SPRINGER BRIEFS IN APPLIED SCIENCES AND TECHNOLOGY · POLIMI SPRINGER BRIEFS

Barbara Pernici *Editor*

Special Topics in Information Technology

POLITECNICO MILANO 1863

Springer Open

SpringerBriefs in Applied Sciences and Technology

PoliMI SpringerBriefs

More information about this subseries at http://www.springer.com/series/11159
http://www.polimi.it

Barbara Pernici
Editor

Special Topics in Information Technology

Editor
Barbara Pernici
DEIB
Politecnico di Milano
Milan, Italy

ISSN 2191-530X ISSN 2191-5318 (electronic)
SpringerBriefs in Applied Sciences and Technology
ISSN 2282-2577 ISSN 2282-2585 (electronic)
PoliMI SpringerBriefs
ISBN 978-3-030-32093-5 ISBN 978-3-030-32094-2 (eBook)
https://doi.org/10.1007/978-3-030-32094-2

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

Preface

This book describes nine of the most promising results from doctoral studies in Information Technology at the Department of Electronics, Information, and Bioengineering of Politecnico di Milano. Information Technology has always been interdisciplinary, as many aspects have to be considered in IT systems. The characteristics of the IT Ph.D. doctoral studies at Politecnico di Milano is an emphasis on this interdisciplinary nature, that is becoming more and more important in the recent technological developments, in collaborative projects and in the education of young researchers. The focus of advanced research is therefore on a rigorous approach to specific research topics starting from a broad background in different aspects of Information Technology, and in particular in the areas of Computer Science and Engineering, Electronics, Systems and Controls, and Telecommunications.

Each year, more than 50 doctors are graduated from the program. The present book collects the nine best results from the defended theses in 2018–19 selected for the IT Ph.D. Award. Each of the nine authors provides a chapter summarizing research results, including an introduction, description of methods, main achievements, and future work on the topic. Hence, this book provides a cutting-edge overview of the newest research trends in Information Technology developed at Politecnico di Milano, in an easy-to-read format for presenting the main results also to nonspecialists in the specific field.

Milan, Italy
July 2019

Barbara Pernici

Introduction

The Ph.D. Program in Information Technology

A fundamental pillar for the research work in the Department of Electronics, Information, and Bioengineering (DEIB) is its Ph.D. program in Information Technology, with more than 50 doctors graduating each year. The program is characterized by a broad approach to information technology, in the areas of Computer Science and Engineering, Electronics, Systems and Controls, and Telecommunications.

The characteristics of the IT Ph.D. doctoral studies in the DEIB department is an emphasis on interdisciplinarity, that is becoming more and more important in the recent technological developments in which the boundaries between disciplines are becoming increasingly fuzzy in collaborative projects and education of young researchers. Therefore, the focus of the program is on providing a broad doctoral-level educational approach and research environment and at the same time, a rigorous approach to the specific research topics developed by the Ph.D. candidates in different aspects of Information Technology, based on both on theoretical approaches and on the validation of new ideas in selected application domains.

Starting with the present volume, we present the ongoing research work in doctoral studies in the department, through a collection of summaries illustrating the results of the best Ph.D. theses defended in the academic year 2018–19. In this chapter, the coordinators of each of the areas are going to introduce the four areas of the Ph.D. program in Information Technology, namely Computer Science and Engineering, Electronics, Systems and Controls, and Telecommunications, and provide a short introduction to the papers selected for the volume.

Telecommunications Area

The research carried out by the Ph.D. graduates within Telecommunications Engineering span heterogeneous disciplines, which include information theory, communication networks, signal processing, and microwaves and photonics.

The two theses selected to be part of this volume address two different aspects and applications of digital signal processing. The contribution on "Advances in Wave Digital Modeling of Linear and Nonlinear Systems" advances the theory of wave digital modeling, that is, the construction of digital representation of analog systems. The thesis re-vitalizes and extends the theory of Wave Digital Filters (WDF) of the late 70s by proposing novel approaches, which can be applied to generic physical systems provided that they can be described by an equivalent electric circuit. The contributions provided in the work have relevant implications not only in the field of numerical simulation of physical phenomena, but also in the field of digital audio signal processing, as it allows to model different kinds of processing structures in a unified fashion.

The second contribution comes from the Ph.D. thesis "Interference Mitigation Techniques in Hybrid Wired-Wireless Communications Systems for Cloud Radio Access Networks with Analog Fronthauling"; the work addresses the fifth generation (5G) of radio mobile networks with the goal of improving the indoor coverage of 5G networks. The key contribution of the work is a novel architecture to provide coverage indoors by leveraging preexisting copper-based cables connections. The thesis first theoretically demonstrates the feasibility and efficiency of the proposed solution, and finally also describes the realization and performance evaluation of a prototype platform, which experimentally proves the capability of the proposed architecture to notably extend indoor coverage.

Electronics Area

The Ph.D. focused in Electronics explores the enabling frontier scenarios of current and future era technologies of information, communication, control, automation, energy, and mobility.

Design and innovation of devices, circuits and electronic systems continue to provide the fundamental building blocks necessary for modern life in all its perspectives, including its most recent declinations "smart-" (smart cyberphysical-systems, smart industries, smart manufacturing, smart living, smart mobility, smart lighting, smart cities, smart aging, etc.) and "autonomous-" (autonomous driving, autonomous vehicles, autonomous-manufacturing, autonomous agents, etc.), so pervasive in modern daily human activities.

For instance, the new concept of smart mobility finds one of the most significant declinations in self and highly assisted driving, where the continuous-wave frequency-modulated (FMCW) radar is a key element. This short-range measuring

radar set capable of determining distance increases reliability by providing distance measurement along with speed measurement, which is essential in modern automotive applications. Dmytro Cherniak has given a strong contribution to the field of integrated FMCW radar. These innovations support the emerging market for autonomous vehicles, which will rely heavily on radar and wireless sensing to achieve reliable all-weather mobility. High performance and low-power consumption meet in the first implemented digital PLL-based FMCW modulator prototype fabricated in 65-nm CMOS technology that demonstrates the above-state-of-the-art performance of fast chirp synthesis, also developed and fabricated in 28-nm CMOS technology focusing on low-fractional spur operation.

From a completely different perspective, the human brain is a marvelous machine, which is capable of solving challenging problems in a very short time and with small energy consumption. Recognizing a face, learning to walk, and taking a quick decision under a large number of stimuli are generally straightforward functions for mankind. However, recreating this type of computation in silicon with the same speed, energy consumption, and hardware resources is practically impossible with the current microelectronic technology. The work by Valerio Milo addresses these challenges, by showing that a new type of devices and architecture is needed if we want to mimic the brain computation. It is shown that synaptic devices made of inorganic materials, such as metal oxides, show the same type of plasticity that is observed in biological synapses. By taking advantage of such biological plasticity rules within artificial spiking neural networks, it is possible to learn, make associations, and take decisions, which are some of the most fundamental cognitive functions of the human brain. Although the work is at the very early stages, the results are very promising for supporting the grand challenge of bio-inspired circuits that can compute as the brain, and even illuminate about the biological processes of learning in our brain.

Computer Science and Engineering Area

The Computer Science and Engineering Area covers five different research lines, ranging from System Architectures, to Data, Web and Society, to Artificial Intelligence and Robotics, to Advanced Software Architectures and Methodologies, and Information Systems. Research on these lines are covering both foundational aspects as well as application-driven aspects. The group of System Architectures is active on several research topics, including security, performance evaluation, dependability, electronic design automation, processor and multiprocessor architectures, operating systems, embedded systems, computational intelligence, wireless sensor networks, enterprise digital infrastructures, and high-performance computing. The research in the area of Data, Web, and Society addresses technologies, design methods, and tools for data management systems, information management and querying on the Web, and multimedia and multichannel communication. The research group on Artificial Intelligence and Robotics also covers

the areas of autonomous agents, computational intelligence, machine learning, autonomous robotics, computer vision, including related philosophical aspects. The research group on Software Architectures and Methodologies addresses topics on dependable evolvable software engineering, compiler technologies, and natural language processing and accessibility. Finally, the research line on Information Systems focuses on the following sectors: design of adaptive information systems, data and information quality, Big Data and data analysis, social media intelligence, information security and security analysis, service quality and green information systems, and design and monitoring of multiparty business processes integrated with smart objects.

The four Ph.D. theses selected in the Computer Science and Engineering Area fall in the System Architectures research line, but they are addressing four different and challenging research problems.

The Ph.D. thesis "Learning and Adaptation to Detect Changes and Anomalies in High-Dimensional Data" by Diego Carrera investigates the problem of monitoring a datastream and detecting whether the data generating process changes, from normal to novel and possibly anomalous conditions, has relevant applications in many real scenarios, such as health monitoring and quality inspection of industrial processes. In the first part, the thesis models data as realization of random vectors, as it is customary in the statistical literature. In this setting, the thesis focuses on the change detection problem, where the goal is to detect whether the datastream permanently departs from normal conditions. The thesis theoretically proves the intrinsic difficulty of this problem when the data dimension increases and propose a novel nonparametric and multivariate change detection algorithm. In the second part, the thesis focuses on data having complex structure and adopts dictionaries yielding sparse representations to model normal data. Novel algorithms are proposed to detect anomalies in such datastreams and to adapt the learned model when the process generating normal data changes.

The Ph.D. thesis titled "Enhancing Video Recommendation Using Multimedia Content" by Yashar Deldjoo is focused on how to improve video recommendation systems by using complex multimedia content and learning form multimodal sources. The Ph.D. thesis investigates the possibility of uncovering relationships between modalities and obtaining an in-depth understanding of natural phenomena occurring in a video. The thesis studies the automated extraction of multimedia information from videos and their integration with video recommender systems. In the thesis, a variety of tasks related to movie recommendation using multimedia content have been studied, implemented, and evaluated. The results of this thesis confirm the fact that recommender system research can benefit from knowledge in multimedia signal processing and machine learning over the past years for solving various recommendation tasks.

The Ph.D. thesis titled "Dynamic Application Autotuning for Self-aware Approximate Computing" by Davide Gadioli addresses the problem of software application autotuning to support self-aware approximate computing in several scenarios, from embedded systems to high-performance computing. To improve computation efficiency, this thesis focuses on a software-level methodology to

enhance a target application with an adaptive layer that provides self-optimization capabilities. The benefits of dynamic autotuning framework, namely mARGOt, has been evaluated in three case studies. The first case study is a probabilistic time-dependent routing application from a navigation system in smart cities, the second use case is a molecular docking application to perform virtual-screening, and the third one is a stereo-matching application to compute the depth of a three-dimensional scene. Experimental results have shown how it is possible to improve computation efficiency by adapting reactively and proactively the application to provide continuously the most suitable software-knobs configuration according to application requirements and system monitoring.

The Ph.D. thesis "CAOS: CAD as an Adaptive Open-Platform Service for High Performance Reconfigurable Systems" by Marco Rabozzi studies the increasing demand for computing power in fields such as genomics, image processing and machine learning, pushing towards hardware specialization and heterogeneous systems to keep up with the required performance level at a sustainable power consumption. Despite the potential benefits of reconfigurable hardware, offering a compelling trade-off between efficiency and flexibility, one of the main limiting factors to the adoption of Field- Programmable Gate Arrays (FPGAs) is complexity in programmability, as well as the effort required to port software solutions to efficient hardware–software implementations. This thesis presents the CAOS platform to guide the application developer in the implementation of efficient hardware–software solutions for high-performance reconfigurable systems. The platform assists the designer from the high-level analysis of the code, towards the optimization and implementation of the functionalities to be accelerated on the reconfigurable nodes. Finally, CAOS is designed to facilitate the integration of external contributions and to foster research on Computer-Aided Design (CAD) tools for accelerating software applications on FPGA-based systems.

Systems and Controls Area

The Systems and Control area covers various fields, ranging from systems theory and control system science, to robotics and industrial automation, nonlinear and networked complex systems, ecosystems and environmental modeling, and operations research, with attention both to basic research and industrial applications. Various classes of dynamical systems are studied, including nonlinear systems, switched and hybrid systems, stochastic systems, discrete event systems, cyber-physical systems, large-scale, and networked systems. Various methodologies are investigated and developed regarding estimation, control (*e.g.*, predictive, distributed, and robust methods), model identification, and data analysis (black box or gray box estimation, learning, filtering, and prediction). Specific application areas of interest are automotive, railway and aerospace control, modeling and control of energy systems (with regard to generation, distribution, and management), mechatronics and motion control, robotics, process modeling, simulation and

control, modeling and control of manufacturing systems. Another area of research focuses on the theoretical and numerical analysis of nonlinear dynamical systems, leading to applications in biology, epidemiology, social sciences, and vehicle dynamics. The quantitative analysis and management of environmental systems at both global and local scales is also pursued with the aim to develop efficient and sustainable decision systems. Finally, the operations research and discrete optimization group investigates complex decision-making problems, for which it develops mathematical models and efficient optimization methods, using various tools, such as mathematical programming, combinatorial optimization, stochastic programming, robust optimization, bilevel programming, and continuous approximation models.

The Ph.D. thesis "A General Framework for Shared Control in Robot Teleoperation with Force and Visual Feedback" by Davide Nicolis falls in the robotics area, and studies the control of robotic systems for teleoperation, employing a control system structured in two levels. A lower level employs sliding mode controllers to robustly guarantee the dynamic behavior of both the master and slave systems. The higher level of the control system performs a performance optimization taking into account various motion-related constraints. The thesis features various original contributions, ranging from the design of the sliding mode controller that assigns the robot's impedance in the task space, to the generation of virtual fixtures with a rendering of the force feedback to the operator, to a stability study of the overall system. A particularly appreciable aspect of the work is the application of servo-visual control techniques on the slave robot, thereby endowing it with a certain degree of autonomy, which delivers some of the cognitive load of the operator. Finally, the thesis also addresses the problem of recognizing the human's intentions using machine learning techniques, especially, designed for assistive robotic applications. All the theoretical results of this work have been experimentally validated on realistic experimental platforms (using single- and double-arm ABB robots).

Barbara Pernici
Matteo Cesana
Angelo Geraci
Cristina Silvano
Luigi Piroddi
DEIB, Politecnico di Milano, Milan, Italy

Contents

Part I
Telecommunications

Chapter 1
Advances in Wave Digital Modeling of Linear and Nonlinear Systems: A Summary

Alberto Bernardini

Abstract This brief summarizes some of the main research results that I obtained during the three years, ranging from November 2015 to October 2018, as a Ph.D. student at Politecnico di Milano under the supervision of Professor Augusto Sarti, and that are contained in my doctoral dissertation, entitled "*Advances in Wave Digital Modeling of Linear and Nonlinear Systems*". The thesis provides contributions to all the main aspects of Wave Digital (WD) modeling of lumped systems: it introduces generalized definitions of wave variables; it presents novel WD models of one- and multi-port linear and nonlinear circuit elements; it discusses systematic techniques for the WD implementation of arbitrary connection networks and it describes a novel iterative method for the implementation of circuits with multiple nonlinear elements. Though WD methods usually focus on the discrete-time implementation of analog audio circuits; the methodologies addressed in the thesis are general enough as to be applicable to whatever system that can be described by an equivalent electric circuit.

1.1 Introduction

My doctoral dissertation presents various contributions to the recent evolution of modeling and implementation techniques of linear and nonlinear systems in the Wave Digital (WD) domain. The overarching goal of WD methods is to build digital implementations of analog systems, which are able to emulate the behavior of their analog counterpart in an efficient and accurate fashion. Though such methods usually focus on the WD modeling of analog audio circuits; the methodologies addressed in the thesis are general enough as to be applicable to whatever physical system that can be described by an equivalent electric circuit, which includes any system that can be thought of as a port-wise interconnection of lumped physical elements. The possibility of describing systems through electrical equivalents has relevant implications not only in the field of numerical simulation of physical phenomena, but also in the field of digital signal processing, as it allows us to model different

A. Bernardini (✉)
Politecnico di Milano, Piazza Leonardo Da Vinci 32, 20133 Milano, Italy
e-mail: alberto.bernardini@polimi.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_1

kinds of processing structures in a unified fashion and to easily manage the energetic properties of their input-output signals. However, digitally implementing nonlinear circuits in the Kirchhoff domain is not straightforward, because dual variables (currents and voltages) are related by implicit equations which make computability very hard. Mainstream Spice-like software, based on the Modified Nodal Analysis (MNA) framework [20], is not always suitable for realizing efficient and interactive digital applications, mainly because it requires the use of iterative methods based on large Jacobian matrices (e.g., Newton–Raphson) for solving multi-dimensional nonlinear systems of equations.

WD Filters (WDFs) are a very attractive alternative. During the seventies, Alfred Fettweis introduced WDFs as a special category of digital filters based on a lumped discretization of reference analog circuits [19]. A WDF is created by port-wise consideration of a reference circuit, i.e., decomposition into one-port and multi-port circuit elements, a linear transformation of Kirchhoff variables to wave signals (incident and reflected waves) with the introduction of a free parameter per port, called reference port resistance, and a discretization of reactive elements via the bilinear transform. Linear circuit elements, such as resistors, real sources, capacitors and inductors, can be described through wave mappings without instantaneous reflections, as they can be all "adapted" exploiting the mentioned free parameter; in such a way that local Delay-Free Loops (DFLs), i.e., implicit relations between wave variables are eliminated. Series and parallel topological connections between the elements are implemented using scattering topological junctions called "adaptors", which impose special adaptation conditions to eliminate global DFLs and ensure computability. It follows that WDFs, as opposed to approaches based on the MNA, allow us to model separately the topology and the elements of the reference circuit. Moreover, WDFs are characterized by stability, accuracy, pseudo-passivity, modularity and low computational complexity, making many real-time interactive applications easy to be realized. Most classical WD structures can be implemented in an explicit fashion, using binary connection trees, whose leaves are linear one-ports, nodes are 3-port adaptors and the root may be a nonlinear element. However, WDFs are also characterized by important limitations. The first main weakness of state-of-the-art WDFs is that WD structures, characterized by explicit input-output relations, can contain only one nonlinear element, as nonlinear elements cannot be adapted. In fact, the presence of multiple nonlinear elements might affect computability, which characterizes classical linear WDFs, as DFLs arise. As a second main limitation of traditional WDFs, up to three years ago, there were no systematic methods for modeling connection networks which embed non-reciprocal linear multi-ports, such as nullors or controlled sources. Finally, very few studies were presented on the use of discretization methods alternative to the bilinear transform and potentially adjustable step-size in WD structures.

The thesis presents various techniques to overcome the aforementioned limitations. After a review of the state of the art on WD methods up to 2015, the thesis is organized in five parts and each part presents original contributions to a specific important aspect of WD modeling. In the following, Sects. 1.2, 1.3, 1.4, 1.5 and 1.6 resume the content of Part I, Part II, Part III, Part IV and Part V of the thesis; each

subsection providing a summary of a specific chapter. Section 1.7 concludes this brief and proposes some possible future developments.

1.2 Part I: Rethinking Definitions of Waves

Part I, containing Chaps. 2 and 3 of the thesis, discusses two families of definitions of wave signals; one based on one free parameter per port, the other based on two free parameters per port.

1.2.1 Mono-Parametric Definitions of Wave Variables

Chapter 2 presents a generalization of the traditional definitions of voltage waves, current waves and power-normalized waves, characterized by one port resistance per port. The generalization is done introducing a scalar parameter ρ in expressions at the exponent of the port resistance. The generalized definition also includes novel definitions of waves never appeared in the literature. Such a generalized definition was firstly presented in [35] and it proved useful for modeling WD structures based on waves with different units of measure in an unified fashion. In fact, it was used in [30] and in [16] for modeling arbitrary reciprocal and non-reciprocal multi-port WD junctions. Chapter 2 also includes the scattering relations of some fundamental circuit elements and the scattering relations of series and parallel adaptors. It is shown that some scattering relations are invariant to the wave type (e.g., voltage waves, current waves, power-normalized waves), while other scattering relations are not invariant to the wave type since they depend on ρ.

1.2.2 Biparametric Definitions of Wave Variables

Chapter 3 presents the content of the published journal article [7], which introduces two dual definitions of power-normalized waves characterized by two free parameters per port, instead of one as it happens in traditional WDFs. It is shown that such dual definitions are two possible generalizations of the traditional definitions based on one port resistance and that when the two free parameters at the same port are set to be equal, they reduce to the traditional definition of waves. The WD structures based on the new definitions of wave variables are called Biparametric WDFs (BWDFs). It is shown that since BWDFs are characterized by more degrees of freedom than WDFs, they enable a higher modeling flexibility. For instance, all ports of adaptors based on biparametric definitions of waves can be made reflection free at the same time. This fact allows us to reduce the number of local DFLs in WD structures and to design adaptors whose behavior is uniform at all ports. Moreover, when the

free parameters are properly set, series adaptors can be described using scattering matrices made of zeros in the diagonal entries and minus ones in the non-diagonal entries, while parallel adaptors can be described using scattering matrices made of zeros in the diagonal entries and ones in the non-diagonal entries. This further property is useful in many situations for reducing the number of multiplies required for implementing WD structures based on power-normalized waves. Finally, a discussion on the implementation of nonlinear circuits with multiple nonlinearities using BWDFs is provided. In particular, it is shown that, despite the use of adaptors with all reflection-free ports and less multipliers simplifies the modeling of nonlinear circuits with multiple nonlinearities in many aspects (e.g., considerably reducing the number of local DFLs and, consequently, the complexity of the implicit equations describing the circuit in the WD domain), not all global DFLs can be eliminated even using BWDFs and iterative methods are still required.

1.3 Part II: Modeling Nonlinear One-Port and Multi-port Elements

Part II, containing Chaps. 4, 5 and 6, is devoted to the modeling of nonlinear one-ports and multi-ports in the WD domain. The common objective in all chapters is searching for the conditions that allow us to use explicit scattering relations for describing nonlinear one-ports and multi-ports in the WD domain. Also the use of the Newton–Raphson (NR) method for finding the solution of implicit scattering relations is discussed.

1.3.1 Canonical Piecewise-Linear Representation of Curves in the Wave Digital Domain

Chapter 4 presents the content of the published article [8], where a method is discussed that, starting from certain parameterized PieceWise-Linear (PWL) i–v curves of one-ports in the Kirchhoff domain, expresses them in the WD domain using a global and explicit representation. Global, explicit representations of nonlinearities allow us to implement their input-output relations without managing look-up tables, performing data interpolation and/or using local iterative solvers. It is shown how certain curves (even some multi-valued functions in the Kirchhoff domain) can be represented as functions in explicit canonical PWL form in the WD domain. A general procedure is also provided that returns the conditions on the reference port resistance under which it is possible to find explicit mappings in the WD domain.

1.3.2 Wave Digital Modeling of Nonlinear Elements Using the Lambert Function

Certain transcendental equations involving exponentials can be expressed in explicit form using the Lambert W function. Chapter 5 presents the content of the published journal article [15], which explores how the W function can be used to derive explicit WD models of some one-ports [32] and multi-ports characterized by exponential nonlinearities, such as banks of diodes in parallel and/or anti-parallel or BJTs in certain amplifier configurations.

1.3.3 Wave Digital Modeling of Nonlinear 3-Terminal Devices for Virtual Analog Applications

Chapter 6 presents the content of the article [11] submitted for publication and currently in peer review. It discusses an approach for modeling circuits containing arbitrary linear or nonlinear 3-terminal devices in the WD domain. Such an approach leads us to the definition of a general and flexible WD model for 3-terminal devices, whose number of ports ranges from 1 to 6. The WD models of 3-terminal devices already discussed in the literature could be described as particular cases of the model presented in this chapter. As examples of applications of the proposed approach, WD models of the three most widespread types of transistors in audio circuitry, i.e., the MOSFET, the JFET and the BJT are developed. These WD models are designed to be used in Virtual Analog audio applications. It follows that the proposed models are derived with the aim of minimizing computational complexity, and avoiding implicit relations between port variables, as far as possible. Proposed MOSFET and JFET models result into third order equations to solve; therefore, closed-form wave scattering relations are obtained. Instead, the Ebers-Moll model describing the BJT is characterized by transcendental equations which cannot be solved in closed-form in the WD domain; consequently, iterative methods for finding their solutions have been studied. The standard NR method recently used in the literature on WDFs [25] do not satisfy all the requirements of robustness and efficiency needed in audio Virtual Analog applications. For this reason, a modified NR method is developed that exhibits a significantly higher robustness and convergence rate with respect to the traditional NR method, without compromising its efficiency. In particular, the behavior of the proposed modified NR method is far less sensitive to chosen initial guesses than the traditional NR method. The proposed method converges for any initial guess and any incident wave signal within reasonable ranges which are deduced from the parameters of the implemented reference circuits.

1.4 Part III: Modeling Connection Networks

Part III, containing Chaps. 7, 8 and 9, focuses on the modeling of arbitrary connection networks in the WD domain. We refer to a *connection network* as a linear multi-port whose ports are connected to arbitrary loads, which might be single circuit elements or other networks. Connection networks in the WD domain are implemented using scattering junctions called adaptors. In turn, such WD junctions are characterized by scattering matrices, whose properties depend on the characteristics of the modeled reference connection network in the Kirchhoff domain and on the chosen definition of wave variables. Connection networks can be classified in two main classes; the class of reciprocal connection networks and the class of non-reciprocal connection networks. In turn, the class of reciprocal connection networks includes "wire connections" of arbitrary complexity (i.e., series/parallel connections or interconnections which are neither series nor parallel) and connection networks embedding reciprocal multi-ports, such as ideal two-winding or multi-winding transformers. On the other hand, non-reciprocal connection networks, embed one or more non-reciprocal multi-ports, such as nullors or controlled sources.

1.4.1 Modeling Sallen-Key Audio Filters in the Wave Digital Domain

In [31, 34] a method based on the MNA analysis of the reference connection network with *instantaneous Thévenin equivalents* connected to all its ports is presented. Ideal sources of instantaneous Thévenin equivalents are set equal to the waves incident to the WD junction and series resistances are set equal to reference port resistances (free parameters). Chapter 7 presents the content of the published article [29], where the MNA-based method is applied for implementing Sallen-Key filter circuits in the WD domain. In particular, the eighteen filter models presented by Sallen and Key in their historical 1955 manuscript [26] are grouped into nine classes, according to their topological properties. For each class the corresponding WD structure is derived.

1.4.2 Modeling Circuits with Arbitrary Topologies and Active Linear Multiports Using Wave Digital Filters

Chapter 8 presents the content of the published journal article [30], where the MNA-based method, presented in [31, 34] for modeling arbitrary connection networks using WD adaptors based on voltage waves, is extended in such a way that it can be applied to all mono-parametric definitions of waves using the generalized definition described in Chap. 2. Moreover, it is shown that, connecting *instantaneous Norton equivalents* to the ports of the connection network (instead of *instantaneous*

Thévenin equivalents), in order to perform the MNA-based method, brings considerable advantages in terms of computational cost because the number of nodes is always reduced.

1.4.3 Generalized Wave Digital Filter Realizations of Arbitrary Reciprocal Connection Networks

Chapter 9 presents the content of the published journal article [16] in which the approach developed by Martens et al. [23] for modeling and efficiently implementing arbitrary reciprocal connection networks using WD scattering junctions based on voltage waves is extended to be used in WDFs based on the generalized monoparametric and biparametric definitions of waves discussed in Part I. The method presented in this chapter is less general than the one presented in Chap. 8 (and published in [30]), since it is limited to the modeling of reciprocal connection networks. However, as far as reciprocal connection networks are concerned, the computational efficiency of the proposed method generally surpasses or at least matches that of the method in [30], both when the cost of scattering and the sizes of linear systems to be inverted in order to form the scattering matrix are considered.

1.5 Part IV: Implementation of Circuits with Multiple Nonlinearities

As outlined in the introduction, circuits with one nonlinear element can be implemented using the trapezoidal discretization method obtaining explicit WD structures, i.e., digital structures without DFLs [24]. This is a considerable advantage of WD modeling over alternative Virtual Analog modeling approaches developed in the domain of voltages and currents, since they are typically characterized by implicit equations and require the use of iterative solvers. Unfortunately, such a great benefit is not preserved when circuits with multiple nonlinearities are considered because all DFLs cannot be eliminated [7]. However, even in those cases, working in the WD domain have proven to bring advantages. For instance, in [33] a method is introduced that allows us to group all the nonlinear elements "at the root" of the WD structure, enabling the possibility of separating the nonlinear part of the circuit from the linear one; then, the multivariate nonlinear system of equations describing the nonlinear part is solved using tabulation [13, 14, 33] or multivariate NR solvers [25]. Another approach is presented in [28], where multidimensional WDFs and the multivariate NR method are combined for solving nonlinear lumped circuits. An alternative approach, described in [27], exploits the contractivity property of certain WD structures for accommodating multiple nonlinearities using fixed point iteration schemes. However, from a theoretical stand point, in [27], contractivity ensuring

convergence of the fixed point algorithm is proven and analyzed only considering linear WD structures.

Part IV, containing Chaps. 10, 11 and 12, is mainly devoted to the WD implementation of circuits with multiple nonlinearities using a further method called Scattering Iterative Method (SIM) that has been recently developed starting from the preliminary results presented in [6]. SIM is a relaxation method characterized by an iterative procedure that alternates a local scattering stage, devoted to the computation of waves reflected from each element, and a global scattering stage, devoted to the computation of waves reflected from a WD junction to which all elements are connected. Similarly to what happens in the implementation methods described in [27], wave signals circulate in the WD structure back and forth up to convergence. A proven theorem guarantees that SIM converges when applied to whichever circuit characterized by a reciprocal connection network and an arbitrary number of linear or nonlinear one-ports whose i-v characteristic is monotonic increasing. It is worth noticing that the absence of guarantee of convergence when it comes to implement a nonlinear circuit using a whichever iterative method, does not necessarily imply that the method cannot be used for solving that circuit anyway; as a matter of fact, this happens in most situations in Spice-like software. The proven theorem not only helps in identifying a class of nonlinear circuits for which the convergence of SIM is theoretically ensured, but it also gives us insights about a strategy for increasing its convergence speed, properly setting the free parameters (port resistances). SIM is able to solve circuits with an arbitrary number N_{nl} of 2-terminal nonlinear elements using N_{nl} independent one-dimensional NR solvers instead of one N_{nl}-dimensional NR solver. This fact implies a number of interesting features that greatly differentiate SIM from techniques based on high-dimensional NR solvers [25], like higher robustness (or even convergence guarantee, when dealing with elements with monotonically increasing i-v curves), higher efficiency and the possibility of solving the nonlinearities in parallel threads of execution.

1.5.1 Wave-Based Analysis of Large Nonlinear Photovoltaic Arrays

Chapter 10 presents the content of the published journal article [4] in which SIM is employed for modeling and efficiently simulating large nonlinear photovoltaic (PV) arrays under partial shading conditions. Given the irradiation pattern and the nonlinear PV unit model (e.g., exponential junction model with bypass diode) with the corresponding parameters, the WD method rapidly computes the current-voltage curve at the load of the PV array [10]. The main features of the WD method are the use of a scattering matrix modeling the arbitrary PV array topology and the adoption of one-dimensional solvers to locally handle the nonlinear constitutive equations of PV units. A rigorous analysis of SIM shows that it can be considered as a fixed-point method that always converges to the PV array solution. Compared with standard

Spice-like simulators, the WD method is up to 35 times faster for PV arrays made of thousands of units. This makes the proposed method palatable for the development of dedicated systems for the real time control and optimization of large PV plants, e.g, maximum power point trackers based on the rapid exploration of the i-v curve at the load.

1.5.2 Wave Digital Modeling of the Diode-Based Ring Modulator

Chapter 11 presents the content of the published article [12] in which SIM is shown to be suitable also for the discrete-time emulation of audio circuits for Virtual Analog applications since it is robust and comparable to or more efficient than state-of-the-art strategies in terms of computational cost. In particular, a WD model of a ring modulator circuit constituted of four diodes and two multi-winding transformers is derived. An implementation of the WD model based on SIM is then discussed and a proof of convergence is provided.

1.5.3 Linear Multi-step Discretization Methods with Variable Step-Size in Nonlinear Wave Digital Structures

Circuit modeling in the WD domain typically entails the implementation of capacitors and inductors employing the trapezoidal discretization method with fixed sampling step. However, in many cases, alternative discretization techniques, eventually based on adaptive sampling step, might be preferable. Chapter 12 presents the content of the journal article, in which an unified approach for implementing WD dynamic elements based on arbitrary linear multi-step discretization methods with variable step-size as time-varying WD Thévenin or Norton equivalents is discussed. Moreover, it is shown that such an approach is particularly suitable to be combined with SIM for solving circuits with multiple nonlinearities.

1.6 Part V: New Applications of WD Principles

Physical systems that do not come in the form of electrical circuits can often be accurately represented by electrical equivalents, and then modeled and implemented as WD structures. In some applications, it is even possible to define an electrical circuit that models a digital signal processing structure, and use its WD model to implement such system more efficiently. As an example of the sort, in Part V containing Chap. 13, it is shown how certain first-order beamforming systems can be represented using electrical equivalents and then implemented in the WD domain.

1.6.1 Wave Digital Implementation of Robust First-Order Differential Microphone Arrays

As a secondary research project during my Ph.D., I worked on the modeling of Differential Microphone Arrays (DMAs) [2, 3, 17, 22]. In this regard, Chap. 13, presents the content of the published journal article [1], in which a novel time-domain WD implementation of robust first-order DMAs with uniform linear array geometry [36] is described. In particular, it is shown that the reference beamforming system, composed of an array of sensors and a bank of filters (one per sensor) designed in the frequency domain, can be exactly represented by a bank of simple electrical circuits. This fact allows us to derive a bank of WDFs, one per sensor, and obtain a time-domain realization of the same beamformer which is less computationally demanding than its counterpart implemented in the frequency domain. The proposed beamforming method is extremely efficient, as it requires at most two multipliers and one delay for each filter, where the necessary number of filters equals the number of physical microphones of the array, and it avoids the use of fractional delays. The update of the coefficients of the filters, required for reshaping the beampattern, has a significantly lower computational cost with respect to the time-domain methods presented in the literature [18]. This makes the proposed method suitable for real-time DMA applications with time-varying beampatterns.

1.7 Conclusions and Future Works

In this brief, I resumed the main contributions to WD modeling of lumped systems presented in my doctoral dissertation. As far as future work is concerned, I think the properties of BWDFs should be explored further. For instance, two parameters per port could be exploited for increasing the speed of convergence of SIM. Moreover, it is worth extending the applicability of SIM to circuits containing multi-port nonlinearities and nonreciprocal linear elements. Such extensions would pave the way towards the realization of new general purpose circuit simulators that are potentially more efficient, robust and parallelizable than mainstream Spice-like software.

Another promising application of the presented WD modeling techniques is the design of inverse nonlinear systems, given direct systems represented as electrical equivalent circuits. In fact, the inverse of a nonlinear circuit system can be obtained by properly adding a nullor to the direct system itself [21]. In this regard, preliminary results published in the conference paper [9] show how the approach for modeling non-reciproca junctions in the WD domain, presented in [30], can be exploited for modeling the inverse of certain single-input-single-output audio circuits.

