MECHANICAL ENGINEERING | PHYSICS | PRESERVATION OF THE ARCHITECTURAL HERITAGE | SPATIAL PLANNING AND URBAN DEVELOPMENT | STRUCTURAL SEISMIC AND GEOTECHNICAL ENGINEERING | TECHNOLOGY AND DESIGN FOR ENVIRONMENT AND BUILDING | TERRITORIAL DESIGN AND GOVERNMENT | AEROSPACE ENGINEERING | ARCHITECTURAL AND URBAN DESIGN | ARCHITECTURAL COMPOSITION | ARCHITECTURE, URBAN DESIGN, CONSERVATION OF HOUSING AND LANDSCAPE | BIOENGINEERING | BUILDING ENGINEERING | DESIGN | DESIGN AND TECHNOLOGIES FOR CULTURAL HERITAGE | ELECTRICAL ENGINEERING | ENERGY AND NUCLEAR SCIENCE AND TECHNOLOGY | ENVIRONMENTAL AND INFRASTRUCTURES ENGINEERING | INDUSTRIAL CHEMISTRY AND CHEMICAL ENGINEERING | **INFORMATION TECHNOLOGY** | INTERIOR ARCHITECTURE AND EXHIBITION DESIGN | MANAGEMENT, ECONOMICS AND INDUSTRIAL ENGINEERING | MATERIALS ENGINEERING | MATHEMATICAL MODELS AND METHODS IN ENGINEERING

PhD Yearbook | 2016

DOCTORAL PROGRAM IN INFORMATION TECHNOLOGY

Chair:
Prof. Andrea Bonarini

Introduction

The PhD program in Information Technology (IT) goes back to the year 2001, when the two traditional programs in Automation-Computer Engineering and Electronics-Telecommunications were merged. As such, the unified course covers the research interests in four scientific areas, namely Computer Science and Engineering, Electronics, Systems and Control, and Telecommunications. This broad variety of research activities is completely focused in the ICT area, and perfectly corresponds to the core mission of the Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB).
However, following the historical development of the Department, and the new trends of the modern society, some cross-related research fields are also developed, such as ecology, environmental modelling, operations research, and transportation systems. The PhD program in IT is the largest at the Politecnico in terms of number of students. There are more than 50 first year students and about 190 in total.
The students are subject to an examination every year to evaluate the progress achieved in their research and course work.

Topics

The research carried out in the Department in the field of Information Technology is supported by 35 laboratories, and is organized in 4 main areas, listed here below.
Computer Science and Engineering (Vice-Coordinator: Prof. Cristiana Bolchini): Information systems, Database management, Information design for the web, Methods and applications for interactive multimedia, Embedded systems design and design methodologies, Dependable systems: performance, security and reliability, Autonomous robotics, Artificial intelligence, Computer vision and image analysis, Machine learning, Dependable Evolvable Pervasive Software Engineering, Compiler Technology, Natural Language Processing and Accessibility.
Electronics (Vice-Coordinator: Prof. Angelo Geraci): Circuits and systems: theory and applications, Single-photon detectors and applications, Radiation detectors and low noise electronics, Electronic circuit design, Electron devices.
Systems and Control (Vice-Coordinator: Prof. Paolo Bolzern): Control systems, Robotics and industrial automation, Optical measurements and laser instrumentation, Dynamics of complex system, Planning and management of environmental systems, Operations research and discrete optimization.
Telecommunications (Vice-Coordinator: Andrea Virgilio Monti Guarnieri): Networking, Applied electromagnetics, Information transmission and radio communications, Optical communications, Wireless and space communications, Remote sensing, Signal processing for multimedia and telecommunications.

Industrial collaborations

Due to its intrinsic technological nature, the PhD curriculum is corroborated by many industrial collaborations. About 40% of the total number of scholarships are funded by industry or by international research projects involving industrial partners. In the school vision, the collaboration between university and industry is ideally based on the challenge of turning invention and scientific research into technological innovation. This shapes new technology frontiers and builds a fertile atmosphere for a mixture of world-class research at universities and in private companies. This also contributes to create a common terrain of friendly culture, to size the risk, and to believe in strong basic research. The external referee board is composed by members of public and private companies, working in industry and in applied research. The board is in charge of monitoring the activities of the PhD program and giving suggestions for its development. The board meets once a year to point out the new emerging research areas worth to be investigated and to monitor the visibility of the course in the industrial world.

