request

The RREQ message is the first message routed whenever a new mesh network has to be established. In particular, whenever a node *A* needs to send data to node *B* it first checks its routing table to find a suitable path to node *B*. If a path doesn't exist yet, a *Route Request* message will be broadcasted to start a *route discovery* phase. Every node, receiving this RREQ message, will check its routing table to find a suitable path to node *B*, and if such path doesn't exist again, it will subsequently broadcast a *Route Request* message to start its *route discovery* to node *B*. This process is usually repeated until node *B* is reached, a path is found through a numbers of

middle-route nodes, or a timeout occurs. Payload of this message consists of one byte only, indicating the number of hops that this message has actually done starting from the source.

#### 11.2.1.4 Route Reply

RREP messages are generated by destination nodes in a *route discovery* procedure. Specifically, this message will be generated if a new RREQ message is received, or if a RREQ message from the same originating source is received and the internal hops counter is less than the previously arrived packets, so allowing the network to know that a better route (with fewer hops) exists. Its payload consists of two fields:

- Hops;
- Hops to go;

where “Hops” contains hops count already done, and “Hops to go” contains the ones to reach RREP destination (which is the source of the RREQ packet). This second field is necessary in order to make sure the internal counters of source, destination and intermediate nodes remain aligned.

#### 11.2.1.5 Route Error

Whenever an intermediate node cannot deliver a message to a neighbor node (e.g. to the next node of the route path from node *A* to node *B*), a RERR message is generated and is sent back to the source, signaling that this particular route toward destination is no longer valid or unreachable. The payload of this packet contains only the address of the node that could not be reached.

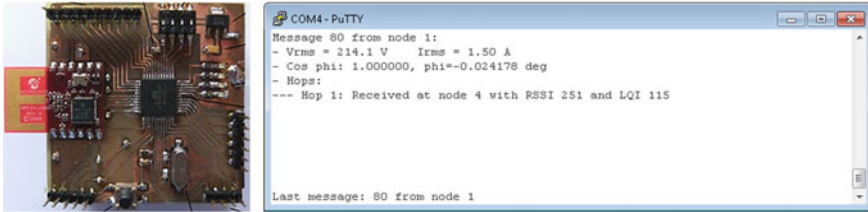
#### 11.2.1.6 Data Message

Data messages are created every time one node needs to send data to another node and it already knows the path. Its payload consists of user data belonging to upper network stack communication layers, and hence are organized from the upper application layer.

### 11.3 Firmware Implementation of the Network Layer

In order to implement this network logic, the microcontroller firmware was designed to manage the network stack by using the following two programming models: *Event based programming* and *Polling based programming*.





**Fig. 11.2** Implemented board and output of the developed wireless mesh network sensor node

The former programming model relies on microcontroller’s interrupts. This is a highly desirable approach since it guarantees that whenever a network packet is received, it will be processed with the lowest possible latency. Moreover if a node is in power-saving mode this approach will natively “wake up” the microcontroller to run the interrupting network operation only, and will end up with a return to sleep mode, thus running a very conservative power-saving policy.

The latter programming model instead relies on an application scheduler responsible to simply poll network stack for messages availability. The main drawback of this approach, in presence of many tasks, is an accumulated delay, eventually ending up to a lack of reactivity.

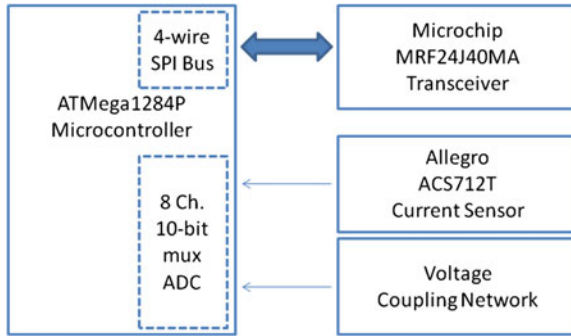
The choice of the programming model heavily depends on the application and the usage scenario. For this reason, our network stack was designed to allow flexible switch between these different choices. We developed, indeed, each network operation as a C++ function, and created a common entry point which is called by interrupt service routine in case of event base programming model, or explicitly by the main loop in the case of manual scheduling schemas.

## 11.4 Case Study: Power Meter

We decided to implement and test our network stack with a simple yet real application. We built a simple wireless power meter by implementing a board with the above mentioned chips. We deployed on our lab three sensor nodes and one base station node, formally identical to the other nodes, which collected all data coming from sensor nodes. Figure 11.2 shows the developed board and the output of the wireless mesh network sensor node.

### 11.4.1 Hardware Setup

ATMega1284P is a high-performance low power 8-bit microprocessor. For our purposes it offers 16-bit timers, 8 multiplexed 10-bit Sample and Hold ADCs, serial



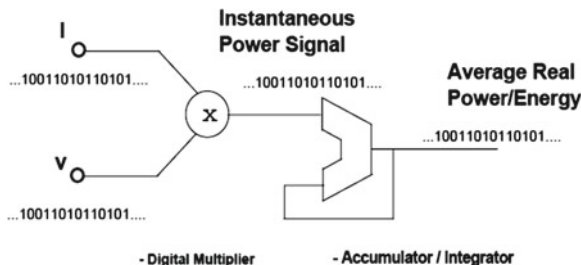
**Fig. 11.3** System architecture: microcontroller communicates with a 802.15.4 transceiver by an SPI bus, and reads AC voltage and current by an appropriate coupling network and an Hall effect sensor respectively

interfaces for communication, 16 Kbytes of internal SRAM and 128 K of Flash memory. It runs at 8 MHz and, besides its low cost, it has enough room for sampling at constant rate and keeping a network layer up and running. The adopted transceiver has a small SMD form factor, an integrated PCB antenna, a crystal, a voltage reference, hardware CSMA-CA mechanism, automatic ACK response and FCS Check. It also features a 4-wire SPI bus used to send and receive custom network packets. The desired power metering application is then realized by reading electric voltage and current flowing from the AC (220 V) power supply and a generic load. In particular voltage is initially connected to a resistor divider to meet ADC range requirements, and then is connected to a complex anti-aliasing filter to be sure that its frequency content is the correct one. Current is sensed through an Hall effect based current sensor [2]. Its output is a voltage proportional to the sensed current, and its output is already anti-aliased by connecting a filter capacitor to the appropriate pin. Thus the output voltage is simply scaled down to meet ADC range and connected directly to the ADC input pin. Figure 11.3 shows the system architecture.

### 11.4.2 Principles of Measurements

Apparent electrical power is defined as the product of instantaneous voltage and current. The ADC gives two digital samples proportional to the analog voltage and current signal read at the inputs. The accumulation of voltage  $V_n$  and current  $I_n$  samples products over a time period will give the average apparent power consumption when divided by the number of samples  $N$ , as shown in Fig. 11.4.

In order to obtain active power measurements from AC systems, a power factor correction, which is the phase relationship between voltage and current signals, must be kept into account. In such case, average power in a period is defined as Active AC Power and specifically as



**Fig. 11.4** Average power and energy in digital domain: voltage and current digital samples are multiplied together, and the result integrated over time using an accumulator

$$AVG\_Power = \frac{\sum_{n=0}^{N-1} V_n \cdot I_n}{N} = V_{rms} \cdot I_{rms} \cdot \cos \varphi \quad (11.1)$$

where  $\varphi$  is the phase angle between current and voltage, corresponding to the time domain delay between current and voltage.

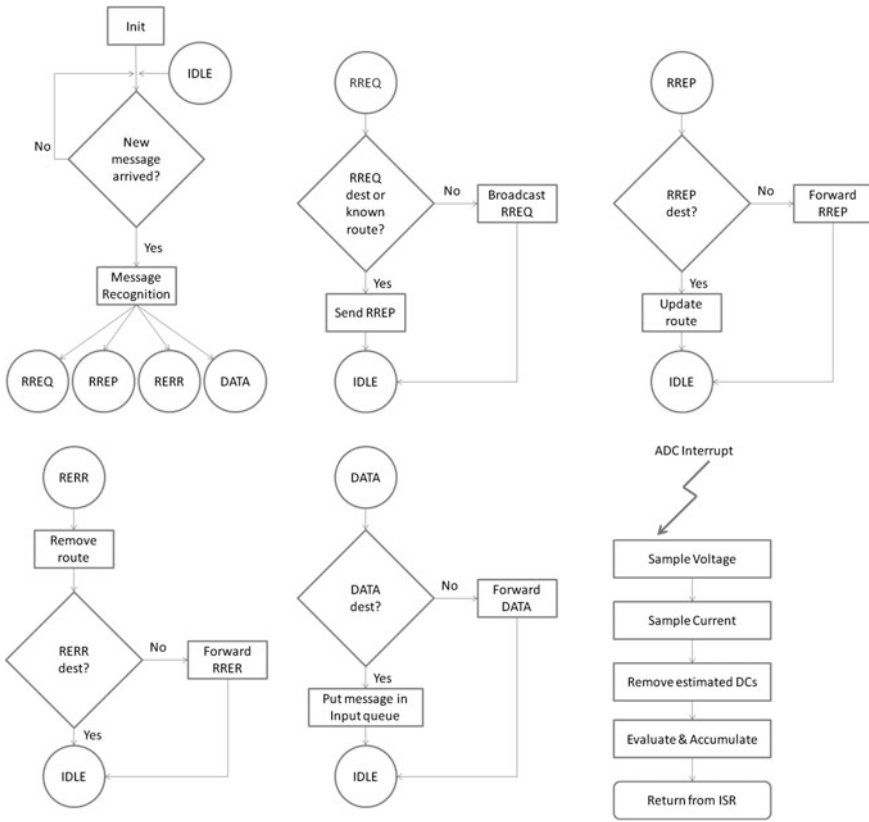
In order to get accurate measure of energy values in an AC system, it is required to make frequent measures and average them, thus greatly increasing consistency. The sampling frequency  $F_s$  of the system should then be much higher than the measured grid frequency. Let be  $N$  the total number of samples acquired in  $N/F_s$  seconds. Energy (in Watt seconds) can be obtained by multiplying  $N/F_s$  with the average power:

$$Energy = \frac{\sum_{n=0}^{N-1} V_n \cdot I_n}{F_s} = \frac{N}{F_s} \cdot V_{rms} \cdot I_{rms} \cdot \cos \varphi \quad (11.2)$$

The power factor measurement can be then obtained by dividing the Average Power with the Apparent Power.

### 11.4.3 Firmware Architecture

The implemented firmware is a mixture of functions aimed to communicate with transceiver (via SPI bus) and functions aimed to constantly sample voltage and current. In particular we selected the *Polling based programming model* for the network tasks, and the *Event based programming model* for the acquisition tasks. We basically run ADC conversions based on interrupts at regular rates, while “the rest of the time” handle network I/Os and other small tasks. This is a typical approach in acquisition nodes, where sampling must occur at fixed intervals and a network operation could delay acquisition by a significant amount of time. Figure 11.5 shows the basic flowchart of the implemented firmware.



**Fig. 11.5** Basic flowchart of the firmware implementation: voltage and current are sampled at constant rate thanks to the interrupts, while in the spare time firmware holds up the network

The ADC has 8 conversion channels, with individual input selection and gain. All inputs are multiplexed, forcing us to sample voltage and current at two different times. We rely on Timer1 Interrupt Service Routing to trigger ADC acquisitions at a constant 4 KHz rate. We start ADC conversion by sampling the voltage channel input and wait until conversion is done. At the end of the conversion, we switch ADC channel to the current channel input, and rest for a small amount of time before making a new acquisition to avoid crosstalk between channels. After sampling the current, we remove the currently estimated DC offset value from each sample.

After the sampling step, we calculate  $V_{rms}$ ,  $I_{rms}$ , Apparent Power and Energy. Since we are using a low-power, low-resources 8-bit microcontroller, we defer all heavy divisions at later time. Indeed, we simply accumulate  $V_{rms}$ ,  $I_{rms}$ , Apparent Power and Energy, all multiplied by  $N$ . We decided to perform division by  $N$  and calculate real values once per second only, yet obtaining a good refresh rate for visualization purposes.

Anti-aliasing filter, current sensor, and multiplexed ADC inputs usually produce an undesirable constant phase shift between voltage and current signals. Then a calibration was performed by correcting the phase angle measured when the system is connected to a known pure resistive load.

To estimate DC voltage and current offsets, we accumulate every sampled value  $V_n$  and  $I_n$  over a period of 1 s, and then divide by the number  $N$  of samples. This is a dirty yet fast converging algorithm which produces very small ripple around real DC offsets. To further reduce this ripple, we may feed these values to a low-pass filter, at the cost of slowing down its convergence. The obtained average values are finally sent to the collecting node by using the network stack, which will deliver the message to destination, as previously described.

## 11.5 Enhanced Wireless Node

The design of the wireless mesh network node imposed some heavy constraints both on the hardware side and on the firmware side. In particular, we looked only at off-the-shelf components, with low current consumptions, and inevitably with low resources available. However, some applications, especially those with very high data rates flows, may need more processing power or higher memory requirements. With these requirements, moving from 8-bit to 32-bit microcontrollers is a must. This architecture offers a new set of devices from different vendors, each with its own strengths and weaknesses, but in order to reuse most of the network-layer code, we choose to develop an enhanced version of the node using the AT32UC3C2256C [4] Atmel microcontroller. This device features 256 Kb of Flash memory, 64 Kb of SRAM, up to 45 I/O pins, up to 66 MHz clock, and 11 ADC channels multiplexed with 12-bit resolution up to 2 Msps, two of which can be sampled in parallel, making it ideal for our power meter application. Indeed, we can now sample voltage and current simultaneously, getting rid of the intrinsic phase shift due to the typical sequential sampling of the multiplexed channel approach, and getting higher sampling rate for better accuracy. We removed also some critical assembly code, explicitly written to allow fast 32-bit operations in an 8-bit environment, as we can now apply native 32-bit math, optimizing the performance of our filters.