References

1. Bernardini A, Antonacci F, Sarti A (2018) Wave digital implementation of robust first-order differential microphone arrays. IEEE Signal Process Lett 25(2):253–257. https://doi.org/10.1109/LSP.2017.2787484
2. Bernardini A, D'Aria M, Sannino R (2017) Beamforming method based on arrays of microphones and corresponding apparatus. US Patent US9913030B2
3. Bernardini A, D'Aria M, Sannino R, Sarti A (2017) Efficient continuous beam steering for planar arrays of differential microphones. IEEE Signal Process Lett 24(6):794–798. https://doi.org/10.1109/LSP.2017.2695082
4. Bernardini A, Maffezzoni P, Daniel L, Sarti A (2018) Wave-based analysis of large nonlinear photovoltaic arrays. IEEE Trans Circuits Syst I Regul Pap 65(4):1363–1376. https://doi.org/10.1109/TCSI.2017.2756917
5. Bernardini A, Maffezzoni P, Sarti A (2019) Linear multi-step discretization methods with variable step-size in nonlinear wave digital structures for virtual analog modeling. IEEE/ACM Trans Audio Speech Language Process 27(11):1763–1776. https://doi.org/10.1109/TASLP.2019.2931759
6. Bernardini A, Sarti A (2016) Dynamic adaptation of instantaneous nonlinear bipoles in wave digital networks. In: Proceedings of the 24th European signal processing conference Budapest, Hungary. https://doi.org/10.1109/EUSIPCO.2016.7760406
7. Bernardini A, Sarti A (2017) Biparametric wave digital filters. IEEE Trans Circuits Syst I Regul Pap 64(7):1826–1838. https://doi.org/10.1109/TCSI.2017.2679007
8. Bernardini A, Sarti A (2017) Canonical piecewise-linear representation of curves in the wave digital domain. In: Proceedings of the 25th European signal processing conference (EUSIPCO), pp 1125–1129. Kos, Greece. https://doi.org/10.23919/EUSIPCO.2017.8081383
9. Bernardini A, Sarti A (2019) Towards inverse virtual analog modeling. In: Proceedings of the 22nd conference digital audio effects, Birmingham, UK. Paper number #8
10. Bernardini A, Sarti A, Maffezzoni P, Daniel L (2018) Wave digital-based variability analysis of electrical mismatch in photovoltaic arrays. In: Proceedings of the IEEE international symposium on circuits and systems (ISCAS), pp 1–5. Florence, Italy. https://doi.org/10.1109/ISCAS.2018.8351026
11. Bernardini A, Vergani AE, Sarti A (2019) Wave digital modeling of nonlinear 3-terminal devices for virtual analog applications. Springer circuits, systems, and signal processing (CSSP) (Submitted)
12. Bernardini A, Werner KJ, Maffezzoni P, Sarti A (2018) Wave digital modeling of the diode-based ring modulator. In: Proceedings of the 144th convention audio engineering society, Milan, Italy. Convention paper #10015
13. Bernardini A, Werner KJ, Sarti A, Smith III JO (2015) Modeling a class of multi-port nonlinearities in wave digital structures. In: Proceedings of the European signal processing conference, Nice, France. https://doi.org/10.1109/EUSIPCO.2015.7362466
14. Bernardini A, Werner KJ, Sarti A, Smith III JO (2015) Multi-port nonlinearities in wave digital structures. In: Proceedings of the IEEE International Symposium on Signals, Circuits and Systems, Iaşi, Romania. https://doi.org/10.1109/ISSCS.2015.7203989
15. Bernardini A, Werner KJ, Sarti A, Smith III JO (2016) Modeling nonlinear wave digital elements using the Lambert function. IEEE Trans Circuits Syst I Regul Pap 63(8):1231–1242. https://doi.org/10.1109/TCSI.2016.2573119
16. Bernardini A, Werner KJ, Smith III JO, Sarti A (2019) Generalized wave digital filter realizations of arbitrary reciprocal connection networks. IEEE Trans Circuits Syst I Regul Pap 66(2):694–707. https://doi.org/10.1109/TCSI.2018.2867508
17. Borra F, Bernardini A, Antonacci F, Sarti A (2019) Uniform linear arrays of first-order steerable differential microphones. IEEE/ACM Trans Audio Speech Language Process 27(12):1906–1918. https://doi.org/10.1109/TASLP.2019.2934567

18. Buchris Y, Cohen I, Benesty J (2016) First-order differential microphone arrays from a time-domain broadband perspective. In: 2016 IEEE international workshop on acoustic signal enhancement (IWAENC), pp 1–5. https://doi.org/10.1109/IWAENC.2016.7602886
19. Fettweis A (1986) Wave digital filters: theory and practice. Proc IEEE 74(2):270–327. https://doi.org/10.1109/PROC.1986.13458
20. Ho CW, Ruehli AE, Brennan PA (1975) The modified nodal approach to network analysis. IEEE Trans Circuits Syst 22(6):504–509. https://doi.org/10.1109/TCS.1975.1084079
21. Leuciuc A (1998) The realization of inverse system for circuits containing nullors with applications in chaos synchronization. Int J Circuit Theory Appl 26(1):1–12. https://doi.org/10.1002/(SICI)1097-007X(199801/02)26:1<1::AID-CTA989>3.0.CO;2-B
22. Lovatello J, Bernardini A, Sarti A (2018) Steerable circular differential microphone arrays. In: 26th European signal processing conference (EUSIPCO), pp 11–15. Rome, Italy. https://doi.org/10.23919/EUSIPCO.2018.8553083
23. Martens GO, Meerkötter K (1976) On N-port adaptors for wave digital filters with application to a bridged-tee filter. In: Proceedings of the IEEE international symposium on circuits and systems, pp. 514–517. Munich, Germany
24. Meerkötter K, Scholz R (1989) Digital simulation of nonlinear circuits by wave digital filter principles. In: IEEE international symposium on circuits and systems, pp 720–723. https://doi.org/10.1109/ISCAS.1989.100452
25. Olsen MJ, Werner KJ, Smith III JO (2016) Resolving grouped nonlinearities in wave digital filters using iterative techniques. In: Proceedings 19th international conference digital audio effects, pp 279–286. Brno, Czech Republic
26. Sallen RP, Key EL (1955) A practical method of designing RC active filters. IRE Trans Circuit Theory 2(1):74–85. https://doi.org/10.1109/TCT.1955.6500159
27. Schwerdtfeger T, Kummert A (2014) A multidimensional approach to Wave Digital Filters with multiple nonlinearities. In: 22nd Proceedings of the European signal processing conference (EUSIPCO), pp 2405–2409. Lisbon, Portugal
28. Schwerdtfeger T, Kummert A (2015) Newton's method for modularity-preserving multidimensional wave digital filters. In: Proceedings of the IEEE international workshop multidimensional system Vila Real, Portugal
29. Verasani M, Bernardini A, Sarti A (2017) Modeling sallen-key audio filters in the wave digital domain. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 431–435. New Orleans, LA. https://doi.org/10.1109/ICASSP.2017.7952192
30. Werner KJ, Bernardini A, Smith III JO, Sarti A (2018) Modeling circuits with arbitrary topologies and active linear multiports using wave digital filters. IEEE Trans Circuits Syst I Regul Pap 65(12):4233–4246. https://doi.org/10.1109/TCSI.2018.2837912
31. Werner KJ, Dunkel WR, Rest M, Olsen MJ, Smith III JO (2016) Wave digital filter modeling of circuits with operational amplifiers. In: Proceedings of the european signal processing conference, pp 1033–1037. Budapest, Hungary. https://doi.org/10.1109/EUSIPCO.2016.7760405
32. Werner KJ, Nangia V, Bernardini A, Smith III JO, Sarti A (2015) An improved and generalized diode clipper model for wave digital filters. In: Proceedings of the 139th convention audio engineering society, New York. Convention paper. #9360
33. Werner KJ, Nangia V, Smith III JO, Abel JS (2015) Resolving wave digital filters with multiple/multiport nonlinearities. In: Proceedings of the 18th international conference on digital audio effects, pp 387–394. Trondheim, Norway
34. Werner KJ, Smith III JO, Abel JS (2015) Wave digital filter adaptors for arbitrary topologies and multiport linear elements. In: Proceedings of the 18th international conference on digital audio effects, pp 379–386. Trondheim, Norway
35. Werner KJ (2016) Virtual analog modeling of audio circuitry using wave digital filters. PhD Dissertation, Stanford University, CA
36. Zhao L, Benesty J, Chen J (2014) Design of robust differential microphone arrays. IEEE/ACM Trans Audio Speech Lang Process 22(10):1455–1466. https://doi.org/10.1109/TASLP.2014.2337844

Chapter 2
Enhancing Indoor Coverage by Multi-Pairs Copper Cables: The Analog MIMO Radio-over-Copper Architecture

Andrea Matera

Abstract Nowadays, the majority of indoor coverage issues arise from networks that are mainly designed for outdoor scenarios. Outdoor networks, somewhat uncontrollably, may penetrate indoors with the consequence of coverage holes and outage issues, hence reducing network performances. Moreover, the ever-growing number of devices expected for 5G worsens this situation, calling for novel bandwidth-efficient, low-latency and cost-effective solutions for indoor wireless coverage. This is the focus of this article, which summarizes the content of my Ph.D. thesis by presenting an analog Centralized Radio Access Network (C-RAN) architecture augmented by copper-cable, possibly pre-existing, to provide dense coverage inside buildings. This fronthaul architecture, referred to as Analog MIMO Radio-over-Copper (AMIMO-RoC), is an extreme RAN functional-split-option: the all-analog Remote Radio Units take the form of tiny, simple and cheap in-home devices, and Base Band Unit includes also signals' digitization. The A-MIMO-RoC architecture is introduced in this article starting from demonstrating its theoretical feasibility. Then, the origin and evolution of A-MIMO-RoC are described step-by-step by briefly going through previous works based on numerical analysis and simulations results. Finally, the overall discussion is complemented by results obtained with a prototype platform, which experimentally prove the capability of A-MIMO-RoC to extend indoor coverage over the last 100–200 m. Prototype results thus confirm that the proposed A-MIMO-RoC architecture is a valid solution towards the design of dedicated 5G indoor wireless systems for the billions of buildings which nowadays still suffer from severe indoor coverage issues.

This work is a summary of the Ph.D. thesis "Interference Mitigation Techniques in Hybrid Wired-Wireless Communications Systems for Cloud Radio Access Networks with Analog Fronthauling" [1] supervised by Prof. Umberto Spagnolini.

A. Matera (✉)
Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milan, Italy
e-mail: andrea.matera@polimi.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_2

2.1 Introduction

The goal of this article is to present a novel network architecture, referred to as Analog Multiple-Input Multiple-Output Radio-over-Copper (A-MIMO-RoC), whose goal is to guarantee pervasive indoor wireless coverage within buildings. The role of the proposed A-MIMO-RoC architecture in the fifth generation (5G) of wireless systems is clarified in the following through four main pillar ideas that inspired this work: *(i)* enhance/enable indoor coverage, *(ii)* Centralized Radio Access Network (C-RAN) architecture, *(iii)* analog fronthauling and, finally, *(iv)* the RoC paradigm. These concepts, which will continually recur in the remainder of the paper, are now introduced one-by-one.

2.1.1 Indoor Propagation and C-RAN Architectures

By 2020 5G networks will be a reality, and 96% of wireless data traffic will originate or terminate within a building, with an exponential increase of indoor wireless market value [2]. However, only 2% of the 30 billion square meters of indoor commercial real estate have dedicated in-building wireless systems, while the remaining are still served by networks originally designed for outdoors. This is the reason why providing enhanced solutions for indoor coverage is the main task that both industries and academia are facing for the deployment of upcoming 5G systems.

Network deployment densification is almost mandatory to achieve the 5G performance targets. However, it will cause challenging interference scenarios, thus calling for an actual revolution in the way RAN resources are handled. According to this context, the C-RAN architecture [3], already adopted in 4G networks, will surely play a key role also in the next generation 5G networks, albeit it will need to be substantially redesigned to accommodate the more demanding 5G requirements.

2.1.2 Analog Fronthauling Architectures

In 4G networks, Remote Radio Units (RRUs) and BaseBand Units (BBUs) communicate over the so-called FrontHaul (FH) link, which is a high-capacity link conventionally based on fiber optics. This architecture is designed to support the streaming of digitized RF signals according to specific digital streaming protocols, such as the Common Public Radio Interface (CPRI) [4].

In perspective, it is widely agreed that today's CPRI-based mobile FH will hardly scale to the increased 5G radio signal bandwidths, especially for multiple-antenna RRUs. This is mainly due to signal digitization that, beside the uncontrolled end-to-end latency, would also cause a severe bandwidth expansion, thus exceeding the capacity of current FH links. In order to overcome these limitations, numerous RAN

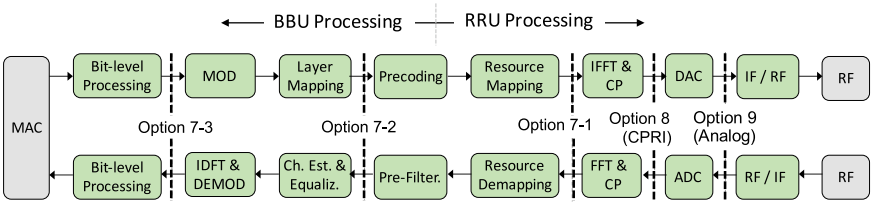


Fig. 2.1 Functional split options proposed for 5G New Radio (NR) and analog fronthaul

architectures have been recently proposed with the aim of making the redistribution of RAN functionalities between BBU and RRU more flexible [5] (see Fig. 2.1).

In contrast with these digital RAN architectures, an effective solution to relax the FH requirements is the overtaking of conventional digital FH connections in favor of a fully analog relay of RF signals between RRUs and BBUs. Analog FH links completely by-pass bandwidth/latency issues due to digitization, reduce hardware costs, improve energy efficiency, and easily allow for synchronization among multiple decentralized RRUs, thus enabling any MIMO joint-RRUs processing. As shown in Fig. 2.1, C-RAN architectures with analog FH represent an extreme RAN functional split option whereby even the Analog-to-Digital and Digital-to-Analog Converters (ADC/DAC) are shifted to the BBU, while only the RF functionalities are left to the RRUs, which in turn become extremely low-complexity and analog-based only devices. RRU functionalities are limited to the relaying of Intermediate Frequency (IF) signals to/from the BBUs, after having frequency-converted and scaled down these IF signals in order to comply with FH capabilities. Furthermore, the RRUs used in analog FH are protocol-independent units, hence capable to transparently relay signals belonging to any Radio Access Technology (RAT), which represents an important step towards the wide-range heterogeneity promised by 5G networks.

2.1.3 Radio-over-Copper

Analog C-RAN architectures can be based on different FH technologies. Among these, Analog Radio-over-Fiber (A-RoF), which is based on the analog relaying of optical IF signals over long-range fiber-optic links, is a very attractive FH architecture capable to carry several Gbps in terms of equivalent wireless data-rate [6]. However, the deployment of a pervasive optical infrastructure would be too costly to provide a satisfactory business case for practical indoor applications. In this case, the most intuitive solution is to extend indoor coverage by relaying the analog FH signals over the pre-existing Local Area Network (LAN) cabling infrastructure of buildings leading to the so-called A-RoC paradigm [7]. A-RoC architecture allows for the remotization of RF equipment, which are moved in proximity of the User Equipments (UEs) without the need of a new network infrastructure, hence improving indoor

wireless coverage over the last 100–200 m. Moreover, by leveraging the Power-over-Ethernet (PoE) technology, A-RoC enables to power the remote RF equipments over the same copper-cables, which simplifies the final architecture as no additional power supply devices are needed. The A-RoC paradigm encompasses both the advantages of analog FH and the economical benefits of reusing the existing copper-cables infrastructures, thus becoming a perfect candidate for extending 5G indoor wireless coverage as pursued by the telecom industry [8–10].

2.1.4 Contribution

This article is intended as a summary of my Ph.D. thesis [1]. Starting from the native A-RoC concept, it presents a more general and flexible analog FH architecture capable to carry multi-RAT/multi-antenna RF signals indoors over 50–200 m copper-cables. This novel FH architecture is referred to as Analog-MIMO-RoC, emphasizing the multiple-links nature of both radio (by antennas) and cable (by twisted-pairs) channels connecting the BBU with the UEs.

In particular, this article briefly presents the A-MIMO-RoC architecture from both theoretical and experimental perspectives. In this direction, firstly, the feasibility of the proposed A-MIMO-RoC architecture is demonstrated by numerical results, showing that LAN cables are suitable candidates for relaying RF signals indoors over more than 100 m. Secondly, the evolution of the A-MIMO-RoC architecture is described by reviewing previous works that exhaustively present the theory behind the A-MIMO-RoC architecture. These works cover both uplink and downlink A-MIMO-RoC channels, single- and multi-UE settings, and also provide useful insights on the performance trade-offs among heterogenous 5G services when they coexist in the A-MIMO-RoC architecture. Lastly, the overall theoretical discussion is complemented by experimental validations of a A-MIMO-RoC prototype for multi-antenna FH indoor applications.

2.2 The Genesis of the A-MIMO-RoC Architecture

By reviewing important milestones in the evolution of copper-based FH architectures, this section describes how A-MIMO-RoC origins from the native A-RoC concept.

The A-RoC concept dates back to [7], where twisted-pairs copper-cables were proposed for femto-cell systems to exchange analog RF signals between a remote location hosting all PHY/MAC functionalities (BBU) and an in-home antenna device performing only the analog relay of signals (RRU).

Afterwards, the A-RoC architecture gained lots of attention becoming the basis of commercial solutions exploiting the pre-existing LAN cables of buildings to extend indoor coverage over distances longer than 100 m [8]. By using all 4 twisted-pairs contained into the LAN cable at low-frequency (characterized by low attenuation

and crosstalk interference [11]), one can serve up to 4 antennas (e.g., 4 × 4 MIMO) per LAN cable.

Still based on the A-RoC concept, Huang et al. [12] proposed an LTE-over-copper architecture based on the colocation of RRU and Digital Subscriber Line (DSL) equipment in the same street cabinets. Authors proved that, by employing a single twisted-pair in the 21–24 MHz cable frequency band (not used by any DSL service), it is feasible to transport a 3GPP compliant LTE radio signal up to 350 m away from the cabinet. Crosstalk mitigation in LTE-over-copper systems is covered in Medeiros et al. [13] for the case of 6 twisted-pairs interfering with each other, still in the 21–24 MHz frequency range.

All the aforementioned works proved the feasibility of A-RoC as an alternative/complementary technology for FH links. However, none of them attempted to push the usage of cable frequency beyond the first tens of MHz, thus not making an efficient usage of the large bandwidth offered by copper-cables. As detailed in the following, this is precisely one of the key characteristics of the A-MIMO-RoC architecture presented here.

2.3 Theoretical Feasibility and Challenges of A-MIMO-RoC

This article presents an extension of the A-RoC paradigm to multiple-antennas RRUs and multiple twisted-pairs copper-cables, e.g., LAN cables (i.e., Cat-5/6/7). As shown in Fig. 2.2, the result is a more flexible and general FH architecture, namely

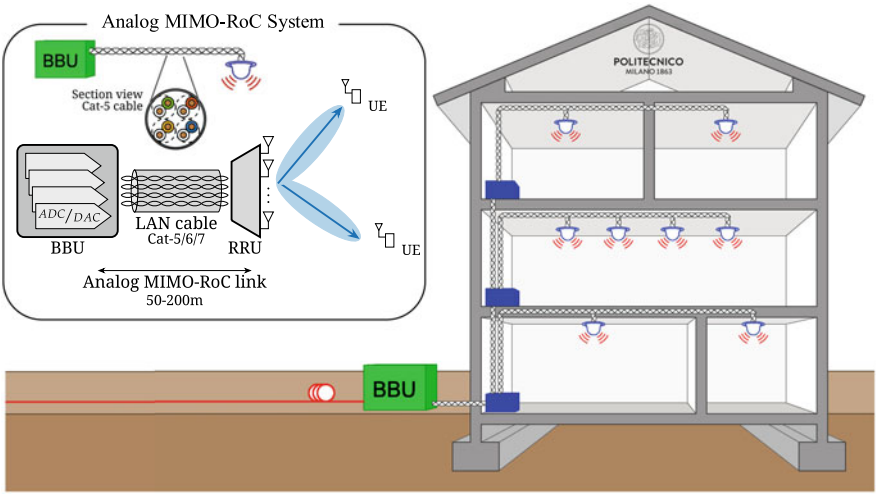


Fig. 2.2 The analog MIMO Radio-over-Copper architecture for indoor wireless coverage

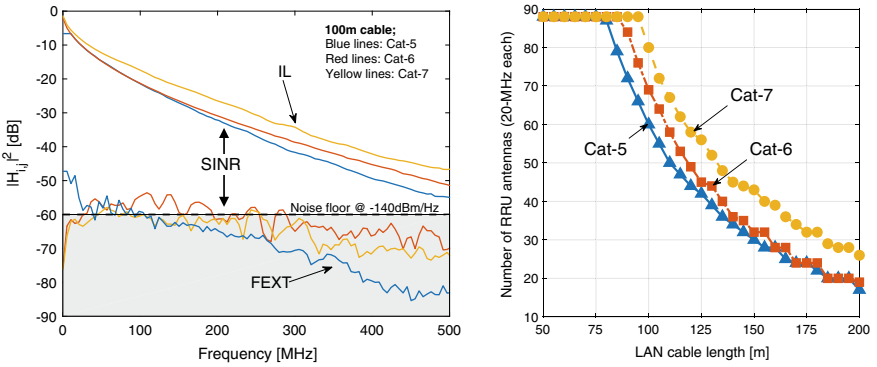


Fig. 2.3 The transport capabilities of LAN cables: IL and FEXT characteristics vs frequency (left) and number of RRU antennas served per LAN cable versus cable type and cable length (right) [11]

A-MIMO-RoC, which is characterized by the cascade of a MIMO wired channel over a MIMO wireless channel. Inspired by the A-RoC concept, an important stepping-stone towards the proposal of the A-MIMO-RoC architecture has been to demonstrate the huge transport capabilities offered by LAN cables, which contain 4 twisted-pairs bonded together offering up to 500 MHz bandwidth/pair (2 GHz overall). To this aim, the high frequency portion of LAN cables for indoor C-RAN applications has been first explored in our previous work [11] by numerical simulations. As shown in the left part of Fig. 2.3, although cable Insertion Loss (IL) and Far-End-Crosstalk (FEXT) among the 4 pairs rapidly increase versus frequency, the Signal-to-Interference-plus-Noise Ratio (SINR) is high enough to allow for transmission up to several hundreds of MHz. This is confirmed by the right portion of Fig. 2.3, which shows that more than 60 RRU antennas carrying a 20 MHz LTE channel/ea. can be served by a 100 m Cat-5 cable with approx. 500 MHz bandwidth/pair, and this number raises up to approx. 70 and 80 RRU antennas for Cat-6 and Cat-7 LAN cables, respectively.

Previous work [11] asserts the theoretical feasibility of the proposed architecture. However, in practice, the analog nature of A-MIMO-RoC poses several technical challenges that complicate its design and optimization: *(i)* as mentioned, FEXT among the 4 pairs and IL, if not properly handled, severely limit the performance of LAN cables, especially at high cable frequency (see the left part of Fig. 2.3), *(ii)* for a large number of RRU antennas, it arises the problem of how to map the signal to/from each antenna onto the available cable resources defined in space dimension (i.e., the 4 twisted-pairs) and frequency dimension (i.e., the frequency bandwidth of each twisted-pair), *(iii)* the all-analog RRU equipment (e.g., home-device) should be as simple/cheap as possible, but in the meanwhile able to handle up to several tens of antennas, *(iv)* LAN cables are subject to strict power constraints that must be carefully taken care of in the system design, especially in downlink direction, *(v)* in case of multiple-UEs, interference cancellation techniques for compound A-MIMO-RoC should be properly designed, but still releasing the RRU from any computationally complex signal processing, and *(vi)* in heterogeneous 5G networks, the coexistence

among different services with different constraints in terms of data-rate, reliability and latency must be carefully investigated.

The goal of my Ph.D. research activity has been to propose effective solutions to the problems above. These are briefly presented in the next section.

2.4 A-MIMO-RoC: Numerical Analysis and Simulations Works

This section describes the evolution of the A-MIMO-RoC architecture by briefly presenting the main works contributing to build the A-MIMO-RoC theory and addressing the critical issues introduced in the previous section. In this regard, the critical aspects *(i), (ii), (iii)* mentioned above have been addressed in our previous works [14, 15], which present a single-UE uplink A-MIMO-RoC scenario based on LAN cables. In particular, [14] proposes a novel and flexible resource allocation strategy between the RF signals at each RRU antenna, received from the UEs over the wireless channel, and the twisted-pair/frequency resources over the cable FH link. This wireless-wired resource mapping, referred to as Space-Frequency to Space-Frequency (SF2SF) multiplexing due to the space-frequency nature of both wireless and wired signals, is shown in the paper to substantially mitigate the impairments introduced by the cable FH, once evaluated in terms of UE throughput. An example is shown in Fig. 2.4, in which the RRU receives the signal from the single-UE, maps it onto the cable by exploiting the SF2SF multiplexing principle, and then relays it

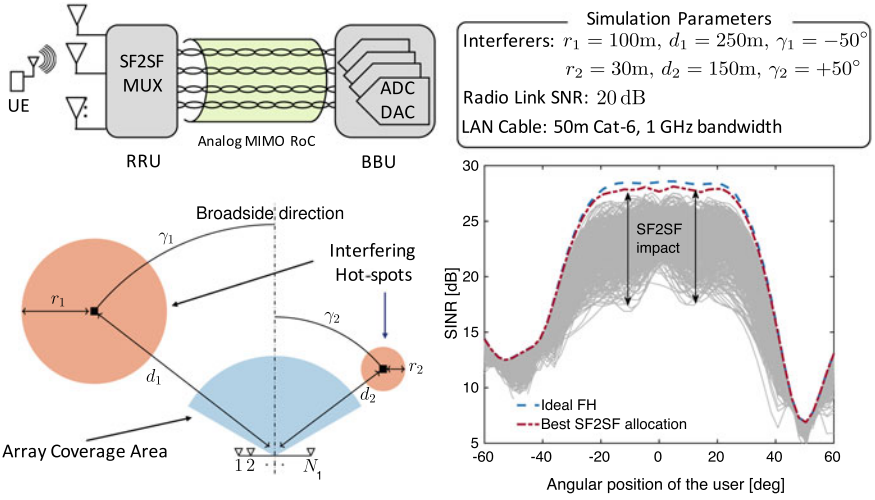


Fig. 2.4 Impact of SF2SF multiplexing on the single-user A-MIMO-RoC performance [14] (color figure online)

to the BBU over the LAN cable. The SINR for the single-UE signal at the BBU is shown as performance metric by varying the angular position of the UE within the array coverage area, and assuming that two adjacent hotspots interfere with the useful signal. Each solid grey line in the plot is obtained by randomly selecting a SF2SF allocation at the RRU. For the purposes of this article, beside the expected performance losses in correspondence of the two interfering hotspots, it is enough to notice the great gain the can be achieved by selecting the best SF2SF at each angle (dashed-dotted red line) which attains the performance of the ideal mobile FH (dashed blue line). More details can be found in [14].

A similar scenario is considered in [15], where the coexistence between FH signals and other services (e.g., DSL, PoE, etc.) over the same LAN cable has been tested for indoor, thus confirming the performance boost provided by SF2SF multiplexing technique. Still focusing on the single-UE A-MIMO-RoC channel, [16] presents an information theoretical study for heterogenous 5G networks, whereby enhanced Mobile Broadband (eMBB) and Ultra Reliable Low Latency Communications (URLLC) services coexist in the same physical resources. The analysis in [16] provides some useful insights on the performance trade-offs between the two services when they coexist in the uplink of the proposed A-MIMO-RoC architecture.

The problem of optimally designing SF2SF multiplexing is tackled for the A-MIMO-RoC downlink channel in single- and multi-UE settings in [17] and [18], respectively. In addition to all the issues due to the analog relaying of signals already mentioned for the uplink channel, the downlink SF2SF problem is complicated by the different power constraints that need to be jointly fulfilled both at the cable input and at the RRU antennas. Furthermore, in multi-UE settings digital precoding at the BBU needs to be properly designed in order to cope with the resulting multi-UE interference. As a first step, [17] confirms the potential of SF2SF also for the single-UE downlink A-MIMO-RoC channel. Then, [18] shows that, in multi-UE settings, the SF2SF multiplexing optimization (performed at the RRU), jointly designed with digital precoding of the overall wired-plus-wireless channels (performed at the BBU) and UE ordering optimization, is able to cope with both multi-UE interference and analog FH impairments, thus providing substantial performance gains in terms of minimum rate guaranteed for all the UEs. Finally, [19] focuses on the precoder design problem proposing a nearly-optimal BBU precoding algorithm and RRU power amplification strategy for the multi-UE A-MIMO-RoC downlink channel.

2.5 A-MIMO-RoC Prototype and Experimental Results

The A-MIMO-RoC prototype platform has been purposely developed in order to validate by real-world experiments the proposed FH architecture. The goal of this section is to describe part of the experiments that have been conducted in order to: *(i)* confirm the transport capabilities offered by the pre-existing copper cabling infrastructure of buildings, and *(ii)* experimentally prove the potential of the proposed SF2SF wireless-wired resource allocation strategy, expected to enable A-MIMO-

RoC to transparently carry multi-antenna RF signals over a single LAN cable. In particular, this is achieved by demonstrating the possibility to relay MIMO LTE signals, in an all-analog fashion, over LAN cables exploiting also high cable-frequency (tested here up to 400 MHz) with negligible performance degradation.

2.5.1 Experimental Settings

The A-MIMO-RoC prototype, shown in Fig. 2.5, is composed by two identical bi-direction LAN-to-Coax Converters (L2CCs). The prototype platform has been tested through the TRIANGLE testbed, the essence of the H2020 TRIANGLE project [20], that allowed us to experimentally measure the end-to-end performance degradation introduced by the analog relaying over the copper-cable that is interposed between the two L2CCs. The LTE signals to be relayed over copper have been generated by the Keysight UXM RAN emulator.

The TRIANGLE testbed has been employed in its typical device-testing configuration, but inserting a 4-pairs RJ45 LAN cable between the RF output ports of the UXM and the RF connections at the UEs. The cable interfaces with the UXM and the UEs by means of the two L2CCs, which represent the core of the A-MIMO-RoC platform. The L2CCs perform impedance adaptation, cable equalization, coax-to-pairs mapping/demapping and RF/IF conversion (i.e., that implement SF2SF functionalities), and include all-passive/all-analog devices along the signal path to ensure fully bi-directional operations.

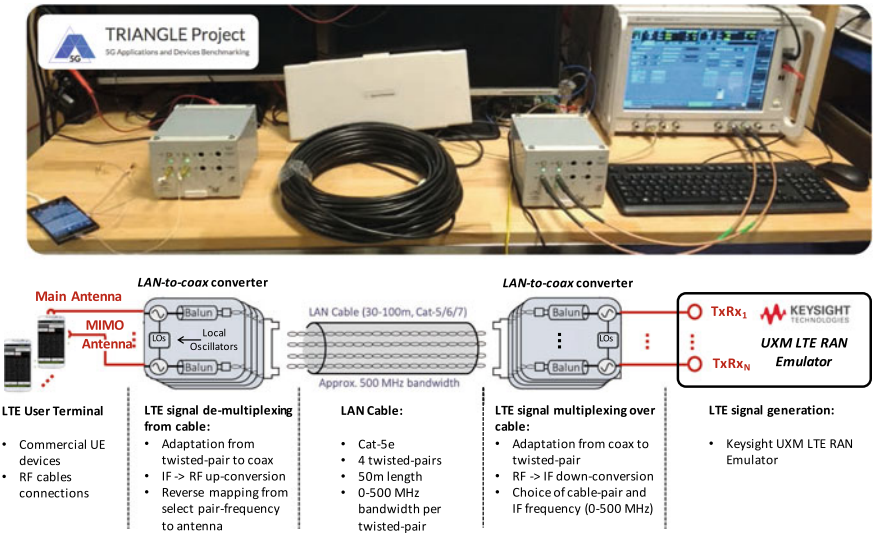


Fig. 2.5 A-MIMO-RoC prototype and experiment setup

For all wireless communications performed by exploiting the A-MIMO-RoC platform, we used a 50 m Cat-5e LAN cable (which is the cable type commonly deployed in buildings) with a bandwidth experimented here up to 400 MHz per each twisted-pair. For the LTE signals, RF cables have been used to connect the second L2CC (i.e., the one on the left of Fig. 2.5) to the UE.

The top of Fig. 2.5 shows a simplified block diagram of the experimental setup, the bottom details the role of each component used for the experiments. In particular, the experiment setup is as follows (only the downlink is described, uplink is symmetrical): *(i)* up to 4 LTE signals are generated by the UXM; *(ii)* RF cables are connected at each RF output of the UXM; *(iii)* the signal carried on each RF cable is IF-converted to match the LAN cable bandwidth (e.g., in the 10–400 MHz range), possibly multiplexed in frequency over cable by the first L2CC; *(iv)* each IF-converted signal is conveyed by one of the 4 twisted-pairs: cable adaptation/equalization, coax-to-pairs mapping and RF-to-IF conversion between coax and twisted-pair are performed by the first L2CC; *(v)* at the other end of the cable, RF cables are connected to the RF connectors of the second L2CC, that performs cable adaptation/equalization, pairs-to-coax de-mapping and IF-to-RF conversion to interface with the UEs under test; *(vi)* the DEKRA Performance Tool and the TestelDroid test suite, integrated into the TRIANGLE testbed, are used to test the UEs performances [20].

2.5.2 A-MIMO-RoC for LTE Versus Cable IF

This experiment concerns the tests about relaying over copper a 2×2 MIMO LTE signal. The goal is two-fold: *(i)* to prove the feasibility of transporting 2 LTE RF bands, corresponding to the MIMO signal, over 2 different twisted-pairs of the LAN cable, but at the same cable IF (f_{IF}) to validate the extreme case of a strong mutual interference between two twisted-pairs (see Fig. 2.6); and *(ii)* to evaluate the performance degradation caused by increasing f_{IF}, i.e., by using f_{IF} for which interference among pairs and attenuation are more severe.

Performance have been evaluated in downlink direction in terms of throughput and BLER for different MCSs, from 0-QPSK to 17-16QAM [22]. Figure 2.7 shows

LTE System Settings	
Signal Generation	Keysight UXM RAN Emulator
LTE Band	B38 TDD, 2570-2620 GHz
Bandwidth (BW_{LTE})	5 MHz
Antenna Config.	2x2 MIMO
Transmission Mode	TM3
Radio Channel	ETSI Static MIMO
Tx Power	-20 dBm

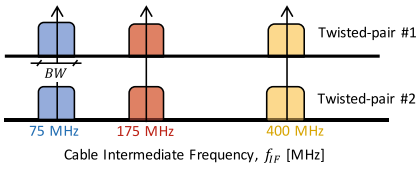


Fig. 2.6 A-MIMO-RoC experiment settings and RF signals mapping over cable

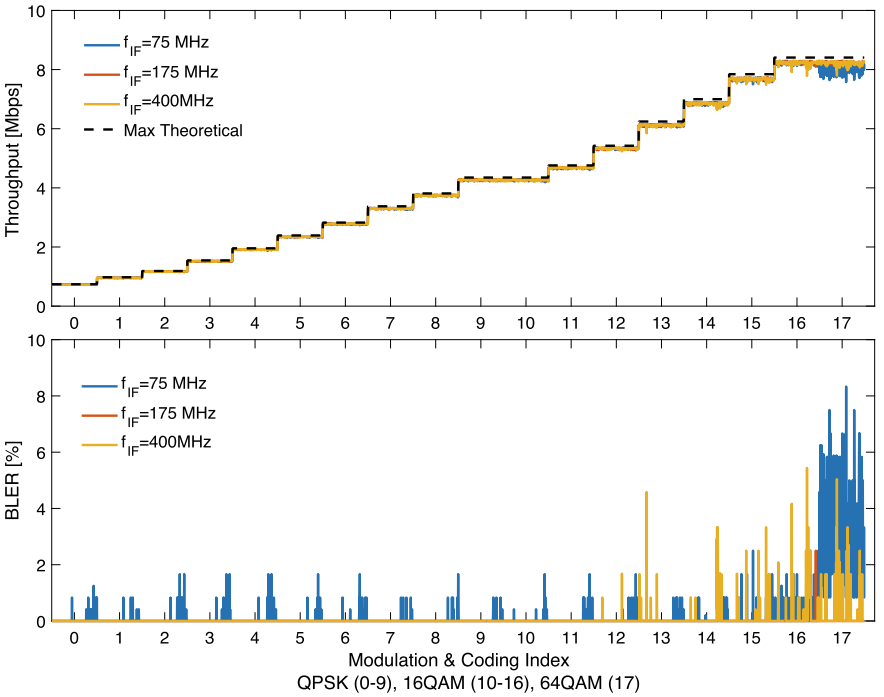


Fig. 2.7 Experiments for a 2 × 2 MIMO LTE signal: BLER and throughput versus MSC and IFs [21]

both throughput and BLER for a Static MIMO channel, $BW_{LTE} = 5$ MHz and three different cable $f_{\text{IF}} = 75{,}175{,}400$ MHz. For brevity, this is the only configuration shown in this article. However, additional results for LTE signals with different MIMO channel models and RF bandwidths are in [21]. The LTE system settings and signal mapping over cable adopted for the tests are reported in Fig. 2.6. The maximum theoretical throughput achievable by each MCS [22] over the considered channel bandwidth BW is shown as reference. Figure 2.7 proves that the performance loss due to the A-MIMO-RoC prototype is almost negligible for all considered MCS and cable f_{IF} selections. As expected, BLER increases for high MCS, but the degradation w.r.t the maximum throughput is still small. Concluding, Fig. 2.7 confirms experimentally the feasibility of relaying MIMO LTE signals over copper-cables at high frequency, even in the worst case of 2 LTE bands carried over 2 twisted-pairs at the same $f_{\text{IF}} = 400$ MHz, and thus maximally interfering with each other.

2.6 Conclusions

This article summarized the content of my Ph.D. thesis by presenting the Analog MIMO Radio-over-Copper (A-MIMO-RoC) architecture: an analog Cloud Radio Access Network (C-RAN) architecture that exploits analog FrontHaul (FH) links based on pre-existing copper-cables, e.g., Local Area Network (LAN) cables, to distribute Radio-Frequency (RF) signals into buildings at low cost, complexity and latency. The article first introduced the A-MIMO-RoC architecture by showing simulation results that assert the feasibility of LAN cables to efficiently relay RF indoors over the last 50–200 m. Then, several previous works have been briefly reviewed in order to present the theory behind A-MIMO-RoC. These works cover multiple aspects of A-MIMO-RoC including wired-wireless resource allocation strategies for uplink and downlink channels, interference mitigation techniques for single- and multi-user settings, and the investigation of performance trade-offs for heterogeneous 5G networks with multiple services coexisting in the same physical resources. The overall theoretical discussion has been finally supported by experimental results obtained with a hardware prototype platform. These confirmed that A-MIMO-RoC is not only an interesting research topic providing numerous theoretical insights, but mainly a practical solution capable to cope with real-world problems that engineers and researchers are facing today in deploying next generation 5G indoor networks.

Acknowledgements Part of this work has been funded by the TRIANGLE project, European Union Horizon 2020 research and innovation programme, grant agreement No 688712. The part of this work about 5G RAN functional split options is the result of a fruitful interaction with Nokia Bell Labs and their support is acknowledged. Last but not least, many thanks go to my Ph.D. supervisor Prof. Umberto Spagnolini for his continuous support and guidance.

References

1. Matera A (2019) Interference mitigation techniques in hybrid wired-wireless communications systems for cloud radio access networks with analog fronthauling
2. CommScope (2016) Wireless in buildings: what building professionals think, Report
3. Checko A et al (2015) Cloud RAN for mobile networks: a technology overview. Commun Surv Tuts 17(1):405–426
4. CPRI Specifications V.6.1 (2014-07-01), September 2014
5. 3GPP TSG RAN (2017) TR 38.801 v14.0.0, Study on new radio access technology: radio access architecture and interfaces (Release 14)
6. Wake D et al (2010) Radio over fiber link design for next generation wireless systems. J Lightw Technol 28(16):2456–2464
7. Gambini J et al (2013) Wireless over cable for femtocell systems. IEEE Commun Mag 51(5):178–185
8. Lu C et al (2014) Connecting the dots: small cells shape up for high-performance indoor radio. Ericsson Rev 91:38–45
9. Weldon MK (2016) The future X network: a Bell Labs perspective. CRC Press, Boca Raton
10. HKT, GSA, Huawei, "Indoor 5G Networks - White Paper," Sept. 2018
11. Naqvi SHR et al (2017) On the transport capability of LAN cables in all-analog MIMO-RoC fronthaul. In: IEEE WCNC. IEEE, pp 1–6

12. Huang Y et al (2015) LTE over copper-potential and limitations. In: IEEE PIMRC. IEEE, pp 1339–1343
13. Medeiros E et al (2016) Crosstalk mitigation for LTE-over-copper in downlink direction. IEEE Commun Lett 20(7):1425–1428
14. Matera A et al (2017) Space-frequency to space-frequency for MIMO radio over copper. In: IEEE ICC. IEEE, pp 1–6
15. Matera A et al (2017) On the optimal space-frequency to frequency mapping in indoor single-pair RoC fronthaul. In: EuCNC. IEEE, pp 1–5
16. Matera A et al (2018) Non-orthogonal eMBB-URLLC radio access for cloud radio access networks with analog fronthauling. Entropy 20(9):661
17. Matera A et al (2018) Analog MIMO-RoC downlink with SF2SF. IEEE Wirel Commun Lett 1–1
18. Matera A et al (2019) Analog MIMO radio-over-copper downlink with space-frequency to space-frequency multiplexing for multi-user 5G indoor deployments. IEEE Trans Wirel Commun 18(5):2813–2827
19. Rizzello V et al (2019) Precoding design for the MIMO-RoC downlink. In: IEEE ICASSP 2019, pp 4694–4698
20. Cattoni AF et al (2016) An end-to-end testing ecosystem for 5G. In: EuCNC. IEEE, pp 307–312
21. Matera A et al (2019) Analog MIMO radio-over-copper: prototype and preliminary experimental results. In: ISWCS 2019. IEEE
22. 3GPP TS 36.213 group radio access network; Evolved universal terrestrial radio access (E-UTRA), "Physical layer procedure."