Educational aspects

The teaching organization and subject of the courses reflect the scientific interests of DEIB faculties. The curricula include a wide choice of courses (about 30 per year), of different nature offered by this PhD Program, and more than 30 courses for basic skill formation offered by the Polimi PhD School. Access to external courses and summer schools is also encouraged.
The challenge is to promote interdisciplinary research while offering advanced courses that spur innovative and cutting edge research.
Therefore, particular attention is devoted to help each student to make the best choice according to an internal regulation scheme.

Internationalization

Every year, several courses are delivered by foreign professors. Moreover, the PhD program encourages joint curricula through agreements with foreign institutions. We have several Double Degree and Joint Degree agreements with countries in all continents. Every year we have more than 150 applications from foreign countries and about 20% of our PhD candidates come from there.

Conclusions

It is quite clear that the core mission of this PhD Program is to offer an excellent PhD curriculum, through high quality courses, a truly interdisciplinary advanced education, cutting-edge research, and international and industrial collaborations.

PHD BOARD OF PROFESSORS		
Andrea Bonarini	Stefano Ceri	Paolo Giacomazzi
Cesare Alippi	Matteo Cesana	Pier Luca Lanzi
Francesco Amigoni	Giancarlo Ferrigno	Maurizio Magarini
Luciano Baresi	Carlo Ettore Fiorini	Federico Malucelli
Luca Bascetta	Simone Garatti	Andrea Virgilio Monti Guarnieri
Giuseppe Bertuccio	Gian Guido Gentili	Barbara Pernici
Cristiana Bolchini	Angelo Geraci	Ivan Rech
Paolo Bolzern	Carlo Ghezzi	Alessandro Sottocornola Spinelli
Andrea Francesco Castelletti		

PHD ADVISORY BOARD	
Livio Baldi	Micron Semiconductors Italia
Claudio Bartolini	HP Labs (USA)
Riccardo De Gaudenzi	European Space Agency
Giuseppe Fogliazza	MCM S.p.A.
Renato Marchi	Gruppo PAM
Fabrizio Renzi	IBM Italia S.p.A.
Stefano Signani	Unicredit S.p.A.
Massimo Valla	Telecom
Stefano Verzura	Huawei

Prizes and awards

In 2015 the following awards have been obtained by PhD students:

Chorafas Foundation Award: **Giovanni Marucci, Marco Rocco**
Best Ph.D. in Information Technology Award: **Stefano Grillanda**
CHEST Grant (an European competition on social event): **Roman Fedorov**
2015 AIxIA (Italian Association for the Artificial Intelligence) best Italian PhD thesis on Artificial Intelligence: **Fabio Panozzo**
IEEE/RSJ IROS 2015 best student paper award finalists: **Matteo Parigi Polverini and Roberto Rossi**
IEEE NSS best student oral presentation award: **Paolo Trigilio**
Rocca Fellowship: **Mirko Gelsomini**

SOFTWARE LEVEL ADAPTATION IN CYBER PHYSICAL SYSTEMS

Mikhail Afanasov - Supervisor: Prof. Luca Mottola

CPSs are defined as systems, where physical processes are tightly integrated with the computational processes. This integration makes CPS to interact with and possibly taking actions on the real world.

The real world exhibits multiple dimensions that influence the behavior of CPS software continuously and independently. These require the CPS software to cope with uncertainty of sensoric input and to produce corresponding reactions. This close interaction between the real world and CPSs reveals several challenges.

In CPSs, the events that occur in the physical world have to be reflected in the CPSs software, and the decisions taken by the software influence the physical world. Due to this tight integration, the CPSs software is continuously confronted with a range of largely unpredictable environment dynamics and changing requirements. This demands CPSs software to *adapt* to a range of different situations.

The adaptation problem is even more difficult under time constraints. CPSs also include such systems as: aerial drones, and sensor networks for automobiles or railways. All these systems have soft or hard real-time requirements, and in all these cases CPSs are required to handle both periodic and aperiodic tasks within *time constraints*.