## 11.6 Conclusions

We designed an innovative wireless mesh network layer based on AODV routing protocol schemas. We implemented it on low-power devices and built a custom power meter wireless sensor node. We tested successfully the application on a small environment, and then moved toward a slightly more expensive and slightly less power-consumption friendly 32-bit microcontroller to allow better results for the proposed application.

## References

1. Ad hoc On-Demand Distance Vector (AODV)—Rfc 3561 IETF. <http://www.ietf.org/rfc/rfc3561.txt>
2. Allegro ACS712T 20A Datasheet. <http://www.allegromicro.com>
3. Atmel ATmega1284P Datasheet. <http://www.atmel.com>
4. Atmel AT32UC3C2256C Datasheet. <http://www.atmel.com>
5. Giaconia, G.C., et al.: Integration of distributed on site control actions via combined photovoltaic and solar panels system. IEEE 2nd International Conference on Clean Electrical Power, pp. 171–177, Capri, Italy (2009)
6. Giaconia, G.C., et al.: Combined photovoltaic/solar prototype. [http://beywatch.eu/list/file/BeyWatch\\_D3.4\\_UNIPA\\_FF\\_2010-07-12.pdf](http://beywatch.eu/list/file/BeyWatch_D3.4_UNIPA_FF_2010-07-12.pdf) (2010). Cited 30 Oct 2013
7. Microchip MRF24J40MA 2.4 GHz IEEE Std. 802.15.4 RF Transceiver Module Datasheet. <http://www.microchip.com>
8. The Dynamic Source Routing Protocol (DSR)—Rfc 4728 IETF. <http://www.ietf.org/rfc/rfc4728.txt>
9. TinyOS Official. <http://www.tinyos.net>
10. ZigBee Alliance. <http://www.zigbee.org>
11. ZigBee Alliance "New ZigBee PRO feature: Green Power". <https://docs.zigbee.org/zigbee-docs/dcn/12/docs-12-0646-01-0mwg-new-zigbee-pro-feature-green-power.pdf>

# Chapter 12

## High Performance Bit-Stream Decompressor for Partial Reconfigurable FPGAs

Gian Carlo Cardarilli, Marco Re and Ilir Shuli

**Abstract** In Digital Signal Processing (DSP), Field Programmable Gate Arrays (FPGAs) are becoming ubiquitous for their capability to process massive amount of data in parallel maintaining the flexibility of the software approach. FPGA chips of major vendors also support partial dynamic programming, namely the ability to change the functionality of portions of FPGA while the rest of the functionalities remain active. In this way, partial reconfiguration of the FPGA requires a fast reload of a partial bitstream. To this purpose, an improvement of the reconfiguration speed (with the contemporary reduction of the memory occupancy) is obtained by compressing the bitstreams. High performance on board decompressors are required to speed-up the reconfiguration operation. In this paper a new hardware oriented technique for the bitstream compression and decompression is proposed. This technique maintains good compression factors and correspond to a very simple and fast hardware architecture for the compressor block.

### 12.1 Introduction

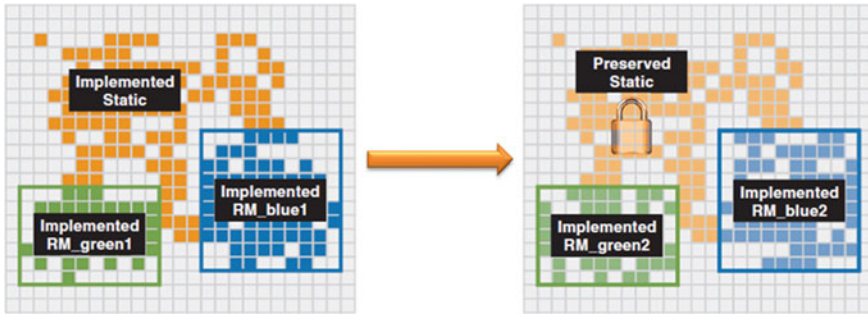
Nowadays digital systems need more and more computing power in order to fulfill strict requirements on overall system throughput and latency. In the digital signal processing field, Field Programmable Gate Arrays (FPGAs) are becoming ubiquitous for their capabilities to process massive amount of data in parallel. FPGAs have continuously improved their performance during the years, increasing available

---

G. C. Cardarilli (✉) · M. Re (✉) · I. Shuli (✉)  
Department of Electronic Engineering, University of Rome “Tor Vergata”,  
Via del Politecnico 1, 00173 Rome, Italy  
e-mail: g.cardarilli@uniroma2.it

M. Re  
e-mail: marco.re@uniroma2.it

I. Shuli  
e-mail: shuli@ing.uniroma2.it



**Fig. 12.1** Dynamic partial reconfiguration process [5]

resources, area and pin count, and working frequency. These advancements come at the expense of increasing power consumption, although the use of different techniques to lower the power consumption. FPGA chips of major vendors have also the ability to be reconfigured (totally or partially) at run-time. The dynamic partial reconfiguration capability allows to change the functionality of portions of FPGA while the rest of the device remains active.

The FPGA resources can be classified into “regions”. One of them, called “static region”, contains all the circuitry that needs to remain always operative. Besides the static region, there exist many Partial Reconfiguration Regions (PRRs) where the designer can implement many reconfigurable modules (RM). The PRRs can be reconfigured at run-time with new functionalities. In this way, FPGA functions that need to be always active reside in the static region while functions that are mutually exclusive can share the same FPGA region (and resources) that are time-multiplexed.

Figure 12.1 shows the reconfiguration of 2 PRRs where the static region functionality remains the same, while the partial reconfigurable functionality (green region) changes from RM\_green1 to RM\_green2 (the same happens in the blue region).

The partial reconfiguration can be exploited for applications that can be divided into modules or subsystems where the modules are not all active at the same time (frequently this constraint is fulfilled because some functionalities are required only in some phases of the operations). In this way, by using dynamic partial reconfiguration we can time-multiplex part of the FPGA resources.

A fast partial reconfiguration of the FPGA is at this point required to obtain an efficient execution of those algorithms. Since the different configuration bitstreams are stored in external (and relatively slow) nonvolatile memory, great attention must be paid to reduce the partial reprogramming time. A possible approach to improve the reconfiguration speed (reducing at the same time the configuration memory) is based on the compression of the partial bitstreams.

Bitstream compression for FPGA implementations exploiting dynamic partial reconfiguration has been proposed in different papers [1, 2].

For example in [1] an LZSS [3] variant is used for a XCV2000E device. In [2] several compression techniques have been analyzed and compared and, together with



a state-of-the-art LZSS hardware decompressor. By using compressed bitstreams and a hardware decompressor we are able to stream data at the maximum data-rate allowed by the Internal Configuration Access Port (ICAP) interface of the FPGA.

Typically, there are two ways to transfer the bitstream to the ICAP:

- Direct ROM transfer
- Bitstream caching.

In the first technique the bitstream is read from an external ROM and written to the ICAP port directly. With this approach the reconfiguration time is highly dependent on the data rate sustainable by the external Flash memory. On the other hand, energy consumption to perform the reconfiguration is minimal (in the sense that just one bitstream reading is required). In the latter technique, at the system start the bitstream is transferred to a high-speed volatile memory, such as a Dynamic RAM, and then it is written into the ICAP. This technique permits lower reconfiguration times because volatile memories are usually much faster than non-volatile memories. The speedup is achieved at the expense of greater energy dissipation (the bitstream has to be transferred twice) and of double memory occupation (because it is stored both in the non-volatile and the volatile memories, at the same time).

Bitstream compression helps to mitigate the undesired effects of both techniques

- Reducing the transfer times from the non-volatile memories
- Reducing the memory occupation.

## 12.2 Bitstream Compression Techniques

An efficient implementation of the bitstream compression technique requires a low-complexity decompression hardware, in order to save FPGA area and power.

In our system we implemented the compression technique, developing a fast and simple decompression hardware. In this way we halved the size of the bitstream, obtaining the maximum reconfiguration speed (it is limited by the ICAP itself) using a direct Flash transfer approach.

LZ77 [3] compression algorithm is a good choice for a wide variety of bit sequences and can be used with good results in different fields. In the case of FPGA configuration files (bit-streams), the situation is a bit different. In fact, bit-streams usually contain a high number of zeros. This happens because each bit of the bit sequence represents the configuration of an FPGA element such as for example (routing nodes (PIPs) or LUT content) that are usually sparse. This happens not only for full bit-streams, but also for partial configuration files. In fact, in this case, the resources used for implementing a functionality are greater than those required for a static implementation because of the needs of additional structures for signals routing. This resource redundancy justifies the presence of a lot of zeros also in partial bit-streams: additional zones are needed to propagate the signals routing through the PR module, but frequently a great percentage of available resources in these zones remain unused

(especially for little PR modules, but also for the biggest one). Considering this peculiarity, a new compression algorithm, specifically designed for this particular application, has been developed: its name is “Zero Compression Algorithm” and it is derived directly from another widely diffused compression algorithm, known as “Run-length Encoding” [4].

This algorithm is based on a very simple idea: why storing long zero-sequences, when it is possible to reconstruct them inside the FPGA? Starting from this idea, different aspects must be analyzed for the definition of an efficient implementation of the compression algorithm, for example, the number of bits used to indicate the presence of coded words (these extra-bits are frequently known as “headers”), the length of each coded word, the position (inside the coded bit-stream) where headers will be stored. It’s not easy to get a good balance between those parameters. For example, longer coded words could seem a good solution to mitigate negative effects of information overhead introduced by headers but, if this choice leads to a smaller number of coded words, the opposite effect could be obtained.

The very good results have been obtained after an optimization process of the above mentioned parameters, getting compression ratios that are sometimes better than those obtained with LZ77 [3] algorithm and also characterized by lower reconfiguration times.

### ***12.2.1 Zero Compression Technique***

Run-length encoding (see [4]) is a compression technique that allows to efficiently compress long sequences of identical words in a data stream. The sequence is encoded as a tuple  $(w, l)$  where  $w$  indicates the word and  $l$  the number of occurrences of the word. This way of compressing data is very effective in applications where data changes occur in a predictable way from one word to another e.g. audio streams. It would seem that this technique could be very effective even for FPGA bitstream compression because a great number of zeroes is present in the bitstreams. This is due to the presence of the configuration switches that have obviously only one “active” position which corresponds to a logic “1” and the other positions correspond to logic “0”. However, analyzing an FPGA configuration bitstream it can be easily noted that there are many “0” sequences of less than 8 bits. This could be a potential limit of the bitstream compression effectiveness because the algorithm works by storing tuples of data for every change of sequence words. If the bitstream is analyzed by groups of 8 bits a.k.a. bytes the sequence of following “0” words quickly decreases, also other value sequences become very short. If the traditional run-length compression technique is applied, the resulting compressed bitstream is larger than the uncompressed bitstream. For these reasons we propose a novel bitstream compression technique based only on compressing 0 values.

The zero compression technique analyzes the bitstream in words of 8 bits. This kind of analysis permits a good level of granularity (allowing the compression of a significant number of zeroes) but also limiting the overhead related to the headers introduced by uncompressed words.

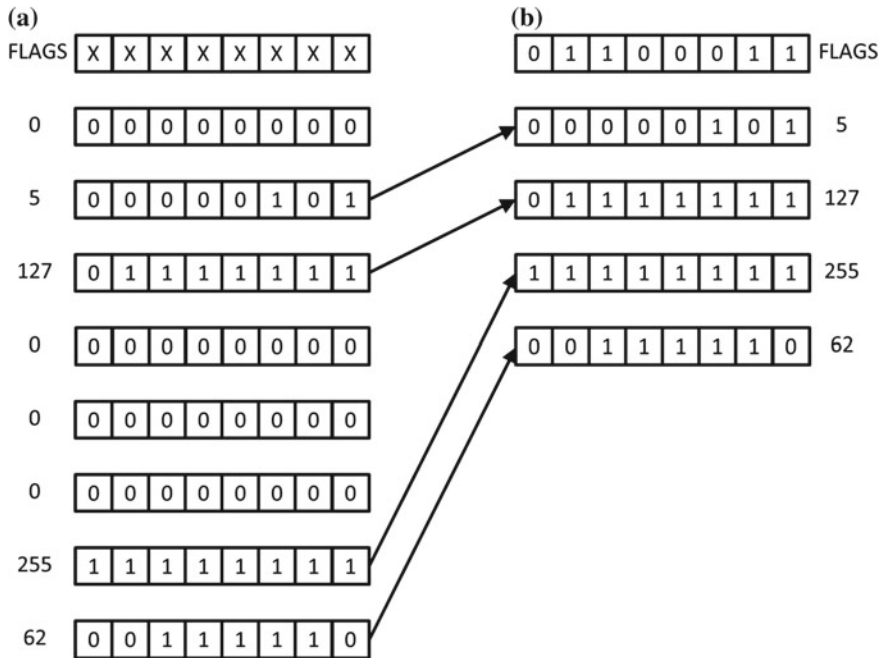


Fig. 12.2 a First picture; b second picture

The algorithm works as explained below

1. Initialize an empty flag word
2. Take 8 words from the uncompressed bitstream to compress
3. Evaluate every bit of the flag word “0” if the word is zero and “1” if the word is not zero
4. Prepare the compressed output buffer
5. Put the flag word in the output buffer
6. Put the non zero words in the output buffer discarding the zero words
7. Jump to 1 if the bitstream length is not reached

The algorithm is shown in Fig. 12.2.