Part II
Electronics

Chapter 3
Chirp Generators for Millimeter-Wave FMCW Radars

Dmytro Cherniak and Salvatore Levantino

Abstract The vast number of radar applications generates the demand for highly-linear, low-noise, fast chirp generators implemented in nanoscale CMOS technologies. Off-the-shelf chirp synthesizers are realized in BiCMOS technologies and demonstrate excellent phase noise performances, though, at high cost and limited modulation speed. This chapter describes a new class of fast and reconfigurable chirp generators based on digital bang-bang phase-locked loops, suitable for integration in modern CMOS processes. After analyzing the impact of the chirp generator impairments on a frequency-modulated continuous-wave (FMCW) radar system, a novel pre-distortion scheme for the linearization of critical blocks is introduced to achieve at the same time low phase noise and fast linear chirps. The chirp generator fabricated in 65-nm CMOS technology demonstrates above-state-of-the-art performance: It is capable of generating chirps around 23-GHz with slopes up to 173 MHz/μs and idle times of less than 200 ns with no over or undershoot after an abrupt frequency step. The circuit consuming 19.7 mA exhibits a phase noise of −100 dBc/Hz at 1 MHz offset from the carrier and a worst case in-band fractional spur level below −58 dBc.

Keywords CMOS · Phase-locked loops · Radar

3.1 Introduction

Cost reduction is the cornerstone in nowadays radar systems. The radar technology, which has been for a long time, since the invention in early 1930s, exclusive to military and defence, is now being adopted in a wide spectrum of applications

This work has been supported by Infineon Technologies, Austria.

D. Cherniak (✉)
Infineon Technologies, Siemensstrasse 2, 9500 Villach, Austria
e-mail: dmytro.cherniak@infineon.it

S. Levantino
Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy
e-mail: salvatore.levantino@polimi.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_3

including automotive, industrial and consumer market [1, 2]. In the context of advance driver assistance system (ADAS) and autonomous driving, the radar technology has been employed for about two decades for object detection as well as range, relative velocity and azimuth sensing. In the future, the combination of radar sensors with machine learning will enable the vision of the vehicle. In comparison with other environmental perception sensors such as cameras and light detection and ranging (LIDAR), radars operate under foggy, dusty, snowy and badly lighted environment, which is essential for the automotive applications [3, 4]. Initial automotive radar sensors were extremely expensive as they were based on discrete circuit elements. The next-generation radar systems were implemented with several monolithic microwave integrated circuits (MMICs) in high-performance GaAs technology [1]. Further cost reduction and increased level of integration was achieved by moving to SiGe bipolar or BiCMOS technology [5]. Most of the nowadays radar systems are realized in BiCMOS technology, but the fast development of the automotive industry demand for a single-chip radar solution in modern CMOS technology [6].

In a frequency-modulated continuous-wave (FMCW) radar system, an FM-modulated carrier is transmitted and the delay between the reflected and the transmitted carriers is taken as a measure of the distance, i.e. the range, between the radar and the target. Using triangular or saw-tooth waveforms to frequency-modulate the carrier, the delay between the reflected and transmitted signals can be easily and accurately measured by detecting the frequency offset between the two signals, as shown in Fig. 3.1. A carrier with a linear FM modulation is referred to as a *chirp* signal. The performance of an FMCW radar is mainly determined by the speed, linearity and phase noise of the chirp generator [7]. Different radar applications require different chirp configurations as well as different noise levels. In general, enlarging the modulation bandwidth, i.e. the peak-to-peak amplitude of the chirp modulation signal, improves the range resolution and lowering the phase noise increases the signal-to-noise ratio (SNR). However, reducing the period of the chirp (enabling fast chirps) would be one of the keys to improve the radar system performances. In fact, fast chirps allow:

- a larger separation in frequency of the targets,
- increasing the beat frequency beyond the flicker noise corner,
- increasing the maximum unambiguous velocity,
- improving the velocity resolution,
- averaging the detected signal, thus improving the SNR [8].

3.2 Digital PLL with Two-Point Modulation Scheme

The architecture of the DPLL-based modulator with the two-point injection technique is presented in Fig. 3.2a. The modulation signal $mod[k]$ is simultaneously applied with inverted sign to the feedback path of the phase-locked loop as well as added to the output of the digital loop filter. The above configuration allows to overcome

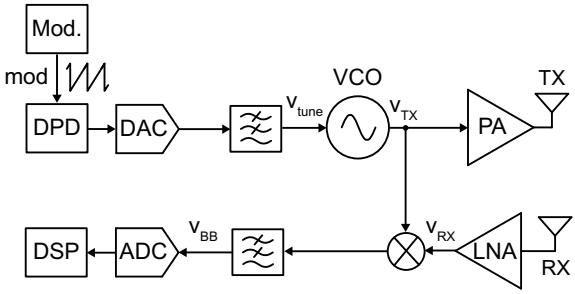


Fig. 3.1 FMCW radar architecture

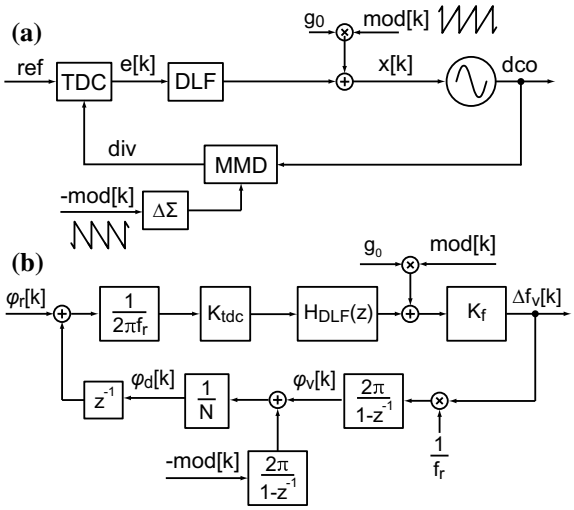


Fig. 3.2 DPLL-based modulator with two-point injection **a** architecture and its **b** phase-domain model

the bandwidth limitation of the conventional PLL-based modulators [9]. Figure 3.2b presents the simplified linearized phase-domain model of the DPLL-based modulator architecture in Fig. 3.2a. The signals $\varphi_r[k]$, $\varphi_d[k]$ and $\varphi_v[k]$ in Fig. 3.2b represent the variable phase of the reference, divided and output clocks, respectively. The signal $\Delta f_v[k]$ is the variation of the DCO output frequency and f_r is the reference frequency. The transfer function of the DLF is the standard proportional-integral one: $H_{DLF}(z) = \beta + \alpha/(1 - z^{-1})$.

The expression of the loop gain is given by

$$G_{loop}(z) = K \cdot H_{DLF}(z) \cdot \frac{z^{-1}}{1 - z^{-1}}, \tag{3.1}$$

where $K = K_f K_{tdc}/(N f_r^2)$, being N the feedback divider ratio, K_{tdc} the TDC gain expressed in [s/bit], K_f the DCO gain expressed in [Hz/bit].

Denoting as $mod(z)$ and $\Delta f_v(z)$ the z-transforms of the time-domain signal $mod[k]$ and the output frequency variation $\Delta f_v[k]$, and being $G_{loop}(z)$ the PLL

loop gain, the transfer function from $mod(z)$ injected into the feedback to $\Delta f_v(z)$ is:

$$H_{lp}(z) = f_r \cdot \frac{G_{loop}(z)}{1 + G_{loop}(z)}, \tag{3.2}$$

that is the standard low-pass transfer function of a PLL multiplied by f_r.

The transfer function from the other point of injection, at the DLF output, is instead the following one:

$$H_{hp}(z) = \frac{g_0 K_f}{1 + G_{loop}(z)}, \tag{3.3}$$

that is a high pass one.

If the value of the gain g_0 is exactly equal to f_r/K_f, the linear superposition of the two transfer functions results in an all-pass transfer function:

$$\begin{aligned} H_{mod}(z) &= H_{lp}(z) + H_{hf}(z) = \\ &= f_r \cdot \left(\frac{G_{loop}(z)}{1 + G_{loop}(z)} + \frac{1}{1 + G_{loop}(z)} \right) = f_r. \end{aligned} \tag{3.4}$$

In practice, the DCO gain K_f in (3.3) is not only variable over process, temperature and voltage (PVT), but also dependent on the output frequency, given the nonlinearity of a typical DCO tuning characteristic. Thus, the two-point injection technique requires an accurate estimation of g_0 coefficient in order to guarantee that it is always equal to f_r/K_f. Any inaccuracy in DCO gain estimation would alter the $H_{mod}(z)$ transfer function from its ideal value and give rise to the following frequency error:

$$\Delta f_{error} = mod(z) \cdot [f_r - H_{mod}(z)], \tag{3.5}$$

where $H_{mod}(z)$ is the transfer function from $mod(z)$ to $\Delta f_v(z)$, given by (3.4).

Intuitive considerations may suggest that the error made in the estimation of K_f is suppressed by the loop to some extent. However, how this error impacts the output Δf_{error} is not obvious and has never been addressed in the literature. In the following, the analysis of chirp linearity errors in the presence of DCO gain errors is carried out.

Replacing the DCO gain in (3.3) with the estimate $\hat{K}_f$ and computing again the expression of $H_{mod}(z)$, we derive the following equation:

$$\begin{aligned} H_{mod}(z) &= f_r \cdot \left(\frac{K_f}{\hat{K}_f} \cdot \frac{1}{1 + G_{loop}(z)} + \frac{G_{loop}(z)}{1 + G_{loop}(z)} \right) = \\ &= f_r \cdot \frac{1 + (1 + \epsilon_{K_f}) \cdot G_{loop}(z)}{(1 + \epsilon_{K_f}) \cdot [1 + G_{loop}(z)]}, \end{aligned} \tag{3.6}$$

where the last expression follows after introducing the relative gain error ϵ_{K_f}, such that $\hat{K}_f = K_f \cdot (1 + \epsilon_{K_f})$.

The chirp error ϵ_{chirp} normalized to the modulation amplitude BW can be derived (3.5) by replacing the expression of H_{mod} obtained in (3.6):

$$\epsilon_{chirp}(z) = \frac{\Delta f_{error}}{BW} = \frac{mod(z) \cdot f_r}{BW} \cdot \frac{\epsilon_{K_f}}{1 + \epsilon_{K_f}} \cdot \frac{1}{1 + G_{loop}(z)}. \tag{3.7}$$

Equation (3.7) reveals that, in the presence of DCO gain estimation error, the spectrum of the chirp error is given by the spectrum of the modulation signal after high-pass shaping with dominant pole located at around the closed-loop bandwidth of the PLL [the last factor in (3.7)]. This means that the PLL is able to reject chirp errors induced by the gain error, as long as the modulation speed is slow enough with respect to the bandwidth of the closed loop. If, instead, a portion of the spectrum of $mod[k]$ falls outside PLL bandwidth, the loop plays no role and the error propagates to the output causing chirp distortion.[1]

3.3 Adaptive Digital Pre-distortion

In FM modulators based on DPLLs, the nonlinearity of the DCO is the main source of chirp distortion. In practical DCOs, there are several sources of nonlinearity. On top of the intrinsic $1/\sqrt{LC}$ nonlinearity, DCOs may also exhibit a random or periodic nonlinearity due to mismatch within the digitally-controlled capacitor bank.

Figure 3.3 presents a typical class-B LC oscillator where the resonant tank includes a fixed inductor L_T and a digitally-controlled capacitor C_{Tx}, which is typically implemented using high-quality metal-metal capacitors with near-zero temperature and voltage coefficients [10]. However, the parasitic capacitance C_p of the MOS devices (e.g. cross-coupled differential pair, switch, etc.) exhibits a non-zero temperature and voltage coefficients which make the frequency-tuning curve dependent on those parameters.

In principle, DCO tuning characteristic can be linearized by applying the pre-distortion concept depicted in Fig. 3.4. The compensation can be implemented in either analog or digital way. Reference [11] presents a transformer-based LC-oscillator which mitigates the effect of the variable inductance. However, this approach requires complex electromagnetic (EM) simulations for an accurate design of the transformer and of the digitally-controlled capacitors. A more common approach is to implement pre-distortion in the digital domain by means of a look-up table (LUT) or a polynomial which is the inverse of the tuning characteristic of the

[1] More rigorously, the loop gain, $G_{loop}(z)$, in (3.7) should be slightly different from the nominal one, since it is also affected by the DCO gain errors itself ϵ_{K_f}. This correction, however, implies only a negligible variation in the loop gain value, and does not alter appreciably the final result given in (3.7).

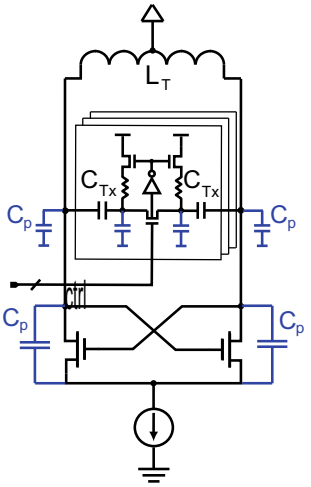


Fig. 3.3 Class-B DCO with annotated parasitic capacitance

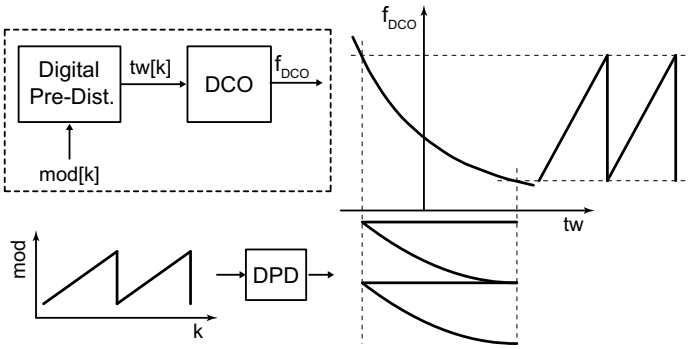


Fig. 3.4 The concept for DCO pre-distortion

DCO. In [12], the pre-distortion coefficients are stored in a 24 kbit SRAM LUT which is filled during the calibration phase that takes 4 s at the start-up. The main issue of this approach is that the pre-distortion coefficients are identified in foreground. So, the calibration is not robust against voltage and temperature variations.

A background calibration of the digital pre-distorter (DPD) was proposed in [13] which describes a DPLL-based phase modulator for wireless communications. The gain required to match the characteristics of each capacitor bank of the DCO is estimated in background by means of the LMS algorithm. The time constant of the implemented background calibration running at 40 MHz clock is in the range of 100 μs which allows effective tracking of slow temperature and voltage variations. To implement an ideal DPD for an N-bit DAC (e.g. like the DCO) a LUT with 2^N fields is required [12]. This kind of complexity is acceptable for the foreground DPD

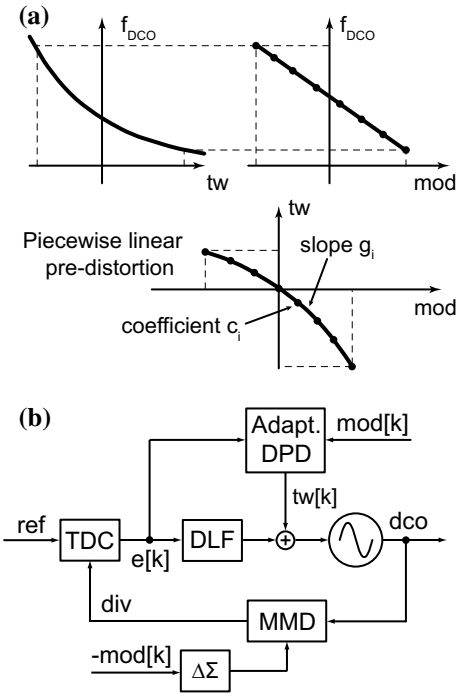


Fig. 3.5 The conceptual drawing of **a** a piecewise-linear DPD and **b** two-point modulator with an adaptive DPD

implementation. However, if the background calibration is required, the complexity of the digital part significantly increases. An LMS-based implementation would require a multiplexer and an accumulator for each LUT field.

The concept of a piecewise-linear DPD is aimed to reduce the complexity of the DPD and make an adaptive implementation feasible. In the context of DPLLs, an adaptive piecewise-linear DPD was originally proposed in [14] and later employed in [15, 16] for the correction of the digital-to-time converter (DTC) and time-to-digital converter (TDC) nonlinearity. The general idea of the piecewise linear DPD is illustrated in Fig. 3.5a. The inverse characteristic of the nonlinear block is approximated with a piecewise-linear curve. The DPD block is intended to remap the modulation signal $mod[k]$ to the tuning word $tw[k]$ in order to achieve a linear tuning characteristic of the output frequency f_{DCO} over the modulation signal $mod[k]$. The piecewise linear characteristic is constructed with a finite set of $\{c_i\}$ and $\{g_i\}$ coefficients representing the position of the segments and the connecting slopes, respectively.

Figure 3.5b presents the concept of the two-point digital PLL-based modulator which incorporates the adaptive DPD. The modulation signal $mod[k]$ is pre-distorted at the high-pass injection point before being applied to the phase-locked loop. The TDC output $e[k]$ is a digital representation of the phase error in the two-point modulator architecture. As it was presented in the previous section, any mismatch between the two injection points would appear as a phase error at the input of the TDC. Thus, the error signal $e[k]$ can be utilized to adapt the DPD characteristic in the background.

3.4 Implemented Prototype

A 23-GHz DPLL-based FMCW modulator was designed and implemented in an existing 65 nm CMOS technology [17]. The prototype features an adaptive piecewise-linear DPD with multiple slope estimation which is applied for DCO nonlinearity correction.

The block diagram of the implemented DPLL-based FMCW modulator is depicted in Fig. 3.6. The DPLL is based on a binary phase detector (BPD) (or a single-bit TDC) which operates in a random noise regime. The regime of the BPD is enabled by means of a DTC similar to the architecture originally proposed in [18]. The feedback path starts with a prescaler-by-four implemented in current-mode-logic (CML), which reduces the frequency of the DCO output to about 5.8 GHz and allows low-power implementation of the multi-modulus divider (MMD) in true-single-phase-clock (TSPC) logic. The output of the prescaler is also used to feed the pad driver.

The single-bit output of the BPD is fed into a digital loop filter which is implemented as a conventional proportional-integral (PI) filter with programmable coefficients. The digital core of the PLL is realized in a single-clock domain which simplifies its design. The clock used for the digital core is the divided clock after the DTC (div). The DPLL is fed with a 52-MHz reference clock derived from an on-chip crystal oscillator with an off-chip resonator.

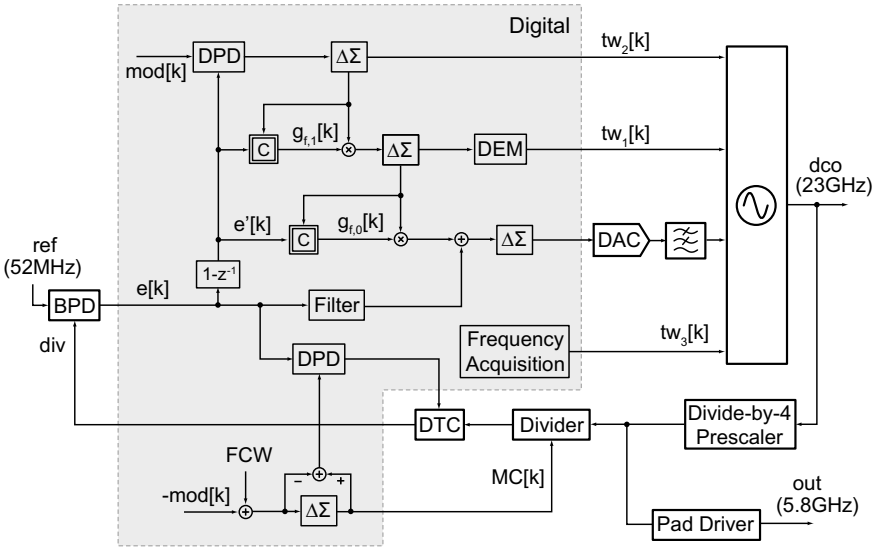


Fig. 3.6 Block diagram of the implemented DPLL-based FMCW modulator prototype

3.4.1 DTC Design and Control

The DTC is intended to cancel the quantization noise introduced by the MMD driven by a second-order $\Delta\Sigma$ modulator. The residual of the MMD quantization noise at the input of a single-bit TDC is within one LSB of the DTC which shall be designed to satisfy the random noise operation [18]. The nonlinearity of the DTC is suppressed by a DPD block [14] which ensures low spur and low phase noise operation of the implemented digital PLL.

The DTC is implemented as buffer with digitally-controlled capacitor load which is followed by a slope-regeneration buffer. Both of the buffers are implemented in CML to improve the rejection to supply disturbances. To realize a large delay-insertion range and fine LSB while maintaining low-area and parasitics, the digitally-control capacitor of the DTC is segmented in two 5-bit thermometer-coded capacitor banks, namely a $coarse$ and a $fine$ bank. Each capacitor bank is implemented using a digitally controlled MOS varactors. The DTC covers about 150-ps range with fine resolution of about 300 fs.

The control scheme of the DTC is based on cancellation of the quantization error which is introduced by the step of the coarse bank at the fine bank. A 16-segment digital piecewise-linear DPD scheme is applied to correct the nonlinearity of the coarse bank, to ensure low-noise and low-fractional spur operation. To match the characteristics of the coarse and fine capacitor banks, the quantization noise of the coarse bank $r[k]$ is scaled by an adaptive gain $\hat{g}_r[k]$ gain. The estimation of $\hat{g}_r[k]$ is implemented based on the LMS algorithm utilizing the quantization noise of the $\Delta\Sigma$ modulator $r[k]$ as a training sequence [18].

3.4.2 DCO Design and Control

To ensure a robust start-up without any additional circuitry, a class-B oscillator topology with an nMOS cross-coupled differential pair has been preferred over a class-C implementation. The main tank inductor is implemented as a single turn coil with a center-tap connected to the supply voltage and with five turns of a top metal layer connected in parallel. A tail filter made of a 75 pH spiral inductor and a 10 pF capacitor is employed to filter the noise of the tail current source, as well as to provide higher impedance at the second harmonic frequency similar to the concept proposed in [19].

To achieve a wide tuning range and fine frequency resolution, the DCO tuning characteristic is segmented in four different digitally-controlled capacitor banks. The coarsest bank is implemented using switched metal-metal capacitors. This capacitor bank is dedicated to coarse frequency tuning only and is designed to achieve about 13% of the tuning range. The finer banks, dedicated to modulation, are implemented using digitally-controlled MOS varactors. The finest tuning bank is implemented using an analog-tuned MOS varactor which is driven by a resistor-string 5-bit DAC.

The finest frequency resolution of the varactor-DAC cascade is about 150 kHz. The achieved complete tuning range is about 16%. The worst-case DCO phase noise, referred to 23 GHz carrier, is −106 dBc/Hz at 1 MHz offset frequency, at 10 mW power consumption. The flicker corner is located at about 200 kHz frequency offset. The control scheme of the multi-bank DCO features an adaptive piecewise-linear DPD applied to the coarsest capacitor bank dedicated to modulation since it contributes the most of the DCO nonlinearity.

The driving scheme for the finer banks is based on canceling the quantization noise of the coarser capacitor banks [13]. To match the characteristics of two consecutive banks, the quantization noise from the coarser banks has to be multiplied by a scaling coefficient (i.e. $\hat{g}_{f1}$ and $\hat{g}_{f0}$). The estimation of the required $\hat{g}_{fi}$ coefficients are done utilizing an LMS algorithm as depicted in Fig. 3.6. A digital delta-sigma modulator $\Delta\Sigma$ is used as a quantizer for each of the modulation capacitor banks since the statistical properties of its quantization noise are well suited for the LMS-based calibration [14]. The quantization noise of the finest bank is filtered by a second-order analog low-pass filter after the DAC.

The two-point modulation technique is implemented completely in the digital domain. The modulation signal $mod[k]$ is simultaneously added to the frequency control word and applied to the coarse tuning bank of the DCO. The adaptive piecewise-linear DPD scheme with multiple slopes estimation is introduced only at the coarse modulation bank, since its mismatch and nonlinearity are the most significant ones. The DPD is implemented with 16 segments. The number of segments is selected as a compromise between the accuracy of the nonlinearity correction and the digital hardware complexity.

3.5 Measurements

The prototype has been fabricated in a standard 65-nm LP CMOS process with no ultra-thick metal layer option. The analog and the digital portions of the chip whose photograph is shown in Fig. 3.7 occupy approximately the same share of the total 0.42 mm^2 area. The output of the frequency divider by four is used to carry out all the measurements. The output of the pad driver is wire bonded to the test board and used for testing. This setup reducing the output frequency to around 5.8 GHz simplifies both phase-noise measurement, which does not require external mixers, and modulation-analysis measurement, as it scales down by four the modulation depth.

The phase noise spectra in integer- and fractional-N modes measured by a R&S FSWP phase-noise analyzer are shown in Fig. 3.8. In both cases, the in-band noise plateau is at about −102 dBc/Hz and the phase noise at 1 MHz offset from the carrier at 5.928 GHz is about −112 dBc/Hz. The latter corresponds to about −100 dBc/Hz at 1-MHz offset when referred to the actual DCO output at 23.712 GHz. In fractional-N mode, the far-out phase noise exhibits a slight increment, increasing the absolute jitter from 213 fs (in integer-N mode) to about 242 fs (in fractional-N mode). Thanks

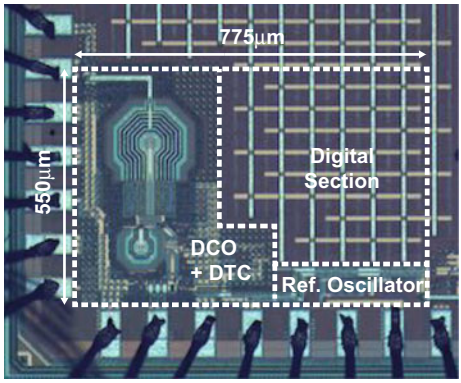


Fig. 3.7 Die photograph

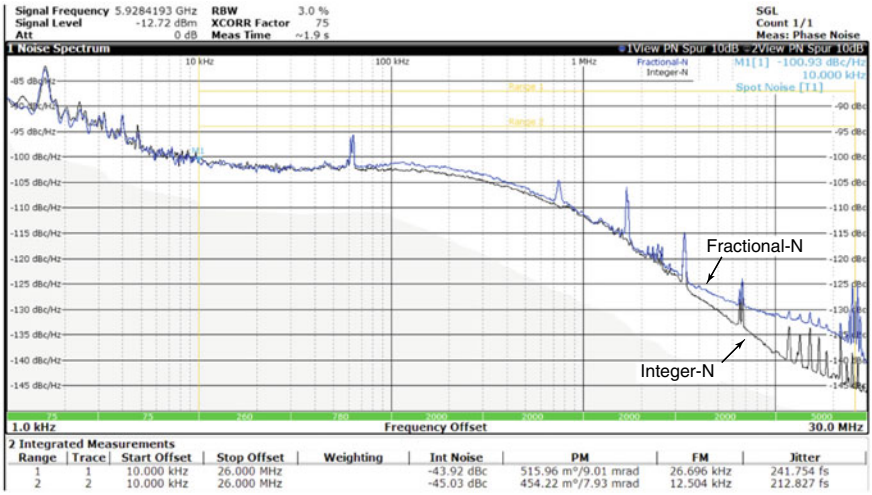


Fig. 3.8 Measured phase noise spectra in both integer-N and fractional-N mode at divider-by-four output

to the DPD of the DTC block, the worst case in-band fractional spur at the offset frequency of 173.8 kHz is below −58 dBc and the out-of-band spur at the offset frequency of 1.62 MHz is below −70 dBc.

An Anritsu MS2850A featuring 1-GHz demodulation bandwidth has been employed as vector signal analyzer (VSA). To assess the efficacy of the DPD in the DCO control, the modulator has been at first tested without enabling the adaptive DPD. A single gain for each bank of the DCO is estimated which corrects for mismatches between coarse and fine banks, but leaves the mismatches among the elements of the coarse bank and the systematic nonlinearity uncorrected. Figure 3.9a shows the demodulated frequency signal for a saw-tooth chirp with 40 µs period and

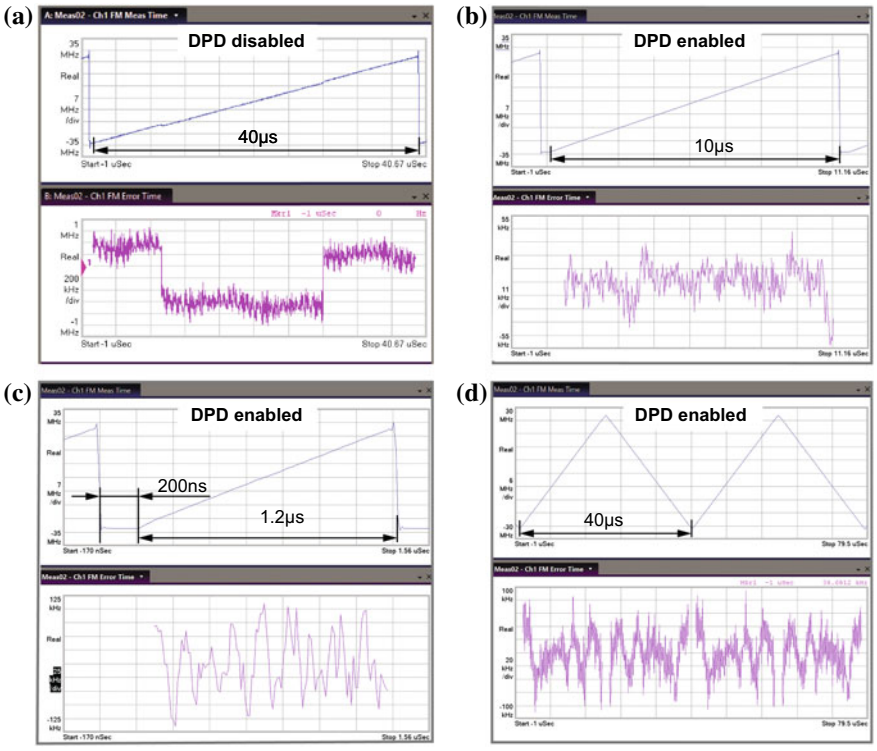


Fig. 3.9 Chirp measurements at divider-by-four output: **a** 40-µs saw-tooth with single-gain calibration (no DPD), **b** 10-µs-saw-tooth with DPD, **c** 1.2-µs saw-tooth with DPD, **d** 40-µs triangular with DPD

52-MHz frequency deviation at the output of the divide-by-four block (equivalent to 208 MHz at DCO output) and the corresponding frequency error. The effects of DCO nonlinearity are clearly visible in the scope, and the peak chirp error is about 1.058 MHz, that is about 2% with respect to the peak-to-peak chirp deviation.

The following figures, Fig. 3.9b–d, present instead the measurements when the adaptive DPD is enabled. The peak-to-peak frequency deviation is 52 MHz in both cases (equivalent to 208 MHz at DCO output). In Fig. 3.9b the chirp rise time is 10 µs, the resulting peak frequency error is below 50 kHz (equivalent to 200 kHz at DCO output), that is less than 0.1% of the maximum frequency deviation, and the RMS chirp error is 0.06%. Figure 3.9c shows the same frequency deviation covered in only 1.2 µs, that results in a state-of-the-art chirp slope of 173 MHz/µs at the DCO output. In this case, the peak error is below 100 kHz (i.e. below 0.2%). The circuit can generate also triangular chirps: Fig. 3.9d shows the demodulated waveform for a 40 µs period. The idle time required to match the 0.1% error specification is lower than 2.4 ns with no over or undershoot, thanks to the introduced DPD circuit which allows to exploit completely the two-point modulation technique.

Table 3.1 Performance comparison

	This Work	Wu JSSC14	Yeo ISSCC16	Vovnoboy JSSC18	Weyer ISSCC18	Ginsburg ISSCC18
Architecture	BBPLL + TPM	ADPLL + TPM	ADPLL + TPM	Analog PLL	DPLL	Analog cascaded PLL
Freq. range (GHz)	20.4–24.6	56.4–63.4	8.4–9.4	75–83	36.3–38.2	76–81
Technology	65 nm CMOS	65 nm CMOS	65 nm CMOS	130 nm BiCMOS	40 nm CMOS	45 nm CMOS
Ref. freq. (MHz)	52	40	276.8	125	120	40
Chirp type	Saw-tooth/Triangular	Triangular	Triangular	Saw-tooth	Triangular	Saw-tooth
Chirp duration (μs)	1.2–315	420–8200	5–220	50–225	50–2000	40
Saw-tooth idle time (μs)	0.2	n/a	n/a	15	n/a	15
Max. chirp Δf_{pp} (MHz)	208	1220	956	8000	500	4000
Max. chirp slope (MHz/μs)	173	4.76	32.6	100	9.1	100
RMS freq. error[b] (kHz)	124/112 (0.06%/0.05%)	384 (0.03%)	1900 (0.12%)	3200 (0.04%)	820 (0.16%)	n/a
Phase noise[a] (dBc/Hz)	−90	−87.7	−86.2	−97	−73.7	−91
Spur level (dBc)	−58	−62	n/a	n/a	−55	n/a
Power (mW)	19.7	48	14.8	590	68	n/a
Area (mm^2)	0.42	2.2	0.18	4.42	0.18	n/a

[a] At 1-MHz offset referred to a 79 GHz carrier
[b] At max. chirp slope

Comparing the presented chirp generator with other state-of-the-art CMOS and BiCMOS implementations in Table 3.1, it can be concluded that the presented DPLL with two point injection and DCO pre-distortion is able to generate fast chirps with the largest maximum slope at better than 0.1% linearity, at competitive phase-noise and power consumption levels.

3.6 Conclusions

Future radar sensors for autonomous vehicles and consumer applications will require fast chirp generators in low-cost CMOS processes. This chapter presented a new class of fast, linear chirp modulators based on digital PLLs and two-point injection of the modulation signal. To mitigate the impact of the nonlinearity of the DCO tuning characteristic, a novel digital piecewise-linear pre-distorter with reduced hardware resources is introduced. The pre-distorter automatically tracks any process and environmental variations. The chirp modulator fabricated in 65-nm CMOS technology demonstrated chirp signals around 23-GHz with slopes up to 173 MHz/μs, less than 0.1% error, and idle times of less than 200 ns.

Acknowledgements The authors would like to acknowledge Prof. Carlo Samori, Dr. Luigi Grimaldi, Dr. Luca Bertulessi for useful discussions and Infineon Technologies for supporting this work.

References

1. Steinbaeck J, Steger C, Holweg G, Druml N (2017) Next generation radar sensors in automotive sensor fusion systems. In: Proceedings of 2017 sensor data fusion: trends, solutions, applications (SDF), pp 1–6
2. Nasr I, Jungmaier R, Baheti A, Noppeney D, Bal JS, Wojnowski M, Karagozler E, Raja H, Lien J, Poupyrev I, Trotta S (2016) A highly integrated 60 GHz 6-channel transceiver with antenna in package for smart sensing and short-range communications. IEEE J Solid-State Circuits 51(9):2066–2076
3. Peynot T, Underwood J, Scheding S (2009) Towards reliable perception for unmanned ground vehicles in challenging conditions. In: Proceedings of 2009 IEEE/RSJ international conference on intelligent robots, pp 1170–1176
4. Brooker G, Hennessey R, Lobsey C, Bishop M, Widzyk-Capehart E (2007) Seeing through dust and water vapour: millimetre wave radar sensors for mining applications. J Field Robot 24(7):527–557
5. Dielacher F, Tiebout M, Lachner R, Knapp H, Aufinger K, Sansen W (2014) SiGe BiCMOS technology and circuits for active safety systems. In: Proceedings of 2014 international symposium on VLSI technology, systems and application (VLSI-TSA), pp 1–4
6. Evans RJ, Farrell PM, Felic G, Duong HT, Le HV, Li J, Li M, Moran W, Morelande M, Skafidas E (2013) Consumer radar: technology and limitations. In: Proceedings of 2013 international conference on radar, pp 21–26
7. Ayhan S, Scherr S, Bhutani A, Fischbach B, Pauli M, Zwick T (2016) Impact of frequency ramp nonlinearity, phase noise, and SNR on FMCW radar accuracy. IEEE Trans Microw Theory Tech 64(10):3290–3301
8. Barrick DE (1973) FM/CW Radar Signals and Digital Processing. U.S, Department of Commerce
9. Cherniak D, Samori C, Nonis R, Levantino S (2018) PLL-based wideband frequency modulator: two-point injection versus pre-emphasis technique. IEEE Trans Circuits Syst I: Regul Pap 65(3):914–924
10. Hussein AI, Saberi S, Paramesh J (2015) A 10 mW 60GHz 65nm CMOS DCO with 24 percent tuning range and 40 kHz frequency granularity. In: Proceedings of 2015 IEEE custom integrated circuits conference (CICC), pp 1–4

11. Wu W, Long JR, Staszewski RB, Pekarik JJ (2012) High-resolution 60-GHz DCOs with reconfigurable distributed metal capacitors in passive resonators. In: Proceedings of 2012 IEEE radio frequency integrated circuits symposium, pp 91–94
12. Wu W, Staszewski RB, Long JR (2014) A 56.4-to-63.4 GHz multi-rate all-digital fractional-N PLL for FMCW radar applications in 65 nm CMOS. IEEE J Solid-State Circuits vol 49(5), 1081–1096 (2014)
13. Marzin G, Levantino S, Samori C, Lacaita AL (2012) A 20 Mb/s phase modulator based on a 3.6 GHz digital PLL with -36 dB EVM at 5 mW power. IEEE J. Solid-State Circuits 47(12):2974–2988
14. Levantino S, Marzin G, Samori C (2018) An adaptive pre-distortion technique to mitigate the DTC nonlinearity in digital PLLs. IEEE J. Solid-State Circuits 49(8):1762–1772
15. Markulic N, Raczkowski K, Martens E, Filho PEP, Hershberg B, Wambacq P, Craninckx J (2016) A DTC-based subsampling PLL capable of self-calibrated fractional synthesis and two-point modula. IEEE J Solid-State Circuits 51(12):3078–3092
16. Yao C-W, Ni R, Lau C, Wu W, Godbole K, Zuo Y, Ko S, Kim N-S, Han S, Jo I, Lee J, Han J, Kwon D, Kim C, Kim S, Son SW, Cho TB (2017) A 14-nm 0.14-ps rms fractional-N digital PLL with a 0.2-ps resolution ADC-assisted coarse/fine-conversion chopping TDC and TDC nonlinearity calibration. IEEE J Solid-State Circuits 52(12):3446–3457
17. Cherniak D, Grimaldi L, Bertulessi L, Samori C, Nonis R, Levantino S (2018) A 23-GHz low-phase-noise digital bang-bang PLL for fast triangular and sawtooth chirp modulation. IEEE J Solid-State Circuits 53(12):3565–3575
18. Tasca D, Zanuso M, Marzin G, Levantino S, Samori C, Lacaita AL (2011) A 2.9-4.0-GHz fractional-N digital PLL with bang-bang phase detector and 560-fs rms integrated jitter at 4.5-mW power. IEEE J Solid-State Circuits 46(12):2745–2758
19. Iotti L, Mazzanti A, Svelto F (2017) Insights into phase-noise scaling in switch-coupled multi-CoreLCVCOs for E-band adaptive modulation links. IEEE J Solid-State Circuits 52(7):1703–1718
20. Tasca D, Zanuso M, Levantino S, Samori C (2010) An automatic retiming system for asynchronous fractional frequency dividers. In: Proceedings of 2010 6th conference on Ph.D. research in microelectronics and electronics (PRIME), pp 1–4

Chapter 4
Modeling and Simulation of Spiking Neural Networks with Resistive Switching Synapses

Valerio Milo

Abstract Artificial intelligence (AI) has recently reached excellent achievements in the implementation of human brain cognitive functions such as learning, recognition and inference by running intensively neural networks with deep learning on high-performance computing platforms. However, excessive computational time and power consumption required for achieving such performance make AI inefficient compared with human brain. To replicate the efficient operation of human brain in hardware, novel nanoscale memory devices such as resistive switching random access memory (RRAM) have attracted strong interest thanks to their ability to mimic biological learning in silico. In this chapter, design, modeling and simulation of RRAM-based electronic synapses capable of emulating biological learning rules are first presented. Then, the application of RRAM synapses in spiking neural networks to achieve neuromorphic tasks such as on-line learning of images and associative learning is addressed.