There are two major classes of systems we focus in this thesis. *Wireless Sensor Networks* are typical platforms where adaptivity is required, since the very functionality of WSNs is entangled with the highly dynamic environment. *Adaptive Time-Critical Systems* are also highly adaptive, but in addition to that, the adaptation is performed with time boundaries.

Taking into explicit account every possible situation in the design and implementation of CPSs software is a challenge. Crucially, multiple combined dimensions concurrently determine how the software should adapt its operation. Moreover, these operations may have to be preformed under time constraints. The challenge increases even more when developers battle against the resource limitations of many existing CPS platforms. Using available approaches, this typically results in entangled implementations that are difficult to debug, to maintain, and to evolve. As the number of dimensions affecting the execution (and their combinations) grows, the implementations quickly turn into “spaghetti code”.

We address this challenge by providing a handful of concepts that provide a time-critical adaptation mechanisms and help developers to implement adaptive CPSs software under resource and time constraints.

The first part of this thesis is intended to solve the *adaptivity* problem. To do so, we adapt the Context-Oriented Programming [64] to WSNs – a paradigmatic example of resource constrained CPSs. In doing so, we provide full support for developing adaptive software for WSNs. Then we outline a handful of concepts that greatly simplify the design of the adaptive software for WSNs. We argue that two main notions in the WSN software are: *i) con- text* that represents a single environmental situation the software executes in, and *ii) context group* – a collection of the contexts sharing common characteristics. These concepts are implemented in our own language CONESC. As the software model needs to be continuously verified during the development, we elaborate on a verification algorithm for the context-oriented models. Our tools allow a designer and a programmer to utilize our concepts in a real development process. With our tool GREVECOM the de- signer can build a model of the adaptive software and exhaustively verify it against environmental evolutions. Based on this model, the CONESC tem- plates are automatically generated allowing the programmer to implement the actual functionality of the application. Finally, our dedicated translator generates the binary based on CONESC sources.

We describe our early experience in developing the adaptive WSN software using our concepts and tools. We also noticed particular recurring patterns that are used in some application. Our evaluation has shown that our concepts make the software components much more decoupled and more simple, which directly influences the ease of debugging, maintaining, and evolving the code. These benefits come with a cost of memory overhead and verification time. We have shown, however, that our approach has a little price: less that 2.5% memory overhead and up to 200ms of the verification time.

The second part is devoted to time-critical systems. We show that the adaptation problems are similar to the ones we observed in WSNs, but in this part we focused on the context activation time aspect. In addition to the COP design concepts, we provided different types of context activation: *i) fast* and *ii) lazy*. The main difference between the two is that the *fast* activation requires less time, but more efforts from a programmer’s perspective, while the *lazy* one allows the programmer not to spend much efforts with the cost of increase activation time. As in time-critical systems tasks may have deadlines, in our solution, the programmer can also add a deadline to the activation command: whenever the deadline is not met, the programmer will be notified about the failure. Our prototype described in Section 5.4 implements these concepts and shows how our concepts can be used in an implementation of adaptive software for time-constrained CPSs. Our measurements reveal that our concepts cover a significant part of the time-critical adaptation routine. The benefits of our approach come with a cost of MCU overhead: $2\mu s$ of additional CPU time is required to execute the adaptation routine.

MODELING OF RELIABILITY AND NEUROMORPHIC APPLICATION OF RESISTIVE SWITCHING DEVICES

Stefano Ambrogio - Supervisor: Prof. Daniele Ielmini

Nowadays, the ever-increasing market for memories due to the widespread of portable electronics, smartphones, tablets, has led to an enormous expansion of non-volatile memories, namely memories that maintain the stored information even without voltage supply. Flash technology represents the main leader, which has dominated the market in the last decades and has now reached the 16 nm scaling node. However, Flash memory is reaching its scaling limits due to issues related to charge and dopant discretization. In addition, there is an urgent need in developing a memory capable of filling the gap that actually exists between the high-performance, low area density and expensive SRAM and DRAM, and the low cost, high density, but low-performance HDD.