### 12.3 Implementation

The hardware implementation of the decompressor and the test bed of the overall system is illustrated in Fig. 12.3.

The decompressor is part, of the Partial Reconfiguration Controller (PRC). This controller is required to allow a fast dynamic partial reconfiguration. The PRC has

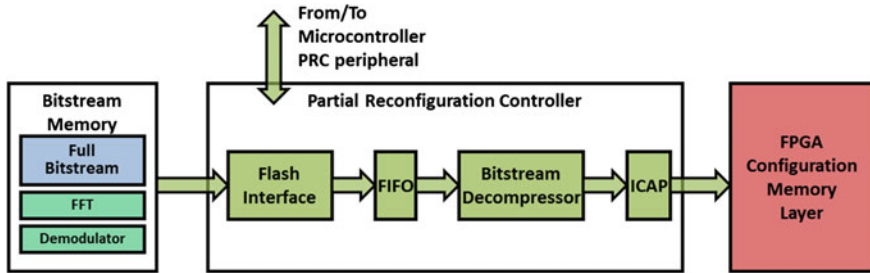


Fig. 12.3 Partial reconfiguration controller

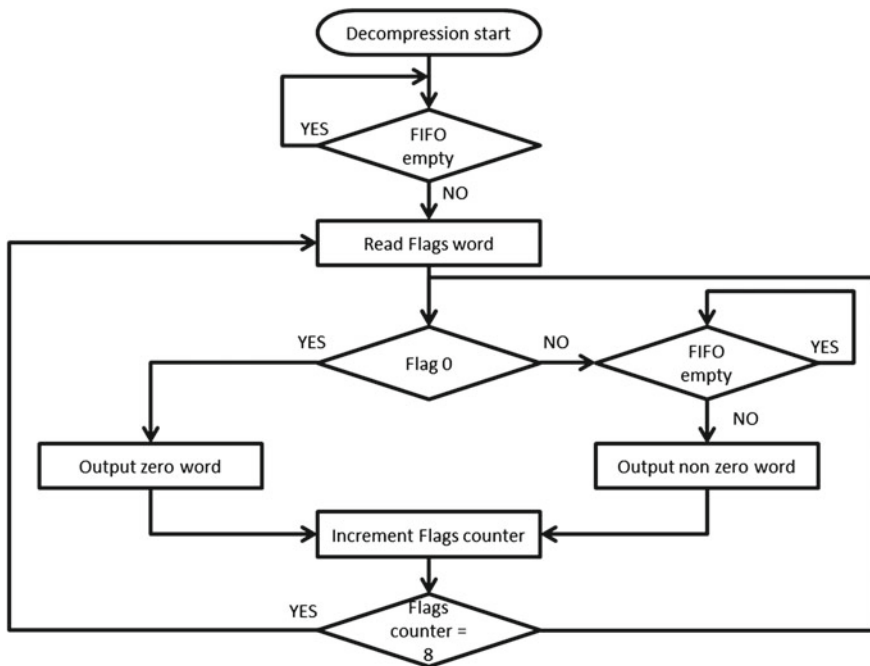


Fig. 12.4 Decompression flow chart

a built-in flash interface for direct access to the non-volatile flash memory thus enabling direct data access to the bitstreams. A FIFO register is used between the flash interface and the decompressor. The decompressed data are sent to the ICAP.

The bitstream decompressor operates reading the data from its input FIFO and performs the exact reverse of the operations shown in Fig. 12.2. The decompression steps are shown in Fig. 12.4.

Table 12.1 shows the FPGA resources needed for the LZSS8 and for the zero decompressor proposed by the authors. It is clear that both the decompressors uses just a few number of FPGA LUTs. The most interesting difference between the

**Table 12.1** Hardware implementation resources

Decompressor HW	LZSS8	Zero decompressor
LUTs	83	72
$F_{MAX}$ [MHz]	198	284

**Table 12.2** Zero decompression times for different FIFO sizes and FPGA ICAP bandwidth/memory bandwidth

	$\lambda$	Opt.						$\lambda$	Opt.	FIFO size (bytes)					
		1	4	8	16	128	1			4	8	16	128		
Cyclone II	2	54.8	71.7	66.6	63.6	61.7	60.3	Virtex-II	2	50.0	70.1	67.6	67.0	66.4	63.8
	3	54.8	62.2	58.4	56.7	55.9	55.2		3	48.2	59.0	57.4	56.8	56.5	55.2
	4	54.8	58.5	55.6	55.1	54.8	54.6		4	48.2	54.3	53.3	53.0	52.8	52.1
	5	54.8	56.4	54.8	54.6	54.4	54.3		5	48.2	51.7	51.2	51.1	50.9	50.5
Spartan-3	2	64.4	80.1	77.9	77.1	76.6	75.1	Virtex-V	2	50.2	70.7	67.7	67.1	66.6	65.0
	3	64.4	72.2	71.0	70.6	70.4	69.7		3	50.2	60.1	58.6	58.1	57.8	56.9
	4	64.4	68.9	68.1	68.0	67.9	67.5		4	50.2	55.8	54.9	54.6	54.4	54.0
	5	64.4	67.1	66.7	66.6	66.5	66.3		5	50.2	53.5	53.0	52.8	52.7	52.4

**Table 12.3** LZSS8 decompression times for different FIFO sizes and FPGA ICAP bandwidth/memory bandwidth

	$\lambda$	Opt.						$\lambda$	Opt.	FIFO size (bytes)					
		1	4	8	16	128	1			4	8	16	128		
Cyclone II	2	63.1	78.1	74.1	69.6	67.1	65.0	Virtex-II	2	50.0	67.5	63.7	62.3	61.5	59.5
	3	63.1	128.9	67.9	65.3	64.4	63.7		3	43.5	95.5	54.7	53.7	53.1	51.9
	4	63.1	67.5	65.3	64.1	63.7	63.2		4	43.5	51.8	50.8	50.2	49.8	49.0
	5	63.1	65.7	64.1	63.6	63.4	63.1		5	43.5	49.2	48.6	48.3	48.1	47.4
Spartan-3	2	59.1	75.8	72.9	71.6	70.9	69.2	Virtex-V	2	50.0	69.4	65.6	64.2	63.5	62.1
	3	59.1	123.1	66.5	65.7	65.3	64.2		3	49.3	105.5	58.2	57.5	57.0	56.3
	4	59.1	64.6	63.7	63.2	62.9	62.3		4	49.3	56.1	55.1	54.7	54.4	53.9
	5	59.1	62.7	62.2	61.8	61.6	61.3		5	49.3	54.0	53.5	53.1	53.0	52.6

presented decompressor and that presented in Koch et al. [2], is the maximum speed that the decompressor can reach. This is an important factor for our application, since the decompression time must be taken into account for evaluating the execution time of the algorithm.

## 12.4 Results

In the Tables 12.2 and 12.3 the reconfiguration time for different ratios  $\lambda = d_{FPGA}/d_{MEM}$  where  $d_{FPGA}$  represents the FPGA reconfiguration interface data rate and  $d_{MEM}$  the external FLASH memory data rate are presented. High  $\lambda$  values

mean low data rates for the external memory supposing that the FPGA configuration data rate remains the same and that the decompressor can achieve the running speed needed to match that output data rate. From simple inspection of the tables it is clear that the performances of the proposed compression algorithm and decompressor architecture are quite good. In the Cyclone II case our algorithm performs even better than the one presented in Koch et al. [2]. One thing that does not emerge from the tables is that our decompressor runs at a much faster speed than its counterpart [2].

## References

1. Huebner, M., Ullmann, M., Weissel, F., Becker, J.: Real-time configuration code decompression for dynamic FPGA self-reconfiguration. In: Parallel and Distributed Processing Symposium, International, p. 138b, 18th International Parallel and Distributed Processing Symposium (IPDPS'04)—Workshop 3, 2004
2. Koch, D., Beckhoff, C., Teich, J.: Hardware decompression techniques for FPGA-based embedded systems. *ACM Trans. Reconfigurable Technol. Syst. (TRET)* **2**, 9 (2009)
3. Ziv, J., Lempel, A.: A universal algorithm for sequential data compression. *IEEE Trans. inf. theory* **23**(3), 337–343 (1977)
4. Salomon, D.: *Data Compression—The Complete Reference*. Springer, New York (2004)
5. Xilinx: Partial reconfiguration user guide (2011). <http://www.xilinx.com>

# Chapter 13

## A Reconfigurable Functional Unit for Modular Operations

Gian Carlo Cardarilli, Luca Di Nunzio, Rocco Fazzolari,  
Salvatore Pontarelli and Marco Re

**Abstract** The efficiency of standard microprocessors decreases when operations on short data are performed because they are optimized to perform operations on fixed size data. Short data processing and bit manipulation can be accelerated integrating a Reconfigurable Functional Unit (RFU) in parallel with the ALU. An RFU is a tightly coupled integrated Reconfigurable Array used to speed-up the computation of a set of operations for which standard microprocessors are not optimized. In this paper we show the benefit of using the Adder-based Dynamic Architecture for Processing Tailored Operators (ADAPTO RFU) [1–3] (a full adder based RFU) on modular operations. In particular we describe how to speed up the modular addition and the Montgomery Multiplication by using the ADAPTO RFU.

### 13.1 Introduction

A Reconfigurable Functional Unit (RFU) is a tightly coupled integrated Hardware accelerator used for the computation of particular operations. Processing performance of conventional embedded processors and DSP can degrade when short data operations are involved. Applications that require short data manipulation are, for

---

G. C. Cardarilli (✉) · L. Di Nunzio (✉) · R. Fazzolari (✉) · S. Pontarelli (✉) · M. Re (✉)  
University of Rome “Tor Vergata”, Via del Politecnico 1, 00191 Rome, Italy  
e-mail: g.cardarilli@uniroma2.it

L. Di Nunzio  
e-mail: di.nunzio@uniroma2.it

R. Fazzolari  
e-mail: salsano.ing@uniroma2.it

S. Pontarelli  
e-mail: salvatore.pontarelli@uniroma2.it

M. Re  
e-mail: marco.re@uniroma2.it

example, the bit reversal or the bit packing/unpacking operations in which different sub-words coming from different words are concatenated to create a new word. All these operations are very simple but require lots of clock cycles to be performed by standard microprocessor because they are designed to perform sequential operations on data having the same native wordlength of the CPU. To solve this problem several solutions have been proposed in the literature, both software [8] and hardware. Among the hardware solutions the most interesting ones in terms of flexibility and performance are the RFUs. Usually these architectures are similar to small FPGA (array of LUTs and pass-transistor for programmable interconnect) and are inserted directly in the datapath of the microprocessor in parallel with the ALU. An RFU can be considered an hardware Instruction Set expansion. Normal operations are performed by the ALU meanwhile non standard operations are executed by the RFU. A good RFU has to be characterized by two main aspects:

1. Fast reconfiguration
2. Low cost.

The first one is important because if the reconfiguration time is too slow it could be greater than the execution time and acceleration could be compromised. The second one is essential in embedded system because cost is an important aspect. Fast reconfiguration time and few number of transistors cannot be satisfied simultaneously in a LUT based RFU, because:

- LUT reconfiguration is a sequential operation requiring a number of clock cycle proportional to the LUT size.
- In order to obtain a multicontext architecture the entire reconfigurable array has to be replicated for every context augmenting the cost in terms of silicon area.

In order to overcome these limitations in [2, 3] a new architecture called ADAPTO has been proposed. Modular operations are not efficiently managed by standard microprocessors even if they are widely used in a large set of applications starting from cryptography [10], error detection and correction codes [12], digital signal processing [4] and so on. The bottleneck caused by the modular operations has been faced by using specialized hardware [9], modified ALU [5], or optimizing the software routines [7]. In this paper we show how ADAPTO can be used to speeding up modular addition and Montgomery Multiplication reducing the number of operations to perform and increasing the parallelism. The gain that we try to obtain by using ADAPTO is twofold:

1. we try to perform different operations with different moduli in the same clock cycle
2. we try to perform each operation in only one clock cycle

The first target is useful, for example, when Residue Number Systems [4] are used. Instead, the second target is useful to all the applications that can use the ADAPTO RFU.



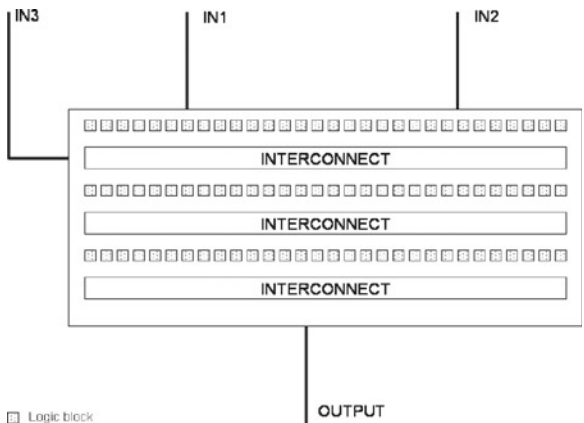


Fig. 13.1 Reconfigurable array structure

## 13.2 The ADAPTO Architecture

The ADAPTO RFU consist in three alternated stripes of Logic Block (LB) and interconnect. It has three 32 bit input and one 32 output connected with the Register File (RF) of the main processor (Fig. 13.1). LBs are based on Full Adders that perform both logical and arithmetical operation meanwhile interconnects are based on pass transistors devices. Multicontext is realized using context memories linked to the LBs and to the gates of the pass transistors used for the interconnect.