4.1 Introduction

In recent years, artificial intelligence (AI) has achieved outstanding performance in a wide range of machine learning tasks including recognition of faces [1] and speech [2] which now play a crucial role in many fields such as transportation and security. To obtain such achievements, AI has first exploited the availability of very large datasets for training deep neural networks (DNNs) in software according to deep learning [3]. Moreover, the maturity of high-performance computing hardware such as the graphics processing unit (GPU) [4] and the tensor processing unit (TPU) [5] has further contributed to accelerate DNN training, enabling to outperform human ability in certain tasks such as image classification [6] and playing board game of Go [7]. However, efficient implementation of AI tasks on modern digital computers based on von Neumann architecture and complementary metal-oxide-semiconductor

V. Milo (✉)
Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano and Italian Universities Nanoelectronics Team (IU.NET), Piazza Leonardo da Vinci 32, Milano 20133, Italy
e-mail: valerio.milo@polimi.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_4

(CMOS) technology has been recently challenged by fundamental issues such as the excessive power consumption and latency due to the looming end of Moore's law [8] and physical separation between memory and processing units [9].

To overcome the so-called von Neumann bottleneck of conventional hardware, novel non-von Neumann computing paradigms have been intensively explored with a view of bringing data processing closer to where data are stored [10]. In this wide range, neuromorphic computing has emerged as one of the most promising approaches since it aims at improving dramatically computation mainly in terms of energy efficiency taking inspiration from how the human brain processes information via biological neural networks [11].

In the last decade, strong research efforts in the field of neuromorphic engineering have led to build medium/large-scale analog/digital neuromorphic systems using CMOS technology [9, 12, 13]. However, the use of bulky CMOS circuits to closely reproduce synaptic dynamics has proved to be a major issue toward hardware integration of massive synaptic connectivity featuring human brain. This limitation has thus led to the exploration of novel memory device concepts, such as resistive switching random access memory (RRAM) and phase change memory (PCM), which display features suitable for synaptic application such as nanoscale size, fast switching behavior, low power operation, and tunable resistance by application of electrical pulses enabling to emulate synaptic plasticity at device level [14].

This chapter covers the application of hafnium-oxide (HfO_2) RRAM devices as plastic synapses in spiking neural networks (SNNs) to implement brain-inspired neuromorphic computing. After reviewing the main features of brain-inspired computation, physical mechanisms and operation principle of RRAM devices are described. Then, the scheme and operation of two hybrid CMOS/RRAM synapse circuits capable of implementing bio-realistic learning rules such as spike-timing dependent plasticity (STDP) and spike-rate dependent plasticity (SRDP) are presented. Finally, SNNs with resistive synapses are explored in simulation demonstrating their ability to achieve fundamental cognitive primitives such as on-line learning of visual patterns and associative memory.

4.2 Brain-Inspired Computing

Brain-inspired computing is considered a promising approach capable of tackling issues challenging today's digital processors thanks to its ability to replicate the massive parallelism and high energy efficiency of the human brain. To achieve such performance, human brain first relies on a high-density layered architecture consisting of large networks of biological processing units referred to as neurons where each neuron is connected with other neurons via 10^4 synapses on average [14]. In addition to the architecture, another feature playing a key role in brain computation is the spike-driven information processing. As illustrated in Fig. 4.1a, in biological neural networks the neurons interact with the next neuron by propagation of voltage spikes along the axon and their transmission through the synapses, namely the nanoscale gaps between the axon terminals and the dentrites. As a result of spike transmis-

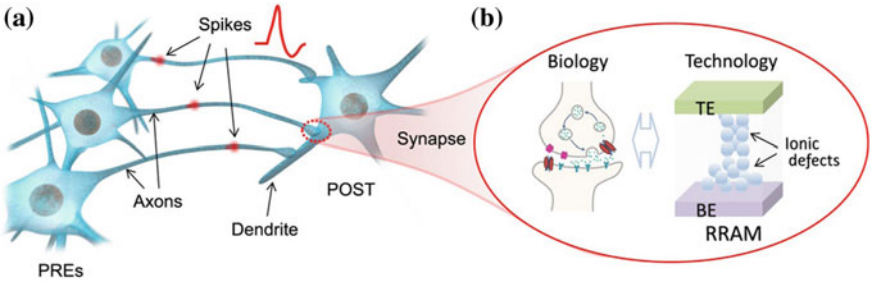


Fig. 4.1 **a** Illustration of some PREs connected to a single POST by synapses in a bio-realistic neural network. PREs emit spikes that are sent to the POST causing an increase of its membrane potential. **b** Sketch evidencing the strong analogy between a biological synapse where conductivity is tuned by voltage-induced ion migration and a RRAM device where conductivity is controlled by ionic defect migration resulting in formation/retraction of conductive paths. Reprinted from [18]

sion, calcium ions diffuse in the neuron activating the release of neurotransmitters within the synaptic gap where they diffuse eventually binding to sites of sodium ion channels of the post-synaptic neuron. This induces the opening of sodium ion channels and consequently the diffusion of sodium ions in the cell, which results in an increase of membrane potential leading to the generation of a spike by post-synaptic neuron as it crosses an internal threshold [14]. These biological processes at synaptic level thus suggest that information is processed by generation and transmission of spikes whose short duration of the order of 1 ms combined with typical neuron spiking rate of 10 Hz leads to a power consumption of only 20 W that is dramatically lower than power dissipated by modern computers [9, 14]. Therefore, brain computing relies on neurons integrating input signals sent by other neurons and firing spikes after reaching the threshold, and synapses changing their weight depending on spiking activity of pre-synaptic neuron (PRE) and post-synaptic neuron (POST). In particular, two biological rules like STDP [15] and SRDP [16] are considered two fundamental schemes controlling synaptic weight modulation, which is in turn believed to underlie learning ability in the brain.

To faithfully replicate the brain-inspired computing paradigm in hardware, in recent years neuromorphic community has intensively investigated novel material-based devices such as RRAM which, as shown in Fig. 4.1b, can mimic biological processes governing synaptic plasticity by exploiting its ability to change resistance via creation/rupture of conductive filaments in response to application of voltage pulses thanks to the resistive switching phenomenon [17, 18].

4.3 Resistive Switching in RRAM Devices

RRAM is a two-terminal nanoscale memory device displaying resistive switching, namely the ability to change the device resistance via the creation and disruption of filamentary conductive paths under the application of voltage pulses. The RRAM

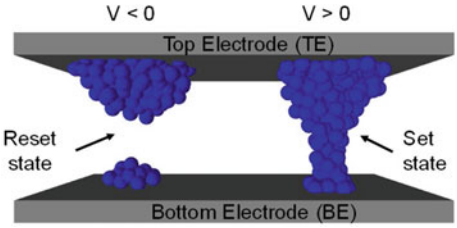


Fig. 4.2 Sketch of RRAM states, namely the reset state (left) and the set state (right). The application of a positive voltage to the TE causes the migration of defects toward the depleted gap, thus recreating the continuous filamentary path. The application of a negative voltage to the TE causes the retraction of defects back to the top reservoir, and the formation of a depleted gap

structure relies on a metal-insulator-metal (MIM) stack where a transition metal oxide layer, such as HfO_x, TiO_x, TaO_x and WO_x, is sandwiched between two metal electrodes referred to as top electrode (TE) and bottom electrode (BE), respectively [19, 20]. To initiate the filamentary path across the MIM stack, a soft breakdown process called forming is first operated, which locally creates a high concentration of defects, e.g., oxygen vacancies or metallic impurities, enhancing conduction. After forming, RRAM can exhibit bipolar resistive switching, where the application of a positive voltage can induce an increase of conductance, or set transition, whereas the application of a negative voltage can lead to reset transition, or a decrease of conductance [20].

Figure 4.2 schematically shows the 2 states of a RRAM device, namely the reset state (left), or high resistance state (HRS), where the filamentary path of defects is disconnected because of a previous reset transition. Under a positive voltage applied to the TE, defects from the top reservoir migrate toward the depleted gap, thus leading to the set state (right), or low resistance state (LRS), via a set transition. Applying a negative voltage to the TE causes the retraction of defects from the filament toward the top reservoir, and the formation of the depleted gap [20]. Note that the TE and BE generally differ by material and/or structure, with the TE being generally chemically active, e.g., Ti, Hf, or Ta, which enables the creation of an oxygen exchange layer with a relatively high concentration of oxygen vacancies [21]. On the other hand, the BE is chemically inert to prevent set transition when a positive voltage is applied to the BE [22].

4.4 Synapse Circuits with RRAM Devices

RRAM devices exhibit features as nanoscale size and low current operation making them attractive for realization of hardware synapses. The major challenge thus consists of implementing the synaptic plasticity schemes considered at the basis of biological learning such as STDP [15] and SRDP [16] at device level. To achieve synaptic

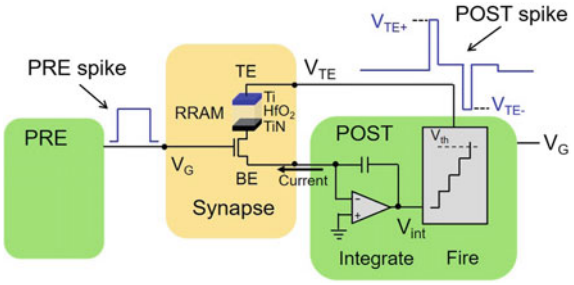


Fig. 4.3 Schematic of a 1T1R RRAM structure serving as synaptic connection between a PRE circuit and a POST circuit with I&F architecture. Application of a PRE spike at FET gate induces a synaptic current in 1T1R cell which is integrated by POST leading its membrane potential V_{int} to become more positive. As V_{int} hits the threshold V_{th}, a POST spike is backward sent at TE to update synaptic weight according to STDP rule. Adapted with permission from [24]. Copyright 2016 IEEE

plasticity in hardware, the combination of RRAM devices and field-effect transistors (FETs) serving as both cell selectors and current limiters has been widely used leading to the design of hybrid synaptic structures such as the one-transistor/one-resistor (1T1R) structure [23, 24] and the four-transistors/one-resistor (4T1R) structure [24, 25].

4.4.1 1T1R Synapse

Figure 4.3 shows a circuit schematic where a hybrid structure based on serial connection of a Ti/HfO_2/TiN RRAM cell and a FET, referred to as 1T1R structure, works as electronic synapse connecting a PRE to a POST with an integrate-and-fire (I&F) architecture. This building block was designed to achieve STDP rule which is considered one of the key mechanisms regulating learning in mammals. According to STDP rule, synaptic weight can change depending on the relative time delay Δt between spikes emitted by PRE and POST. If the PRE spike precedes the POST spike, Δt is positive resulting in an increase of synaptic weight or long-term potentiation (LTP) of the synapse. Otherwise, if PRE spike generation takes place slightly after the POST spike, Δt has negative value leading to a decrease of synaptic weight, or long-term depression (LTD) of the synapse [15]. The STDP implementation in 1T1R synapse was achieved as follows. When the PRE sends a 10-ms-long voltage pulse at FET gate, a current proportional to RRAM conductance flows across the synapse since its TE is biased by a continuous voltage with low amplitude used for communication phase. This current thus enters POST where it is integrated, causing an increase of POST membrane/internal potential V_{int}. As this integral signal crosses a threshold V_{th}, the POST sends both a forward spike toward next neuron layer and a suitably-designed spike including a 1-ms-long pulse with positive amplitude

followed, after 9 ms, by a 1-ms-long negative pulse, which is backward delivered at TE to activate a synaptic weight update according to STDP rule. Here, as the PRE spike precedes the POST spike ($0 < \Delta t < 10$ ms), PRE voltage overlaps only with the positive pulse of the POST spike, thus causing a set transition within RRAM device resulting in LTP of 1T1R synapse. Otherwise, if the PRE spike follows the POST spike (-10 ms $< \Delta t < 0$), overlap occurs between PRE spike and the negative pulse in the POST spike, thus activating a reset transition in RRAM device resulting in LTD of 1T1R synapse [23, 24]. Note that STDP implementation in 1T1R RRAM synapse was demonstrated in both simulation using a stochastic Monte-Carlo model of HfO_2 RRAM [23] and hardware as reported in [26].

4.4.2 4T1R Synapse

In addition to STDP implementation via 1T1R synapse, a novel hybrid CMOS/RRAM synapse was designed to replicate another fundamental biological learning rule called SRDP which states that high-frequency stimulation of PREs leads to synaptic LTP whereas a low-frequency stimulation of PREs leads to synaptic LTD [16]. As shown in Fig. 4.4a, PRE and POST are connected by a synapse circuit based on a 4T1R structure including a HfO_2 RRAM device and 4 FETs shared into M_1–M_2 and M_3–M_4 branches. PRE circuit includes both a signal channel transmitting a Poisson-distributed spike train with average frequency f_{PRE} at the M_1 gate and its copy delayed by a time Δt_D at the M_2 gate, and a noise channel driving M_3 by application of noise spikes at low frequency f_3, whereas POST circuit relies on an I&F stage with fire and random noise outputs which alternatively control RRAM TE and M_4 gate by a multiplexer. If $f_{PRE} > \Delta t_D{}^{-1}$, the high probability that M_1 and M_2 are simultaneously enabled by overlapping PRE spikes at gate terminals leads to the activation of synaptic currents within M_1/M_2 branch causing, after integration by

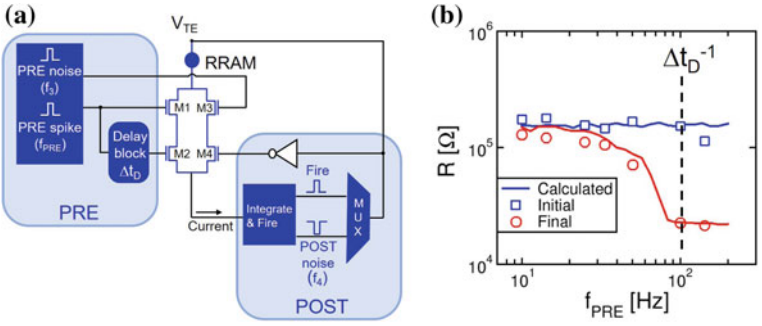


Fig. 4.4 **a** Schematic of 4T1R RRAM circuit operating as synaptic connection between a PRE circuit and a POST circuit with I&F architecture. **b** Measured and calculated resistance of 4T1R synapse for increasing f_{PRE} supporting high-frequency synaptic potentiation. Adapted from [25]

I&F stage, the generation of a fire pulse which is delivered at TE. As a result, overlap between spikes applied to M_1 gate, M_2 gate, and TE induces a set transition within RRAM device, namely LTP. Note that M_1/M_2 branch is the only active branch during LTP operation since fire pulse disables M_4 because of the inversion by NOT gate in the POST. Otherwise, if $f_{PRE} < \Delta t_D{}^{-1}$, no spike overlap between channels driving M_1 and M_2 takes place, thus making LTP branch disabled. To achieve LTD at low f_{PRE}, stochastic overlap among random noise voltage spikes provided at gate of M_3 and M_4 and a negative pulse applied at TE is exploited. In fact, these overlapping spikes cause a reset transition within RRAM device resulting in LTD at the synapse. Therefore, the operation principle of 4T1R synapse supports its ability to replicate SRDP rule implementing LTP at high f_{PRE} and a noise-induced stochastic LTD at low f_{PRE} [24, 25].

The ability to reproduce SRDP in 4T1R synapse was validated testing separately LTP and LTD by use of 2T1R integrated structures. Figure 4.4b shows measured and calculated RRAM resistance as a function of f_{PRE}, evidencing that a resistance change from initial HRS to LRS can be activated in a 2T1R structure serving as M_1/M_2 branch only when f_{PRE} is greater or equal than the reciprocal of $\Delta t_D = 10$ ms used in the experiment, namely $f_{PRE} \geq 100$ Hz. Similar to LTP, LTD was also experimentally demonstrated in [24, 25] evidencing that RRAM resistance initialized in LRS can reach HRS provided that PRE noise frequency f_3 is higher than POST noise frequency f_4 to achieve a sufficiently high overlap probability among random noise spikes controlling LTD branch.

4.5 Spiking Neural Networks with RRAM Synapses

4.5.1 Unsupervised Pattern Learning by SRDP

The fundamental ability of biological brain consists of learning by adaptation to environment with no supervision. This process, which is referred to as unsupervised learning, relies on schemes such as STDP and SRDP that adjust synaptic weights of neural networks according to timing or rate of spikes encoding information such as images and sounds. In particular, unsupervised learning of visual patterns has recently attracted increasing interest leading to many simulation/hardware demonstrations of SNNs with RRAM synapses capable of STDP [18, 23, 24, 26–30] and SRDP [25, 31].

Figure 4.5a illustrates a SNN inspired to perceptron network model developed by Rosenblatt in the late 1950s [32] consisting of 64 PREs fully connected to a single POST. This network was simulated using 4T1R RRAM structures as synapses to demonstrate unsupervised on-line learning of images with SRDP by conversion of image pattern (the 'X') and its surrounding background in high-frequency spiking activity of PREs and low-frequency spiking activity of PREs, respectively. To achieve this goal, after random initialization of synaptic weights between HRS and LRS

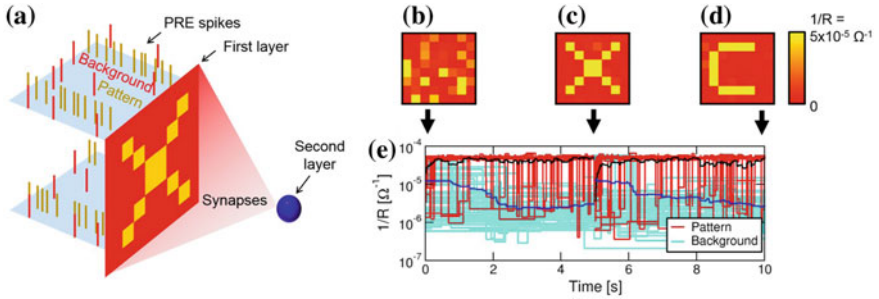


Fig. 4.5 **a** Sketch of a perceptron neural network where PREs are fully connected to a single POST. PREs receive spike sequences at high/low frequency upon pattern/background presentation which results in potentiation/depression of corresponding synaptic weights according to SRDP rule. Color plots of synaptic weights at **b** t = 0, **c** t = 5 s and **d** t = 10 s during a calculated on-line learning of two 8 × 8 sequential images. **e** Time evolution of calculated conductance supporting the ability of perceptron with 4T1R SRDP synapses to learn a new pattern after forgetting the previously stored pattern. Adapted from [25]

(Fig. 4.5b), an image including an 'X' pattern was submitted for 5 s to the input layer leading to high-frequency stimulation of PREs ($f_{PRE} = 100$ Hz) within the 'X' and low-frequency stimulation of PREs ($f_{PRE} = 5$ Hz) outside the 'X'. As described in Sect. 4.4.2, this resulted in a selective LTP of 'X' synapses and stochastic LTD of background synapses, thus leading synaptic weights to adapt to submitted image within 5 s (Fig. 4.5c). After t = 5 s, 'X' image was replaced by a 'C' image with no overlap with 'X' for other 5 s resulting in a high/low frequency stimulation of PREs within/outside 'C'. As shown by color map in Fig. 4.5d, external stimulation causes potentiation of 'C' synapses and depression of all the other synapses, which evidences network ability to learn a new pattern erasing the previously stored pattern [25]. This result is also supported by calculated time evolution of synaptic conductance during learning shown in Fig. 4.5e, which displays a fast conductance increase (LTP) of pattern synapses and a slower conductance decrease (LTD) of background synapses achieved using PRE and POST noise frequencies equal to $f_3 = 50$ Hz and $f_4 = 20$ Hz, respectively [25]. This simulation study corroborated the ability of perceptron SNNs to learn on-line sequential images thanks to 4T1R RRAM synapses capable of SRDP, and provided a solid basis for its experimental demonstration presented in [25].

4.5.2 Associative Memory with 1T1R RRAM Synapses

In addition to unsupervised image learning, another brain-inspired function receiving strong interest is the associative memory, namely the ability to retrieve past memories by their partial stimulation. To achieve this cognitive primitive, pioneering works by Hopfield [33] and Amit [34] focused on a type of neural network, called recurrent

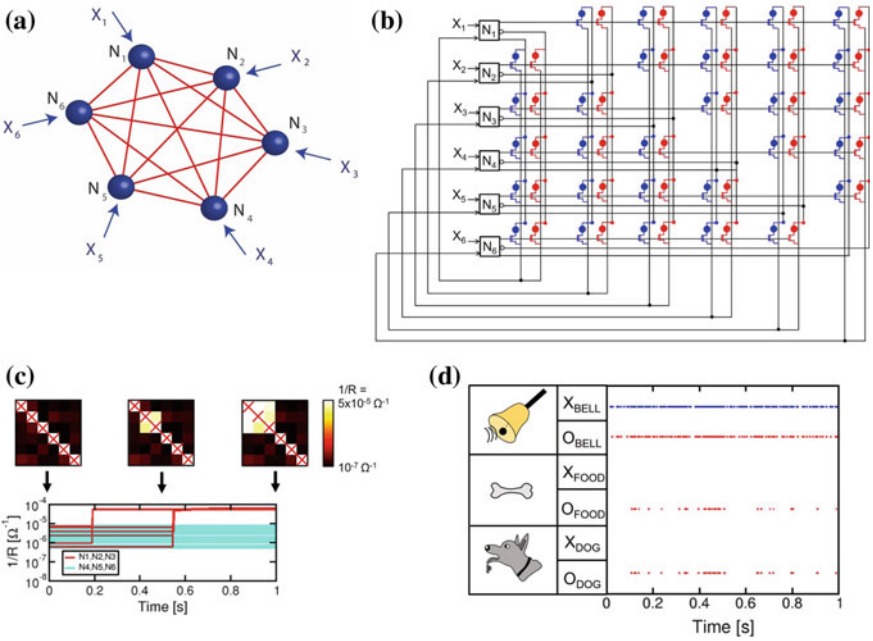


Fig. 4.6 **a** Sketch of a Hopfield recurrent neural network with 6 fully connected neurons receiving external stimulation X_i, i = 1:6. **b** Circuit schematic of a Hopfield network with 6 neurons fully connected via bidirectional excitatory (blue) and inhibitory (red) 1T1R RRAM synapses. **c** Calculated attractor learning process evidencing storage of the attractor state linking N_1, N_2 and N_3. **d** Illustration of associative memory implementation by recall process in an attractor network with RRAM synapses taking inspiration from Pavlov's dog experiments. Adapted with permission from [35]. Copyright 2017 IEEE

neural network or attractor neural network, where the neurons are fully connected with each other by bidirectional excitatory/inhibitory synapses.

Figure 4.6a shows a sketch of a Hopfield recurrent neural network with 6 neurons evidencing all-to-all synaptic connections, external stimulation X_i provided to each neuron, and the absence of self-feedback to prevent divergence of network dynamics. Based on this configuration, the circuit schematic of Hopfield neural network with bidirectional excitatory/inhibitory 1T1R RRAM synapses shown in Fig. 4.6b was designed and tested in simulation [35]. In this network, each neuron is implemented by an I&F block that provides signal to other neurons as a PRE but also receives signal by other neurons as a POST. Specifically, each I&F neuron has two current inputs, namely the external current spikes X_i and the sum of weighted currents activated by other neurons, and three outputs driving the gate of row 1T1R synapses, the TE of column excitatory synapses (blue 1T1R cells) and the TE of column inhibitory synapses (red 1T1R cells) by voltage pulses, respectively [35].

To demonstrate associative memory in this Hopfield network implementation, the network was first operated in learning mode stimulating a subset of neurons (N_1, N_2,

and N_3) with high-frequency Poisson distributed spike trains to store their attractor state. This process led each attractor neuron to fire three output spikes, namely (i) a positive voltage pulse at gate of all its row 1T1R synapses, (ii) a positive voltage pulse at TE of its column excitatory 1T1R synapses and (iii) a negative voltage pulse at TE of its column inhibitory 1T1R synapses, by causing a stochastic co-activation of attractor neurons at certain times. As this event occurs, voltage overlap at gate/TE of synapses shared by pairs of attractor neurons leads mutual excitatory 1T1R cells to undergo a set transition, thus LTP, whereas the corresponding inhibitory 1T1R cells undergo a reset transition, thus LTD. It means that learning phase in this Hopfield SNN with 1T1R synapses consists of the storage of attractor state associated with externally stimulated neurons via potentiation/depression of mutual excitatory/inhibitory synapses by a STDP-based learning scheme inspired to well-known Hebb's postulate stating that neurons that fire together, wire together [36].

Figure 4.6c shows a calculated learning process with N_1, N_2, and N_3 as attractor neurons evidencing the gradual storage of corresponding attractor state via potentiation of mutual excitatory synapses (top) due to the convergence of corresponding RRAM devices to high conductance values (bottom). After implementing attractor learning, the network was operated in another mode referred to as recall mode. During recall process, a high-frequency stimulation of a part of stored attractor state is applied, e.g., only one out of 3 attractor neurons, leading to activation of high currents across high conductance synapses shared with other attractor neurons which are transmitted at their inputs. As a result, attractor neurons with no external stimulation start spiking, thus retrieving the whole attractor state via a sustained spiking activity able to persist even after removing external stimulation [35]. Importantly, the ability of Hopfield networks to recall an attractor state by its incomplete stimulation was exploited to replicate in simulation a fundamental mammalian primitive referred to as associative memory taking inspiration from the Pavlov's dog experiments [37]. Indeed, as illustrated in Fig. 4.6d, after repeatedly stimulating a dog combining the ring of a bell with the presentation of food leading it to salivate, an external stimulation only with the ring of bell is able to reactivate the bell-food-salivation attractor in the dog [35]. Finally, Hopfield neural network with 1T1R synapses was also successfully used to explore in simulation pattern completion task, namely the reconstruction of an image stimulating its small features [38], which supports the strong potential of Hopfield neural networks with resistive synapses in computational tasks.

4.6 Conclusions

This chapter covers design, modeling and simulation of SNNs with CMOS/RRAM synapses capable of implementing brain-inspired neuromorphic computing. First, unsupervised learning at synaptic level has been addressed by development of 1T1R synapse and 4T1R synapse capable of replicating STDP and SRDP, respectively. Then, applications of these resistive synapses in SNNs have been investigated. A perceptron network with 4T1R synapses has been simulated demonstrating its ability to

achieve on-line learning of sequential images via SRDP-based adaptation of synaptic weights. In addition, attractor learning and recall processes have been achieved in a Hopfield recurrent neural network with excitatory/inhibitory 1T1R synapses by simulations supporting its ability to implement associative memory. These results support RRAM as promising technology for future development of large-scale neuromorphic systems capable of emulating unrivaled energy and computational efficiency of biological brain.

References

1. Taigman Y, Yang M, Ranzato M, Wolf L (2014) DeepFace: Closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1701–1708
2. Xiong W et al (2017) The Microsoft 2017 conversational speech recognition system. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 5934–5938
3. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
4. Coates A et al (2013) Deep learning with COTS HPC systems. In: Proceedings of the 30th International Conference on Machine Learning vol 28(3), pp 1337–1345
5. Jouppi NP et al (2017) In-datacenter performance analysis of a Tensor Processing Unit^{TM}. In: Proceedings of the 44th Annual International Symposium on Computer Architecture (ISCA), pp 1–12
6. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In: IEEE ICCV, pp 1026–1034
7. Silver D et al (2016) Mastering the game of Go with deep neural networks and tree search. Nature 529:484–489
8. Theis TN, Wong H-SP (2017) The end of Moore's law: a new beginning for information technology. Comput Sci Eng 19(2):41–50
9. Merolla PA et al (2014) A million spiking-neuron integrated circuit with a scalable communication network and interface. Science 345(6197):668–673
10. Wong H-SP, Salahuddin S (2015) Memory leads the way to better computing. Nat Nanotechnol 10(3):191–194
11. Mead C (1990) Neuromorphic electronic systems. Proc IEEE 78(10):1629–1636
12. Furber SB, Galluppi F, Temple S, Plana LA (2014) The SpiNNaker project. Proc IEEE 102(5):652–665
13. Qiao N et al (2015) A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses. Front Neurosci 9:141
14. Kuzum D, Yu S, Wong H-SP (2013) Synaptic electronics: materials, devices and applications. Nanotechnology 24:382001
15. Bi G-Q, Poo M-M (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and post synaptic cell type. J Neurosci 18(24):10464–10472
16. Bear MF (1996) A synaptic basis for memory storage in the cerebral cortex. Proc Natl Acad Sci USA 93(24):13453–13459
17. Yu S et al (2011) An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation. IEEE Trans Electron Devices 58(8):2729–2737
18. Wang W et al (2018) Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses. Sci Adv 4:eaat4752
19. Wong H-SP et al (2012) Metal-oxide RRAM. Proc IEEE 100(6):1951–1970

20. Ielmini D (2016) Resistive switching memories based on metal oxides: mechanisms, reliability and scaling. Semicond. Sci Technol 31(6):063002
21. Lee HY et al (2008) Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM. In: IEEE IEDM Tech Dig 297–300
22. Balatti S et al (2015) Voltage-controlled cycling endurance of HfO_x-based resistive-switching memory (RRAM). IEEE Trans Electron Devices 62(10):3365–3372
23. Ambrogio S et al (2016) Neuromorphic learning and recognition with one-transistor-one-resistor synapses and bistable metal oxide RRAM. IEEE Trans Electron Dev 63(4):1508–1515
24. Milo V et al (2016) Demonstration of hybrid CMOS/RRAM neural networks with spike time/rate-dependent plasticity. In: IEEE IEDM Tech Dig 440–443
25. Milo V et al (2018) A 4-transistors/one-resistor hybrid synapse based on resistive switching memory (RRAM) capable of spike-rate dependent plasticity (SRDP). IEEE Trans Very Large Scale Integration (VLSI) Syst 26(12):2806–2815
26. Pedretti G et al (2017) Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity. Sci Rep 7(1):5288
27. Yu S et al (2012) A neuromorphic visual system using RRAM synaptic devices with sub-pJ energy and tolerance to variability: experimental characterization and large-scale modeling. In: IEEE IEDM Tech Dig 239–242
28. Suri M et al (2012) CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: auditory (cochlea) and visual (retina) cognitive processing applications. In: IEEE IEDM Tech Dig 235–238
29. Serb A et al (2016) Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses. Nat Commun 7:12611
30. Prezioso M et al (2018) Spike-timing-dependent plasticity learning of coincidence detection with passively integrated memristive circuits. Nat Commun 9:5311
31. Ohno T et al (2011) Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. Nat Mater 10(8):591–595
32. Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65(6):386–408
33. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79:2554–2558
34. Amit DJ (1989) Modeling brain function: the world of attractor neural networks. Cambridge University Press, Cambridge
35. Milo V, Ielmini D, Chicca E (2017) Attractor networks and associative memories with STDP learning in RRAM synapses. In: IEEE IEDM Tech Dig 263–266
36. Hebb DO (1949) The organization of behavior: a neurophysiological theory. Wiley, New York
37. Pavlov IP (1927) Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. Oxford University Press, London
38. Milo V, Chicca E, Ielmini D (2018) Brain-inspired recurrent neural network with plastic RRAM synapses. In: IEEE International Symposium on Circuits and Systems (ISCAS), pp 1–5

Part III
Computer Science and Engineering

Chapter 5
Learning and Adaptation to Detect Changes and Anomalies in High-Dimensional Data

Diego Carrera

Abstract The problem of monitoring a datastream and detecting whether the data generating process changes from normal to novel and possibly anomalous conditions has relevant applications in many real scenarios, such as health monitoring and quality inspection of industrial processes. A general approach often adopted in the literature is to learn a model to describe normal data and detect as anomalous those data that do not conform to the learned model. However, several challenges have to be addressed to make this approach effective in real world scenarios, where acquired data are often characterized by high dimension and feature complex structures (such as signals and images). We address this problem from two perspectives corresponding to different modeling assumptions on the data-generating process. At first, we model data as realization of random vectors, as it is customary in the statistical literature. In this settings we focus on the change detection problem, where the goal is to detect whether the datastream permanently departs from normal conditions. We theoretically prove the intrinsic difficulty of this problem when the data dimension increases and propose a novel non-parametric and multivariate change-detection algorithm. In the second part, we focus on data having complex structure and we adopt dictionaries yielding sparse representations to model normal data. We propose novel algorithms to detect anomalies in such datastreams and to adapt the learned model when the process generating normal data changes.

5.1 Introduction

The general problem we address here is the monitoring of datastreams to detect in the data-generating process. This problem has to be faced in several applications, since the change could indicate an issue that has to be promptly alarmed and solved. In particular, we consider two meaningful examples. Figure 5.1a shows a Scanning Electron Microscope (SEM) image acquired by an inspection system that monitors

D. Carrera (✉)
Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, Italy
e-mail: diego.carrera@polimi.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_5

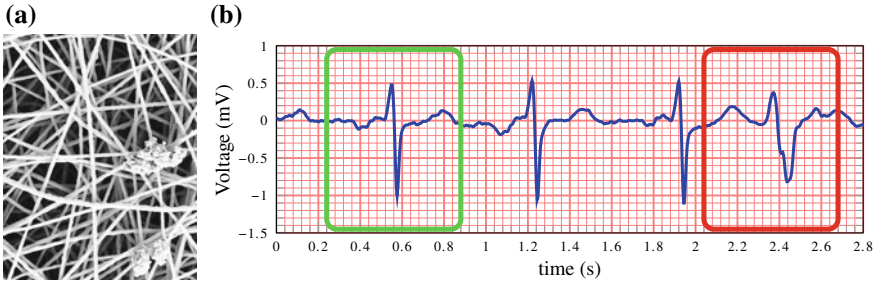


Fig. 5.1 **a** A detail of a SEM image acquired by the considered inspection system depicting nanofibrous material. The two small *beads*, i.e., fiber clots, on the right reduce the quality of the produced material. **b** Example of few seconds of ECG signal containing 4 normal heartbeats (one of which highlighted in green) and an anomalous one (highlighted in red) (Color figure online)

the quality of nanofibrous materials produced by a prototype machine. In normal conditions, the produced material is composed of tiny filaments whose diameter ranges up to 100 nanometers. However, several issues might affect the production process and introduce small defects among the fibers, such as the clots in Fig. 5.1a. These defects have to be promptly detected to improve the overall production quality.

The second scenario we consider is the online and long-term monitoring of ECG signals using wearable devices. This is a very relevant problem as it would ease the transitioning from hospital to home/mobile health monitoring. In this case the data we analyze are the heartbeats. As shown in Fig. 5.1b, normal heartbeats feature a specific morphology, while the shape of anomalous heartbeats, that might be due to potentially dangerous arrhythmias, is characterized by a large variability. Since the morphology of normal heartbeats depends on the user and the position of the device [16], the anomaly-detection algorithm has to be configured every time the user places the device.

Monitoring this kind of datastream raises three main challenges: at first data are characterized by complex structure and high dimension and there is no analytical model able to describe them. Therefore, it is necessary to learn models directly from data. However, only normal data can be used during learning, since acquiring anomalous data can be difficult if not impossible (e.g., in case of ECG monitoring acquiring arrhythmias might be dangerous for the user). Secondly, we have to careful design indicators and rules to assess whether incoming data fit or not the learned model. Finally, we have to face the domain adaptation problem, since normal condition might changes during time and the learned model might not be able to describe incoming normal data, thus it has to be adapted accordingly. For example, in ECG monitoring the model is learned over a training set of normal heartbeats acquired at low heart rate, but the morphology of normal heartbeats changes when the heart rate increases, see Fig. 5.2b.

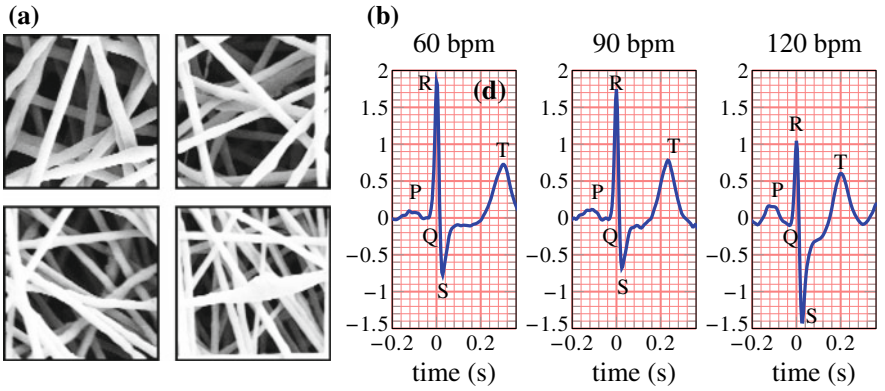


Fig. 5.2 **a** Details of SEM images acquired at different scales. The content of these images is different, although they are perceptually similar. **b** Examples of heartbeats acquired at different heart rate. We report the name of the waveforms of the ECG [13]

We investigate these challenges following two directions, and in particular, we adopt two different modeling assumptions on data-generating process. At first, we assume that data can be described by a smooth probability density function, as customary in the statistics literature. In these settings we focus on the change-detection problem, namely the problem of detecting permanent changes in the monitored datastreams. We investigate the intrinsic difficulty of this problem, in particular when the data dimension increases. Then, we propose QuantTree, a novel change detection algorithms based on histograms that enables the non-parametric monitoring of multivariate datastreams. We theoretically prove that any statistic compute over histograms defined by QuantTree does not depend on the data-generating process.

In the second part we focus on data having a complex structure, and address the anomaly detection problem. We propose a very general anomaly-detection algorithm based on a dictionary yielding sparse representation learned from normal data. As such, it is not able to provide sparse representation to anomalous data, and we exploit this property to design low dimensional indicators to assess whether new data conform or not to the learned dictionary. Moreover, we propose two domain adaptation algorithms to make our anomaly detector effective in the considered application scenarios.

The chapter is structure as follows: Sect. 5.2 presents the most relevant related literature and Sect. 5.3 formally state the problem we address. Section 5.4 focuses on the contribution of the first part of the thesis, where we model data as random vectors, while Sect. 5.5 is dedicated to the second part of the thesis, where we consider data having complex structures. Finally, Sect. 5.6 presents the conclusions and the future works.

5.2 Related Works

The first algorithms [22, 24] addressing the change-detection problem were proposed in the statistical process control literature [15] and consider only univariate datastreams. These algorithms are well studied and several properties have been proved due to their simplicity. Their main drawback is that they require the knowledge of the data generating distributions. Non-parametric methods, namely those that can operate when this distribution is unknown, typically employ statistics that are based on natural order of the real numbers, such as Kolmogorov–Smirnov [25] and Mann-Whitney [14]. Extending these methods to operate on multivariate datastreams is far from being trivial, since no natural order is well defined on $\mathbb{R}^d$. The general approach to monitor multivariate datastreams in is to learn a model that approximates ϕ_0 and monitor a univariate statistic based on the learned model. One of the most popular non-parametric approximation of ϕ_0 is given by Kernel Density Estimation [17], that however becomes intractable when the data dimension increases.

All these methods assume that data can be described by a smooth probability density function (pdf). However, complex data such as signal and images live close to a low-dimensional manifold embedded in a higher dimensional space [3], and do not admit a smooth pdf. In these cases, it is necessary to learn meaningful representations to data to perform any task, from classification to anomaly detection. Here, we consider dictionary yielding sparse representations that have been originally proposed to address image processing problems such as denoising [1], but they were also employed in supervised tasks [19], in particular classification [20, 23]. The adaptation of dictionaries yielding sparse representations to different domain were investigated in particular to address the image classification problem [21, 27]. In this scenario training images are acquired under different conditions than the test ones, e.g. different lightning and view angles, and therefore live in a different domain.

5.3 Problem Formulation

Let as consider a datastream $\{\mathbf{s}_t\}_{t=1,\dots}$, where $\mathbf{s}_t \in \mathbb{R}^d$ is drawn from a process $\mathcal{P}_N$ in normal condition. We are interest in detect whether $\mathbf{s}_t \sim \mathcal{P}_A$, i.e., $\mathbf{s}_t$ is drawn from an alternative process $\mathcal{P}_A$ representing the anomalous conditions. Both $\mathcal{P}_N$ and $\mathcal{P}_A$ are unknown, but we assume that a training set of normal data is available to approximate $\mathcal{P}_N$. In what follows we describe in details the specific problems we consider in the thesis.

Change Detection. At first, we model $\mathbf{s}_t$ as a realization of a continuous random vector, namely we assume that $\mathcal{P}_N$ and $\mathcal{P}_A$ admit smooth probability density functions ϕ_0 and ϕ_1, respectively. This assumption is not too strict, as it usually met after a feature extraction process.

We address the problem of detecting abrupt change in the data-generating process. More precisely, our goal is to detect if there is a change point $\tau \in \mathbb{N}$ in the

datastreams such that $\mathbf{s}_t \sim \phi_0$ for $t \leq \tau$ and $\mathbf{s}_t \sim \phi_1$ for $t > \tau$. For simplicity, we analyze the datastream in batches $W = \{\mathbf{s}_1, \ldots, \mathbf{s}_\nu\}$ of ν samples and detect changes by performing the following hypothesis test:

$$H_0 : W \sim \phi_0, \qquad H_1 : W \nsim \phi_0 \tag{5.1}$$

The null hypothesis is rejected whether $\mathcal{T}(W) > \gamma$, where $\mathcal{T}$ is a statistic typically defined upon the model that approximate the density ϕ_0, and γ is defined to guarantee a desired probability of false positive rate α, i.e. $P_{\phi_0}(T(W) > \gamma) \leq \alpha$.

Anomaly Detection. The anomaly-detection problem is strictly related to change detection. The main difference is that in anomaly detection we analyze each $\mathbf{s}_t$ independently to whether is draw from $\mathcal{P}_N$ or $\mathcal{P}_A$, without taking into account the temporal correlation (for this reason we will omit the subscript $_t$). In this settings we consider data having complex structure, such as heartbeats or patches, i.e., small region extracted from an image. We adopt dictionaries yielding sparse representation to describe $\mathbf{s} \sim \mathcal{P}_N$, namely we assume that $\mathbf{s} \approx D\mathbf{x}$, where $D \in \mathbb{R}^{d \times n}$ is a matrix called *dictionary*, that has to be learned from normal data and the coefficient vector $\mathbf{x} \in \mathbb{R}^n$ is *sparse*, namely it has few of nonzero components. To detect anomalies, we have to define a decision rule, i.e., a function $\mathcal{T} : \mathbb{R}^d \to \mathbb{R}$ and a threshold γ such that

$$\mathbf{s} \text{ is anomalous} \iff \mathcal{T}(\mathbf{s}) > \gamma, \tag{5.2}$$

where $\mathcal{T}(\mathbf{s})$ is defined using the sparse representation $\mathbf{x}$ of $\mathbf{s}$.

Domain Adaptation. The process generating normal data $\mathcal{P}_N$ may change over time. Therefore, a dictionary D learned on training data (i.e., in the source domain), might not be able to describe normal data during test (i.e., in the target domain). To avoid degradation in the anomaly-detection performance, D has to be adapted has soon as $\mathcal{P}_N$ changes. For example, in case of ECG monitoring the morphology of normal heartbeats changes when the heart rate increases, while in case of SEM images, the magnification level of the microscope may change, and this modify the qualitative content of the patches, as shown in Fig. 5.2.

5.4 Data as Random Vectors

In this section we consider data modeled as random vectors. At first we investigate the detectability loss phenomenon, showing that the change-detection performance are heavily affected by the data dimension. Then we propose a novel change-detection algorithm that employs histograms to describe normal data.

5.4.1 Detectability Loss

As described in the Sect. 5.3, we assume that both $\mathcal{P}_N$ and $\mathcal{P}_A$ admit a smooth pdf $\phi_0, \phi_1 \colon \mathbb{R}^d \to \mathbb{R}$. For simplicity, we also assume that ϕ_1 that can be expressed as $\phi_1(\mathbf{s}) = \phi_0(Q\mathbf{s} + \mathbf{v})$, where $\mathbf{v} \in \mathbb{R}^d$ and $Q \in \mathbb{R}^{d \times d}$ is an orthogonal matrix. This is quite a general model as it includes changes in the mean as well as in the correlations of components of $\mathbf{s}_t$. To detect changes, we consider the popular approach that monitor the loglikelihood w.r.t. the distribution generating normal data ϕ_0:

$$\mathcal{L}(\mathbf{s}_t) = \log(\phi_0(\mathbf{s}_t)). \tag{5.3}$$

In practice, we reduce the multivariate datastream $\{\mathbf{s}_t\}$ to a univariate one $\{\mathcal{L}(\mathbf{s}_t)\}$. Since ϕ_0 is unknown, we should preliminary estimate $\widehat{\phi}_0$ from data, and use it in (5.3) in place of ϕ_0. However, in what follows we will consider ϕ_0 since it make easier to investigate how the data dimension affect the change-detection performance.

We now introduce two measures that we will use in our analysis. The *change magnitude* assesses how much ϕ_1 differs from ϕ_0 and is defined as $\mathrm{sKL}(\phi_0, \phi_1)$, namely the symmetric Kullback–Leibler divergence between ϕ_0 and ϕ_1 [12]. In practice, large values of $\mathrm{sKL}(\phi_0, \phi_1)$ makes the change very apparent, as proved in the Stein's Lemma [12]. The *change detectability* assesses how the change is perceivable by monitoring the datastream $\{\mathcal{L}(\mathbf{s}_t)\}$ and is defined as the signal-to-noise ratio of the change $\phi_0 \to \phi_1$:

$$\mathrm{SNR}(\phi_0, \phi_1) := \frac{\left(\underset{\mathbf{s}\sim\phi_0}{E}[\mathcal{L}(\mathbf{s})] - \underset{\mathbf{s}\sim\phi_1}{E}[\mathcal{L}(\mathbf{s})] \right)^2}{\underset{\mathbf{s}\sim\phi_0}{\mathrm{var}}[\mathcal{L}(\mathbf{s})] + \underset{\mathbf{s}\sim\phi_0}{\mathrm{var}}[\mathcal{L}(\mathbf{s})]}, \tag{5.4}$$

where E and var denote the expected value and the variance, respectively.

The following theorem proves *detectability loss* on Gaussian datastreams (the proof is reported in [2]).

Theorem 1 *Let $\phi_0 = \mathcal{N}(\mu_0, \Sigma_0)$ be a d-dimensional Gaussian pdf and $\phi_1 = \phi_0(Q\mathbf{s} + \mathbf{v})$, where $Q \in \mathbb{R}^{d \times d}$ is orthogonal and $\mathbf{v} \in \mathbb{R}^d$. Then, it holds*

$$SNR(\phi_0, \phi_1) \leq \frac{C}{d} \tag{5.5}$$

where the constant C depends only on $sKL(\phi_0, \phi_1)$.

The main consequences of Theorem 1 is that the change detectability decreases when the data dimension increases, as long as the change magnitude is kept fixed. Remarkably, this results is independent on how the changes affected the datastream (i.e., it is independent on Q and $\mathbf{v}$), but only on the change magnitude. Moreover, it does not depend on estimation error, since in (5.4) we have considered the true and unknown distribution ϕ_0. However, the detectability loss becomes more severe when the loglikelihood is computed w.r.t. the estimated $\widehat{\phi}_0$, as we showed in [2].

Finally, we remark that the detectability loss holds also for more general distributions ϕ_0, such as Gaussian Mixture, and on real data. In that case the problem cannot treated analytically, but we empirically show similar results using Controlling Change Magnitude (CCM) [6], a framework to inject changes of a given magnitude in real world datastreams.

5.4.2 *QuantTree*

In this section we present QuantTree, a novel change detection algorithm that computes a histogram to approximate ϕ_0. A histogram h is defined as $h = \{(B_k, \widehat{\pi}_k)\}_{k=1,\dots,K}$, where the bins $\{B_k\}$ identify a partition of the data space, while the $\{\widehat{\pi}_k\}$ are the probabilities associated to the bins. In practice, we estimate h to approximate ϕ_0 and π_k is an estimate of the probability of $\mathbf{s} \sim \phi_0$ to fall inside B_k. As described in Sect. 5.3, we monitor the datastream in a batch-wise manner and we consider statistics $\mathcal{T}_h(W)$ defined over the histogram h, namely statistics that depend only on the number y_k of samples of W that fall in the bin B_k, for $k = 1, \dots, K$. Examples of such statistics are the Total Variation distance and the Pearson's statistic.

The proposed QuantTree algorithm takes as input target probability values $\{\pi_k\}$ and generates a partitioning $\{B_k\}$ such that the corresponding probability $\{\widehat{\pi}_k\}$ are close to $\{\pi_k\}$. QuantTree is an iterative splitting scheme that generate a new bin at each iteration k. The bin is defined by splitting along a component chosen at random among the d available. The splitting point is selected to contain exactly round$(\pi_k N)$ samples of the training set, where S is the number of training samples. An example of partitioning computed by QuantTree is show in Fig. 5.3.

The main property of QuantTree is that its peculiar splitting scheme makes the distribution of any statistics $\mathcal{T}_h$ independent on the data-generating distribution ϕ_0. This properly is formally stated in the following theorem, that we proved in [4].

Theorem 2 *Let $\mathcal{T}_h(\cdot)$ be defined over an histogram h computed by QuantTree. When $W \sim \phi_0$, the distribution of $\mathcal{T}_h(W)$ depends only on v, N and $\{\pi_k\}_k$.*

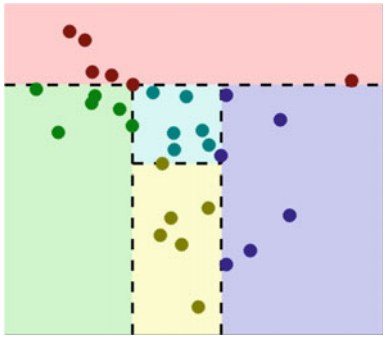


Fig. 5.3 A partitioning with $K = 5$ bins computed by QuantTree over a training set of $N = 30$ samples. We set $\pi_k = 1/K$ to yield a uniform density histograms, thus all the bins contain 6 samples each

In practice, the distribution of any statistic $\mathcal{T}_h$ depends only on the cardinalities of the training set and the window W and on the target probabilities $\{\pi_k\}$. The main consequence of Theorem 2 is that the threshold γ that guarantees a given false positive rate α does not depend on ϕ_0. Therefore, we can precompute γ through by estimating the distribution of $\mathcal{T}_h$ over synthetically generated samples though Montecarlo simulations. To the best of our knowledge, QuantTree is one of the first algorithms that performs non-parametric monitoring of multivariate datastreams.

We compare the histograms computed by QuantTree with other partitioning in the literature in [4] through experiments on Gaussian datastreams and real world datasets. Histograms computed by QuantTree yield a larger power and are the only ones that allows to properly control the false positive rate. We remark that also QuantTree suffers of the detectability loss, confirming the generality of our results on detectability loss.

5.5 Data Featuring Complex Structures

In this section we employ dictionaries yielding sparse representations to model data having complex structures. At first we present our general anomaly detection algorithm, then we introduce our two domain-adaptation solutions, specifically designed for the application scenarios described in Sect. 5.1.

5.5.1 Sparsity-Based Anomaly Detection

Our modeling assumption is that normal data $\mathbf{s} \sim \mathcal{P}_N$ can be well approximated as $\mathbf{s} \approx D\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$ is sparse, namely it has only few nonzero components and the dictionary $D \in \mathbb{R}^{d \times n}$ approximates the process $\mathcal{P}_N$. The sparse representation $\mathbf{x}$ is computed by solving the sparse coding problem:

$$\mathbf{x} = \underset{\widetilde{\mathbf{x}} \in \mathbb{R}^n}{\arg\min} \frac{1}{2} \|\mathbf{s} - D\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \tag{5.6}$$

where the ℓ^1 norm $\|\mathbf{x}\|_1$ is used to enforce sparsity in $\mathbf{x}$, and the parameter $\lambda \in \mathbb{R}$ controls the tradeoff between the ℓ^1 norm and the reconstruction error $\|\mathbf{s} - D\mathbf{x}\|_2^2$. The dictionary D is typically unknown, and we have to learn it from a training set of normal data by solving the dictionary learning problem:

$$D, X = \underset{\widetilde{D} \in \mathbb{R}^{d \times n}, \widetilde{X} \in \mathbb{R}^{n \times m,}}{\arg\min} \frac{1}{2} \|S_0 - \widetilde{D}\widetilde{X}\|_2^2 + \lambda \|X\|_1, \tag{5.7}$$

where $S_0 \in \mathbb{R}^{d \times m}$ is the training set that collects normal data column-wise.

To assess whether $\mathbf{s}$ is generated or not from $\mathcal{P}_N$, we define a bivariate indicator vector $\mathbf{f} \in \mathbb{R}^2$ collecting the reconstruction error and the sparsity of the representation:

$$\mathbf{f}(\mathbf{s}) = \begin{bmatrix} \|\mathbf{s} - D\mathbf{x}\|_2 \\ \|\mathbf{x}\|_1 \end{bmatrix}. \tag{5.8}$$

In fact, we expect that when $\mathbf{s}$ is anomalous it deviates from normal data either in the sparsity of the representation or in the reconstruction error. This means that $\mathbf{f}(\mathbf{s})$ would be an outlier w.r.t. the distribution ϕ_0 of $\mathbf{f}$ computed over normal data. Therefore, we detect anomalies as in (5.2) by setting $\mathcal{T} = -\log(\phi_0)$, where ϕ_0 is estimate from a training set of normal data (different from the set S_0 used in dictionary learning) using Kernel Density Estimation [5].

We evaluate our anomaly-detection algorithm on a dataset containing 45 SEM images. In this case $\mathbf{s} \in \mathbb{R}^{15\times 15}$ is a small squared patch extracted from the image. We analyze each patch independently to determine if it is normal or anomalous. Since each pixel of the image is contained in more than one patch, we obtain several decisions of each pixel. To detect defects at pixel level, we aggregate all the decisions through majority voting. An example of the obtained detections is shown in Fig. 5.4: our algorithm is able to localize all the defects by keeping the false positive rate small. More details on our solution and experiments are reported in [8].

In case of ECG monitoring, the data to be analyzed are the heartbeats, that are extracted from the ECG signal using traditional algorithms. Since our goal is to perform ECG monitoring directly on the wearable device, that has limited computational capabilities, we adopt a different sparse coding procedure, that is based on the ℓ^0 "norm" and it is performed by means of greedy algorithms [11]. In particular, we proposed a novel variant of the OMP algorithm [9], that is specifically designed for dictionary $D \in \mathbb{R}^{d\times n}$ where $n < d$, that is settings we adopt in ECG monitoring. Dictionary learning, that is required every time the device is positioned as the shape of normal heartbeats depends both of the users and device position, is performed on a host device, since the computational resources of wearable devices are not sufficient.

5.5.2 *Multiscale Anomaly-Detection*

In the quality inspection through SEM images, we have to face the domain adaptation problem since the magnification level of the microscope may change during monitoring. Therefore, we improve our anomaly-detection algorithm to make it scale-invariant. The three key ingredients we use are: *(i)* a multiscale dictionary that is able to describe patch extracted from image at different resolution, *(ii)* a multiscale sparse representation that captures the structure in the patch at different scales and *(iii)* a triviate indicator vector, that is more powerful that the one in (5.8).

We build the multiscale dictionary as $D = [D_1|\ldots|D_L]$. Each subdictionary D_j is learned by solving problem (5.7) over patched extracted from synthetically rescaled

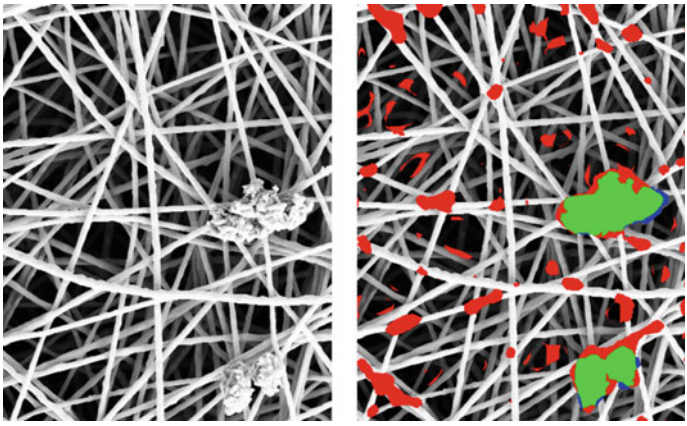


Fig. 5.4 Detail of defect detection at pixel level. Green, blue and red pixel identify the true positive, false negative and false positive, respectively. The threshold γ in (5.2) was set to yield a false positive rate equal to 0.05 (Color figure online)

version of training images. Therefore, we can learn a multiscale dictionary even if the training images are acquired at a fixed scale. To compute the multiscale representation $\mathbf{x}$ we solve the following sparse coding problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{s} - D\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 + \xi \sum_{j=1}^{L} \|\mathbf{x}_j\|_2, \tag{5.9}$$

where each $\mathbf{x}_j$ refers to the subdictionary D_i and $\mathbf{x}$ collects all the $\mathbf{x}_j$. The group sparsity term $\sum_{j=1}^{L} \|\mathbf{x}_j\|_2$ ensures that each patch is reconstructed using coefficients of $\mathbf{x}$ that refers only to few scales. Finally, we define a trivariate indicator vector $\mathbf{f}$ that includes the reconstruction error and the sparsity as in (5.8) and also the group sparsity term $\sum_j \|\mathbf{x}_j\|_2$. We detect anomalies using the same approach described in Sect. 5.5.1. Our experiments [7] show that employing a multiscale approach in all phases of the algorithm (dictionary learning, sparse coding and anomaly indicators) achieves good anomaly detection performance even when training and test images are acquired at different scales.

5.5.3 Dictionary Adaptation for ECG Monitoring

To make our anomaly-detection algorithm effective in long-term ECG monitoring, we have to adapt the user-specific dictionary. In fact, to keep the training procedure safe for the user, this dictionary is learned from heartbeats acquired in resting conditions, and it would be not able to describe normal heartbeats acquired during daily activities,

when the heart rate increases and normal heartbeats get transformed, see Fig. 5.2b. These transformations are similar for every user: the T and P waves approach the QRS complex and the support of the heartbeats narrows down. Therefore, we adopt user-independent transformations $F_{r,r_0} \mathbb{R}^{d_r \times d_{r_0}}$ to adapt the user-specific dictionary $D_{u,r_0} \in \mathbb{R}^{d_{r_0} \times n}$:

$$D_{u,r} = F_{r,r_0} D_{u,r_0}, \tag{5.10}$$

where r_0 and r denotes the heart rates in resting condition and during daily activities, respectively, and u indicates the user. The transformation F_{r,r_0} depends only on the heart rates and is learned from collections $S_{u,r}$ of training sets of normal heartbeats of several users acquired at different heart rates extracted from publicly available datasets of ECG signals. The learning has to performed only ones by solving the following optimization problem:

$$F_{r,r_0} = \underset{F,\{X_u\}_u}{\arg\min} \frac{1}{2} \sum_{u=1}^{L} \|S_{u,r} - F D_{u,r_0} X_u\|_2^2 + \mu \sum_{u=1}^{L} \|X_u\|_1 + \frac{\lambda}{2} \|W \odot F\|_2^2 + \xi \|W \odot F\|_1,$$

where the first term ensures that the transformed dictionaries $F D_{u,r_0}$ provide good approximation to the heartbeats of the user u. Moreover, we adopt three regularization terms: the first one is based on ℓ^1 norm and enforces sparsity in the representations of heartbeats at heart rate r for each user u. The other two terms represent a weighted elastic net penalization over F to improves the stability of the optimization problem, and the weighting matrix W introduces regularities in F_{r,r_0}.

The learned F_{r,r_0} (for several values of r and r_0) are then hard-coded in the wearable device to perform online monitoring [18]. We evaluate our solution on ECG signals acquired using the Bio2Bit Dongle [18]. Our experiments [10] show that the reconstruction error is kept small when the heart rate increases, implying that our transformations effectively adapt the dictionaries to operate at higher heart rate. Moreover, our solution is able to correctly identify anomalous heartbeats even at very large heart rate, when the morphology of normal heartbeats undergoes severe transformations.

5.6 Conclusions

We have addressed the general problem of monitoring a datastream to detect whether the data generating process departs from normal conditions. Several challenges have to be addressed in practical applications where data are high dimensional and feature complex structures. We address these challenges from two different perspectives by making different modeling assumption on the data generating process.

At first we assume that data can be described by a smooth probability density function and address the change detection problem. We prove the detectability loss phenomenon, that relates the change-detection performance and the data dimension, and we propose QuantTree, that is one of the first algorithms that enables non-

parametric monitoring of multivariate datastream. In the second part, we employ dictionary yielding sparse representation to model data having featuring complex structures and propose a novel anomaly-detection algorithm. To make this algorithm effective in practical applications, we propose two domain-adaptation solutions that turn to be very effective in long-term ECG monitoring and in a quality inspection of an industrial process.

Future works include the extension of QuantTree to design a truly multivariate change-detection algorithm, and in particular to control the average run length instead of the false positive rate. Another relevant directions is the design of models that provide good representation for detection tasks. In fact, our dictionaries are learned to provide good reconstruction to the data, and not to perform anomaly detection. Very recent works such as [26] show very promising results using deep learning, but a general methodology is still not available.

References

1. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
2. Alippi C, Boracchi G, Carrera D, Roveri M (2016) Change detection in multivariate datastreams: likelihood and detectability loss. In: Proceedings of the international joint conference on artificial intelligence (IJCAI), vol 2, pp 1368–1374
3. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828
4. Boracchi G, Carrera D, Cervellera C, Maccio D (2018) Quanttree: histograms for change detection in multivariate data streams. In: Proceedings of the international conference on machine learning (ICML), pp 638–647
5. Botev ZI, Grotowski JF, Kroese DP et al (2010) Kernel density estimation via diffusion. Ann Stat 38(5):2916–2957
6. Carrera D, Boracchi G (2018) Generating high-dimensional datastreams for change detection. Big Data Res 11:11–21
7. Carrera D, Boracchi G, Foi A, Wohlberg B (2016) Scale-invariant anomaly detection with multiscale group-sparse models. In: Proceedings of the IEEE international conference on image processing (ICIP), pp 3892–3896
8. Carrera D, Manganini F, Boracchi G, Lanzarone E (2017) Defect detection in sem images of nanofibrous materials. IEEE Trans Ind Inform 13(2):551–561
9. Carrera D, Rossi B, Fragneto P, Boracchi G (2017) Domain adaptation for online ecg monitoring. In: Proceedings of the IEEE international conference on data mining (ICDM), pp 775–780
10. Carrera D, Rossi B, Fragneto P, Boracchi G (2019) Online anomaly detection for long-term ecg monitoring using wearable devices. Pattern Recognit 88:482–492
11. Carrera D, Rossi B, Zambon D, Fragneto P, Boracchi G (2016) Ecg monitoring in wearable devices by sparse models. In: Proceedings of the European conference on machine learning and knowledge discovery in databases (ECML-PKDD), pp 145–160
12. Cover TM, Thomas JA (2012) Elements of information theory. Wiley, Hoboken
13. Felker GM, Mann DL (2014) Heart failure: a companion to Braunwald's heart disease. Elsevier Health Sciences
14. Hawkins DM, Deng Q (2010) A nonparametric change-point control chart. J Qual Technol 42(2):165–173
15. Hawkins DM, Qiu P, Chang WK (2003) The changepoint model for statistical process control. J Qual Technol 35(4):355–366

16. Hoekema R, Uijen GJ, Van Oosterom A (2001) Geometrical aspects of the interindividual variability of multilead ecg recordings. IEEE Trans Biomed Eng 48(5):551–559
17. Krempl G (2011) The algorithm apt to classify in concurrence of latency and drift. In: Proceedings of the intelligent data analysis (IDA), pp 222–233
18. Longoni M, Carrera D, Rossi B, Fragneto P, Pessione M, Boracchi G (2018) A wearable device for online and long-term ecg monitoring. In: Proceedings of the international conference on artificial intelligence (IJCAI), pp 5838–5840
19. Mairal J, Bach F, Ponce J (2012) Task-driven dictionary learning. IEEE Trans Pattern Anal Mach Intell 34(4):791–804
20. Mairal J, Ponce J, Sapiro G, Zisserman A, Bach F (2009) Supervised dictionary learning. In: Advances in neural information processing systems (NIPS), pp 1033–1040
21. Ni J, Qiu Q, Chellappa R (2013) Subspace interpolation via dictionary learning for unsupervised domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 692–699
22. Page ES (1954) Continuous inspection schemes. Biometrika 41(1/2):100–115
23. Ramirez I, Sprechmann P, Sapiro G (2010) Classification and clustering via dictionary learning with structured incoherence and shared features. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 3501–3508
24. Roberts S (1959) Control chart tests based on geometric moving averages. Technometrics 1(3):239–250
25. Ross GJ, Adams NM (2012) Two nonparametric control charts for detecting arbitrary distribution changes. J Qual Technol 44(2):102
26. Ruff L, Goernitz N, Deecke L, Siddiqui SA, Vandermeulen R, Binder A, Müller E, Kloft M (2018) Deep one-class classification. In: Proceedings of the international conference on machine learning (ICML), pp 4390–4399
27. Shekhar S, Patel VM, Nguyen HV, Chellappa R (2013) Generalized domain-adaptive dictionaries. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 361–368

Chapter 6
Enhancing Video Recommendation Using Multimedia Content

Yashar Deldjoo

Abstract Video recordings are complex media types. When we watch a movie, we can effortlessly register a lot of details conveyed to us (by the author) through different multimedia channels, in particular, the audio and visual modalities. To date, majority of movie recommender systems use collaborative filtering (CF) models or content-based filtering (CBF) relying on metadata (e.g., editorial such as genre or wisdom of the crowd such as user-generated tags) at their core since they are human-generated and are assumed to cover the 'content semantics' of movies by a great degree. The information obtained from multimedia content and learning from muli-modal sources (e.g., audio, visual and metadata) on the other hand, offers the possibility of uncovering relationships between modalities and obtaining an in-depth understanding of natural phenomena occurring in a video. These discerning characteristics of heterogeneous feature sets meet users' differing information needs. In the context of this Ph.D. thesis [9], which is briefly summarized in the current extended abstract, approaches to automated extraction of multimedia information from videos and their integration with video recommender systems have been elaborated, implemented, and analyzed. Variety of tasks related to movie recommendation using multimedia content have been studied. The results of this thesis can motivate the fact that recommender system research can benefit from knowledge in multimedia signal processing and machine learning established over the last decades for solving various recommendation tasks.

6.1 Introduction and Context

Users base their decision making about which movie to watch typically on its content, whether expressed in terms of metadata (e.g., genre, cast, or plot) or the feeling experienced after watching the corresponding movie trailer in which the visual content (e.g., color, lighting, motion) and the audio content (e.g., music or spoken dialogues)

Y. Deldjoo (✉)
SisInf Lab, Department of Electrical Engineering and Information Technology, Polytechnic University of Bari, Via Orabona, 4, 70125 Bari, Italy
e-mail: deldjooy@acm.org; yashar.deldjoo@poliba.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_6

play a key role in users' perceived affinity to the movie. The above examples underline that human interpretation of media items is intrinsically content-oriented.

Recommender systems support users in their decision making by focusing them on a small selection of items out of a large catalogue. To date, most video recommendation models use collaborative filtering (CF), content-based filtering on metadata (CBF-metadata), or a combination thereof at their core [33, 35, 42]. While CF models exploit the correlations encoded in users' preference indicators—either implicit (clicks, purchases) or explicit (ratings, votes)—to predict the best-matching user-item pairs, CBF models use the preference indications of a single target user and content information available about the items in order to build a user model (aka user profile) and compute recommendations. CBF approaches typically leverage *metadata* as a bridge between items and users, effectively disregarding a wealth of information encoded in the actual audio visual signals [13]. If we assume the primary role of a recommender system is to help people make choices that they will ultimately be satisfied with [7], such systems should thus take into account multiple source of information driving users' perception of media content and make a rational decision about their relative importance. This would in turn offer users the chance to learn more about their multimedia taste (e.g., their visual or musical taste) and their semantic interests [10, 13, 44].

The above are underlying ideas about why multimedia content can be useful for recommendation of *warm items* (videos with sufficient interactions). Nevertheless, in most video streaming services, new videos are continuously added. CF models are unable to make predictions in such scenario, since the newly added videos lack interactions, technically known as *cold-start* problem [5] and the associated items are referred by *cold items* (videos with few interactions) or *new items* (videos with no interactions). Furthermore, metadata can be rare/absent for cold/new videos, making it difficult to provide good quality recommendations [43]. Despite much research conducted in the field of RS for solving different tasks, the cold start (CS) problem is far from solved and most existing approaches suffer from it. Multimedia information *automatically extracted* from the audio-visual signals can serve as a proxy to solve the CS problem; in addition, it can act as a complementary information to identify videos that "look similar" or "sound similar" in warm start (WS) settings. These discerning characteristics of multimedia meet users' different information needs. As a branch of recommender systems, my Ph.D. thesis [9] investigates a particular area in the design space of recommender system algorithm in which the generic recommender algorithm needs to be optimized in order to use a wealth of information encoded in the actual image and audio signals.

6.2 Problem Formulation

In this section, we provide a formal definition of *content-based filtering video recommendation systems* (CBF-VRS) exploiting *multimedia content information*. In particular, we propose a general recommendation model of videos as composite

media objects, where the recommendation relies on the computation of distinct utilities, associated with image, audio, and textual modalities and a final utility, which is computed by aggregating individual utility values [22].[1]

A CBF-VRS based on multimedia content information is characterized by the following components:

1. Video Items: A video item s is represented by the triple: $s = (s_V, s_A, s_T)$ in which s_V, s_A, s_T refer to the *visual*, *aural*, and *textual* modalities, respectively. s_V encodes the *visual information* represented in the video frames; s_A encodes the *audio information* represented in sounds, music, spoken dialogues of a video; finally s_T is the metadata (e.g., genre labels, title) or natural language (e.g., sub-captions or speech spoken by humans and transcribed as text).

A video is a *multi-modal (composite)* media type (using s_A, s_V and s_T). This is while an audio item that represents performance of a classical music piece can be seen as an *uni-modal (atomic)* media type (using only s_A). A pop song with lyrics can be regarded as composite (using s_A and s_T), while an image of a scene or a silent movie atomic as well (using s_V). Multi-modal data supplies the system with rich and diverse information on the phenomenon relevant to the given task [27].

Definition 6.1 A CBF-VRS exploiting multimedia content is characterized as a system that is able to store and manage video items $s \in \mathcal{S}$, in which $\mathcal{S}$ is a repository of video items.

2. Multimedia Content-Based Representation: Developing a CBF-VRS based on multimedia content relies on content-based (CB) descriptions according to distinct modalities (s_V, s_A, s_T). From each modality, useful features can be extracted to describe the information of that modality. Different features can be classified based on several dimensions, e.g., the semantic expressiveness of features, level of granularity among others [4]. As for the former for instance, it is common to distinguish three levels of expressiveness, with increasing extent of semantic meaning: *low-level*, *mid-level*, and *high-level* features with respect to which features are categorized as shown in Table 6.1.

Over the last years, a large number of CB descriptors have been proposed to quantify various type of information in a video as summarized in Table 6.1. These descriptors are usually extracted by applying some form of signal processing or machine learning specific to a modality, and are described based on specific feature vectors. For example, in the visual domain, a rich suite of of low-level visual features are proposed by research in communities of multimedia, machine learning and computer vision for the purpose of image understanding, which we deem important for a CBF-VRS. The most basic and frequently used low-level features are *color*, *texture*, *edge* and *shape*, which are used to describe the "visual contents" of an image [26]. Besides, in the last two decades the need for devising descriptors that reduce or eliminate sensitivity to variations such as illumination, scale, rotation, and view point was

[1] Note that in this section, although definition of utilities are based on CBF model, in practice they can include a combination of CBF and CF models at their core.

Table 6.1 Categorization of different multimedia features based on their semantic expressiveness. Low-level features are close to the raw signal (e.g., energy of an audio signal, contrast in an image, motion in a video, or number of words in a text), while high-level features are close to the human perception and interpretation of the signal (e.g., motif in a classical music piece, emotions evoked by a photograph, meaning of a particular video scene, story told by a book author). In between, mid-level features are more advanced than low-level ones, but farther away from being semantically meaningful as high-level ones. They are often expressed as a combination or transformations of low-level features, or they are inferred from low-level features via machine learning

Hierarchy/Modalities	Visual	Audio	Textual
High-level (semantic)	Events, story	Structure, mood, message	Story, writing style
Mid-level (syntactic)	Objects, people, their interaction	Note onsets, rhythm patterns	Sentence, term-frequency
Low-level (stylistic)	Motion, color, texture, shape	Pitch, timbre, loudness	Tokens, n-grams

recognized in the community of computer vision. This gave rise to the development of a number of popular computer vision algorithms for image understanding [46]. They include for instance scale invariant feature transform (*SIFT*) [36], speeded up robust features (*SURF*) [6], local binary patterns (*LBP*) [41], discrete Wavelet transform (*DWT*), such as Gabor filters [37], discrete Fourier transform (*DFT*), and histogram of oriented gradients (*HOG*) [8]. The peak of these developments was reached in the early 2010s, when deep convolutional neural networks (*CNNs*) achieved groundbreaking accuracy for image classification [34]. One of the most frequently stated advantages of the Deep Neural Networks (*DNNs*) is that, they leverage the representational power of high-level semantics encoded in DNNs to narrow the semantic gap between the visual contents of the image and high-level concepts in the user's mind when consuming media items.

Definition 6.2 A CBF-VRS exploiting multimedia content is a system that is able to process video items and represent each modality in terms of a feature vector $\mathbf{f_m} = [f_1,\ f_2,\ \ldots f_{|\mathbf{f_m}|}] \in \mathbb{R}^{|\mathbf{f_m}|}$ where $m \in \{V, A, T\}$ represents the visual, audio or textual modality.[2]

3. Recommendation Model: A recommendation model provides suggestions for items that are most likely of interest to a particular user [42].

Let $\mathcal{U}$ and $\mathcal{S}$ denote a set of users and items, respectively. Given a target user $u \in \mathcal{U}$, to whom the recommendation will be provided, and a repository of items $s \in \mathcal{S}$, the general task of a personalized recommendation model is to identify the video item s^* that satisfies

$$\forall u \in \mathcal{U}\ ,\ s_u^* = \arg\max_{s \in \mathcal{S}}\ R(u, s) \tag{6.1}$$

[2]Note that here we step our attention outside the end-to-end learning approaches often performed by deep neural networks where the intermediate step of feature extraction is not done explicitly, and instead feature extraction and the final machine learning task are jointly performed.

where $R(u, s)$ is the *estimated utility* of item s for the user u on the basis of which the items are ranked [3]. The utility is infact a measure of *usefulness* of an item to a user and is measured by the RS to judge how much an item is *worth* being recommended. For example, some examples of such a utility function include a utility represented by a *rating* or a *profit* function [1].

Definition 6.3 A **multi-modal CBF-VRS** is a system that aims to improve learning performance using the knowledge/information aquired from different data sources of different video modalities. The utility of recommendations in a multi-modal CBF-VRS can be specified with respect to several specific utilities computed across each modality, thus

$$\forall u \in \mathcal{U}\ ,\ s_u^* = \arg\max_{s \in \mathcal{S}}\ R(u, s) = F(R_m(u, s)) \tag{6.2}$$

where $R_m(u, s)$ denotes the utility of item s *for user* u with regards to modality $m \in \{V, A, T\}$, and F is an aggregation function of the estimated utilities for each modality.

Based on the semantics of the aggregation, different functions can be employed, each implying a particular interpretation of the affected process. A standard and simplest form of aggregation functions are *conjunctive* (such as min operator), *disjunctive* (such as max operator), and *averaging* [39, 45]. As an example of the latter, and the one used in the field of multimedia information retrieval (MMIR), the weighted average linear combination is commonly used thus

$$R(u, s) = \sum_m w_m R_m(u, s) \tag{6.3}$$

where w_m is a weight factor indicating the importance of modality m, known as modality weights. The weights can be chosen as fixed weights or learned via machine learning. For example, recently studies based on *dictionary learning*, *co-clustering* and *multi-modal topic modeling* have become increasingly popular paradigms for the task of multi-modal inference [30]. For instance, multi-modal topic modelling (commonly methods based on latent semantic analysis or latent Dirichlet allocation) [30] models visual, audio and textual words with an underlying latent topic space.