### 13.2.1 Computational Unit

In order to guarantee flexibility, high reconfiguration speed and low area overhead, the computational unit has been based on full adders instead of LUTs. The full adder can be easily configured for performing the following operations: one bit addition, NOT and PASS, 2 input AND, 2 input OR, 2 input XOR. The sum output R is the 3 input XOR while the carry output C performs the majority voting of the inputs. The different functions can be selected by forcing one or more input pins of the FA to a fixed value as shown in Fig. 13.2. For instance, a 2 input AND is obtained by putting the Cin input to 0 and taking as output the Cout pin. Clearly, this structure is less flexible than that based on LUT but, as shown in [1] a multicontext LB based on FA require a small context memory.

Another advantage of this solution is the lower number of MOSFETs required for its implementation (for example [14] and [6] use respectively 14 and 10 MOSFETs).

The basic computational element of the new architecture, shown in Fig. 13.3, is the Logic Block (LB). It is based on two multiplexers, a selector (realized by a multiplexer with a suitable coding of selection bits) and a full adder. The input

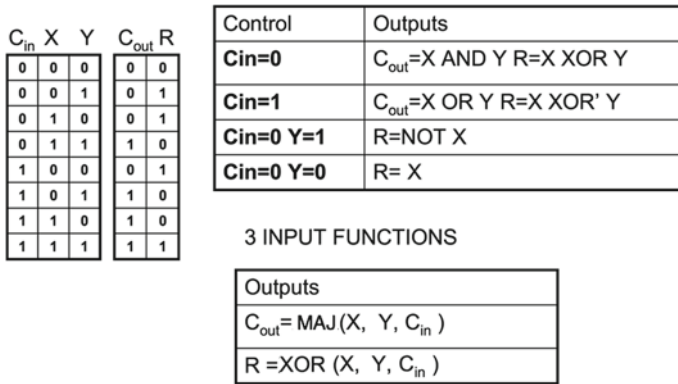


Fig. 13.2 Functions implemented by full-adder

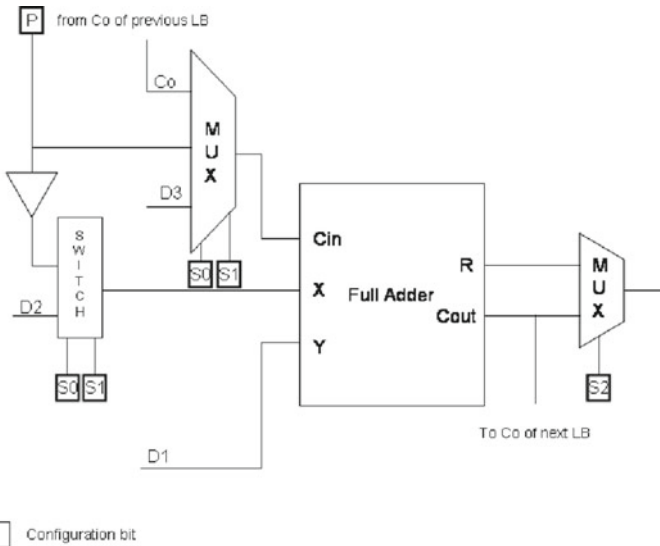


Fig. 13.3 Logic block architecture

and the output multiplexer with the selector are used for programming purposes. In particular inputs S0, S1, S2, and P are used to select the operation to be performed and the operands (Fig. 13.2). The signal  $C_O$  is directly connected to the signal  $C_{out}$  of the previous LB. If a zero carry is required (for example in the case of the LSB of a multibit adder) 0 input is selected by the configuration bits  $S_0$  and  $S_1$ . As shown in [1] an equivalent multicontext LUT based LB requires  $160 * N + 20$  transistors where N is the number of context meanwhile our LB requires  $65 + N_{FA} + 24 * N$  transistor where N is the number of context meanwhile and  $N_{FA}$  is the number of transistor of the Full Adder. For a 16 context array of  $32 \times 3$  Logic Elements

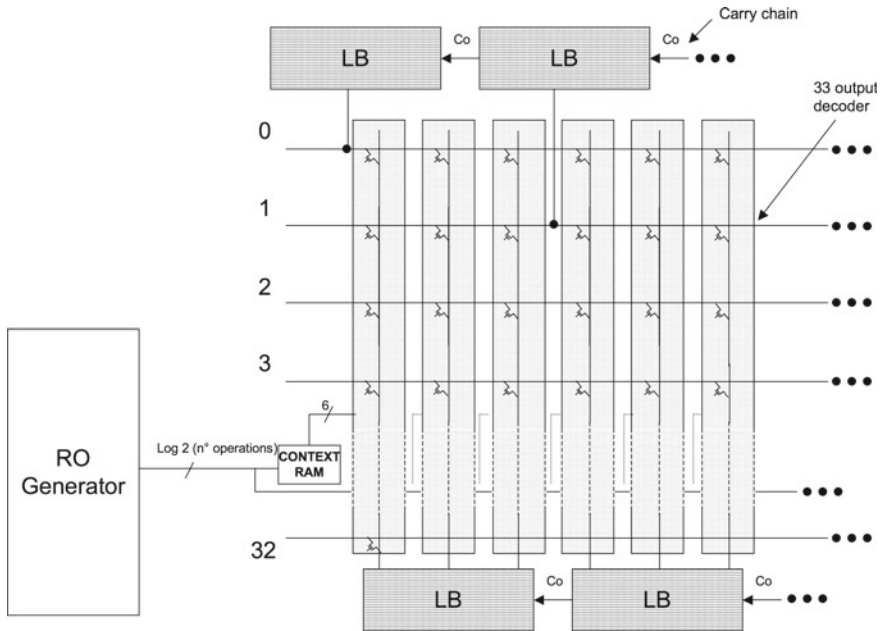


Fig. 13.4 Structure of the reconfigurable interconnect network

the LUT based architecture will require about 250,000 transistors, while the Full Adder implementation requires less than 44,000. Moreover an additional penalty for the LUT based approach is a greater bitstream size. In fact for the same above configuration, LUT based architecture requires 24,800 configuration bits meanwhile the adder based architecture uses only 6,200 bits.

### 13.2.2 Interconnect Structure

The interconnection structure is shown in Fig. 13.4. This structure is based on a multicontext approach (in order to allow high reconfiguration speed). Each LB output can be linked with any inputs of the LBs belonging to the stripe on the bottom row. In addition to the 32 inputs coming from the upper LBs, we have also added two additional line for interconnect configuration linked to the 0 and the 1 logic value in order to realize shifting operations with 1 or 0 insertion and operations on constant value.

Interconnect is based on pass-transistor devices. Multicontext configuration bits are stored in local memories. Consequently, the control of interconnect reconfiguration requires only few lines for carrying the address of the multicontext memories. For  $N$  contexts, only  $\log_2(N)$  addressing lines are required. Moreover, since only one of 34 transistors of the interconnect column is activated at each time (an input pin

must be connected to a single signal source), it is possible to use a column decoder in order to reduce the size of the multicontext memory. Carry chain uses a direct interconnect (linking adjacent LBs) for the speeding-up of the carry propagation in multibit adders.

### 13.3 Modular Operations

In this section two examples of how to implement modular operation by using ADAPTO are presented. The first operation is the modular addition of two numbers, i.e.  $R = X + Y \bmod M$ . The second operation is the Montgomery multiplication, i.e.  $R = A \cdot B \cdot 2^{-n} \bmod M$ . The first example show how the intrinsic parallelism of ADAPTO can be exploited to perform different modular additions in parallel, as we will show in the next subsection. Instead, the implementation of a step of the Montgomery algorithm as a single ADAPTO operation how to exploit the three operands architecture to accumulate the partial results of a complex algorithm.

#### 13.3.1 Modular Addition

Usually the addition is defined between operand that are yet in the range  $[0 \dots M-1]$  and the result belongs to the same range. A simple way to perform this operation is to add the two numbers  $X$  and  $Y$  and, if the result is greater than the modulo  $M$ ,  $M$  is subtracted to the result of the sum. This approach is reported in the following Algorithm 1.

---

#### Algorithm 1: Modular Addition

---

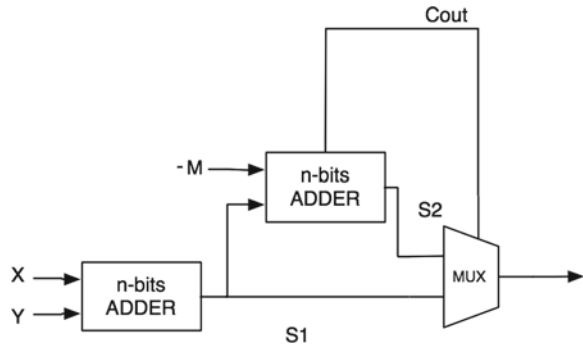
**Input:**  $M, X < M, Y < M$

**Output:**  $(X + Y) \bmod M$ .

- 1:  $R \leftarrow X + Y$
  - 2: **if**  $R \geq M$  **then**
  - 3:    $R \leftarrow R - M$
  - 4: **end if**
  - 5: **return**  $R$
- 

This algorithm can be implemented both in software and in hardware. A typical hardware implementation is reported in Fig. 13.5.

In the adder implementation the possible results  $S1 = X + Y$  and  $S2 = X + Y - M$  are computed and the mux selects between  $S1$  and  $S2$  depending on the value of the signal Cout. But Cout is the carry result of the operation  $S2 = X + Y - M$  and therefore is equal to 1 only if  $X + Y - M < 0$ . This condition corresponds to select  $S1$  if  $X + Y < M$ ,  $S2$  otherwise. The **if** condition in Algorithm 1 corresponds to the

Fig. 13.5 Modulo  $M$  adder

mux in the hardware implementation of Fig. 13.5. Now we discuss how to implement this operation by using ADAPTO. The first adders slice can compute  $S1 = X + Y$  by using  $X$  and  $Y$  as the two input operands. The second adder slice computes  $S2 = S1 - M$ . It must be noticed that  $M$  is seen as a constant and therefore is given as input to the adder by programming the first interconnect slice in order to provide  $-M$  as the second input. The input corresponding to  $-M$  is the binary representation of  $2^{n+1} - M$ , where  $n$  is the number of bits used to represent  $M$ . These two stages are identical to the adders presented in Fig. 13.5. In the third stage the adder should select between  $S1$  and  $S2$  depending from the value of  $Cout$ . Unfortunately, ADAPTO can not send to the third adders slice both the results of the first and the second stage ( $S1$  and  $S2$ ) and therefore the mux in Fig. 13.5 cannot be directly implemented. To supersede this limitation we configure the third slice to perform the following operation:

---

```

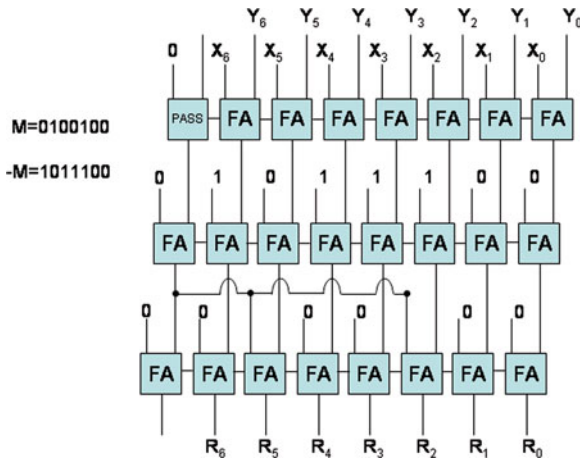
if  $Cout = 1$  then
   $R \leftarrow S2 + M$ 
else
   $R \leftarrow S2$ 
end if

```

---

In this way the final result is  $R = X + Y - M$  if  $X + Y \geq M$ , otherwise  $R = (X + Y - M) + M = X + Y$ . Now we take the binary representation of  $M = m_{n-1} \dots m_0$  and compute the bitwise AND between the bits of  $M$  and  $Cout$  obtaining  $CM = Cout \cdot m_{n-1} \dots Cout \cdot m_0$ . The conditional operation described before can be performed by the third stage of ADAPTO as  $R \leftarrow S2 + CM$ .  $M$  is constant and therefore the operand  $CM$  corresponds to give as second input of the  $i$  FA  $Cout$  if  $m_i = 1$ , zero if  $m_i = 0$ . In Fig. 13.6 the configuration of ADAPTO performing the modular addition is presented.

The adder performs addition modulo 45 (101101 in binary), the first stage is a standard adder and the second stage is an adder with a constant input ( $-M$ ). We



**Fig. 13.6** Configuration of ADAPTO performing the modular addition

can see that in this stage the *Cout* of the second stage is presented as input of the third stage by using a FA of the second stage as a routing element. In fact, ADAPTO does not allow to directly provides as output both the carry and the output of the FA, but only one of the two outputs. So, the *Cout* is given as input by the next FA with  $a_{n+1} = 0$  and  $b_{n+1} = 0$  and the sum result is taken as output. In this way *Cout* has been routed to the next stage. The use of a logic resource as a routing element is widely used in FPGA, in which the LUT is configured as pass-thru, to saving routing resources or to form a shortest path between two points of the FPGA. In our case the use of the FA is mandatory to route both the sum and the carry of a FA to the next stage. Instead, a modification of the interconnect matrix of the architecture of ADAPTO will require a doubling in the number of transistors need to route both the outputs. The third stage is a standard adder that has as inputs  $S2$ , i.e. the result of the second stage, and  $CM$ . The final result is therefore the modular addition between the two inputs. We notice that the 32 bits width of ADAPTO allows performing different one modular additions in parallel. For example the ADAPTO engine can be configured to implement four different modulo additions with moduli of 6 bits or three modular additions with two 9 bits moduli and one 8 bit modulus.

### 13.3.2 Montgomery Multiplication

In [11] Peter Montgomery proposed a method for avoiding expensive reductions modulo  $M$  after multiplication modulo  $p$ . It uses the so-called Montgomery representation for integers. The Montgomery representation of an integer  $a \in [0, M - 1]$  is  $A \cdot Z^{-1} \bmod M$  where  $Z > M$  such that  $\gcd(Z, M) = 1$ . The Montgomery mul-

multiplication is defined as  $R = A \cdot B \cdot Z^{-1} \bmod M$  and its computation is particularly simple if  $Z = 2^{-n}$ , where  $n$  is the number of bits needed to represent  $M$ .

The Montgomery representation does not give any computational advantage for a single multiplication. Instead, when several multiplication should be performed, the Montgomery representation can give a gain due to the few computing resources needed to perform the Montgomery multiplication. Hence, Montgomery representation is useful in modular exponentiation, operations performed on ECC etc. In these complex operations the integers are firstly converted in the Montgomery representation, after the sequence of operations are performed in the Montgomery domain, and finally, the result is reconverted in the traditional integer representation.

The conversion between the traditional integer representation of a number  $A$  and the Montgomery representation is the Montgomery multiplication between  $A$  and 1, i.e.

$$R = A \cdot 1 \cdot Z^{-1} \bmod M = A \cdot Z^{-1} \bmod M$$

Instead, the reverse conversion is the Montgomery multiplication between  $A$  and  $Z^2$ , i.e.

$$R = A \cdot Z^2 \cdot Z^{-1} \bmod M = A \cdot Z \bmod M$$

The algorithm that computes the Montgomery multiplication is presented in the following Algorithm 2.

---

**Algorithm 2:** Montgomery Multiplication
 

---

**Input:**  $M, A < M, B < M, n$

**Output:**  $A \cdot B \cdot 2^{-n} \bmod M$ .

```

1:  $R \leftarrow 0$ 
2: for ( $i = 0; i < n; i++$ ) do
3:    $R \leftarrow R + B \cdot a_i$ 
4:   if  $R$  is odd then
5:      $R \leftarrow R + M$ 
6:   else
7:      $R \leftarrow R$ 
8:   end if
9:    $R \leftarrow R/2$ 
10: end for
11: return  $R$ 

```

---

The core of the algorithm is represented by lines 3–9, i.e. the context of the **for** loop. We will show how to configure the ADAPTO engine to perform the operations inside the loop in one clock cycle. Similarly at the previous case, we rewrite the algorithm in order to avoid the use of the **if then** construct. The rewritten version of the algorithm is proposed as Algorithm 3.

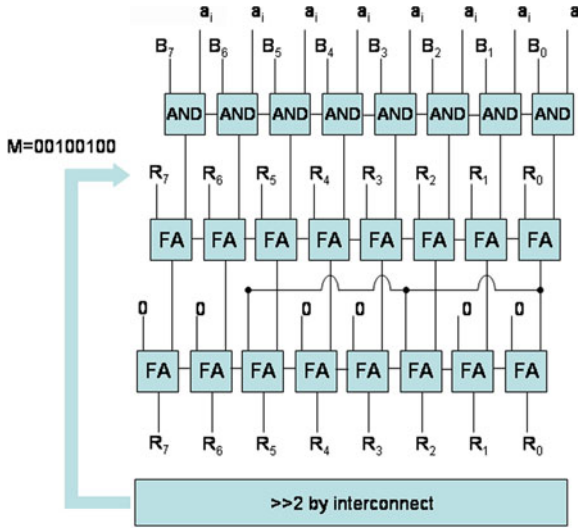


Fig. 13.7 Configuration of ADAPTO performing a step of the montgomery multiplication

---

**Algorithm 3:** Montgomery Multiplication for ADAPTO

---

**Input:**  $M, A < M, B < M, n$

**Output:**  $A \cdot B \cdot 2^{-n} \bmod M$ .

- 1:  $R \leftarrow 0$
  - 2: **for** ( $i = 0; i < n; i ++$ ) **do**
  - 3:  $S1 \leftarrow B \cdot a_i$
  - 4:  $S2 \leftarrow S1 + R$
  - 5:  $S3 \leftarrow S2 + M \cdot S2[0]$
  - 6:  $R \leftarrow S3/2$
  - 7: **end for**
  - 8: **return**  $R$
- 

Line 3 of the revised algorithm can be implemented by the first slice of ADAPTO simply configuring the FAs as an AND between the bits of the operand B and the  $i$  bit of the operand A. The second slice of FAs uses the third input of ADAPTO in order to accumulate the partial results of the Montgomery operation. In particular the third input corresponds to the partial result obtained at the end of the previous loop cycle. The third slice of adders implements the conditional sum expressed by lines 4–8 of Algorithm 2 in a way that is similar to the one used for the modular addition. The addition is performed masking a constant value (the modulo  $M$ ) with a bit corresponding to the control value of the conditional statement. In this case this bit is the least significant bit of  $S2$  and allows identifying if  $S2$  is even or odd. If  $S2$  is odd  $M$  is added to  $S2$ , else  $S2$  remains unchanged. Finally, the interconnect matrix after the third adder slice implements the right shift of the result  $S3$ , diving  $S3$  by two. The



output  $R$  of this operation is given to a register of the register file that will be provided as the third input of ADAPTO at the next iteration of the loop. At the end of the loop the result stored in  $R$  corresponds to the result of the Montgomery multiplication. In Fig. 13.7 the configuration of ADAPTO performing the Montgomery multiplication is presented.

In Fig. 13.7 we can see the configuration of the first slice as an array of AND gates that takes as inputs the operand  $B$  and the bit  $a_i$  of the operand  $A$ . This slice performs line 3 of Algorithm 3. The second slice performs line 4 of Algorithm 3, while the third slice performs the line 5 of Algorithm 3. We notice that, differently from the modular addition, in this case the  $M$  value is masked by the less significant bit of the result of the previous slice (i.e.  $S2[0]$ ). Finally, the last interconnect stage perform the division by 2 shifting right its input.

## 13.4 Conclusions

This paper describe how to use a dynamic reconfigurable architecture (ADAPTO) that can be embedded in microprocessors or low cost DSPs to accelerate the execution of modular arithmetic operations. This architecture is based on the use of a Full Adder instead of using a LUT, as the basic configurable logic element of the RFU. The combined use of slices of Full Adders and interconnection matrices allows realizing a flexible architecture using a limited silicon area with respect to other solutions. The use of ADAPTO allows an high speed execution of some basic modular arithmetic operations such as the modular addition and the Montgomery multiplication algorithms. In particular, in this paper we shown that ADAPTO can performs in a clock cycle, different modular additions, with different moduli, in parallel. This is very useful when RNS [4] representation is used. Instead, for the Montgomery multiplication the proposed architecture is able to reduce the execution time performing in only one clock cycle the core of the routine computing the Montgomery multiplication.

## References

1. Cardarilli, G.C., Di Nunzio, L., Re, M.: Arithmetic/logic blocks for fine-grained reconfigurable units. In: IEEE International Symposium on Circuits and Systems (2009)
2. Cardarilli, G.C., Di Nunzio, L., Re, M.: High performance reconfigurable blocks for real-time reconfigurable unit (ADAPTO) ReCoSoc., 9–11 (2008)
3. Cardarilli, G.C., Di Nunzio, L., Re, M.: A full-adder based reconfigurable architecture for fine grain applications: ADAPTO. In: IEEE International Conference on Electronics, Circuits, and Systems 2008
4. Di Claudio, E.D., Piazza, F., Orlandi, G.: Fast combinatorial RNS processors for DSP applications. IEEE Trans. Comput. **44**(5), 624–633 (1995)
5. Daly, A., Marnane, W., Kerins, T.: An FPGA implementation of a GF(p) ALU for encryption processors. Microprocess. Microsyst. **28**(5—6), 253–260 (2004)

6. Fayed, A.A. Bayoumi, M.A.: A low power 10-transistor full adder cell for embedded architectures. In: The 2001 IEEE International Symposium on Circuits and Systems, ISCAS
7. Knuth, D.E.: The Art of Computer Programming. Reading, MA (1973)
8. Li, B., Gupta, R.: Bit section instruction set extension of ARM for embedded applications. In: International Conference on Compilers, Architecture, and Synthesis of Embedded Systems (CASES) 2002
9. Mamidi, S. Blem, E.R. Schulte, M.J. Glossner, J. Iancu, D. Iancu, A. Moudgill, M. Jinturkar, S.: Instruction set extensions for software defined radio on a multithreaded processor. In: ACM Proceedings of the 2005 International Conference on Compilers, Architectures and Synthesis for Embedded Systems, 266–273 2005.
10. Menezes, A.J., Van Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press, Boca Raton (1997)
11. Montgomery, P.L.: Modular multiplication without trial division. *Math. Comput.* **44**(170), 519–521 (1985)
12. Peterson, W.W., Weldon, E.J.: Error-Correcting Codes. The MIT Press, Cambridge (1972)
13. Razdan, M.D., Smith, R.: A high-performance microarchitecture with hardware programmable functional units. In: Proceedings of MICRO-27, 172–180 Nov 1994
14. Vesterbacka, M.: A 14-transistor CMOS full adder with full voltageswing nodes, SiPS 99. In: IEEE Workshop on Signal Processing Systems 713–722

# Chapter 14

## Wireless and Ad Hoc Sensor Networks: An Industrial Example Using Delay Tolerant, Low Power Protocols for Security-Critical Applications

Claudio S. Malavenda, Francesco Menichelli and Mauro Olivieri

**Abstract** This paper introduces an industrial example of a WSN developed for security-critical applications, in particular, the focus is on the analysis, implementation and experimental testing of a delay tolerant and energy aware protocol for a WSN oriented to security applications. The solution proposed takes advantage from different domains considering as guideline low power consumption and facing the problem of lossy connectivity offered by wireless medium along with very limited resources provided by wireless network nodes. The paper is organized as follows: first we introduce the industrial example that has been used as test-platform, then we give an overview on delay tolerant wireless sensor networking (DTN) and describe the delay tolerant protocol developed. We perform a simulation-based comparative analysis of state of the art DTN approaches and illustrate the improvement offered by the proposed protocol; finally we present experimental data gathered from the implementation of the proposed protocol on a proprietary hardware node.

### 14.1 An Industrial WSN Example

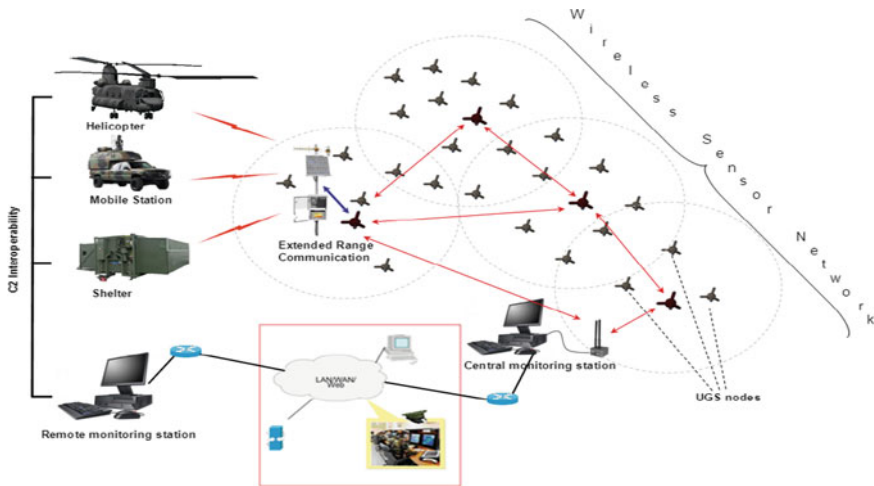
In recent time Wireless Sensor Networks (WSN) [1] have moved rapidly from concept and research to industrialization and actual application in real-life scenarios. As a practical example, we describe an industrial example of WSN used for homeland security known as MasterZone, produced by SELEX Electronic System [14].

---

C. S. Malavenda (✉)  
Selex E.S., via Tiburtina Km. 12,400, 00131 Rome, Italy  
e-mail: claudiosanto.malavenda@selex-es.com

F. Menichelli · M. Olivieri  
Sapienza University of Rome, via Eudossiana 18, 00184 Rome, Italy  
e-mail: menichelli@diet.uniroma1.it

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



**Fig. 14.1** Masterzone architecture overview and configurations

MasterZone, through the use of proprietary Wireless Sensor Nodes and proprietary communication protocols, is designed to guarantee situational awareness and early warning. It supports force protection requirements and civil security needs, through the surveillance of target areas and the detection of hazards in different operational scenarios. The need of situational awareness calls for advanced solutions in support of surveillance and identification in order to derive an accurate common operational picture (COP) for decision makers. This objective is currently achieved by deploying personnel equipped with expensive and sophisticated platforms. These deployments are risky, particularly for the personnel, and expensive, in terms of maintenance cost.