6.3 Brief Overview of Ph.D. Research

The main processing stages involved in a CBF-VRS exploiting multimedia content are shown in Fig. 6.1. The input information are videos (movies) and the preference indications of a single user on them, and the output is a rank list of recommended videos (movies) tailored to target user's preference on the content

- **Temporal segmentation**: The goal of temporal segmentation is to partition the video item—infact the audio and image signals—into smaller structural units that

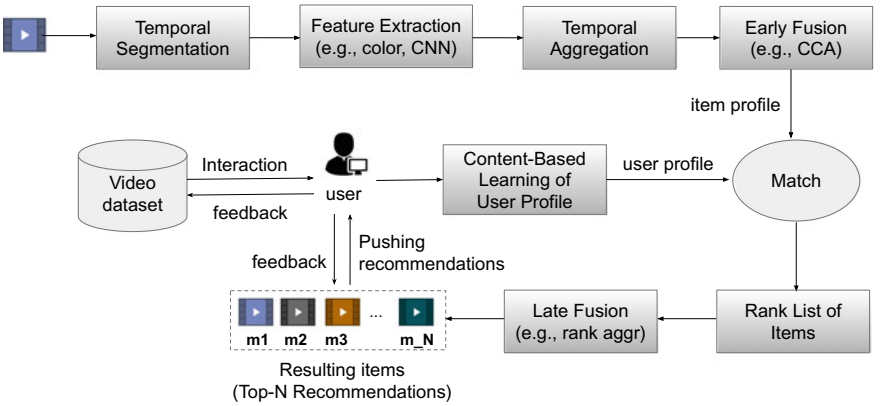


Fig. 6.1 The general framework illustrating the main processing steps involved in a CBF-VRS exploiting multimedia content. The framework focuses on the multimedia processing stages required to build a CBF system, however it can be extended to incorporate CF knowledge (e.g., in a hybrid CBF+CF system) and contextual factor (e.g., in context-aware CBF system)

have similar semantic content [29]. For the audio domain, segmentation approaches commonly operate at the *frame-level* or the *block-level*, the latter sometimes referred to as *segment-level* [25, 31]. For the visual domain, temporal segmentation segments the video into shots based on visual similarity between consecutive frames [15]. Some works consider scene-based segmentation, where a scene is semantically/hierarchically a higher-level video unit compared to a shot [15, 32]. Yet a simpler approach relies on capturing video at a fixed frame rate e.g., 1 fps and use all the resulting frames for processing [40].

- **Feature Extraction**: Feature extraction algorithms aim at encoding the content of the multimedia items in a concise and descriptive way, so to represent them for further use in retrieval, recommendation or similar systems [38]. An accurate feature representation can reflect item characteristics from various perspectives and can be highly indicative of user preferences.
  In the context of this Ph.D., a wide set of audio and visual features has been used to solve different movie recommendation tasks. As for the visual domain, they include *mise-en-scène visual features* (average short length, color variation, motion and lighting key) [12, 15], visual features based on the *MPEG-7 standard* and *pre-trained CNNs* [17] and the most stable and recent datasets, named Multifaceted Movie Trailer Feature dataset (MMTF-14K) and Multifaceted Video Clip Dataset (MVCD-7K) [11, 21], which we made publicly available online. In particular, MMTF-14K[3] provides state-of-the-art audio and visual descriptors for approximately 14K Hollywood-type *movie trailers* accompanied with metadata and user preference indicators on movies that are linked to the ML-20M dataset. The visual descriptors consist of two categories of descriptors: *aesthetic features*

[3] https://mmprj.github.io/mtrm_dataset/index.

and *pre-trained CNN* (AlexNet) features, each of them including different aggregation schemes for the two types of visual features. The audio descriptors consist of two classes of descriptors: *block-level features* (BLF) and *i-vector features* capturing spectral and timbral features of audio signal. To the best of our knowledge, MMTF-14K is the only large scale **multi-modal** dataset to date providing a rich source for devising and evaluating movie recommender systems.

A criticism of the MMTF-14K dataset however is that its underlying assumption relies on the fact that *movie trailers are representative of full movies*. Movie trailers are human-edited and artificially produced with lots of thrills and chills as their main goal is to motivate users to come back (to the cinema) and watch the movie. For this reason, the scenes in trailers are usually taken from the most exciting, funny, or otherwise noteworthy parts of the film,[4] which is a strong argument against the representativeness of trailers for the full movie. To address these shortcomings, in 2019 we introduced a new dataset of video clips, named Multi-faceted Video Clip Dataset (MFVCD-7K).[5] Each movie in MFVCD-7K can have several associated video clips, each focused on a particular scene, displaying it at its natural pace. Thus, video clips in MFVCD-7K can serve as a more *realistic* summary of the movie story than trailers.

- **Temporal aggregation**: This step involves creating a video-level descriptor by aggregating the features temporally. The following approaches are widely used in the field of multimedia processing: (i) *statistical summarization*: it is the simplest approach using the operators mean, standard deviation, median, maximum, or combinations thereof, e.g., means plus covariance matrix to build an item-level descriptor; (ii) *probabilistic modeling*: it is an alternative approach for temporal aggregation, which summarizes the local features of the item under consideration by a probabilistic model. Gaussian mixture models (GMMs) are often used for this purpose; (iii) *other approaches*: other feature aggregation techniques include vector quantization (VQ), vectors of locally aggregated descriptors (VLAD) and Fisher vectors (FV), where the last two were originally used for aggregating image key-point descriptors. They are used as a post-processing step for video representation, for example within a convolutional neural network (CNN) [28]. For different movie recommendation tasks, we used different temporal aggregation functions [12, 13, 15].
- **Fusion**: This step is the main step toward building a multi-modal VRS. Early fusion attempts to combine feature extracted from various unimodal streams into a single representation. For instance, in [13, 16] we studied adoption of an effective early fusion technique named *canonical correlation analysis* (CCA) to combine visual, textual and/or audio descriptors extracted from movie trailers and better exploit complementary information between different modalities. Late fusion approaches combine outputs of several system run on different descriptors. As an example of this approach, in [11, 13], we used a novel late fusion strategy based on a weighted variant of the Borda rank aggregation strategy to combine heterogeneous feature

[4] https://filmshortage.com/the-art-of-the-trailer/.

[5] https://mmprj.github.io/MFVCD-7K.

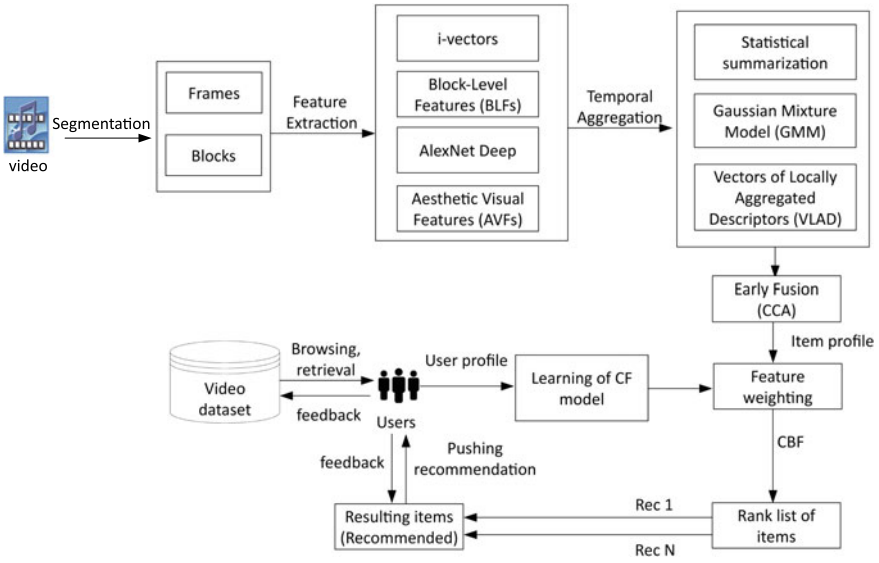


Fig. 6.2 The proposed collaborative-filtering-enriched content-based filtering (CFeCBF) movie recommender system framework proposed in [13] to solve the new/cold-item problem

sets into a unified ranking of videos and showed promising improvements of the final ranking. Note that a different level of hybridization—from a recommendation point of view—can involve fusing CBF system with a CF model, which we considered e.g., in [14, 17].

- **Content-based learning of user profile and Recommendation**: The goal of this step is to learn a user-specific model which is used to predict the target user's interest in (multimedia) items based on her past history of interaction with the items [2]. The learned user profile model is compared to representative item features (or item profiles) in order to make recommendations tailored to target user's preference on the content.

For instance, in [10] a *multi-modal content-based movie recommender system* is presented that exploits rich content descriptors based on state-of-the-art multimedia descriptors: *block-level and i-vector features for audio* and *aesthetic and deep visual features*. For multi-modal learning, a novel late fusion strategy based on an extended version of the Borda rank aggregation strategy was proposed which resulted in an improved ranking of videos. Evaluation was carried out on a subset of MovieLens-20M and multimedia features extracted from 4,000 movie trailers, by (i) a *system-centric study* to measure the offline quality of recommendations in terms of accuracy-related (MRR, MAP, recall) and beyond-accuracy (novelty, diversity, coverage) performance, and (ii) a *user-centric online experiment*, measuring different subjective metrics (relevance, satisfaction, diversity). Results of empirical evalua-

Table 6.2 Comparison of different research works carried out in the context of this Ph.D. thesis. Abbreviations: Focus: Focus of Study, WS: Warm Start, CS: Cold Start, Meta Type: Metadata Type, Ed: Editorial Metadata (e.g., genre labels), UG: User-generated metadata (e.g., tags), M-mod Fusion: Multi-modality Fusion Type, Rec Model: Recommendation Model, CF: Collaborative Filtering, CBF: Content-based Filtering, CA: Context-Aware, Acc: Accuracy Metric (e.g., MAP, NDCG), Beyond: Beyond Accuracy Metric (novelty, diversity, coverage)

Res	Year	Focus		Content modality			Meta type		M-mod fusion		Rec model			Eval type		Eval metric	
		WS	CS	Audio	Visual	Textual	Ed	UG	Early	Mid/Late	CBF	CF	CA	Offline	User-study	Acc	Beyond
[18]	2015	✓			✓						✓			✓		✓	
[15]	2016	✓			✓						✓			✓		✓	
[14]	2016	✓			✓						✓	✓		✓		✓	
[12]	2017	✓			✓						✓			✓		✓	
[24]	2017	✓			✓	✓	✓			✓	✓			✓	✓	✓	✓
[19]	2017	✓			✓								✓				
[20]	2017	✓			✓								✓				
[10]	2018	✓		✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓
[11]	2018		✓	✓	✓	✓	✓	✓		✓	✓			✓		✓	
[17]	2018	✓			✓		✓		✓		✓	✓		✓		✓	
[13]	2019	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓	✓

tion indicates that multimedia features can provide a good alternative to metadata (as baseline), with regards to both accuracy measures and beyond accuracy measures.

In [13] a novel *movie multi-modal recommender system* is proposed that specifically addresses the *new item cold-start problem* by: (i) integrating state-of-the-art audio and visual descriptors, which can be automatically extracted from video content and constitute what we call the *movie genome*; (ii) exploiting an effective data fusion method named *canonical correlation analysis* (CCA) to better exploit complementary information between different modalities; (iii) proposing a two-step hybrid approach which trains a CF model on warm items (items with interactions) and leverages the learned model on the movie genome to recommend cold items (items without interactions). The recommendation method is thus named collaborative-filtering enriched CBF (CFeCBF), which has a different functioning concept compared with a standard CBF system (compare Figs. 6.1 and 6.2). Experimental validation is carried out using a system-centric study on a large-scale, real-world movie recommendation dataset both in an absolute cold start and in a cold to warm transition; and a user-centric online experiment measuring different subjective aspects, such as satisfaction and diversity. Results from both the offline study as well as a preliminary user-study confirm the usefulness of their model for new item cold start situations over current editorial metadata (e.g., genre and cast).

Finally, in Table 6.2, we provide a brief comparison of a selected number of research works completed in the course of this Ph.D. thesis by highlighting their main aspects. Readers are referred to the comprehensive literature review on recommender system leveraging multimedia content in which I describe many domains where multimedia content plays a key role in human decision making and are considered in the recommendation process [23].

6.4 Conclusion

This extended abstract briefly discusses the main outcomes of my Ph.D. thesis [9]. This Ph.D. thesis studies video recommender systems using multimedia content in detail—a particular area in the design space of recommender system algorithms where the generic recommender algorithm can be configured in order to integrate a rich source of information extracted from the actual audio-visual signals of video. I believe different systems, techniques and tasks for movie recommendation, which were studied in this Ph.D. thesis can pave the path for a new paradigm of video (and in general multimedia) recommender system by designing recommendation models built on top of rich item descriptors extracted from content.

References

1. Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749. https://doi.org/10.1109/TKDE.2005.99
2. Aggarwal CC (2016) Content-based recommender systems. Recommender systems. Springer, Berlin, pp 139–166
3. Aggarwal CC (2016) An introduction to recommender systems. Recommender systems. Springer, Berlin, pp 1–28
4. Al-Halah Z, Stiefelhagen R, Grauman K (2017) Fashion forward: forecasting visual style in fashion. In: IEEE international conference on computer vision, ICCV 2017, Venice, Italy, October 22–29, 2017, pp 388–397. https://doi.org/10.1109/ICCV.2017.50.
5. Asmaa Elbadrawy GK (2015) User-specific feature-based similarity models for top-n recommendation of new items. ACM Trans Intell Syst, 6. https://doi.org/10.1145/2700495
6. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: European conference on computer vision, pp 404–417. Springer
7. Chen L, De Gemmis M, Felfernig A, Lops P, Ricci F, Semeraro G (2013) Human decision making and recommender systems. ACM Trans Interact Intell Syst (TiiS) 3(3):17
8. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, CVPR 2005, vol 1, pp 886–893. IEEE
9. Deldjoo Y (2018) Video recommendation by exploiting the multimedia content. PhD thesis, Italy
10. Deldjoo Y, Constantin MG, Eghbal-Zadeh H, Ionescu B, Schedl M, Cremonesi P (2018) Audio-visual encoding of multimedia content for enhancing movie recommendations. In: Proceedings of the 12th ACM conference on recommender systems, RecSys 2018, Vancouver, BC, Canada, October 2–7, 2018, pp 455–459. https://doi.org/10.1145/3240323.3240407
11. Deldjoo Y, Constantin MG, Ionescu B, Schedl M, Cremonesi P (2018) MMTF-14K: a multifaceted movie trailer feature dataset for recommendation and retrieval. In: Proceedings of the 9th ACM multimedia systems conference, MMSys 2018, Amsterdam, The Netherlands, June 12–15, 2018, pp 450–455. https://doi.org/10.1145/3204949.3208141
12. Deldjoo Y, Cremonesi P, Schedl M, Quadrana M (2017) The effect of different video summarization models on the quality of video recommendation based on low-level visual features. In: Proceedings of the 15th international workshop on content-based multimedia indexing, CBMI 2017, Florence, Italy, June 19–21, 2017, pp 20:1–20:6. https://doi.org/10.1145/3095713.3095734
13. Deldjoo Y, Dacrema MF, Constantin MG, Eghbal-zadeh H, Cereda S, Schedl M, Ionescu B, Cremonesi P (2019) Movie genome: alleviating new item cold start in movie recommendation. User Model User-Adapt Interact 29(2):291–343. https://doi.org/10.1007/s11257-019-09221-y
14. Deldjoo Y, Elahi M, Cremonesi P (2016) Using visual features and latent factors for movie recommendation. In: Proceedings of the 3rd workshop on new trends in content-based recommender systems co-located with ACM conference on recommender systems (RecSys 2016), Boston, MA, USA, September 16, 2016, pp 15–18. http://ceur-ws.org/Vol-1673/paper3.pdf
15. Deldjoo Y, Elahi M, Cremonesi P, Garzotto F, Piazzolla P, Quadrana M (2016) Content-based video recommendation system based on stylistic visual features. J Data Semant 5(2):99–113. https://doi.org/10.1007/s13740-016-0060-9
16. Deldjoo Y, Elahi M, Cremonesi P, Moghaddam FB, Caielli ALE (2016) How to combine visual features with tags to improve movie recommendation accuracy? In: International conference on electronic commerce and web technologies, pp 34–45. Springer
17. Deldjoo Y, Elahi M, Quadrana M, Cremonesi P (2018) Using visual features based on MPEG-7 and deep learning for movie recommendation. IJMIR 7(4):207–219. https://doi.org/10.1007/s13735-018-0155-1

18. Deldjoo Y, Elahi M, Quadrana M, Cremonesi P, Garzotto F (2015) Toward effective movie recommendations based on mise-en-scène film styles. In: Proceedings of the 11th biannual conference on Italian SIGCHI chapter, CHItaly 2015, Rome, Italy, September 28–30, 2015, pp 162–165. https://doi.org/10.1145/2808435.2808460
19. Deldjoo Y, Frà C, Valla M, Cremonesi P (2017) Letting users assist what to watch: an interactive query-by-example movie recommendation system. In: Proceedings of the 8th Italian information retrieval workshop, Lugano, Switzerland, June 05–07, 2017, pp 63–66. http://ceur-ws.org/Vol-1911/10.pdf
20. Deldjoo Y, Frà C, Valla M, Paladini A, Anghileri D, Tuncil MA, Garzotta F, Cremonesi P et al (2017) Enhancing children's experience with recommendation systems. In: Workshop on children and recommender systems (KidRec'17)-11th ACM conference of recommender systems, pp N–A
21. Deldjoo Y, Schedl M (2019) Retrieving relevant and diverse movie clips using the mfvcd-7k multifaceted video clip dataset. In: Proceedings of the 17th international workshop on content-based multimedia indexing
22. Deldjoo Y, Schedl M, Cremonesi P, Pasi G (2018) Content-based multimedia recommendation systems: definition and application domains. In: Proceedings of the 9th Italian information retrieval workshop, Rome, Italy, May, 28–30, 2018. http://ceur-ws.org/Vol-2140/paper15.pdf
23. Deldjoo Y, Schedl M, Cremonesi P, Pasi G (2020) Recommender systems leveraging multimedia content. ACM Comput Surv (CSUR)
24. Elahi M, Deldjoo Y, Moghaddam FB, Cella L, Cereda S, Cremonesi P (2017) Exploring the semantic gap for movie recommendations. In: Proceedings of the Eleventh ACM conference on recommender systems, RecSys 2017, Como, Italy, August 27–31, 2017, pp 326–330. https://doi.org/10.1145/3109859.3109908
25. Ellis DP (2007) Classifying music audio with timbral and chroma features. ISMIR 7:339–340
26. Flickner M, Sawhney HS, Ashley J, Huang Q, Dom B, Gorkani M, Hafner J, Lee D, Petkovic D, Steele D, Yanker P (1995) Query by image and video content: the QBIC system. IEEE Comput 28(9):23–32. https://doi.org/10.1109/2.410146
27. Geng X, Wu X, Zhang L, Yang Q, Liu Y, Ye J (2019) Multi-modal graph interaction for multi-graph convolution network in urban spatiotemporal forecasting. arXiv:1905.11395
28. Girdhar R, Ramanan D, Gupta A, Sivic J, Russell B (2017) Actionvlad: Learning spatio-temporal aggregation for action classification. arXiv:1704.02895
29. Hu W, Xie N, Li L, Zeng X (2011) Maybank S (2011) A survey on visual content-based video indexing and retrieval. IEEE Trans Syst Man Cybern Part C (Applications and Reviews) 41(6):797–819
30. Irie G, Liu D, Li Z, Chang S (2013) A bayesian approach to multimodal visual dictionary learning. In: 2013 IEEE conference on computer vision and pattern recognition, Portland, OR, USA, June 23–28, 2013, pp 329–336. https://doi.org/10.1109/CVPR.2013.49
31. Knees P, Schedl M (2013) A survey of music similarity and recommendation from music context data. ACM Trans Multimed Comput Commun Appl (TOMCCAP) 10(1)
32. Koprinska I, Carrato S (2001) Temporal video segmentation: a survey. Signal Process Image Commun 16(5):477–500
33. Koren Y, Bell R (2015) Advances in collaborative filtering. In: Recommender systems handbook, pp 77–118. Springer
34. Liu L, Chen J, Fieguth P, Zhao G, Chellappa R, Pietikainen M (2018) A survey of recent advances in texture representation. arXiv:1801.10324
35. Lops P, De Gemmis M, Semeraro G (2011) Content-based recommender systems: state of the art and trends. In: Recommender systems handbook, pp 73–105. Springer, Berlin
36. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
37. Manjunath BS, Ma WY (1996) Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell 18(8):837–842
38. Marques O (2011) Practical image and video processing using MATLAB. Wiley, New York

39. Marrara S, Pasi G, Viviani M (2017) Aggregation operators in information retrieval. Fuzzy Sets Syst 324:3–19
40. Ng JY, Hausknecht MJ, Vijayanarasimhan S, Vinyals O, Monga R, Toderici G (2015) Beyond short snippets: deep networks for video classification. In: IEEE conference on computer vision and pattern recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015, pp 4694–4702. https://doi.org/10.1109/CVPR.2015.7299101
41. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
42. Ricci F, Rokach L, Shapira B (2015) Recommender systems: introduction and challenges. In: Recommender systems handbook, pp 1–34. Springer, Berlin
43. Roy S, Guntuku SC (2016) Latent factor representations for cold-start video recommendation. In: Proceedings of the 10th ACM conference on recommender systems, pp 99–106. ACM
44. Swearingen K, Sinha R (2002) Interaction design for recommender systems. Des Interact Syst 6:312–334
45. Tzeng GH, Huang JJ (2011) Multiple attribute decision making: methods and applications. CRC Press, Boca Raton
46. Vedaldi A, Fulkerson B (2008) VLFeat: an open and portable library of computer vision algorithms. http://www.vlfeat.org/

Chapter 7
Dynamic Application Autotuning for Self-aware Approximate Computing

Davide Gadioli

Abstract The energy consumption limits the application performance in a wide range of scenarios, ranging from embedded to High-Performance Computing. To improve computation efficiency, this Chapter focuses on a software-level methodology to enhance a target application with an adaptive layer that provides self-optimization capabilities. We evaluated the benefits of dynamic autotuning in three case studies: a probabilistic time-dependent routing application from a navigation system, a molecular docking application to perform virtual-screening, and a stereo-matching application to compute the depth of a three-dimensional scene. Experimental results show how it is possible to improve computation efficiency by adapting reactively and proactively.

7.1 Introduction

The increasing demand for computation power shifted the optimization focus towards efficiency in a wide range of energy-constrained systems, from embedded platforms to High-Performance Computing (HPC) [1]. A promising way to improve energy efficiency is approximate computing [2], which aims at finding a *good enough* solution, avoiding the unnecessary computation effort. It is possible to approximate the computation at different levels: from approximate hardware [3] to software techniques such as loop perforation [4]. Moreover, a large class of applications exposes software parameters that define an accuracy-throughput trade-off, especially in the multimedia field [5]. In this chapter, we focus at the software-level, where an application exposes *software-knobs* [6] that can alter its extra-functional behaviour. In this context, the definition of application performance includes several extra-functional properties (EFPs) in conflict to each other.

The value of an EFP might depend on the underlying architecture configuration, the system workload, and the features of the current input. Since this information usually changes at runtime, it is not trivial to find a one-fits-all configuration of

D. Gadioli (✉)
Politecnico di Milano, Piazza Leonardo da Vinci, 32, Milan, Italy
e-mail: davide.gadioli@polimi.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_7

the application software-knobs at design-time. This is a well-known problem in the autonomic computing field [7], where researchers investigate approaches to provide self-optimization capabilities to a target application, for identifying and seizing optimization opportunities at runtime.

In this context, we propose a methodology to enhance an application with an adaptation layer that exposes reactive and proactive mechanisms to provide continuously the most suitable software-knobs configuration according to application requirements. The decision process is based on the application knowledge, which describes the relationship between software-knob configurations and EFPs. It is possible to learn the application knowledge either at design-time, by using well-known Design Space Exploration techniques [8], or at runtime by using an external component that coordinates a distributed DSE [9]. The benefits of the latter approach are the following: (1) the application can observe its behaviour with the same execution environment of the production run, (2) it can leverage features of the production input, and in the context of a distributed computation, (3) it can leverage the number of application instances to lower the learning time.

We evaluate the benefits of the methodology implementation, named *mARGOt*, in three real-world case studies. In particular, we use a stereo-matching application [5] to assess the benefits of reactive adaptation mechanisms, with respect to changes of both application requirements and performance. We use a probabilistic time-dependent routing stage in a navigation car system [10] to evaluate the benefits of the proactive adaptation mechanisms. Finally, we use a molecular docking application for virtual screening, to assess the benefits of learning the application knowledge at runtime.

The remainder of the Chapter is organized as follows. First, Sect. 7.2 provides an overview of autonomic computing, focusing on application autotuning and highlighting the main contributions of the proposed methodology. Then, in Sect. 7.3, we formalize the problem and describe the *mARGOt* framework. Section 7.4 discusses the benefits of the proposed methodology. Finally, Sect. 7.5 concludes the chapter.

7.2 Autonomic Computing and Application Autotuning

In the context of autonomic computing [7], we perceive a computing system as an ensemble of autonomous elements capable of self-management. The main idea is to enable the system to perform autonomously a task which is traditionally assigned to a human, to cope with the increasing complexity of computation platforms and applications. For example, if the system is able to incorporate new components whenever they become available, the system has the self-configuration ability. To qualify for the self-management ability, a system must satisfy all the related self-* properties. How to provide these properties is still an open question and previous surveys [11, 12] summarize the research effort in this area.

In this chapter, we focus on the self-optimization property at the software-level, which is the ability to identify and seize optimization opportunities at runtime. The methodologies to provide self-optimization properties are also known in the literature

as *autotuners*. It is possible to categorize autotuners in two main categories: static and dynamic.

Static autotuners aim at exploring a large space of software-knobs configuration space to find the most suitable software-knob configuration, assuming a predictable execution environment and targeting software-knobs that are loosely input-dependent. Among static autotuners, we might consider the following works. AutoTune [13] targets multi-node applications and it leverages the Periscope framework [14] to measure the execution time. It targets application-agnostic parameters exposed by the computation pipeline such as communication buffers and OpenHMPP/MPI parameters. QuickStep [15] and Paraprox [16] perform code transformations to automatically apply approximation techniques for enabling and leveraging an accuracy-throughput trade-off. OpenTuner [17] and the ATF framework [18] explicitly address the exponential growth in the complexity of exploring the parameters space, by using an ensemble of DSE techniques and by taking into account dependencies between software-knobs. Although very interesting, these approaches work at design-time and they usually target a different set of software-knobs with respect to dynamic autotuners.

The methodology proposed in this chapter belongs to the category of dynamic autotuners, which aim at changing the software-knobs configuration during the application runtime according to the system evolution. Therefore, they focus on providing adaptation mechanisms, typically based on application knowledge. Among dynamic autotuners, we might consider the following works. The Green framework [19] and PowerDial [6] enhance an application with a reactive adaptation layer, to change the software-knob configurations according to a change on the observed behaviour. The IRA framework [20] and Capri [21] focus instead on providing proactive adaptation mechanisms to select the software-knob configuration according to the features of the current input. On the other hand, Petabricks [22] and Anytime Automaton [23] are capable to adapt the application reactively and proactively. However, they require a significant integration effort from the application developers. Moreover, they are capable to leverage only accuracy-throughput trade-off. All these previous works are interesting and they have significantly contributed to the field, according to their approach on how to provide the self-optimization capabilities. The main contributions of *mARGOt* is to provide a single methodology to provide an adaptation layer with the following characteristics:

- The flexibility to express the application requirements as a constrained multi-objective optimization problem, addressing an arbitrary number of EFPs and software-knobs. Moreover, the application might change the requirements at runtime.
- The capability to adapt the application reactively and proactively, where the user might observe the application behaviour continuously, periodically or sporadically. Moreover, the reaction policies are not only related to the throughput, but they are agnostic about the observed EFP.
- To minimize the integration effort, we employ the concept of separation of concerns. In particular, application developers define extra-functional aspects in a

configuration file and the methodology is capable to generate an easy-to-use interface to wrap the target region of code.

Furthermore, being possible to define the application knowledge at runtime, *mARGOt* can use an external component to orchestrate a distributed DSE at runtime, during the production phase.

7.3 The mARGOt Autotuning Framework

This section describes how the proposed methodology can enhance the target application with an adaptation layer. At first, we formalize the problem, the application requirements and how the *mARGOt* interacts with the application. Then, we describe the main components of the framework and how they can adapt the application reactively and proactively.

7.3.1 Problem Definition

Figure 7.1 shows the overview of the *mARGOt* and how it interacts with an application. To simplify, we assume that the application is composed of a single computation kernel g that reads a stream of input i to produce the required stream of output o. However, *mARGOt* is capable to manage several regions of code independently. Moreover, the target kernel exposes a set software-knobs $\overline{x}$ that alter the extra-functional behaviour. Since a change of the configuration of these knobs might lead to a different quality of the results, we might define the application as $o = g(i, \overline{x})$.

Given this definition, the application requirements are expressed as a constrained multi-objective optimization problem described in Eq. 7.1:

$$\begin{aligned} \max(\min)\ & r(\overline{x}; \overline{m} \mid \overline{f}) \\ \text{s.t. } & C_1 : \omega_1(\overline{x}; \overline{m} \mid \overline{f}) \ \propto \ k_1 \ \ with\ \alpha_1\ confidence \\ & C_2 : \omega_2(\overline{x}; \overline{m} \mid \overline{f}) \ \propto \ k_2 \\ & \dots \\ & C_n : \omega_n(\overline{x}; \overline{m} \mid \overline{f}) \ \propto \ k_n \end{aligned} \tag{7.1}$$

where r denotes the objective function to maximize (minimize), $\overline{m}$ is the vector of metrics of interest (i.e. the EFPs), and $\overline{f}$ is the vector of input features. Let C be the set of constraints, where each constraint C_i is expressed as the function ω_i over $\overline{m}$ or $\overline{x}$, that must satisfy the relationship $\propto \in \{<, \leq, >, \geq\}$, with a confidence α_i, if ω_i targets a statistical variable. If the application is input-dependent, ω_i and r depend on its features.

In this context, the goal of the autotuner is to solve the optimization problem by inspection, using the application knowledge. The application always needs a

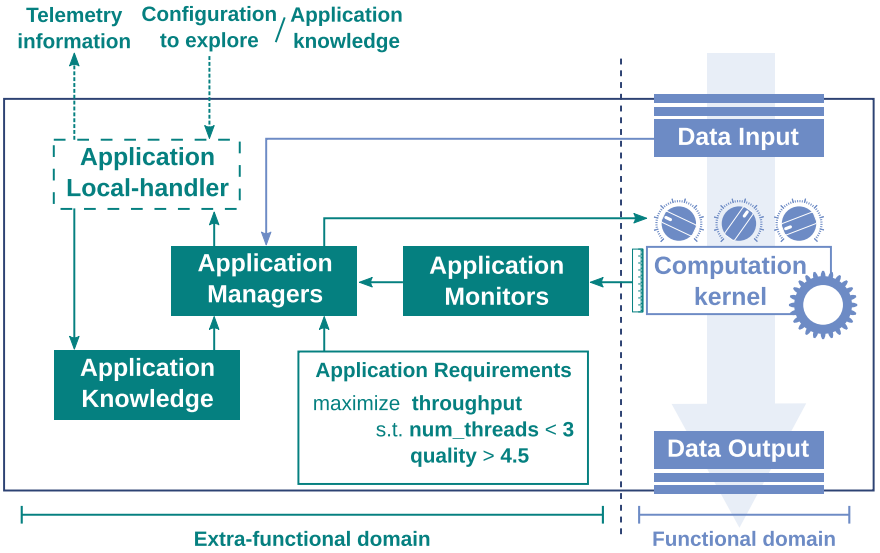


Fig. 7.1 The overview of the proposed autotuning framework. Green elements represent framework components, while blue elements represent application components (Color figure online)

software-knob configuration, therefore if the problem is unfeasible, *mARGOt* can relax the constraints, according to their priority, until it finds a valid solution.

We implemented *mARGOt* as a standard C++ library that should be linked to the target application. Being the time spent by the framework to select the most suitable software-knob configuration stolen from the application, the *mARGOt* implementation has been designed to be lightweight and to minimize the introduced overhead. We publicly released the framework source code [24].

7.3.2 Application-Knowledge

To solve the optimization problem we need a model of the application behaviour. However, the relationship between software-knobs, EFPs, and input features is complex and unknown. Therefore, *mARGOt* defines the application-knowledge as a list of *Operating Points* (OPs). Each OP θ corresponds to a software-knob configuration, together with the reached EFPs. If the application is input-dependent, the OP includes also the information on the related input features: $\theta = \{\overline{x}, \overline{m}, \overline{f}\}$.

This representation has been chosen for the following reasons: it provides a high degree of flexibility to describe the application behaviour, *mARGOt* can solve efficiently the optimization problem described in Eq. 7.1 and it prevents the possibility to select an invalid software-knob configuration.

The application-knowledge is considered an input of *mARGOt*, and there are several tools that can explore the Design Space efficiently. In particular, *mARGOt* uses the XML format of Multicube Explorer [25] to represent the application-knowledge. Moreover, it is possible to learn it at runtime, as described in Sect. 7.3.4.

7.3.3 Interaction with the Application

The core component of *mARGOt* is the application manager, which is in charge of solving the optimization problem and of providing to the application the most suitable software-knob configuration. The application is able to change the application requirements at runtime according to the system evolution. Moreover, if the EFPs are input-dependent, the application can provide features of the actual input to adapt proactively [9].

To enable the reactive adaptation, *mARGOt* must observe the actual behaviour of the application and compare it with the expected one. In particular, we compute a coefficient error for each observed EFP as $e_{m_i} = \frac{expected_i}{observed_i}$, where e_{m_i} is the error coefficient for the i-th EFP. Under the assumption of linear error propagation among the OPs, *mARGOt* is able to adapt reactively.

The methodology implementation provides to application developers a suite of monitors to observe the most common metrics of interest, such as throughput or performance events using PAPI [26]. The customization of a monitor to observe an application-specific EFP, such as the accuracy, is straightforward.

7.3.4 Runtime Application-Knowledge Learner

The *mARGOt* flexibility enables the possibility to change the application knowledge at runtime. Therefore, it is possible to learn the application knowledge directly at runtime, during the production phase. Although this approach is platform independent, it focuses on the High-Performance Computing Scenario. The idea is to perform a distributed Design Space Exploration, leveraging the parallelism level of the platform to lower the exploration time. In particular, when we spawn an application, it notifies its existence to a central coordinator, along with information about the exploration, such as the software-knobs domains, and the list of EFPs. According to a Design of Experiments, the central coordinator dispatches software-knobs configuration to evaluate at each application instance, which provides as feed-back telemetry information. Once the central coordinator collects the required observations, it uses learning techniques to derive the application-knowledge to broadcast to the application instances.

Figure 7.2 shows an overview of the central coordinator. In particular, it uses a thread pool of *application remote-handlers* to interact with *application local-handler* of the application instances. The communication uses the lightweight MQTT or

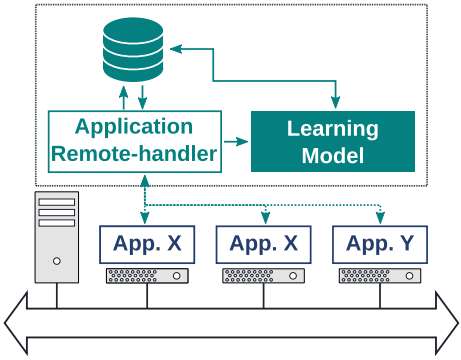


Fig. 7.2 The proposed approach to perform a distributed on-line Design Space Exploration, using a dedicated server outside of the computation node

MQTTs protocols, while we use the Cassandra database to store the required information. The learning module leverages a well-known approach [27] to interpolate application performance, implemented by the state-of-the-art R package [28].

7.4 Experimental Results

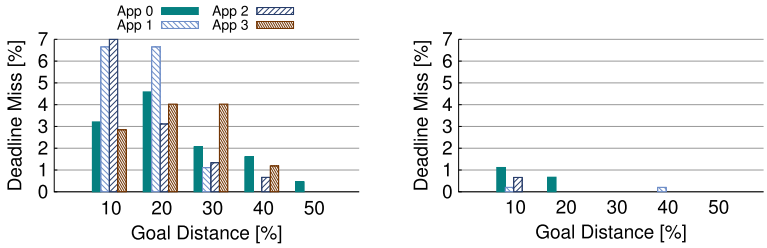
This section assesses the benefits of the proposed adaptation layer, by deploying the *mARGOt* framework in three different case studies. Each case study highlights a different characteristic of *mARGOt*.

7.4.1 Evaluation of the Reactive Adaptation

This experiment aims at assessing the benefits of the reactive adaptation mechanisms. We focus on a Stereo-matching application deployed in a quad-core architecture [29]. The application takes as input a pair of stereo images of the same scene and it computes the disparity map as output. The algorithm exposes application-specific software-knobs that define an accuracy-throughput trade-off. Moreover, it is possible to change the number of software threads that the application can use to carry out the computation.

We create the application-knowledge at design-time, evaluating each configuration in isolation. In this experiment, we execute four instances of stereo-matching, overlapping their execution: each application has an execution time of 200 s, and we spawn them with a delay of 50 s between each other. The application designer would like to maximize the accuracy of the elaboration, provided that the application must sustain a throughput of 2 fps.

In this experimental setup, we compare two adaptation strategies. On one hand, we consider as *baseline* the reaction policy that monitors the throughput of the



(a) Baseline approach (b) *mARGOt* approach

Fig. 7.3 Distribution of deadline misses with respect to the constraint on the throughput, according to the distance from the target value (2 fps)

application and that reacts accordingly, since it is commonly used in literature. In particular, if the monitors observe that the actual throughput of the application is lower than the expected one, *mARGOt* will choose a configuration with an higher expected throughput to compensate. On the other hand, we exploit the *mARGOt* flexibility to consider also the CPU usage in the requirements. In particular, we define a constraint on the available resource usage on top of the one on the throughput, to limit the number of software threads according to the available resources. The value of the constraint is continuously updated at runtime as follows:

$$CPU_{available} = \Gamma - \gamma + \pi_{measured}$$

where Γ is the maximum CPU quota available in the platform, γ is the monitored CPU quota used by the system, and $\pi_{measured}$ is the monitored CPU quota assigned to the application by the Operating System. The *mARGOt* capability to consider an arbitrary number of EFPs enables this adaptation strategy.

From the experimental results, we can observe how the two strategies are capable to satisfy the application requirements on average. However, Fig. 7.3 shows the distribution of the deadline misses for the two strategies. The baseline strategy relies on the scheduler for a fair assignment of the CPU quota, therefore the contention on the resources reduces the predictability of the throughput. Conversely, the *mARGOt* approach avoids this contention by observing the CPU usage. However, we are not able to guarantee a fair allocation of the CPU quota by using this approach.

7.4.2 Evaluation of the Proactive Adaptation

This experiment aims at assessing the benefits of the proactive adaptation mechanisms. We focus on a phase of a car navigation system: the probabilistic time-dependent routing (PTDR) application [10]. It takes as input the starting time of the travel and the speed profiles of all the segments that compose the target path. The

Table 7.1 Number of independent route traversal simulations by varying the requested maximum error and the statistical properties of interest

Approach	Error (%)	Simulations 50th percentile	Simulations 75th percentile	Simulations 95th percentile
Baseline	3	1000	3000	3000
	6	300	1000	1000
Adaptive	3	632	754	1131
	6	153	186	283

speed profiles of a segment vary during the week, with a fifteen minutes granularity. To estimate the arrival time distribution, the PTDR application uses a Monte Carlo approach that simulates multiple times an independent route traversal. The output of the application is a statistical property of the arrival time distribution such as the 50th or 75th percentile. This application has been already optimized to leverage the target HPC node [30], therefore it exposes as software-knob the number of route traversal simulations.