MasterZone is a WSN solution that meets the need of low-cost, low-power consumption, and miniature sensors to ensure easy mass deployment, extended mission lifetime, and hand portability. A possible scenario of interest is depicted in Fig. 14.1. A large quantity of sensor nodes can be deployed to cover a wide area and can routinely collect and report field information to command posts and personnel. MasterZone applications include battlefield and force protection, critical infrastructure protection (airports and runways, industrial sites, utilities), access and border control, and illegal activity monitoring.

The modular sensor node consists of a CPU, a communication board implementing a proprietary communication protocol stack, and a sensor board to be configured in order to host one or more sensing capabilities, according to the context. Within the network, short-range sensor nodes interact with each other, thus creating an ad hoc wireless network. Nodes can automatically aggregate into clusters (short-range communication) and groups of clusters into a network (long-range communication). Within each cluster, a cluster head is elected and is responsible of the data fusion activity. Neighbour nodes are used as routers to convey data and information to the central monitoring station.

In the following sections we present the development of a Delay Tolerant protocol for the Masterzone nodes starting from state of the art research, its implementation on a simulator and performance measures on the final nodes.

## 14.2 DTN Design and Initial Protocol Screening

Delay Tolerant Networking responds to the need of delivering messages in networks characterized by statistical lack of end-to-end connection paths, either proactively available [11] or reactively established with conventional routing protocols. These networks must operate without the assumption that there is a permanent connection or instantaneous end-to-end path between the source and the destination node [3], since disconnections occurring dynamically among nodes are quite common. The main causes of node disconnections can be attributed to Mobility of Nodes and Sparse Network.

Performance metrics are not easy to define in WSN due to its unique properties. Common metrics used in wireless communication, like fairness and throughput, might be not meaningful because WSN nodes can co-operate and because raw data transmission is a rare application in WSN [5]. We used latency, as defined in [7], in order to provide an initial selection among protocols.

The most widely used DTN protocols reported in literature are Direct Diffusion, First Contact, Fuzzy Spray, Prophet, Rapid, MaxProp, Spray and Wait (and variants), Scar, FAD, Epidemic [2, 4, 9, 10, 12, 13]. We performed a comparative analysis of the above protocols based on a commercially available simulator [9].

The results of these simulations are reported in Fig. 14.2 and they can be interpreted selecting as reference three protocols, one for each class of adopted communication scheme: MaxProp, Prophet and Spray and Wait. PROPHET is representative of protocols implementing only the Data Forwarding scheme, Spray and Wait only the controlled replication scheme, MaxProp both. From latency performance we can also note that the selected protocols mark two extremes of a range of latency values, while other protocols are positioned between them according to the scheme implemented. Other protocols having performance outside this range are considered out of interest.

Considering other metrics, as overhead and the delivery ratio [7] and considering the trade-off between performance and power consumption, the Spray and Wait protocol comes out to be the one with lowest overhead while maintaining average results on delivery ratio and delay, in the target application domain. As a consequence of the preliminary analysis, the newly developed protocol has been an optimization of Spray and Wait.

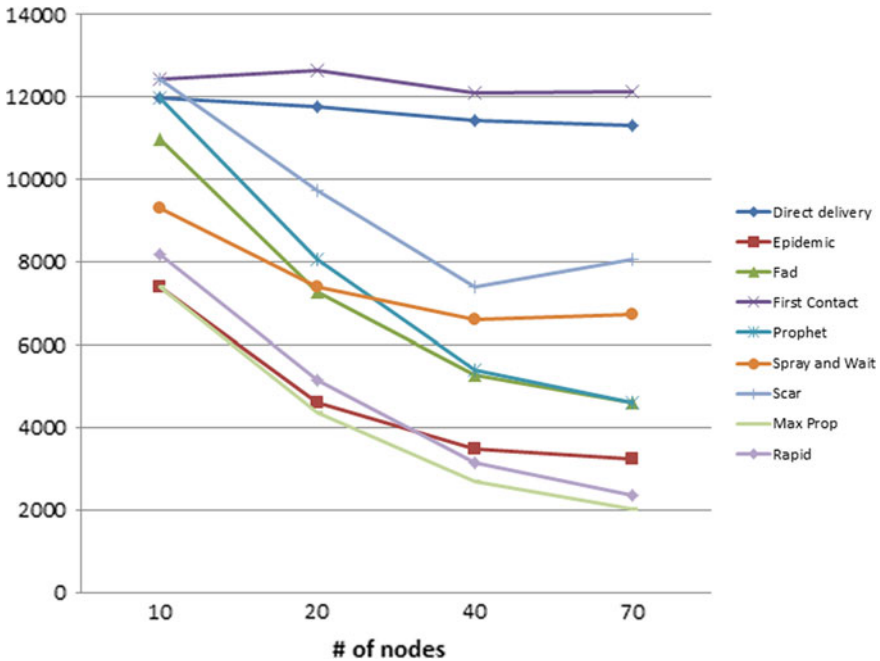


Fig. 14.2 Latency results comparison

### 14.3 Analysis on a Dedicated State Accurate Simulator for DTN Protocols

In order to get a deeper control on the protocol to develop, with state-level accuracy, and in order to produce a better energy model, a custom simulator framework for DTN protocols has been developed. OMNET++ 4.2 [8] has been chosen as starting framework for the new simulator. The simulator aims at modelling, with state-level accuracy, hardware of a WSN node with particular regard to the radio and micro-controller states, in order to produce accurate results on their power consumption. It has been designed to provide a dynamic positioning of WSN nodes over a simulated area. Connections among nodes are dynamically established according to physical parameter relatives to each node, which is modelled with an antenna gain and receive sensitivity. Working frequency is used to model the communication range achievable from each node according to the mutual position of the nodes.

The first use of the simulator consisted in verifying timings on packet delivery and model packet exchange among nodes with a preliminary version of the selected protocol. As it is possible to see from Fig. 14.3 it never occurs that a node starts transmitting while another one in its visibility range is yet in transmission phase. Moreover it is possible to see the packet relay period of 1 s when no collisions occurs, which correctly models the protocol used. Since the protocol used is DTN

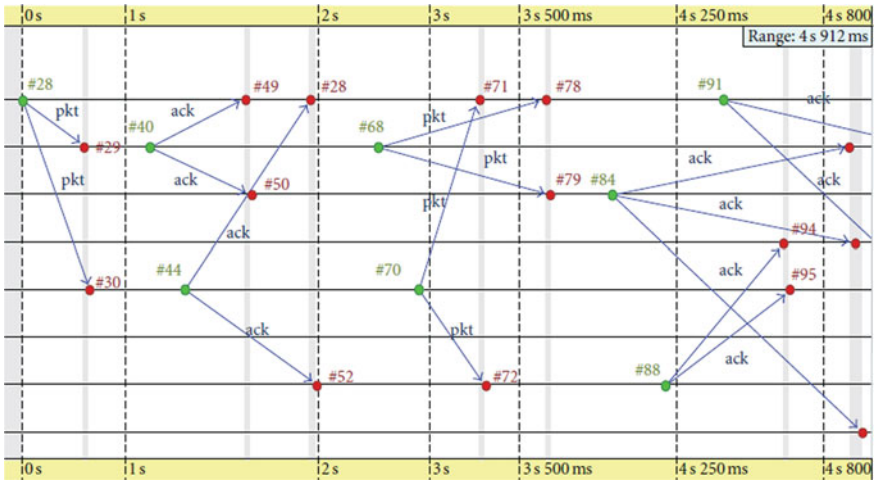


Fig. 14.3 Network timing monitoring view

and well fits for communication among mobile nodes [6], a mobile node modelling feature has been developed and introduced in the framework as well.

### 14.4 Experimental Testing

In this section we report the results and validation of the protocol directly on the physical MasterZone nodes. The tests have the scope of verifying the timing of the node, measuring power consumption and validating the basic handshake. The tests have been divided into different sets.

The first set deals with node power consumption, by analyzing duty cycle and power consumed during different transmission phases. Measurements of current absorbed by the node are performed with a current probe. As from Fig. 14.4, every 800 ms the power consumed shows a high raise due to the state that change from idle/sleep to receiver. Figure 14.5 shows a detail where we can observe the background power consumption of 200 uA in sleep mode and a peak of 22 mA in active receiving mode. Figure 14.6 illustrates the corresponding test for a transmission phase. We can see a first peak of 30 ms with a power consumption of 30 mA for completing the CSMA/CA action at the beginning of the transmission phase, followed by a 900 ms phase during which the node is actually transmitting, consuming 23 mA at -15 dBm of Tx power. It is possible to observe that the whole transmission phase is characterized by ripple in energy consumption. It is due to the fast change of states in the transmitter radio (from idle to transmitter).

The second set of tests targets sensitivity of the node radio receiver and confirms the correct functioning of the CSMA strategy adopted. The status of the radio channel

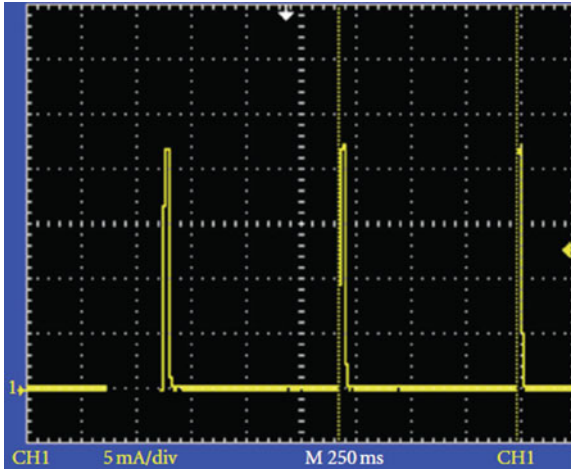


Fig. 14.4 Receiver WOR period—power consumption test result

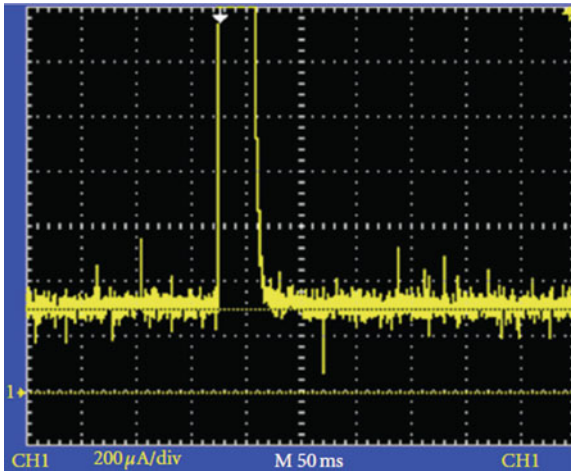


Fig. 14.5 Sleep power consumption test result

in use is monitored and defined as busy or free, according to a predefined threshold on received power. The threshold has been set to the minimum detected value declared on the receiver datasheet, in this way a control output will signal as a logic state when the minimal energy is detected in received channel. The antenna plug of the node has been connected directly to a RF generator. The collected results show a sensitivity of  $-90$  dBm for the node.

The third set of test has been setup using two nodes and the scope is measuring transmission range achievable with a point to point connection. The first node is configured to periodically transmit a packet, the second node is configured to



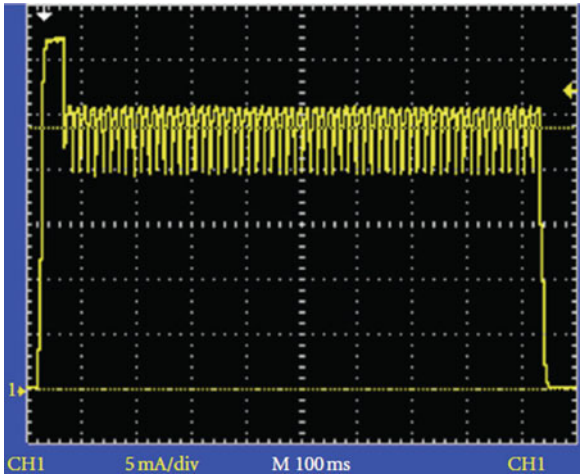


Fig. 14.6 Transmitter power consumption test result

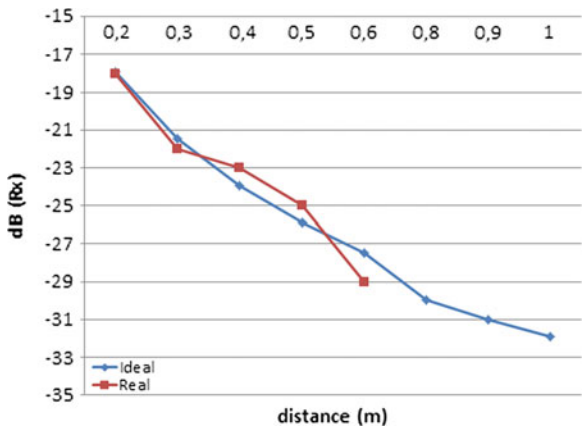
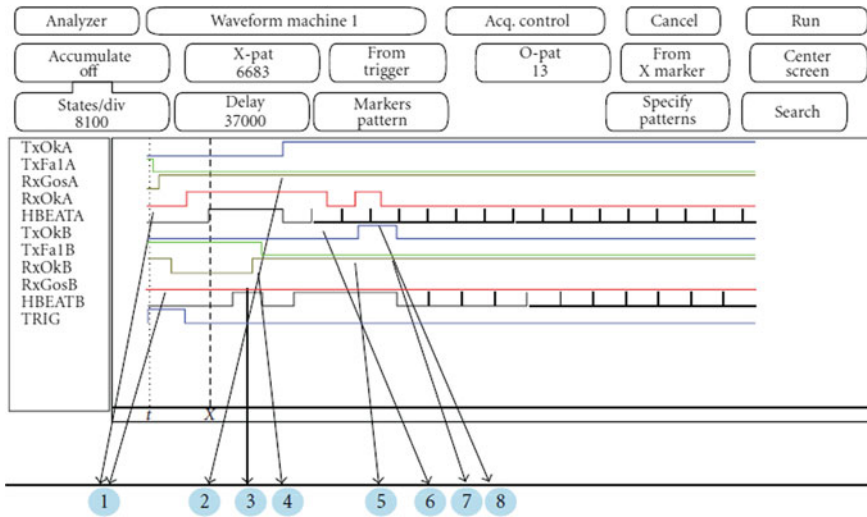


Fig. 14.7 Obtained versus expected measures

stay in reception state, read the RSSI level of received packets and translate it in a dBm values. This translation has been tuned in advance using reference values from datasheet. The node sends the data to a PC via a RS232 serial connection, where they are timestamped and logged. Figure 14.7 shows a plot of the actual measurements towards ideal values. The ideal values (in blue) log expected dBm power at receiver according to Free Path Loss law with a Tx power of 5dBm and a gain antenna value of  $-6$  dBm at a working frequency of 420 MHz. As we can see from the figure, ideal and actual values map almost 1:1 with a few dBm difference. Assuming that the measurements follow this trend, the sensitivity level of  $-90$  dBm may be reached at 800 m distance between transmitter and receiver. More tests should be conducted



**Fig. 14.8** Multi hop signal test

with greater distances between nodes to confirm the trend with distances next to the maximum one achievable.