The application designer would like to minimize the number of route traversal simulations, given that the difference between the value computed with the selected configuration and with 1 M samples are below a given threshold. The threshold value might vary according to the type of user that generates the request. In this experiment, we consider 3% for premium users and 6% for free users.

To evaluate the benefits of the proactive adaptation, we compare two adaptation strategies. On one hand, we fix the number of samples according to the worst path in a representative input set [30]. We use this strategy as a baseline. On the other hand, we use an adaptive strategy that extracts a feature from the actual input to select the number of simulations according to the target path [10].

Table 7.1 shows the experimental results of a large experimental campaign with randomly selected pairs of Czech Republic routes and starting times. The adaptive approach can significantly reduce the number of required simulations according to the target statistical property and the maximum error level. Moreover, we modelled the car navigation system with the simulation environment Java Modeling Tools to measure the benefits of the adaptive strategy at the system level. With a load of 100 k requests every 2 min, an error threshold of 6%, and assuming that we are interested in the 95th percentile, the adaptive strategy can reduce the number of nodes by 26%. These parameters estimate the requested generated by a Smart City with the size of the Milan urban area.

7.4.3 Evaluation of the Runtime Learner

This experiment aims at evaluating the benefits of learning the application knowledge at runtime. In the context of the early stages of the drug discovery process, we focus

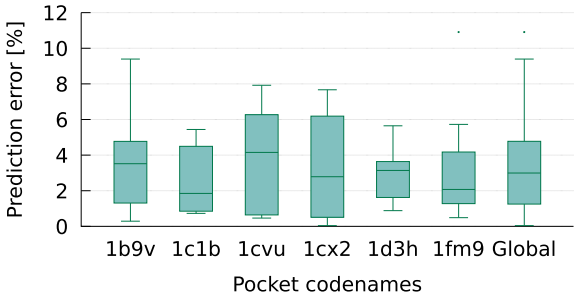


Fig. 7.4 Distribution of the prediction error on the time-to-solution, by varying the target pocket

on a molecular docking application for virtual screening [31]. It takes as input two information. On one hand, the docking site of the target molecule, named *pocket*. On the other hand, a huge library of possible solutions, named *ligands*. The output of the application is a small set of ligands which may have a strong interaction with the target pocket, to forward to the later stages of the process. The application exposes application-specific software-knobs that expose an accuracy-throughput trade-off.

Due to the complexity of estimating a ligand-pocket interaction, and due to the embarrassingly parallel nature of the problem, this application is a perfect match for HPC computation. In this scenario, the application designer would like to maximize the quality of the elaboration, given an upper bound on the time-to-solution. The relationship between the throughput and the software-knobs configuration depends on the characteristic of the actual input, especially of the pocket. Therefore, in this experiment we use *mARGOt* to learn at runtime such relationship, to be exploited in the remainder of the production run.

Figure 7.4 shows the distributions of the prediction error on the time-to-solution with six pocket from the RCSB Protein Databank (PDB) [32]. We use a chemical library with heterogeneous ligands [9]. For example, the number of their atoms range from 28 to 153. Since the prediction error is limited ($<10\%$), *mARGOt* is able to improve the computation efficiency.

7.5 Conclusion

This chapter focuses on a methodology to enhance the target application with an adaptation layer, based on the application-knowledge, that provides the most suitable software-knobs configuration according to the application requirements.

We assessed the benefits of *mARGOt* in three case studies that belong to different application domains. Experimental results show how it is possible to improve drastically the computation efficiency by adapting reactively and proactively. Moreover, it is possible to learn the relationship between EFPs, software-knobs, and input features using the input of the production run, identifying and seizing optimization opportunities.

References

1. Duranton M, De Bosschere K, Coppens B, Gamrat C, Gray M, Munk H, Ozer E, Varganega T, Zendra O (2019) Hipeac vision 2018
2. Sasa M, Stelios S, Henry H, Martin R (2010) Quality of service profiling. In: Proceedings of the 32nd ACM/IEEE international conference on software engineering, vol 1. ACM, pp 25–34
3. Hadi E, Adrian S, Luis C, Doug B (2012) Neural acceleration for general-purpose approximate programs. In: Proceedings of the 2012 45th annual IEEE/ACM international symposium on microarchitecture. IEEE Computer Society, pp 449–460
4. Henry H, Sasa M, Stelios S, Anant A, Martin R (2009) Using code perforation to improve performance, reduce energy consumption, and respond to failures
5. Edoardo P, Davide G, Gianluca P, Vittorio Z, Cristina S (2014) Evaluating orthogonality between application auto-tuning and run-time resource management for adaptive openCL applications. In: Application-specific Systems, architectures and processors (ASAP). IEEE, pp 161–168
6. Henry H, Stelios S, Michael C, Sasa M, Anant A, Martin R (2011) Dynamic knobs for responsive power-aware computing. In: ACM SIGPLAN notices, vol 46. ACM, pp 199–212
7. Jeffrey O Kephart and David M Chess (2003) The vision of autonomic computing. Computer 36(1):41–50
8. Bergstra J, Pinto N, Cox D (2012) Machine learning for predictive auto-tuning with boosted regression trees. In: 2012 innovative parallel computing (InPar), pp 1–9, May 2012
9. Gadioli D, Vitali E, Palermo G, Silvano C (2018) Margot: a dynamic autotuning framework for self-aware approximate computing. IEEE transactions on computers
10. Emanuele V, Davide G, Gianluca P, Martin G, João B, Pedro P, Jan M, Kateřina S, João MPC, Cristina S (2019) An efficient monte carlo-based probabilistic time-dependent routing calculation targeting a server-side car navigation system. IEEE transactions on emerging topics in computing
11. Markus CH, Julie AMcC (2008) A survey of autonomic computing—degrees, models, and applications. ACM Comput Surv (CSUR) 40(3):7
12. Sara M-H, Vinicius HSD, Danny W, Paris A (2017) A systematic literature review on methods that handle multiple quality attributes in architecture-based self-adaptive systems. Inf Softw Technol 90:1–26
13. Renato M, Gilles C, Anna S, Eduardo C, Michael G, Houssam H, Carmen N, Siegfried B, Martin S, Laurent M et al (2012) Autotune: a plugin-driven approach to the automatic tuning of parallel applications. In: International workshop on applied parallel computing. Springer, pp 328–342
14. Shajulin B, Ventsislav P, Michael G (2010) Periscope: an online-based distributed performance analysis tool. In: Tools for high performance computing 2009. Springer, pp 1–16
15. Misailovic S, Kim D, Rinard M (2013) Parallelizing sequential programs with statistical accuracy tests. ACM Trans Embed Comput Syst (TECS) 12(2s):88
16. Mehrzad S, Davoud AJ, Janghaeng L, Scott M (2014) Paraprox: pattern-based approximation for data parallel applications. ACM SIGPLAN Not 49(4):35–50
17. Jason A, Shoaib K, Kalyan V, Jonathan R-K, Jeffrey B, Una-May O, Saman A (2014) Opentuner: an extensible framework for program autotuning. In: 2014 23rd international conference on parallel architecture and compilation techniques (PACT). IEEE, pp 303–315
18. Ari R, Michael H, Sergei G. Atf: a generic auto-tuning framework. In IEEE 19th international conference on high performance computing and communications; IEEE 15th international conference on smart city; IEEE 3rd international conference on data science and systems (HPCC/SmartCity/DSS). IEEE, pp 64–71
19. Woongki B, Trishul MC (2010) Green: a framework for supporting energy-conscious programming using controlled approximation. In: ACM Sigplan Notices, vol 45. ACM, pp 198–209
20. Michael AL, Parker H, Mehrzad S, Scott M, Jason M, Lingjia T (2016) Input responsiveness: using canary inputs to dynamically steer approximation. ACM SIGPLAN Not 51(6):161–176

21. Xin S, Andrew L, Donald SF, Keshav P (2016) Proactive control of approximate programs. ACM SIGOPS Oper Syst Rev 50(2):607–621
22. Yufei D, Jason A, Kalyan V, Xipeng S, Una-May O, Saman A (2015) Autotuning algorithmic choice for input sensitivity. In: ACM SIGPLAN notices, vol 50. ACM, pp 379–390
23. Joshua SM, Natalie EJ (2016) The anytime automaton. In: ACM SIGARCH computer architecture news, vol 44. IEEE Press, pp 545–557
24. mARGOt framework git repository (2018). https://gitlab.com/margot_project/core
25. Vittorio Z, Gianluca P, Fabrizio C, Cristina S, Giovanni M (2010) Multicube explorer: an open source framework for design space exploration of chip multi-processors. In: 23th international conference on architecture of computing systems 2010. VDE, pp 1–7
26. Dan T, Heike J, Haihang Y, Jack D (2010) Collecting performance data with papi-c. In: Matthias SM, Michael MR, Alexander S, Wolfgang EN (eds) Tools for high performance computing 2009, Berlin, Heidelberg, 2010. Springer, Berlin, Heidelberg, pp 157–173
27. Benjamin CL, David MB (2006) Accurate and efficient regression modeling for microarchitectural performance and power prediction. In: ACM SIGOPS operating systems review, vol 40. ACM, pp 185–194
28. Zhenghua N, Jeffrey SR (2012) The crs package: nonparametric regression splines for continuous and categorical predictors. R J 4(2)
29. Davide G, Gianluca P, Cristina S (2015) Application autotuning to support runtime adaptivity in multicore architectures. In: 2015 international conference on embedded computer systems: architectures, modeling, and simulation (SAMOS). IEEE, pp 173–180
30. Radek T, Lukáš R, Jan M, Kateřina S, Ivo V (2015) Probabilistic time-dependent travel time computation using monte carlo simulation. In: International conference on high performance computing in science and engineering. Springer, pp 161–170
31. Claudia B, Andrea RB, Carlo C, Simone L, Gabriele C (2013) Use of experimental design to optimize docking performance: the case of ligendock, the docking module of ligen, a new de novo design program
32. Helen MB, John W, Zukang F, Gary G, Bhat TN, Helge W, Ilya NS, Philip EB (2000) The protein data bank. Nucleic Acids Res 28:235–242

Chapter 8
CAOS: CAD as an Adaptive Open-Platform Service for High Performance Reconfigurable Systems

Marco Rabozzi

Abstract The increasing demand for computing power in fields such as genomics, image processing and machine learning is pushing towards hardware specialization and heterogeneous systems in order to keep up with the required performance level at sustainable power consumption. Among the available solutions, Field Programmable Gate Arrays (FPGAs), thanks to their advancements, currently represent a very promising candidate, offering a compelling trade-off between efficiency and flexibility. Despite the potential benefits of reconfigurable hardware, one of the main limiting factor to the widespread adoption of FPGAs is complexity in programmability, as well as the effort required to port software solutions to efficient hardware-software implementations. In this chapter, we present CAD as an Adaptive Open-platform Service (CAOS), a platform to guide the application developer in the implementation of efficient hardware-software solutions for high performance reconfigurable systems. The platform assists the designer from the high-level analysis of the code, towards the optimization and implementation of the functionalities to be accelerated on the reconfigurable nodes. Finally, CAOS is designed to facilitate the integration of external contributions and to foster research on Computer Aided Design (CAD) tools for accelerating software applications on FPGA-based systems.

8.1 Introduction

Over the last 40 years, software performance has benefited from the exponential improvement of General Purpose Processors (GPPs) that resulted from a combination of architectural and technological enhancements. Despite such achievements, the performance measured on standard benchmarks in the last 3 years only improved

M. Rabozzi (✉)
Dipartimento di elettronica, informazione e bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: marco.rabozzi@polimi.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_8

at a rate of about 3% per year [11]. Indeed, after the failure of Dennard scaling [21], the current diminishing performance improvements of GPP reside in the difficulty to efficiently extract more fine-grained and coarse-grained parallelism from software. Considering the shortcomings of GPP, in current years we are assisting at a new era of computer architectures in which the need for energy-efficiency is pushing towards hardware specialization and the adoption of heterogeneous systems. This trend is also reflected in the High Performance Computing (HPC) domain that, in order to sustain the ever-increasing demand for performance and energy efficiency, started to embrace heterogeneity and to consider hardware accelerators such as Graphics Processing Units (GPUs), FPGAs and dedicated Application-Specific Integrated Circuits (ASICs) along with standard CPU. Albeit ASICs show the best performance and energy efficiency figure, they are not cost-effective solutions due to the diverse and ever-evolving HPC workloads and the high complexity of their development and deployment, especially for HPC. Among the available solutions, FPGAs, thanks to their advancements, currently represent the most promising candidate, offering a compelling trade-off between efficiency and flexibility. Indeed, FPGAs are becoming a valid HPC alternative to GPUs, as they provide very high computational performance with superior energy efficiency by employing customized datapaths and thanks to hardware specialization. FPGA devices have also attained renewed interests in recent years as hardware accelerators within the cloud domain. The possibility to access FPGAs as on-demand resources is a key step towards the democratization of the technology and to expose them to a wide range of potential domains [2, 6, 24].

Despite the benefits of embracing reconfigurable hardware in both the HPC and cloud contexts, we notice that one of the main limiting factor to the widespread adoption of FPGAs is complexity in programmability as well as the effort required to port a pure software solution to an efficient hardware-software implementation targeting reconfigurable heterogeneous systems [1]. During the past decade, we have seen significant progress in High-Level Synthesis (HLS) tools which partially mitigate this issue by allowing to translate functions written in a high-level language such as C/C++ to a hardware description language suitable for hardware synthesis. Nevertheless, current tools still require experienced users in order to achieve efficient implementations. In most cases indeed, the proposed workflows require the user to learn the usage of specific optimization directives [22], code rewriting techniques and, in other cases, to master domain specific languages [10, 13]. In addition to this, most of the available solutions [9, 10, 13] focus on the acceleration of specific kernel functions and leave to the user the responsibility to explore hardware/software partitioning as well as to identify the most time-consuming functions which might benefit the most from hardware acceleration. To tackle these challenges, we propose the CAOS platform bringing the following contributions:

- A comprehensive design flow guiding the designer from the initial software to the final implementation to a high performance FPGA-based system.
- Well-defined Application Programming Interfaces (APIs) and an infrastructure allowing researchers to integrate and test their own modules within CAOS.

- A general method for translating high-level functions into FPGA-accelerated kernels by matching software functions to appropriate architectural templates.
- Support for three different architectural templates allowing to target software functions with different characteristics within CAOS.

Section 8.2 describes the overall CAOS platform, its design flow and infrastructure, while Sect. 8.3 presents an overview of the supported architectural. Section 8.4 discuss the experimental results on case studies targeting different architectural templates. Finally, Sect. 8.5 draws the conclusions.

8.2 The CAOS Platform

The CAOS platform has been developed in the context of the Exploiting eXascale Technology with Reconfigurable Architectures (EXTRA) project and shares with it the same vision [19]. CAOS targets both application developers and researches while its design has been conceived focusing on three key principles: usability, interactivity and modularity. From a usability perspective, the platform supports application designers with low expertise on reconfigurable heterogeneous systems in quickly optimizing their code, analyzing the potential performance gain and deploying the resulting application on the target reconfigurable architecture. Nevertheless, the platform does not aim to perform the analysis and optimizations fully automatically, but instead interactively guides the users towards the design flow, providing suggestion and error reports at each stage of the process. Finally, CAOS is composed of a set of independent modules accessed by the CAOS flow manager that orchestrates the execution of the modules according to the current stage of the design flow. Each module is required to implement a set of well-defined APIs so that external researchers can easily integrate their implementations and compare them against the ones already offered by CAOS.

8.2.1 CAOS Design Flow

The platform expects the application designer to provide the application code written in a high-level language such as C/C++, one or multiple datasets to be used for code profiling and a description of the target reconfigurable system. In order to narrow down and simplify the set of possible optimizations and analysis that can be performed on a specific algorithm, CAOS allows the user to accelerate its application using one of the available architectural templates. An *architectural template* is a characterization of the accelerator both in terms of its computational model and the communication with the off-chip memory. As a consequence, an architectural template constrains the architecture to be implemented on the reconfigurable hardware and poses restrictions on the application code that can be accelerated, so that the

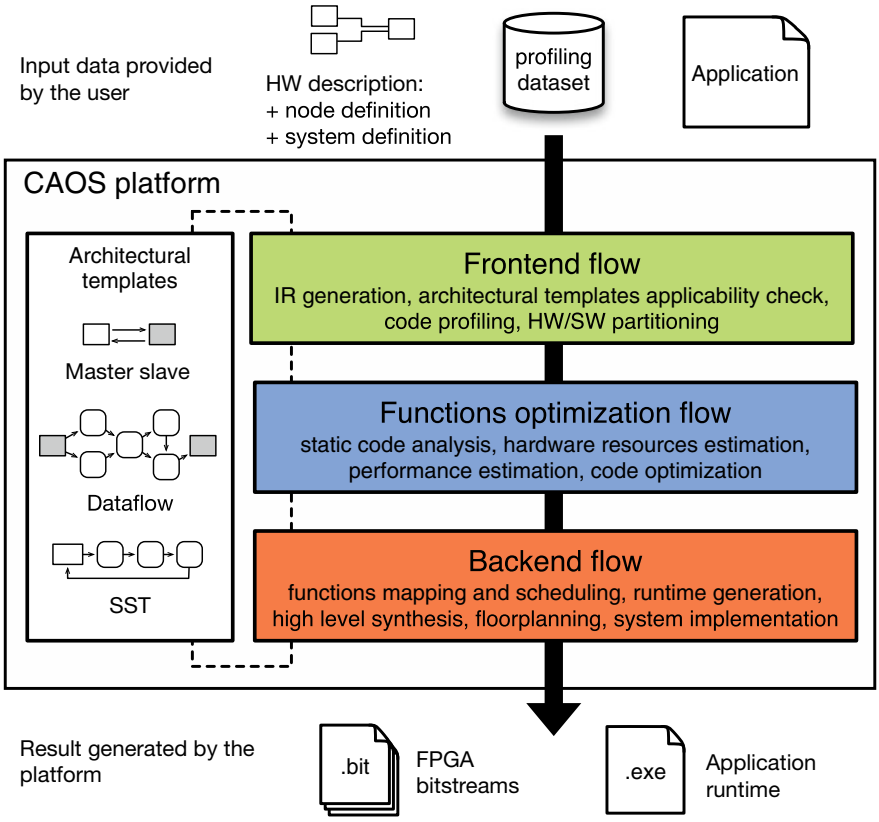


Fig. 8.1 High-level overview of the CAOS platform. The overall design flow can be divided in three main parts: the *frontend*, the *functions optimization* and the *backend* flow. The application code, datasets to profile the application and an HW description constitute the input data provided by the designer. The final outputs generated by the platform are the bitstreams, that the user can use to configure the FPGAs, and the application runtime, needed to run the optimized version of the application

number and types of optimizations available can be tailored for a specific type of implementation. Furthermore, CAOS is meant to be orthogonal and build on top of tools that perform High-Level Synthesis (HLS), place and route and bitstream generation. Code transformations and optimizations are performed at the source code level while each architectural template has its own requirements in terms of High-Level Synthesis (HLS) and hardware synthesis tools to use.

As shown in Fig. 8.1, the overall CAOS design flow is subdivided into three main flows: the frontend flow, the function optimization flow and the backend flow. The main goal of the frontend is to analyze the application provided by the user, match the application against one or more architectural templates available within the platform, profile the user application against the user specified datasets and, finally, guide

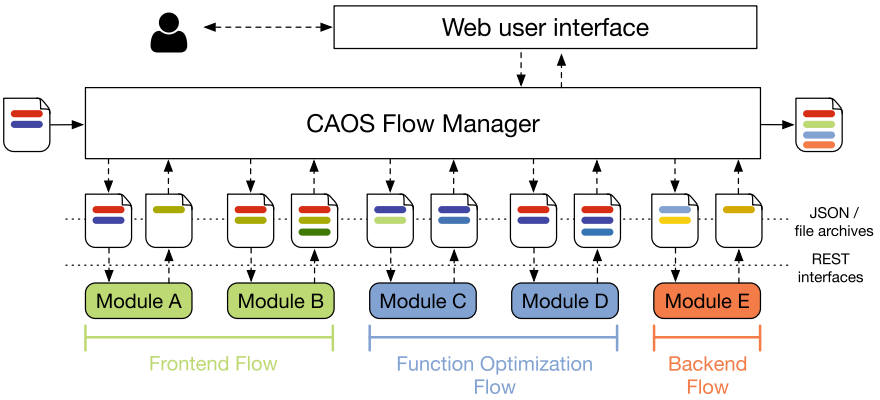


Fig. 8.2 CAOS infrastructure in terms of its main components and their interaction

the user through the hardware/software partitioning of the application to define the functions of the application that should be implemented on the reconfigurable hardware. The function optimization flow performs static analysis and hardware resource estimation of the functionalities to be accelerated on the FPGA. Such analyses are dependent upon the considered architectural template and the derived information are used to estimate the performance of the hardware functions and to derive the optimizations to apply (such as loop pipelining, loop tiling and loop unrolling). After one or more iterations of the function optimization flow, the resulting functions are given to the backend flow in which the desired architectural template for implementing the system is selected and the required High-Level Synthesis (HLS) and hardware synthesis tools are leveraged to generate the final FPGA bitstreams. Within the backend, CAOS takes care of generating the host code for running the FPGA accelerators and optionally guides the place and route tools by floorplanning the system components.

8.2.2 CAOS Infrastructure

In order to simplify the adoption of the platform while being open to contributions, the CAOS infrastructure is organized as a microservices architecture which can be accessed by application designers through a web user interface. The infrastructure, shown in Fig. 8.2, leverages on Docker [8] application containers to isolate the modules and to provide scalability. Each module is deployed in a single container, and several implementations of the same module can coexist to provide different functionalities to different users. Moreover, each module's implementation can be replicated to scale horizontally depending on system load. The modules are connected together and driven by the *CAOS Flow manager* which serves the User Interface (UI) and provides the glue logic that routes each request to the proper module.

The interaction between the flow manager and the CAOS modules is performed by means of data transfer objects defined with a JSON Domain Specific Language (DSL). The user can specify at each phase the desired module implementation and the platform will take care of routing the request to the proper module automatically. Moreover, the platform supports modules deployed remotely by simply specifying their IP address. Another advantage of the proposed infrastructure is that, thanks to Docker containers, it can also be easily deployed on cloud instances, possibly featuring FPGA boards, such as the Amazon EC2 F1 instances. This allows a complete design process in the cloud in which the user can optimize the application through a web UI, while the final result of the CAOS design flow can be directly tested and run on the same cloud instance. CAOS supports the integration of new implementations of the modules described in Sect. 8.2.1. Researchers are free to adopt the preferred tools and programming languages that fit their needs, as long as the module provides the implementation of the REST APIs prescribed by the CAOS flow manager.

8.3 Architectural Templates

The core idea for devising efficient FPGA-based implementations in CAOS revolves around matching software functions to an architectural template suitable for its acceleration. CAOS currently supports three architectural templates: Master/Slave, Dataflow and Streaming architectural templates. In the next sections, we provide an overview of the templates, describing the supported software functions and hardware platforms, the proposed optimizations and the tools on which the templates rely.

8.3.1 Master/Slave Architectural Template

The Master/Slave architectural template [7] targets systems with a shared Double Data Rate (DDR) memory that can be accessed both by the host running the software portion of the application and by the FPGA devices on which we implement the accelerated functionalities (also referred as kernel). The template also requires that the communication between the accelerator and the DDR memory is performed via memory mapped interfaces. Such requirements allow to standardize the data transfer as well as to support random memory accesses to pointer arguments of the target C/C++ function. Currently, the template supports two target systems: Amazon F1 instances in the cloud and Xilinx Zynq System-on-Chips (SoCs).

The generality of the communication model of the Master/Slave architectural template allows supporting a wide range of C/C++ functions. In particular, the function has to abide to quite general constraints for High-Level Synthesis (HLS) such as no dynamic creation of objects/arrays, no syscalls and no recursive function calls. The Master/Slave architectural template currently leverages on Vivado HLS [23]

for the High-Level Synthesis (HLS) of C/C++ code to Hardware Description Language (HDL). Hence, in the CAOS frontend, we verify the applicability of the template to a given function by running Vivado HLS on it and verify that no errors are generated. In addition to the Vivado HLS constraints, we also require the size of the function arguments to be known. This is needed by CAOS to properly estimate the kernel performance throughout the function optimizations flow. During the CAOS functions optimization flow, the template performs static code analysis by leveraging on custom Low-Level Virtual Machine (LLVM) [12] passes. In particular, it identifies loop nests with their trip counts, local arrays and information on the input and output function arguments. Furthermore, the template collects hardware and performance estimations of the current version of the target function directly from Vivado HLS. Such information is then used to identify the next candidate optimizations among loop pipelining, loop unrolling, on-chip caching and memory partitioning. The user can then either select the suggested optimization or conclude the optimization flow if he/she is satisfied with the estimated performance. After having optimized the kernel, the design proceeds to the CAOS backend flow. Here the optimized C/C++ function is translated to HDL and, according to the target system, the template leverages either on the Xilinx SDAccel toolchain [22] or Xilinx Vivado [23] for the implementation and bitstream generation. In both cases, CAOS takes care of modifying the original application and inserts the necessary code and APIs calls to offload the computation of the original software function to the generated FPGA accelerator.

8.3.2 Dataflow Architectural Template

The dataflow architectural template trades off the generality of codes supported by the Master/Slave architectural template in order to achieve higher performance. In a dataflow computing model, the data is streamed from the memory directly to the chip containing an array of Processing Elements (PEs) each of which is responsible for a single operation of the algorithm. The data flow from a PE to the next one in a statically defined directed graph, without the need for any kind of control mechanism. In such a model, each PE performs its operation as soon as the input data is available and forwards the result to the next element in the network as soon as it is computed.

The target system for the dataflow architectural template consists in a host CPU and the dataflow accelerator deployed on a FPGA connected via PCIe to the host. Both the host CPU and the FPGA logic have access to the host DDR memory containing the input/output data. The FPGA accelerator is organized internally as a Globally Asynchronous Locally Synchronous (GALS) architecture divided into the actual accelerated kernel function and a manager. The manager handles the asynchronous communication between the host and the accelerator, whereas the kernel is internally organized as a set of synchronous PEs that perform the computation in parallel.

The architectural template leverages on the OXiGen toolchain [18] and its extension [17] to translate C/C++ functions into optimized dataflow kernels defined with the MaxJ language. In order to efficiently perform the translation from sequential C/C++ code to a dataflow representation, the target function has to satisfy certain requirements detailed in [18]. An exemplary code supported by the template is shown in Listing 8.1. The code requirements are validated in the CAOS frontend flow in order to identify functions that can be optimized with the dataflow architectural template. Within the CAOS function optimization flow, OXiGen performs the dataflow graph construction directly from the LLVM Intermediate Representation (IR) of the target function. Nevertheless, the initial dataflow design might either exceed or underutilize the available FPGA resources. Hence, in order to identify an optimal implementation that fits within the FPGA resources and available data transfer bandwidth, OXiGen supports loop rerolling, to reduce resource consumption, and vectorization, to replicate the processing elements in order to fully utilize the available bandwidth and compute resources. In order to derive the best implementation, OXiGen relies on resource and performance models and runs a design space exploration using an approach based on Mixed-Integer Linear Programming (MILP). Once having generated an optimized kernel, the CAOS backend runs MaxCompiler [13] to synthesize the MaxJ code generated by OXiGen to a Maxeler DFE (Dataflow Engine) that can be accessed by the host system.

Listing 8.1 An exemplary code supported by the dataflow template. The function takes as input a combination of array types and scalar types. The outer loops iterate over the outer dimension of the array types which are translated as streams. Accesses to the streams are linear with constant offsets. The function can have a combination of nesting levels iterating over the inputs or local variables.

```
void foo(type_1* in_1, type_1* in_2, type_2* out_1, int iter) {
    type_1 tmp_vect [15];
    for(int i = const_1; i < iter - const_2; i++) {

        ...statements...

        for(int j = const_3; j < 15; j++)
            tmp_vect[j] = ...expression... ;

        type_1 tmp_scalar = ...expression... ;

        for(int j = const_3; j < 15; j++)
            tmp_scalar = tmp_scalar + tmp_vect[j];
    }
}
```

8.3.3 Streaming Architectural Template

Iterative Stencil Loops (ISLs) represent a class of algorithms that are highly recurrent in many HPC applications such as differential equation solving or scientific simulations. The basic structure of ISLs is depicted in Algorithm 8.1; the outer loop iterates for a given number of times, so-called *time-steps*, while, at each time-step, the inner

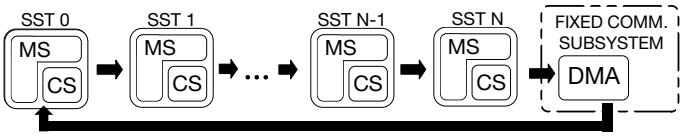


Fig. 8.3 Architecture of an SST-based accelerator for ISL

loop updates each value of the n-dimensional input vector by means of the *stencil* function, computing a weighted sum of the neighbor values in the vector.

Algorithm 8.1 Generic ISL Algorithm.

```
for t ≤ TimeSteps do
  for all points p in matrix M do
    p ← stencil(p)
```

The Streaming architectural template specifically targets stencil codes written in C and leverages the SST architecture proposed by [3] for its implementation. The architectural template of the SST-based accelerator [14] is depicted in Fig. 8.3. The basic SST module performs the computation of a single time-step and is conceptually separated in a *memory system*, responsible for storage and data movement, and a *computing system*, that performs the actual computation. The SST module is designed in order to operate in a streaming mode on the various elements of the input vector; this internal structure is derived by means of the *polyhedral analysis* that allows refactoring the algorithm to optimize on-chip memory resource consumption and implement a dataflow model of computation. Then, the complete SST-based architecture is obtained by replicating N times the basic module to implement a pipeline, where each module computes in streaming a single time-step of the outer loop of the algorithm. Such a pipeline is finally connected with a fixed communication subsystem interfacing with the host machine. Within this context, CAOS offers a design exploration algorithm [20] that jointly maximizes the number of SST processors that can be instantiated on the target FPGA and identifies a floorplan of the design that minimizes the inter-component wire-length in order to allow implementing the system at high frequency. In the frontend, CAOS identifies those functions having the ISL structure shown in Algorithm 8.1, while the CAOS function optimization flow runs the design space exploration algorithm detailed in [20] on the function to accelerate. The approach starts by generating an initial version of the system consisting of a single SST [14] to obtain an initial resource estimate. Subsequently, it solves a maximum independent set problem formulated as an Integer Linear Programming (ILP) to identify the maximum number of SST as well as their floorplan and solves a Traveling Salesman Problem (TSP) to identify the best routing among SSTs. Finally, the CAOS backend generates the accelerator bitstream through Xilinx Vivado while enforcing the identified floorplanning constraints.

8.4 Experimental Results

Within this section we discuss the results achieved by CAOS on different case studies targeting the discussed architectural templates. The first case study we consider is the N-Body simulation, which is a well known problem in physics having applications in a wide range of fields. In particular, we focused on the *all-pair* method: the algorithm alternates a computationally intensive force computation step, in which the pairwise forces between each pair of bodies are computed and a position update step, that updates the velocities and positions of the bodies according to the current resulting forces. CAOS, after profiling and analyzing the application in the frontend, properly identified the force computation as the target function to accelerate and matched it to the Master/Slave architectural template. As we can see from Table 8.1, the CAOS implementation targeting an Amazon F1 instance greatly outperforms, both in terms of performance and energy efficiency, the software implementation from [5] running in parallel on 40 threads on an Intel Xeon E5-2680 v2 and the implementation from [4] on a Xilinx VC707 board. Nevertheless, the bespoke implementation from [5] targeting the same hardware provides 11% higher performance than the CAOS one. However, the CAOS design was achieved semi-automatically in approximately a day of work, while the design from [5] required several weeks of manual effort.

As a second test case, we consider the Curran approximation algorithm [16] for calculating the pricing of Asian options. More specifically, we tested two flavors of the algorithm using 30 and 780 averaging points. Within CAOS, we started from a C implementation and we targeted a host machine featuring an Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz connected via PCIe gen1 x8 to a MAX4 Galava board equipped with an Altera Stratix V FPGA. Since the algorithms operate on subsequent independent data items, the overall computations are easily expressed in C using the structure in Listing 8.1. Hence CAOS identified and optimized the computations leveraging on the dataflow architectural template. As the initial designs did not fit within the device, CAOS applied the rerolling optimization for both cases. As shown in Table 8.2, both CAOS implementations achieve speedups over 100x against the single thread execution on the host system only. Moreover, we compared

Table 8.1 Performance and energy efficiency of the all-pairs N-Body algorithm accelerated via CAOS and the results achieved by bespoke designs proposed in [4, 5]

Reference	Platform	Type	Frequency (MHz)	Performance (MPairs/s)	Performance/Power (MPairs/s/W)
[5]	Intel Xeon E5-2680 v2	CPU	–	2,642	22.98
[4]	Xilinx VC707	FPGA	100	2,327	116.36
[5]	Xilinx VU9P	FPGA	154	13,441	672.06
CAOS	Xilinx VU9P	FPGA	126	12,072	603.61

Table 8.2 Results achieved by CAOS on the Curran approximation algorithm with 30 and 780 averaging points compared against CPU and the bespoke FPGA-based implementations from [15]

Averaging points	Rerolling factor	Speedup w.r.t. CPU	Speedup w.r.t. [15]	Input bandwidth (MByte/s)	Output bandwidth (MByte/s)
30	4	118.4x	1.23x	1,767.64	196.40
780	98	101.0x	0.5x	75.11	8.35

Table 8.3 Results achieved by the CAOS streaming architectural template compared to [14]

Algorithm	#SSTs		Design frequency (MHz)		Performance improvement w.r.t. [14] (%)	Design time reduction w.r.t. [14]
	CAOS	[14]	CAOS	[14]		
Jacobi2D	90	88	228	206	13.20	15.84x
Heat3D	25	25	228	206	10.68	1.69x
Seidel2D	19	19	183	183	0	1.08x

the results against the DFE execution times reported from the designs in [15]. For the version with 30 averaging points, we achieved a speedup of 1.23x. For the version with 780 averaging points, our implementation shows a speed down of about 0.5x. Nevertheless, it was obtained in less than a day of work.

As a final test case, we evaluated the streaming architectural template on three representative ISL computations (Jacobi2D, Heat3D and Seidel2D) targeting a Xilinx Virtex XC7VX485T device [20]. Table 8.3 reports the performance improvement and the design time reduction compared to the methodology in [14]. Thanks to FPGA floorplanning we are able to increase the target frequency for the Jacobi2D and Heat3D algorithms of approximately 11%. Additionally, for the Jacobi2D case, the floorplanner is also able to allocate two additional SSTs improving the performance up to 13%. Nevertheless, the Seidel2D algorithm does not provide the same improvement figure. Indeed, since the total number of SSTs that can be placed into the design is small, the floorplanning reduces its impact on the overall design by leaving more room to the place and route algorithm. Regarding the design time, our approach allows to greatly reduce the number of trial synthesis required, thus leading to an execution time saving of 15.84x for Jacobi2D.

8.5 Conclusions

In this chapter we presented CAOS, a platform whose main objective is to improve productivity and simplify the design of FPGA-based accelerated systems, starting from pure high-level software implementations. Currently, the slowing rate at which

general purpose processors improve performance is strongly pushing towards specialized hardware. We expect FPGAs to have a more prominent role in the upcoming years as a technology to achieve efficient and high performance solutions both in the HPC and cloud domains. Hence, by embracing this idea, we designed CAOS in a modular fashion, providing well-defined APIs that allow external researchers to integrate extensions or different implementations of the modules within the platform. Indeed, the second, yet not less important, objective of CAOS, is to foster research on tools and methods for accelerating software on FPGA-based architectures.

References

1. Bacon DF, Rabbah R, Shukla S (2013) FPGA programming for the masses. Commun ACM 56(4):56–63
2. Cardamone S, Kimmitt JR, Burton HG, Thom AJ (2018) Field-programmable gate arrays and quantum Monte Carlo: power efficient co-processing for scalable high-performance computing. arXiv:1808.02402
3. Cattaneo R, Natale G, Sicignano C, Sciuto D, Santambrogio MD (2016) On how to accelerate iterative stencil loops: a scalable streaming-based approach. ACM Trans Archit Code Optim (TACO) 12(4):53
4. Del Sozzo E, Di Tucci L, Santambrogio MD (2017) A highly scalable and efficient parallel design of n-body simulation on FPGA. In: 2017 IEEE international parallel and distributed processing symposium workshops (IPDPSW), pp 241–246. IEEE
5. Del Sozzo E, Rabozzi M, Di Tucci L, Sciuto D, Santambrogio MD (2018) A scalable FPGA design for cloud n-body simulation. In: 2018 IEEE 29th international conference on application-specific systems, architectures and processors (ASAP), pp 1–8. IEEE
6. Di Tucci L, O'Brien K, Blott M, Santambrogio MD (2017) Architectural optimizations for high performance and energy efficient Smith-Waterman implementation on FPGAs using OpenCL. In: 2017 design, automation and test in Europe conference and exhibition (DATE), pp 716–721. IEEE
7. Di Tucci L, Rabozzi M, Stornaiuolo L, Santambrogio MD (2017) The role of CAD frameworks in heterogeneous FPGA-based cloud systems. In: 2017 IEEE international conference on computer design (ICCD), pp 423–426. IEEE
8. Docker. https://www.docker.com
9. Fort B, Canis A, Choi J, Calagar N, Lian R, Hadjis S, Chen YT, Hall M, Syrowik B, Czajkowski T et al (2014) Automating the design of processor/accelerator embedded systems with legup high-level synthesis. In: 2014 12th IEEE international conference on embedded and ubiquitous computing (EUC), pp 120–129. IEEE
10. Hegarty J, Brunhaver J, DeVito Z, Ragan-Kelley J, Cohen N, Bell S, Vasilyev A, Horowitz M, Hanrahan P (2014) Darkroom: compiling high-level image processing code into hardware pipelines. ACM Trans Graph 33(4):144:1–144:11. https://doi.org/10.1145/2601097.2601174
11. Hennessy JL, Patterson DA (2017) Computer architecture: a quantitative approach. Elsevier, Amsterdam
12. Lattner C (2008) LLVM and Clang: next generation compiler technology. In: The BSD conference, pp 1–2
13. Maxeler Technologies: MaxCompiler. https://www.maxeler.com
14. Natale G, Stramondo G, Bressana P, Cattaneo R, Sciuto D, Santambrogio MD (2016) A polyhedral model-based framework for dataflow implementation on FPGA devices of iterative stencil loops. In: Proceedings of the 35th international conference on computer-aided design, p 77. ACM

15. Nestorov AM, Reggiani E, Palikareva H, Burovskiy P, Becker T, Santambrogio MD (2017) A scalable dataflow implementation of Curran's approximation algorithm. In: 2017 IEEE international parallel and distributed processing symposium workshops (IPDPSW), pp 150–157. IEEE
16. Novikov A, Alexander S, Kordzakhia N, Ling T (2016) Pricing of Asian-type and basket options via upper and lower bounds. arXiv:1612.08767
17. Peverelli F, Rabozzi M, Cardamone S, Del Sozzo E, Thom AJ, Santambrogio MD, Di Tucci L (2019) Automated acceleration of dataflow-oriented c applications on FPGA-based systems. In: 2019 IEEE 27th annual international symposium on field-programmable custom computing machines (FCCM), pp 313–313. IEEE
18. Peverelli F, Rabozzi M, Del Sozzo E, Santambrogio MD (2018) OXiGen: a tool for automatic acceleration of c functions into dataflow FPGA-based kernels. In: 2018 IEEE international parallel and distributed processing symposium workshops (IPDPSW), pp 91–98. IEEE
19. Rabozzi M, Brondolin R, Natale G, Del Sozzo E, Huebner M, Brokalakis A, Ciobanu C, Stroobandt D, Santambrogio MD (2017) A CAD open platform for high performance reconfigurable systems in the extra project. In: 2017 IEEE computer society annual symposium on VLSI (ISVLSI), pp 368–373. IEEE
20. Rabozzi M, Natale G, Festa B, Miele A, Santambrogio MD (2017) Optimizing streaming stencil time-step designs via FPGA floorplanning. In: 2017 27th international conference on field programmable logic and applications (FPL), pp 1–4. IEEE
21. Taylor MB (2013) A landscape of the new dark silicon design regime. IEEE Micro 33(5):8–19
22. Wirbel L (2014) Xilinx SDAccel: a unified development environment for tomorrow's data center. Technical report, The Linley Group Inc.
23. Xilinx Inc.: Vivado design suite. http://www.xilinx.com/products/design-tools/vivado.html
24. Zhang C, Li P, Sun G, Guan Y, Xiao B, Cong J (2015) Optimizing FPGA-based accelerator design for deep convolutional neural networks. In: Proceedings of the 2015 ACM/SIGDA international symposium on field-programmable gate arrays, pp 161–170. ACM

Part IV
Systems and Control

Chapter 9
A General Framework for Shared Control in Robot Teleoperation with Force and Visual Feedback

Davide Nicolis

Abstract In the last decade, the topic of human robot interaction has received increasing interest from research and industry, as robots must now interface with human users to accomplish complex tasks. In this scenario, robotics engineers are required to take the human component into account in the robot design and control. This is especially true in telerobotics, where interaction with the user plays an important role in the controlled system stability. By means of a thorough analysis and practical experiments, this contribution aims at giving a concrete idea of the aspects that need to be considered in the design of a complete control framework for teleoperated systems, that are able to seamlessly integrate with a human operator through shared control.