The fourth session has been setup on a multi-hop testbed and the target parameter has been delay. In this testbed we have two nodes and one sink node. All nodes are in visibility among each others. The test aims at verifying a simple relay functioning. This testbed is setup with two MasterZone node [14] suitably programmed (node A and B) and one node interfaced with a PC (TRIG). The TRIG node transmits ping packets under PC control. This node will not take part in any other radio handshaking. The ping packet received by node A is relayed to node B. A logic state analyzer has been linked to nodes to monitor the handshaking occurring between node A and B. Five I/O pins have been configured on each node to monitor events. The events monitored deal with a successful (TxOk) or failed (TxFail) transmission started from the node, a successful reception (RxOk), or reception of a packet yet stored in the reception queue (ghost packet, RxGos). The signal Hbeat reveals the internal timing of the node. The state analyzer will log all control pin on both nodes in order to catch a clear picture of the handshaking. The test aims at examining if the routing with a minimal set of nodes reflects the expected behaviour. Figure 14.8 reports the result of the test conducted with the configuration just described. The cyan balloons highlight the following communication events.

1. Node A and B receive ping command from the sink node (A receive twice in the same slot)
2. A answer to the sink node with a delay of 2.42 s
3. B receive the answer transmitted by A (the signal toggle monitor the end of a transmission)

4. B tries to forward the ping request issued by the sink node but sense the air occupied
5. B forward the sink request
6. A receives the forwarded request from the sink node and filter it because yet received
7. B transmit the answer from A
8. A receive its own answer from B and just drop it.

## 14.5 Conclusions

In this paper we presented an industrial example of WSN using delay tolerant protocol. First we presented a comparison between Delay Tolerant protocols for WSN systems. We compared different Delay Tolerant protocols through available simulators. After a set of simulation results and comparisons, the most promising one has been selected in order to develop a new custom protocol. In order to reach a more accurate control of the simulation and incorporate a wider set of simulation parameters, the code of the custom protocol based on the selected one has been implemented in the a new simulation environment. A first set of simulation results have been collected with fixed and mobile nodes. These tests have confirmed the suitability of the protocol for an actual implementation. Finally the custom protocol has been ported on a proprietary platform: the correct implementation has been validated through a set of tests on timing, handshaking and power consumption on the physical node, confirming the expected results and paving the way to further development.

## References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. *Comput. Netw.* **38**(4), 393–422 (2002)
2. Demmer, M., Brewer, E., Fall, K., Jain, S., Ho, M., Patra, R.: Implementing delay tolerant networking, intel corporation, <http://www.dtnrg.org/docs/papers/demmer-irb-tr-04-020.pdf>, 2004
3. Delay Tolerant Network Research Group (DTNRG). <http://www.dtnrg.org>
4. Harras, K.A., Almeroth, K.C., Belding-Royer, E.M.: Delay tolerant mobile networks (DTMNs): controlled flooding in sparse mobile networks. In: *Proceedings of the 4th IFIP-TC6 NETWORKING'05*, 2005
5. Kim, J., On, J., Kim, S., Lee, J.: Performance evaluation of synchronous and asynchronous MAC protocols for wireless sensor networks. In: *2nd International Conference on Sensor Technologies and Applications - SENSORCOMM*, pp. 500–506. 2008
6. Malavenda, C.S.: Jaguar 44 UGV: an autonomous deployment of a wireless sensor network. In: *International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)* (2012)
7. Malavenda, C.S., Menichelli, F., Olivieri, M.: Delay-tolerant. low-power protocols for large security-critical wireless sensor networks. *Hindawi Sens. J.* (2012)
8. OMNET++, <http://www.omnetpp.org/>

9. ONE simulator web page <http://www.netlab.tkk.fi/tutkimus/dtn/theone/>
10. Pasztor, B., Musolesi, M., Mascolo, C.: Opportunistic mobile sensor data collection with SCAR. In: Proceedings of the IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS), 2007
11. Proactive and Reactive Routing in Wireless Sensor Networking: <http://it.wikipedia.org/wiki/MANET>
12. Spyropoulos, T., Psounis, K., Raghavendra, C.S.: Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In: WDTN '05 Proceedings of ACM SIGCOMM workshop on Delay-tolerant networking, 2005
13. Vahdat, A., Becker, D.: Epidemic routing for partially connected ad hoc networks. Technical report CS-2000-06, Department of Computer Science, Duke University, April 2000
14. [www.selex-si-uk.com/pdf/Masterzone.pdf](http://www.selex-si-uk.com/pdf/Masterzone.pdf)

# Chapter 15

## A Social Serious Game Concept for Green, Fluid and Collaborative Driving

Francesco Bellotti, Riccardo Berta and Alessandro De Gloria

**Abstract** People spend daily a significant amount of time in cars, and vehicular mobility has remarkable social implications (in particular traffic and pollution). In this context, there is room for drivers can improve their own behavior, also in a common good perspective. This position paper presents a new type of serious gaming application based on the cloud. The serious game processes vehicular data in order to reward and coach the driver. Scores and analytics are computed, and displayed on the automotive dashboard and on smartphone screens, keeping into account the simultaneous presence of various vehicles and stimulating behavior enhancement. The SG has been specified and its development has started both on the client and server side. In a user-centered design perspective, the next steps of development of the application will involve early simulations and user tests in lab in order to check fulfilment of requirements and verification of end-user acceptance.

### 15.1 Introduction

In the recent years, we have seen the rapid surge of gaming and social networking. The potential of Serious Games (SGs)—games designed with a primary goal different than pure entertainment [1–3]—is relevant, because a large population is familiar with playing games. Through gaming, the learning is applied and practiced within that context (situated cognition) [SG4]. This is particularly promising for a domain like green/safe driving, where the user (driver/passenger) is in the field and could exploit this experience to improve his driving habits and performance.

---

F. Bellotti · R. Berta · A. De Gloria (✉)  
Department of Naval Architecture, Electrical, Electronic and Telecommunications  
Engineering, Via Opera Pia 11/a, 16145 Genoa, Italy  
e-mail: adg@unige.it

The TEAM EU FP7 project<sup>1</sup> is developing new, cloud-based collaborative transport solutions combining driving and sophisticated Information and Communication Technologies (ICTs). This involves integrating elements such as vehicle electronics and mobile devices, navigation systems, tablet computers and smartphones, to improve the road users' behavior and experience.

Road users will benefit from the new TEAM technologies through real time traffic recommendations where the self interest is balanced with global mobility and environmental aspects. In this way, TEAM turns static into elastic mobility by joining drivers, travelers and infrastructure operators into one collaborative network. Collaboration is the key concept, which extends the cooperative concept of vehicle-2-x systems [4, 5] to include a certain degree of driver interaction and participation towards shared goals.

TEAM involves the design and development of several different user applications, ranging from Collaborative Navigation to Collaborative Parking, from support for public transport intermodality to collaborative Adaptive Cruise Control, etc. One of these new elastic mobility applications is a SG for green and fluid driving. This is a position paper presenting the main features of this new gamified [6] ICT tool for drivers.

## 15.2 Related Work

There is a certain consensus in literature about the instructional potential of games, mostly because they are considered inspiring and motivating [7, 8]. Several games are being successfully used in various application domains (e.g., [9–11]).

Gamberini et al. [12] discuss their experience in designing a persuasive serious game for power conservation. The paper addresses in particular the points of user assessment and provision of feedback and hints for good performance.

There are some examples also in the domain of mobility, such as Chromaroma [13] by the London Underground. The popular term now is gamification, which concerns the use of game design techniques and mechanics to solve problems and engage audiences [6, 14]. Typically, gamification applies to real-world processes and behaviour, in order to foster people to adopt them. Despite the immediate appeal, there are also some concerns about gamification [15]. In any case, proper pedagogical mechanisms are necessary in order to make games effective for actual instruction. Thus, a meaningful gamification of a system is a far from trivial challenge [16].

The main OEMs have assistant systems to inform drivers about optimal gear changes, and acceleration behaviour and have partially related them to incentives and game-like approaches. Examples are Fiat eco:Drive system [17], Ford Fusion and Mercury Milan EcoGuide with the SmartGauge coaching system (including the growing leaves/vines metaphore to show good driver behaviour) [18], BMW Eco Pro [19], Honda insight Eco Assistant [20]. The Chevrolet Volt's dashboard, called Driver Information Center, provides very pleasant real-time feedback on the driver's

---

<sup>1</sup> See Ref. [4]

driving style, which looks particularly suited to behavior change [21]. The system displays a ball that animates and changes color (e.g., yellow for sudden braking) based on a car's acceleration or deceleration.

Nokia has recently presented the "Routine Driving" infotainment app [22], applying gamification to routine driving in order to transform into performance driving, which is driver training focused on developing optimal vehicle handling skills appropriate to the road terrain. The app provides real-time feedback of how well the driver is driving through a role play game interface, named "Driving Miss Daisy". The app collects driving data such as car speed from OBD, accelerometer readings from the smartphone, altitude from smartphone's GPS, and speed limit of the current road from Nokia's maps API service [23]. The game runs inside a Web browser on the smartphone, as the app involves mash-up of data and functionality from the smartphone, the car, and the cloud. The phone is connected to a MirrorLink-enabled head unit via USB [24]. FleetFlot is a smartphone game designed to promote ecodriving skills [25]. With fleetfoot actions such as smoother acceleration or more anticipation allow earning credits to spend in the game world, for instance to upgrade your customer avatars. Races are organized to find out the best skilled drivers.

Fiat eco:drive Mobile connects the in-car software and data with a smart phone, allowing an immediate analysis of the driver's performance. This new version also includes functionalities for social networking, as the possibility to share results through Facebook and Twitter, creating the possibility of rewarding the best "eco:drivers" with virtual badges and real prizes. Communities are expected to emerge for various driver categories. All the eco:drive users are part of the "eco:Ville" online community in which the eco:drivers savings are collected. The community is continuously growing, with 87,000 users, who have saved a total of 4,900 tons of CO<sub>2</sub> by improving the efficiency of their driving style (as of beginning of 2011).

Meschtscherjakov et al. [26] made an interesting study aimed at evaluating the user acceptance of five different persuasive in-car interfaces designed to support a fuel-efficient driving style. Inbar and Tractinsky [27] discusses how to motivate eco-driving through in-car gaming, highlighting the importance of challenging and competitive situations.

The I-GEAR project is studying incentives for drivers, in a game-like environment, for improving traffic conditions by promoting good drivers' behaviors. A preliminary paper, [28], focuses on ethics, privacy and trust aspects. Nunes et al. [29] argues that using social networks for exchanging real time public transport information among travelers (e.g., punctuality, noise levels, schedules, quality of the transport means, etc.) can be very effective. Social networks would provide an easy way of sharing information and also provide a sense of community to the involved travelers. They also propose the concept of a smartphone social application, with a game structure of crowd rating and rewards.

The Sunset project is a 3 year project kicked-off in 2011 [30]. Its main goals concern the study of social services that motivate people to travel more sustainably in urban areas; the study of Intelligent distribution of incentives (rewards) to balance system and personal goals; the development of algorithms for calculating personal mobility patterns using information from mobile and infrastructure sensors.

Concerning driver profiling, some solutions have recently been proposed, in particular in the field of insurances and fleet management. For instance, MyDrive Solutions have developed black boxes and apps that claim to be able to offer fairer, more personal motor insurance, and can even help to improve driving. They build a driver profile which scores a person's driving according to five categories (consistency, pace, anticipation, calmness, smoothness), before providing a weighted average score, namely the Expert Driver Score [31]. Blackbox Telematics offers a similar device and service, the "Green Driving EcoRisk System", which is typically used for fleet management [32]. Telekom Austria Group M2M, Fela Management AG and G4S Security Systems have entered into a partnership, developing the Eco Driving Solution, that aims at enabling companies to reduce costs by gathering and analyzing a broad range of data around both drivers and vehicle fleets [33]. Driving style parameters like harsh breaking or acceleration, out of hour's usage or idling of the engine is being aggregated, analyzed and compared by a central system. The results can be given to drivers as direct feedback and help the fleet manager to determine the need for training.

The expected added value of the TEAM social gaming application with respect to the above is its real-time driver information and feedback (coaching) and the game-based green drive supporting collaboration and competition.

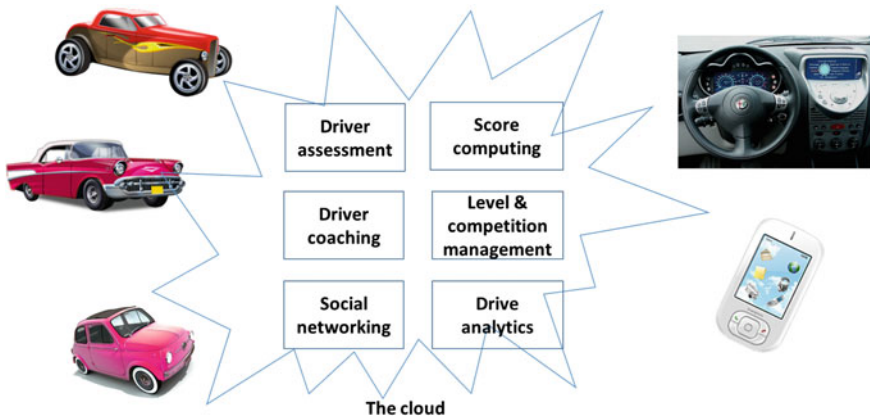
Different to existing solutions like eco:drive Mobile, it enables open communities, not directly related to any OEM. As a limited example of such a huge potential, we can cite Waze [34], a social GPS application that provides free navigation and allows the user to become part of the local driving community in his area, joining forces with other drivers nearby to outsmart traffic, save time and improve everyone's daily commute. Waze use cell phone data. Thus, a major innovation by TEAM is given by the integration and exploitation of the actual vehicle data. The Nokia "Driving Miss Daisy" app already exploits distributed web-services, also including car data, but the TEAM application will make a much more extensive use of vehicle signals and try to devise vehicle-independent assessment techniques, so to the evaluate the real capabilities of the driver, independent of the actual driven car. The black box solutions recently appeared for driver profiling are very interesting and we will develop a similar approach but for real time analysis and feedback for the driver in a pleasant human interaction environment such as gaming, with the added value of real-time coaching.

## 15.3 The Concept

### 15.3.1 Goal

The goal of the application is to promote a proper driver behaviour. The SG should not involve direct competitions among drivers. Rather, it should spur the driver to continuously improve his performance and to cooperate with others in order to keep the





**Fig. 15.1** Sketch of the TEAM serious game

overall levels of pollution and traffic low. In particular, the SG will offer a challenge to stimulate drivers with respect to two main aspects:

- Green driving
- Traffic reduction/avoidance (fluid traffic).

This will be achieved in various ways. The SG application consists of a gamified social network environment where drivers and passengers can share their mobility information and improve their driving performance, in a pleasant and compelling way and featuring a map-based user interface.

The SG exploits vehicle's data in order to create a challenge so that drivers are motivated to collaboratively reach high levels of green driving and low levels of traffic in their zones (typically a city or a city area).

### ***15.3.2 Basic Functioning***

While the user is driving, the application processes vehicular data about the travel in real-time. Each user may be able to insert geo-referenced messages inside the social map environment, when the vehicle is not moving. Other messages could be automatically sent by the car (e.g., windscreen wipers, temperature, airbag, speed) also during the drive, if the user allowed it. This will allow creating and displaying on the map integrated information collected through the vehicles (this is advances the current Waze.com social driving application, which only relies on cell-phone data). Selected (i.e. relevant to the driver/passenger) notifications may appear on the map during the drive.

The green/fluid drive SG will process vehicle data about the travel and provide serious game feedback in real-time, but with a very limited impact on the driver's

cognitive workload in order not to overload nor distract him from his primary task. To this end, the user interface will be very simple and configurable by the user.

In particular, we identify the following levels of data display—data are indexes of green and fluid drive computed through state of the art tools (e.g., [35]):

- an avatar whose look shows the driver's performance level
- indicators showing the contribution of the driver to the current average values at segment, road and city level
- indicators showing the current offset with respect to the average values (at the same times) of the current segment, street and city.

The last one is a cooperative measure, since it is an average resulting from the performance of several vehicles in the area. Moreover, the traffic fluidity is itself a cooperative quantity, as it depends on all the vehicles in the area.

All the presented values can be real time data, averages over a time window, or averages since the beginning of the travel. Information is provided through smart-phones apps and configurable automotive dashboards [36–38]. Standings, charts and detailed analytics are provided in specific tabs of the app or pages of the accompanying social website.

A cumulative score is computed, since the beginning of the trip (or in path-specific competitions), averaging the values of the various indicators. So, the driver's overall performance evaluation will combine personal and cooperative values/aspects. Moreover, social features (e.g., considering teams of friends) can be considered.

Badges will be assigned to good performers, based on various criteria (e.g., time, space/area, friends, common interest, type of vehicle, etc.). Game levels will be introduced in order to consider ever more complex variables and settings. Incentives may be provided in terms of virtual gadgets/facilities and of real-world rewards, such as access to pool lanes, discounts for parking costs, free bus tickets, etc. The system will exploit a user model for driving and a user credibility management system for the information provided in the social environment.

The driver will also be coached in real-time by the system, exploiting expert knowledge and statistically processed information from other drivers, on how to improve his performance.

The application involves also significant privacy and security aspects, that will not be addressed in the first prototyping phase.

## 15.4 Future Work

People spend daily a significant amount of time in cars, and vehicular mobility has remarkable social implications (in particular traffic and pollution). In this context, there is room for drivers can improve their own behavior, also in a common good perspective. This position paper has presented a new type of serious gaming application based on the cloud. The serious game processes vehicular data in order to reward and coach the driver. Scores and analytics are computed, and displayed on the

automotive configurable dashboard and on smartphone screens, keeping into account the simultaneous presence of various vehicles and stimulating behavior enhancement.

After the first year of TEAM the SG has been specified and its development has started both on the client and server side. Given the need to consider a number of vehicles in the traffic, we are creating a tool chain exploiting the OpenDS driving simulator [39] and the SUMO traffic simulator [40]. They are both well established open source tools, which is a key requirement in order to adaptation and flexibility with no license costs. For the social networking management, we are now selecting a framework able to support efficient development.

In a user-centered design perspective, the next steps of development of the application will involve early simulations and user tests in lab in order to check fulfilment of requirements and verification of end-user acceptance.

**Acknowledgments** This work was partly funded by the EC under the FP7-ICT, GA No. 318621.

## References

1. Prensky, M.: Digital game-based learning. *ACM Comput. Entertainment* **1**(1), 21 (2003)
2. Zyda, M.: From visual simulation to virtual reality to games. *IEEE Comput.* **38**(9), 25–32 (2005)
3. Bellotti, F., Berta, R., De Gloria, A.: Designing effective serious games: opportunities and challenges for research. *Int J. Emerg. Technol. Learn. (IJET)*, (Special Issue: Creative Learning with Serious Games) **5**, 22–35 (2010)
4. Martelli, F., Renda, M.E., Santi, P.: Measuring IEEE 802.11p performance for active safety applications in cooperative vehicular systems. In: 73rd IEEE Conference on Vehicular Technology (VTC Spring), vol. 1(5), pp. 15–18 (2011)
5. Ibanez, A.G.; Flores, C.; Reyes, P.D.; Barba, A.; Reyes, A.: A performance study of the 802.11p standard for vehicular applications. In: 7th International Conference on Intelligent Environments (IE), pp. 165, 170, 25–28 July 2011
6. Deterding, S., Bjoerk, S., Dixon, D., Nacke, L.E.: Designing Gamification: Creating Gameful and Playful Experiences, CHI 2013 Workshop. <http://gamification-research.org/chi2013/>
7. Van Eck, R.: Digital game-based learning: it's not just the digital natives who are restless. *EDUCAUSE Rev.* **41**(2), 16 (2006)
8. Iacovides, I.: Exploring the link between player involvement and learning within digital games. In: Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology, Cambridge, UK, Sept 2009
9. Wouters, P., van Oostendorp, H., van Nimwegen, C., van der Spek, E.D.: A meta-analysis of the cognitive and motivational effects of serious games. *Comput. Educ.* **60**(1), (2013)
10. Cole, S.W., Yoo, D.J., Knutson, B.: Interactivity and reward-related neural activation during a serious videogame. *PLoS ONE* **7**(3), e33909 (2012)
11. Alankus, G., May, M., Lazar, A., Kelleher, C.: Towards Customizable Games for Stroke Rehabilitation. In: Proceedings of CHI 2010, Atlanta, GA, USA (2010)
12. Gamberini, L., Corradi, N., Zamboni, L., Perotti, M., Cadenazzi, C., Mandressi, S., Jacucci, G., Tusa, G., Spagnolli, A., Björkskog, C., Salo, M., Aman, P.: Saving is fun: designing a persuasive game for power conservation. In: *Advances in Computer Entertainment Technology*, vol. 16 (2011)
13. <http://www.chromaroma.com/>
14. Deterding, S., Sicart, M., Nacke, L., O'Hara, K., Dixon, D.: Gamification. Using game-design elements in non-gaming contexts. In: Proceedings of CHI Extended Abstracts (2011)

15. Zichermann, G., Cunningham, C.: *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. O'Reilly, Media (2011)
16. Liu, Y., Alexandrova, T., Nakajima, T.: Gamifying intelligent environments. In: *Proceedings of the 2011 International ACM Workshop on Ubiquitous Meta User Interfaces (Ubi-MUI '11)*. ACM, New York, USA (2011)
17. <http://www.fiat.com/ecodrive/>
18. [http://media.ford.com/article\\_display.cfm?article\\_id=29300](http://media.ford.com/article_display.cfm?article_id=29300)
19. <http://www.bmw.com/com/en/newvehicles/7series/sedan/2012/showroom/efficiency/eco-pro-mode.html>
20. <http://automobiles.honda.com/insight-hybrid/fuel-efficiency.aspx>
21. <http://arstechnica.com/gadgets/2011/11/is-the-chevy-volt-the-answer-to-urban-speeding/>
22. Shi, C., Lee, H.J., Kurczak, J., Lee, A.: Infotainment, routine driving, App: gamification and performance driving. In: *4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive UI,12)*, 17–19 Oct 2012, Portsmouth, NH, USA (2012)
23. Nokia Maps APIs. <http://api.maps.nokia.com/en/index.html>
24. Bose, R., Brakensiek, J., Park, K.Y.: Terminal mode: transforming mobile devices into automotive application platforms. In: *Proceedings of Automotive UI*, pp. 148–155 (2010)
25. <http://www.hodosmedia.net/>
26. Meschtscherjakov, A., Wilfinger, D., Scherndl, T., Tscheligi, M.: Acceptance of future persuasive in-car interfaces towards a more economic driving behaviour. In: *Proceedings of AutomotiveUI'09: 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 81–88 (2009)
27. Inbar, O., Tractinsky, N.: Driving the scoreboard: motivating eco-driving through in-car gaming. *Workshop on Gamification and HCI, as part of CHI2011* (2011)
28. Koenig, V., Boehm, F., McCall, R.: Pervasive Gaming as a Potential Solution to Traffic Congestion: New Challenges Regarding Ethics, Privacy and Trust, UL-CONFERENCE-2012-436 (2012)
29. Nunes, A.A., Galvao, T., Falcao e Cunha, J., Pitt, J.V.: Using social networks for exchanging valuable real time public transport information among travellers. In: *13th Conference on Commerce and Enterprise Computing (CEC) 2011*, IEEE, pp. 365–370, 5–7 Sept 2011
30. <http://sunset-project.eu>
31. <http://www.mydrivesolutions.com/blog/driver-behaviour-profiling>
32. <http://www.blackboxtelematics.co.uk/driverbehaviour.html>
33. <http://m2m.telekomaustralia.com/our-offer/solutions/eco-driving-solution/>
34. <http://world.waze.com/>
35. Pandazis, J.-C.: eCoMove: Cooperative ITS for green mobility. In: *18th European Wireless Conference on European Wireless EW, 2012*, pp. 1, 5, 18–20 Apr 2012
36. Osswald, S., Sheth, P., Tscheligi, M.: Hardware-in-the-loop-based evaluation platform for automotive instrument cluster development (EPIC). In: *Proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '13)* (2013)
37. Amditis, L., Andreone, K., Pagle, G., Markkula, E., Deregibus, M., Rue, R., Bellotti, F., Engelsberg, A., Brouwer, R., de Gloria, A.: Towards the automotive HMI of the future. Overview of the AIDE integrated project results. *IEEE Trans. Intell. Transp. Syst.* **11**(3), 567–578 (2010)
38. Bellotti, F., De Gloria, A., Montanari, R., Dosio, N., Morreale, D.: COMUNICAR: designing a multimedia, context-aware human-machine interface for cars. *Cognition, Technology & Work* **7**(1), 36–45 (2005)
39. Math, R., Mahr, A., Moniri, M.M., Mueller, C.: Opens: a new open-source driving simulator for research. *GMM-Fachbericht-AmE 2013* (2013)
40. Krajzewicz, B.: *Traffic Simulation with SUMO: Simulation of Urban Mobility*. Springer, Berlin (2010)