9.1 Introduction

Although many branches of robotics are currently striving for complete autonomy, such as self-driving cars and drones, a parallel research line is instead trying to tie together robots and humans in a integrated way. Indeed, on the one hand, robots are often called to work together with humans as they are inherently designed to interact with their environment, on the other hand, the purpose of robots is that of being useful to its users. Therefore "closing the loop" around this human component is paramount for an optimal system design.

In certain applications demanding high-level reasoning, human presence is not only advised, but it is also mandatory. Nonetheless, the harshness and inaccessibility of the environments where they are called to operate makes the use of remotely controlled robotic systems, essential to complete the required tasks. With its first modern applications in the '50s in the field of nuclear material handling, robot teleoperation is nowadays commonly associated by the general public with surgical robots [19].

D. Nicolis (✉)
Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy
e-mail: davide.nicolis@polimi.it

B. Pernici (ed.), *Special Topics in Information Technology*, PoliMI SpringerBriefs,
https://doi.org/10.1007/978-3-030-32094-2_9

Even so, its use has spread to multiple domains, and it has sought the interest not only of the research community, but also industries. In the frame of the H2020 *RoMaNS* project, innovative teleoperation interfaces and visual servoing are proposed for the sorting of radioactive material [2], while the *WALK-MAN* project has aimed at building a humanoid teleoperated platform for inspection in dangerous unstructured environments damaged by natural events [11].

Massive attention is being given to telerobotics by industries producing pharmaceuticals, to handle toxic substances and customize drug production [18]. Clean room applications of teleoperated robots are now being investigated by ESA and NASA, in the scope of the Mars Sample Return exploration campaign, which aims at bringing back Mars samples for the analysis of life traces. Whereas classical glovebox systems are inadequate for planetary protection protocols, a remotely controlled robot can solve this problem by ensuring a high degree of isolation between the robotized cell and the external environment [21]. Nevertheless, space telerobotics is already a reality for in-orbit servicing, and at an advanced research stage in terms of exploration, with both directly teleoperated systems, such as the ones used in the *Kontur* project, and embedding the concept of supervised autonomy in the scope of *METERON* [17], with experiments already conducted from aboard the International Space Station.

9.1.1 Research Challenges

Historically, research on bilateral teleoperation has considered the closed loop system stability in presence of communication delays and with uncertain dynamics of the environment and the user [8]. The foundations of modern teleoperation were outlined by Anderson and Spong [4] with the scattering approach, highlighting the trade-off between stability and transparency, i.e. the system displayable impedance range. More recent approaches have shifted towards less conservative techniques such as time-domain passivity. However, to ensure practical stability, some feeling of presence at the remote location has to be sacrificed, degrading the user perception of the remote environment [10].

In the unstructured environments where teleoperation is required, the robust control of master and slave robots interaction dynamics are paramount, in order to ensure the accurate execution of the desired task and its stability. The problem of impedance control in teleoperation has been investigated in [15], which proposed tuning guidelines for the intercontinental control of a humanoid robot subject to communication delays.

The development of interfaces and control algorithms for kinesthetic, tactile, and visual feedback in teleoperation has found fertile grounds in research, for applications where it is critical that the user has a strong sense of presence, while exploiting these cues to improve the system usability. In [7] classical visual servoing aids a physician

in the execution of a tele-ecography, by exploiting image features to automatically turn the robot tool while optimizing the ultrasound image quality. Overall, these topics outline a research interest that is increasingly focusing on the interaction between user and control system in the form of shared autonomy architectures. Splitting the burden of operation between user and robot becomes the main design objective, in order to reduce operator fatigue and offer an intuitive experience when using the platform for tasks that are physically and cognitively demanding [3]. Shared human-robot controllers can be the answer for an efficient and seamless human-robot cooperation, with estimation and prediction of user behavior being able to further improve the user experience.

The present paper aims at briefly illustrating a comprehensive control framework for teleoperated robot systems, from the low level motion control aspects, interaction control, and system stability analysis with human-in-the-loop, to the higher level visual-aided control algorithms reducing the operator cognitive load. In Sect. 9.2, a robust controller exploiting sliding mode control techniques is proposed for redundant robot manipulators to arbitrarily and reliably assign the robot impedance in the task space and track the user input in a teleoperation scenario. The enclosing optimization-based high level model predictive controller ensures robustness of the sliding mode to actuation delays and unmodeled system dynamics. System stability is proved using absolute stability criteria, with guidelines to tune the controller parameters. In Sect. 9.3, this architecture is exploited to allow the straightforward definition of virtual fixtures and their rendering to the operator via force feedback. Finally, in Sect. 9.4, the framework is extended to include visual feedback from a camera mounted on a dual-arm slave robot, and visual servoing is employed to improve user interaction with the system, by introducing some autonomy into the slave robot in terms of camera control. Section 9.5 concludes this brief.

9.2 Robust Impedance Shaping of Redundant Teleoperators

At a local control level, robust control theory is used to guarantee a desired dynamic behavior of the robot during interaction. Sliding mode control [20] robustly shapes master and slave manipulators impedances, irrespectively of uncertainties. This formulation is addressed by directly specifying the task sliding manifold, while the manipulator redundancy is considered by successive null-space projections of sliding manifolds defined by lower priority tasks.

A hierarchical optimization outer layer takes into account individual control and motion constraints, such as torque or joint limits, while its model predictive nature also guarantees robust compensation of actuation delays and unmodeled input filter dynamics. The overall stability analysis in presence of variable communication delays during contact is performed thanks to Llewellyn's absolute stability criterion.

9.2.1 Model Predictive Sliding Mode Control

The generic manipulator dynamics can be represented by the following model

$$\boldsymbol{B}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{n}(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \boldsymbol{\tau} - \boldsymbol{J}(\boldsymbol{q})^T \boldsymbol{F}_e \tag{9.1}$$

where $\boldsymbol{q},\ \dot{\boldsymbol{q}},\ \ddot{\boldsymbol{q}} \in \mathbb{R}^n$ are the robot joint positions, velocities and accelerations respectively, $\boldsymbol{B}(\boldsymbol{q}) \in \mathbb{R}^{n\times n}$ is the symmetric positive definite inertia matrix, $\boldsymbol{n}(\boldsymbol{q}, \dot{\boldsymbol{q}}) \in \mathbb{R}^n$ is a vector comprising Coriolis, gravitational, and friction terms, $\boldsymbol{\tau} \in \mathbb{R}^n$ is the actuation torque, $\boldsymbol{F}_e \in \mathbb{R}^m$ is the external generalized force due to interaction with the environment, and finally $\boldsymbol{J}(\boldsymbol{q}) \in \mathbb{R}^{m\times n}$ is the robot end effector Jacobian, with $m \leq n$. The objective of impedance control is to change the displayed end effector dynamics according to the following equation

$$\boldsymbol{M}\ddot{\tilde{\boldsymbol{x}}} + \boldsymbol{D}\dot{\tilde{\boldsymbol{x}}} + \boldsymbol{K}\tilde{\boldsymbol{x}} = \hat{\boldsymbol{F}}_e \tag{9.2}$$

where $\boldsymbol{M}, \boldsymbol{D}, \boldsymbol{K} > 0 \in \mathbb{R}^{m\times m}$ are the desired inertia, damping and stiffness matrices respectively, while $\tilde{\boldsymbol{x}} = \boldsymbol{x}_r - \boldsymbol{x} \in \mathbb{R}^m$ is the end effector tracking error with $\boldsymbol{x}_r$ being the reference trajectory, and $\hat{(\cdot)}$ indicating estimated quantities.

Unfortunately, the simple application of a nominal control law [9] results in a system that is heavily affected by uncertainties and disturbances (the second term on the right-hand side of (9.3)), and that is not able to satisfy the dynamics requirement.

$$\boldsymbol{M}\ddot{\tilde{\boldsymbol{x}}} + \boldsymbol{D}\dot{\tilde{\boldsymbol{x}}} + \boldsymbol{K}\tilde{\boldsymbol{x}} = \hat{\boldsymbol{F}}_e - \boldsymbol{M}\boldsymbol{J}\hat{\boldsymbol{B}}^{-1}(\tilde{\boldsymbol{B}}\ddot{\boldsymbol{q}} + \boldsymbol{\eta}) \tag{9.3}$$

Therefore, the following Second Order sliding mode control law [6] is proposed in order to compensate the uncertainties, while maintaining a negligible control chattering

$$\boldsymbol{v} = \boldsymbol{v}_0 + \boldsymbol{v}_{smc} = \boldsymbol{v}_0 + k_1 \frac{\boldsymbol{\Sigma}_q}{\sqrt{\|\boldsymbol{\Sigma}_q\|}} + k_2 \int_{t_0}^{t} \frac{\boldsymbol{\Sigma}_q}{\|\boldsymbol{\Sigma}_q\|} d\tau \tag{9.4}$$

where $\boldsymbol{v}$ is the auxiliary control to be selected after applying a feedback linearizing controller to (9.1), $\boldsymbol{v}_0$ the nominal control used to obtain the desired impedance, and $\boldsymbol{\Sigma}_q$ the so-called sliding variable. By appropriately choosing $\boldsymbol{\Sigma}_q$, we can prove the following

Theorem 9.1 *Consider the partially feedback linearized version of the robot dynamics (9.1) and the control (9.4) with sufficiently high gains. Let*

$$\boldsymbol{\Sigma}_q(t) = \dot{\boldsymbol{q}}(t) - \dot{\boldsymbol{q}}_0(t) \tag{9.5}$$

with $\dot{\boldsymbol{q}}_0(t) = \int_{t_0}^{t} \boldsymbol{v}_0 d\tau + \dot{\boldsymbol{q}}(t_0)$. On the sliding manifold $\boldsymbol{\Sigma}_q = \boldsymbol{0}$, the system evolves with dynamics free of disturbances and uncertainties. Moreover, this holds beginning from the initial time instant t_0 (integral sliding mode).

Therefore, thanks to (9.4), we can completely disregard the uncertainties, and select the nominal control $\boldsymbol{v}_0$ to ensure the desired end effector impedance as in the nominal case. For redundant manipulators ($m < n$), however, other tasks can be accomplished simultaneously. Hence, we enclose the sliding mode controller with a hierarchical formulation of model predictive control:

- To enforce the desired task hierarchy
- To consider unilateral motion and torque constraints
- To compensate actuation delays and unmodeled input dynamics

The proposed controller results in a cascade of optimizations starting from the higher priority one, enforcing the impedance (9.6a), to the tasks with lower priority (9.6b)

$$\begin{aligned} \boldsymbol{v}_{0,0} &= \arg\min_{\boldsymbol{v}_0} \|\dot{\boldsymbol{\sigma}}\|^2_{\boldsymbol{Q}} \\ s.t. \quad \underline{\boldsymbol{\tau}} &\leq \boldsymbol{\tau}(\boldsymbol{v}_0, \boldsymbol{v}_{smc}) \leq \overline{\boldsymbol{\tau}} \\ \underline{\boldsymbol{q}} &\leq \boldsymbol{q} \leq \overline{\boldsymbol{q}} \\ \boldsymbol{A}\boldsymbol{v}_0 &\leq \boldsymbol{b} \end{aligned} \tag{9.6a}$$

$$\begin{aligned} \boldsymbol{v}_{0,i} &= \arg\min_{\boldsymbol{v}_0} \|\dot{\boldsymbol{\sigma}}_i\|^2_{\boldsymbol{Q}_i} \\ s.t. \quad & \textit{constraints up to task } i-1 \\ \boldsymbol{J}_{i-1}\boldsymbol{v}_0 &= \boldsymbol{J}_{i-1}\boldsymbol{v}_{0,i-1} \end{aligned} \tag{9.6b}$$

where $\dot{\boldsymbol{\sigma}}_i$ represents the derivative of the desired task sliding manifold, and $\boldsymbol{Q}_i$ cost weighting factors.

By taking the predictions of the state and the sliding manifolds, robustness up to d steps of actuation delay is guaranteed by applying the following control law at time k, where $(\cdot)^*$ indicates predicted quantities.

$$\boldsymbol{\tau}_k = \hat{\boldsymbol{B}}(\boldsymbol{q}^*_{k+d})\boldsymbol{v}_k + \hat{\boldsymbol{n}}(\boldsymbol{q}^*_{k+d}, \dot{\boldsymbol{q}}^*_{k+d}) + \boldsymbol{J}^{*T}_{k+d}\hat{\boldsymbol{F}}^*_{e,k+d} \tag{9.7}$$

$$\boldsymbol{v}_k = \boldsymbol{v}_0(\dot{\boldsymbol{\sigma}}^*_{k+d}) + \boldsymbol{v}_{smc}(\boldsymbol{\Sigma}^*_{q,k+d}) \tag{9.8}$$

9.2.2 Robust Impedance Shaping

The application of this Model Predictive Sliding Mode Control scheme (Fig. 9.1) to a teleoperation system is straightforward, if we modify the desired impedance in order to reflect the slave environment force $\boldsymbol{F}_e$ on the master (9.9), and ensure master reference tracking $\tilde{\boldsymbol{x}} = \boldsymbol{x}_s - k_p\boldsymbol{x}^d_m$ for the slave (9.10), where k_f and k_p are force and position scaling factors, and $(\cdot)^d$ indicates a delayed quantity due to the communication delay between the two systems.

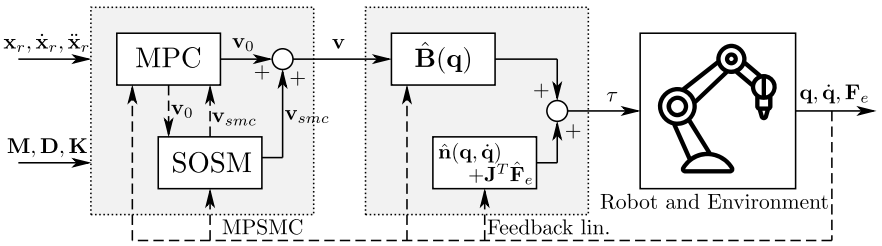


Fig. 9.1 The proposed model predictive sliding mode control scheme

$$\boldsymbol{M}_m\ddot{\boldsymbol{x}}_m + \boldsymbol{D}_m\dot{\boldsymbol{x}}_m + \boldsymbol{K}_m\boldsymbol{x}_m = \boldsymbol{F}_h - k_f\boldsymbol{F}_e^d \tag{9.9}$$

$$\boldsymbol{M}_s\ddot{\tilde{\boldsymbol{x}}} + \boldsymbol{D}_s\dot{\tilde{\boldsymbol{x}}} + \boldsymbol{K}_s\tilde{\boldsymbol{x}} = -\boldsymbol{F}_e \tag{9.10}$$

Tuning guidelines for the master and slave parameters can be obtained by analyzing the hybrid matrix $\boldsymbol{H}$ describing the interconnected system, while guaranteeing teleoperation stability in presence of delays.

$$\begin{bmatrix} F_h \\ -\dot{x}_s \end{bmatrix} = \boldsymbol{H}\begin{bmatrix} \dot{x}_m \\ F_e \end{bmatrix} = \begin{bmatrix} \frac{M_m s^2 + D_m s + K_m}{s} & \frac{k_f}{(1+s\tau)^2}e^{-sd_2} \\ -k_p e^{-sd_1} & \frac{s}{M_s s^2 + D_s * s + K_s} \end{bmatrix}\begin{bmatrix} \dot{x}_m \\ F_e \end{bmatrix} \tag{9.11}$$

where τ is the force feedback filter constant, and d_1, d_2 the communication delays. From (9.11), Llewellyn's absolute stability condition requires that all impedance dynamic parameters are greater than zero, and that the following condition is satisfied, with d_{rt} the round-trip communication delay.

$$\begin{aligned} \Lambda(\omega) = &\frac{2D_m D_s\omega^2}{D_s^2\omega^2 + (M_s\omega^2 - K_s)^2} + \frac{k_p k_f}{1+\omega^2\tau^2}\cdot \\ &\cdot\left(\frac{(1-\omega^2\tau^2)\cos(d_{rt}\omega) - 2\omega\tau\sin(d_{rt}\omega)}{1+\omega^2\tau^2} - 1\right) \geq 0, \quad \forall\omega \geq 0 \end{aligned} \tag{9.12}$$

Given the previous equation, we can show that the optimal tuning that maximizes teleoperation transparency is obtained by solving the optimization problem

$$\begin{aligned} &\max_{D_m, M_s, D_s, K_s} \quad \omega_0 \\ s.t. \quad &0 < D_m \leq \overline{D}_m \\ &0 < \underline{M}_s \leq M_s \\ &0 < D_s \\ &0 < K_s \ll 2\sqrt{D_m D_s/(k_p k_f(2\tau + \bar{d}_{rt})^2)} \\ &\tau = M_s\sqrt{k_p k_f/(D_m D_s)} \\ &D_s = 2\sqrt{M_s K_s} \end{aligned} \tag{9.13}$$

where ω_0 is the largest zero crossing frequency of (9.12) when the force feedback is unfiltered. This formulation and tuning of the teleoperation controller is able to provide better performance transparency without sacrificing stability, compared to modern time-domain passivity approaches [16].

9.3 Virtual Force Feedback

Force feedback information can be provided to the user not only through contact with a real environment, but also via an artificial force feedback generated by *virtual fixtures* [1]. The proposed architecture naturally gives itself to the specification of these robot motion constraints, with the force feedback rendered via the dual solution of the optimization to help and guide the operator. The robot autonomously can control some of the tasks and provides kinesthetic information on the status of the remotely controlled device [13].

In order to understand how the virtual force feedback can be computed, let us rewrite the hierarchical controller of the slave ((9.6a), (9.6b)) in an alternative way, by considering two priority stages and highlighting the slack variables $\boldsymbol{w}_S$ in the second one

$$\begin{aligned} \boldsymbol{v}_{0,0} = \arg\min_{\boldsymbol{v}_0} \; & \|\boldsymbol{M}_s \ddot{\tilde{\boldsymbol{x}}} + \boldsymbol{D}_s \dot{\tilde{\boldsymbol{x}}} + \boldsymbol{K}_s \tilde{\boldsymbol{x}} + \boldsymbol{F}_e\|^2_{\boldsymbol{Q}} \\ s.t. \quad & \boldsymbol{A}_H \boldsymbol{v}_0 \leq \boldsymbol{b}_H \end{aligned} \tag{9.14a}$$

$$\begin{aligned} \{\boldsymbol{v}_{0,1},\, \boldsymbol{w}_S\} = \arg\min_{\boldsymbol{v}_0,\, \boldsymbol{w}_S} \; & \|\boldsymbol{w}_S\|^2_{\boldsymbol{Q}_S} \\ s.t. \quad & \boldsymbol{A}_H \boldsymbol{v}_0 \leq \boldsymbol{b}_H \\ & \boldsymbol{J}_s \boldsymbol{v}_0 = \boldsymbol{J}_s \boldsymbol{v}_{0,0} \\ \dot{\boldsymbol{\sigma}}_S = \boldsymbol{A}_S \boldsymbol{v}_0 - \boldsymbol{b}_S \leq \boldsymbol{w}_S & \end{aligned} \tag{9.14b}$$

where we indicate with subscripts H and S, hard and soft virtual fixtures for reasons that will be now explained.

Let us start by considering the first optimization, and writing down the Karush–Kuhn–Tucker (KKT) optimality conditions

$$\left(\boldsymbol{A}_H \boldsymbol{v}_{0,1} - \boldsymbol{b}_H\right) \boldsymbol{\lambda}_{H,1} = \boldsymbol{0} \tag{9.15a}$$

$$\nabla f(\boldsymbol{v}_{0,1}) + \boldsymbol{A}_H^T \boldsymbol{\lambda}_{H,1} = \boldsymbol{0} \tag{9.15b}$$

where $\boldsymbol{\lambda}_{H,1}$ is the dual optimum associated with the hard constraints of the first stage, and $\nabla f(\boldsymbol{v}_{0,1})$ the gradient of the cost function in the optimum. Whenever a hard fixture is active, (9.15a) has to hold, and the corresponding Lagrange multiplier $\boldsymbol{\lambda}_{H,1}$ may be different from zero. Equation (9.15b) can then be rewritten as follows

$$\boldsymbol{M}_s\ddot{\tilde{\boldsymbol{x}}} + \boldsymbol{D}_s\dot{\tilde{\boldsymbol{x}}} + \boldsymbol{K}_s\tilde{\boldsymbol{x}} = -\boldsymbol{F}_e - \boldsymbol{F}_{v,H} \tag{9.16a}$$

$$\boldsymbol{F}_{v,H} = \boldsymbol{J}_s^{T\dagger}\boldsymbol{A}_H^T\boldsymbol{\lambda}_{H,1} \tag{9.16b}$$

Therefore, with a Lagrangian mechanics interpretation, the second of the KKT conditions represent the slave dynamics with an additional virtual environment force $\boldsymbol{F}_{v,H}$, defined by the virtual fixture Lagrange multiplier. This additional contribution can then be added to the master manipulator force feedback in (9.9).

For the soft fixture stage, application of the same principle gives the following for the second KKT condition

$$\boldsymbol{A}_S^T\boldsymbol{K}_S\boldsymbol{w}_S = -\boldsymbol{A}_H^T\boldsymbol{\lambda}_{H,2} - \boldsymbol{J}_s^T\boldsymbol{\lambda}_{S,2} \tag{9.17}$$

where $\boldsymbol{\lambda}_{H,2}$ and $\boldsymbol{\lambda}_{S,2}$ are the dual optima associated with the constraints of the second stage, and $\boldsymbol{K}_S = 2\boldsymbol{Q}_S$ represents the soft fixture stiffness. Therefore, the soft fixtures contribution to the force feedback is easily computed

$$\boldsymbol{F}_{v,S} = \boldsymbol{J}_s^{T\dagger}\begin{bmatrix}\boldsymbol{A}_H^T & \boldsymbol{J}_s^T\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_{H,2}\\ \boldsymbol{\lambda}_{S,2}\end{bmatrix} \tag{9.18}$$

With the proposed controller formulation, we can avoid the tuning of very large and very small weights to simulate rigid fixture surfaces, since hard constraints are separately defined in the first optimization stage and are never violated. While we can also prevent the appearance of unwanted master-slave coordination errors due to the penetration of soft fixtures in the second stage.

Overall, from (9.9), by applying both environmental and virtual feedbacks, the desired master dynamics equation becomes the following.

$$\boldsymbol{M}_m\ddot{\boldsymbol{x}}_m + \boldsymbol{D}_m\dot{\boldsymbol{x}}_m + \boldsymbol{K}_m\boldsymbol{x}_m = \boldsymbol{F}_h - k_f(\boldsymbol{F}_e + \boldsymbol{F}_{v,H} + \boldsymbol{F}_{v,S})^d \tag{9.19}$$

9.4 Occlusion-Free Visual Feedback

Visual feedback by means of image-based visual servoing is integrated and experimentally validated on a dual-arm platform, where one arm is teleoperated and the other completely autonomous and equipped with a eye-in-hand camera sensor. The proposed extension to the framework helps the user in navigating cluttered environments and keep a clear line of sight with its target by autonomously avoiding occlusions, reducing the operator workload to complete a reaching task by delegating any camera reorientation to the autonomous arm.

The vision system requirements in a remote reaching task can be summarized as follows:

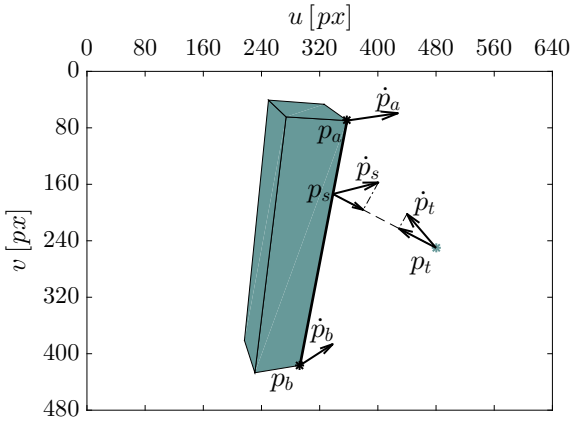


Fig. 9.2 Pictorial representation of the occlusion avoidance constraint in the image plane. The colored area identifies the projection of the occluding object

1. Autonomous and continuous camera positioning. The user should not worry about controlling the camera robot.
2. Teleoperated arm and target always visible in the camera field of view (FoV).
3. Occlusions due to additional objects in the scene must be avoided by the camera.
4. Intuitive mapping of the master device in the camera frame. The user should clearly understand what is happening remotely, and how to influence the system behavior.

The occlusion-free behavior can be easily included in the slave controller as an inequality constraint, by using a minimum distance criterion [12]. Indeed, it is sufficient to notice in Fig. 9.2 that, for a generic point $\boldsymbol{p}_t$ in the FoV, an occlusion only occurs if it enters the polytope area defined by the occluding object projection in the image. The constraint is defined as follows

$$(\boldsymbol{p}_t - \boldsymbol{p}_s)^T(\boldsymbol{p}_t - \boldsymbol{p}_s) - t_b(\boldsymbol{p}_t - \boldsymbol{p}_s)^T(\dot{\boldsymbol{p}}_s - \dot{\boldsymbol{p}}_t) \geq 0, \forall s \in [0, 1] \tag{9.20}$$

where the first term on the left represents the squared distance between the feature of interest $\boldsymbol{p}_t$ and a point $\boldsymbol{p}_s$ on the segment delimiting the polytope, while the second one is proportional to their relative velocity. t_b is a design parameter that relates to the maximum time required by the robot to bring to a halt the features in the image. By making use of the pinhole camera model and the camera Jacobian, it is straightforward to obtain a dependence on the robot joint accelerations as required by the optimization procedure (9.6a).

The constraint can also be made robust to the features relative position and velocity uncertainties (9.21): $\boldsymbol{\Delta}_p \in \mathbb{D}_p$ and $\boldsymbol{\Delta}_{\dot{p}} \in \mathbb{D}_{\dot{p}}$.

$$(\boldsymbol{p}_t - \boldsymbol{p}_s + \boldsymbol{\Delta}_p)^T(\boldsymbol{p}_t - \boldsymbol{p}_s + \boldsymbol{\Delta}_p) - t_b(\boldsymbol{p}_t - \boldsymbol{p}_s + \boldsymbol{\Delta}_p)^T(\dot{\boldsymbol{p}}_s - \dot{\boldsymbol{p}}_t + \boldsymbol{\Delta}_{\dot{p}}) \geq 0, \quad \forall s \in [0, 1], \boldsymbol{\Delta}_p \in \mathbb{D}_p, \boldsymbol{\Delta}_{\dot{p}} \in \mathbb{D}_{\dot{p}} \tag{9.21}$$

It can be shown that this approach has also the benefit of providing robustness against camera interaction matrix uncertainties as well as unmodeled dynamics of the environment.

Although the constraint itself guarantees the absence of occlusions, it does not ensure an optimal camera motion, nor a movement that is perceived as natural by the user. To achieve this, a state machine is defined, where in each state different quantities are regulated and added to the cost function to be minimized in (9.6a). The two main system phases are identified by an *approach* state (S_A), where there is no danger of occlusion occurring, and the camera simply follows the teleoperated arm, and an *occlusion* state (S_O), where an occlusion is likely to happen, and the camera has to promptly react.

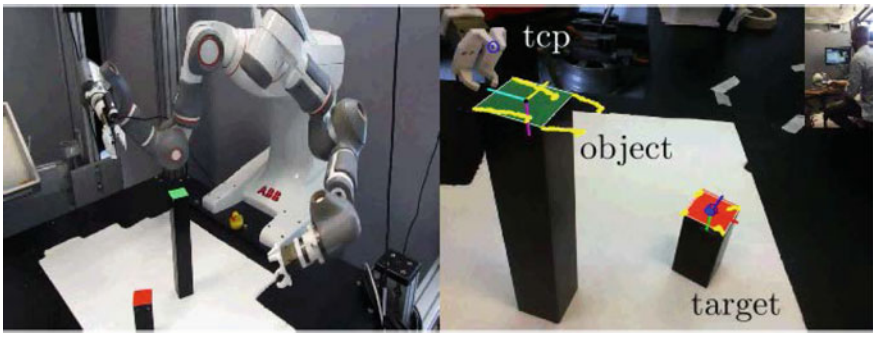
In the approach phase, the main impedance and tracking task are pursued, while regulating the teleoperated arm and target goal to fixed points in the FoV ($\boldsymbol{e}_t, \boldsymbol{e}_{g,com}$), and limiting the camera rotation around its optical axis (e_ϕ). When an occlusion constraint activates, a transition to S_O occurs.

$$\boldsymbol{v}_{0,0} = \arg\min_{\boldsymbol{v}_0} \|\boldsymbol{e}_t\|^2_{\boldsymbol{Q}_t} + \|\boldsymbol{e}_{g,com}\|^2_{\boldsymbol{Q}_{g,com}} + \|\dot{\boldsymbol{\sigma}}\|^2_{\boldsymbol{Q}} + \|e_\phi\|^2_{Q_\phi} \tag{9.22}$$

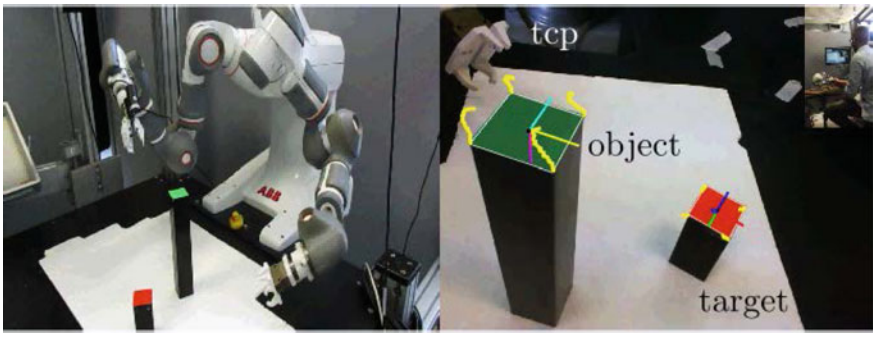
During the occlusion phase, instead, the teleoperated end effector and goal features velocities are minimized ($\dot{\boldsymbol{p}}_t, \dot{\boldsymbol{p}}_{g,com}$), along with the occluding object projected area (a_o) and the camera velocity along its optical axis ($v^c_{c,z}$). This allows to create a camera motion that pushes the occluding object out of the FoV ($\boldsymbol{e}_{o,com}$) and recovers nominal operation, transitioning back to S_A.

$$\begin{aligned}\boldsymbol{v}_{0,0} = \arg\min_{\boldsymbol{v}_0} \|\dot{\boldsymbol{p}}_t\|^2_{\boldsymbol{Q}_t} + \|\dot{\boldsymbol{p}}_{g,com}\|^2_{\boldsymbol{Q}_{g,com}} + \|\dot{\boldsymbol{\sigma}}\|^2_{\boldsymbol{Q}} + \\ + \|\boldsymbol{e}_{o,com}\|^2_{\boldsymbol{Q}_{o,com}} + \|a_o\|^2_{Q_{o,area}} + \|v^c_{c,z}\|^2_{Q_{c,z}}\end{aligned} \tag{9.23}$$

The proposed framework has been experimentally validated on a platform consisting of a dual-arm ABB YuMi robot slave device, and a Novint Falcon master device (Fig. 9.3). The controller is not only able to track the user input, but also to autonomously adjust the camera position to provide an intuitive visual feedback, as well as environmental and virtual force feedback, fully integrating robot control and user intention in a single shared autonomy architecture.



(a) As the user tries to go behind an object, its apparent motion in the FoV activates the constraint and the transition to S_O.



(b) In the occlusion phase, the user can still teleoperate the robot, however the controller moves the camera to reduce the risk of occlusion and return to S_A. The tool and the target CoM displacement from their reference is minimized.

Fig. 9.3 Occlusion-free teleoperation experiment

9.5 Conclusion

We have presented a complete robot teleoperation framework, encompassing robot dynamics control, as well as user interaction through force and visual feedback. We have also shown how operator and robot autonomy can coexist by means of a shared controller experimentally validated on a real industrial dual-arm robot.

Further studies should be conducted on how to improve the user-robot interaction. In human-robot collaboration for industrial applications, the use of machine learning techniques have been employed to infer the user intention during collaborative tasks [5]. Based on human motion limb studies and synthetic data, the model trained in [14] is able to predict the user motion and use this information to actively help the operator via the inclusion of an assistance control component. A similar approach could be investigated to further reduce the user cognitive burden and simplify task execution

in telerobotics: [22] provides one possible implementation, with an assistive behavior learned via demonstration and then integrated with the user input.

References

1. Abbott JJ (2005) Virtual fixtures for bilateral telemanipulation. Mechanical Engineering, The Johns Hopkins University
2. Abi-Farraj F, Pedemonte N, Giordano PR (2016) A visual-based shared control architecture for remote telemanipulation. In: IEEE/RSJ international conference on intelligent robots and systems, IROS'16
3. Ahola JM, Koskinen J, Seppälä T, Heikkilä T (2015) Development of impedance control for human/robot interactive handling of heavy parts and loads. In: Proceedings of the international conference on ASME 2015 international design engineering technical conference and computers and information in engineering conference (IDETC/CIE), p 8. IEEE, Boston, MA
4. Anderson RJ, Spong MW (1989) Bilateral control of teleoperators with time delay. IEEE Trans Autom Control 34(5):494–501
5. Andrea Z, Andrea C, Luigi P, Paolo R (2018) Prediction of human activity patterns for human-robot collaborative assembly tasks. IEEE Trans Ind Inform
6. Bartolini G, Ferrara A, Levant A, Usai E (1999) On second order sliding mode controllers. Variable structure systems, sliding mode and nonlinear control. Springer, Berlin, pp 329–350
7. Chatelain P, Krupa A, Navab N (2015) Optimization of ultrasound image quality via visual servoing. In: 2015 IEEE international conference on robotics and automation (ICRA), pp 5997–6002. IEEE
8. Hirche S, Ferre M, Barrio J, Melchiorri C, Buss M (2007) Bilateral control architectures for telerobotics. Advances in telerobotics. Springer, Berlin, pp 163–176
9. Hogan Neville (1985) Impedance control: An approach to manipulation: Part ii - implementation. J Dyn Syst, Meas, Control 107(1):8–16
10. Lawrence DA (1993) Stability and transparency in bilateral teleoperation. IEEE Trans Robot Autom 9(5):624–637
11. Negrello F, Settimi A, Caporale D, Lentini G, Poggiani M, Kanoulas D, Muratore L, Luberto E, Santaera G, Ciarleglio L et al (2018) Humanoids at work. IEEE Robot Autom Mag
12. Nicolis D, Palumbo M, Zanchettin AM, Rocco P (2018) Occlusion-free visual servoing for the shared autonomy teleoperation of dual-arm robots. IEEE Robot Autom Lett 3(2):796–803
13. Nicolis D, Zanchettin AM, Rocco P (2017) A hierarchical optimization approach to robot teleoperation and virtual fixtures rendering. IFAC-PapersOnLine 50(1):5672–5679, 2017. 20th IFAC World Congress
14. Nicolis D, Zanchettin AM, Rocco P (2018) Human intention estimation based on neural networks for enhanced collaboration with robots. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 1326–1333. IEEE
15. Peer A, Hirche S, Weber C, Krause I, Buss M, Miossec S, Evrard P, Stasse O, Neo ES, Kheddar A, et al (2018) Intercontinental multimodal tele-cooperation using a humanoid robot. In: 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 405–411. IEEE
16. Ryu J-H, Artigas J, Preusche C (2010) A passive bilateral control scheme for a teleoperator with time-varying communication delay. Mechatronics 20(7):812–823
17. Schmaus P, Leidner D, Krüger T, Schiele A, Pleintinger B, Bayer R, Lii NY (2018) Preliminary insights from the meteron supvis justin space-robotics experiment. IEEE Robot Autom Lett 3(4):3836–3843
18. Shibuya Hoppman. Shibuya hoppman webpage, 2012–2016
19. Smartsurg (2017) http://www.smartsurg-project.eu/
20. Utkin VI (2013) Sliding modes in control and optimization. Springer Science & Business Media, Berlin

21. Vrublevskis J, Berthoud L, Guest M, Smith C, Bennett A, Gaubert F, Schroeven-Deceuninck H, Duvet L, van Winnendael M (2018) Description of European space agency (ESA) concept development for a mars sample receiving facility (MSRF). Second international mars sample return, vol 2071
22. Zeestraten MJA, Havoutis I, Calinon S (2018) Programming by demonstration for shared control with an application in teleoperation. IEEE Robot Autom Lett 3:1848–1855