For this reason, in the last years there has been a huge effort in both industrial and academic research groups to overcome this scaling issue by realizing new and revolutionary non-volatile memories with novel materials, new physical concepts and increased scaling capability. The main purpose is to obtain a novel device with the high-performance and endurance of SRAM and DRAM, combined with the high-density integration and robustness of conventional HDD. This new concept has been defined as Storage Class Memory. Although it seems not easy to find a new technology able to fit all the proposed requirements, some novel devices are promising for future non-volatile applications.

In addition to the need of novel non-volatile memory concepts, new devices are also object of active research for novel logic and neuromorphic computation. Some operations like real-world image learning, recognition and decision are extremely expensive for boolean CMOS processors, while, for the human brain, they represent quite easy processes. In this framework, the new devices could help in developing complex, high density and low power neuromorphic networks able to perform efficiently learning and recognition tasks.

Metal-oxide resistive switching memory, or RRAM, represents one of the most promising technologies for both non-volatile memory and neuromorphic computation. This device has become one of the top competitors for new generation memory due to its extreme ease of fabrication in the back-end of a CMOS process, fast switching, relatively high endurance and low power operation. All these properties make RRAM extremely interesting for future Storage Class Memory. In addition, its intrinsic variability, which represents the major concern for industrial memory production, could become the winning feature for neuromorphic network synapses.

This doctoral dissertation focuses on Hafnium Oxide Resistive Random Access Memory (RRAM). The work proposes an understanding of the device working principles, noise issues, stress and resistance instabilities, degradation and, finally, neuromorphic applications. The modeling has been the crucial focus of this work, starting from the experimental characterization and data analysis, from both single device and array.

Chapter 1 presents a brief overview of the status of non-volatile memory technology, starting from Flash technology and the memory hierarchy in a computer and later reviewing phase change, spin transfer torque, oxide based and conductive bridge memories.
Chapter 2 proposes a new analytical model for HfO_2-based RRAM. In this model, set and reset transitions are explained in terms of conductive filament (CF) growth and gap opening by means of ions migration activated by temperature and field driven. In particular, set transition is modeled as a positive thermal feedback leading to abrupt CF growth, while gradual reset transition is explained through a negative thermal feedback. The model is then used to describe the switching mechanism in the complementary resistive switch (CRS). Finally, a study of the impact of the gap resistivity in CRS characteristics is provided, showing a trade-off between off-state leakage and set/reset window.
Chapter 3 addresses set/reset variability, presenting a statistical Monte Carlo model for switching statistics based on a Poisson distribution of injected defects during switching. The model is capable of predicting LRS distribution as a function of the compliance current and HRS distribution as a function of the stop voltage. Finally, numerical modeling results are shown to provide a deeper insight into discrete fluctuation events.
Chapter 4 provides an insight into read noise, and in particular it analyzes random telegraph noise (RTN). RTN is attributed to bistable defects causing the depletion from electrons of a portion of the CF. This explanation is supported by a numerical 3D model for electron transport. Finally, it discusses the bias dependence of RTN switching times and amplitude, explained by Joule heating and Poole-Frenkel barrier variations caused by the electrostatics of the RTN bistable defect.
Chapter 5 starts with a study on cycling failure, evidencing as the major issue a negative set event due to defects injection from the bottom electrode. Then, a detailed study on the dependence of failure on V_{stop} is provided, also with the support of an Arrhenius model. After that, we focus on degradation of switching voltages and LRS resistance during cycling, which is caused by an acceleration of the switching process. Finally, a model capable of describing the degradation dependence on V_{stop} is provided.
Chapter 6 gives a physical understanding of the asymmetry between set and reset transitions in conductive bridge memory, providing both numerical and analytical models to explain it. The CF formation is accompanied with the buildup of a mechanical stress field, which tends to erase the CF, accounting for the switching asymmetry. The role of the structural relaxation is then addressed, which is responsible for the CF stabilization and data retention.
Chapter 7 provides a reliability study of a 1 kb HfO_2 array, analyzing data retention of resistance states. A new method for studying retention statistics is shown and it is applied to study resistance drift as a function of program time and verify levels. Finally, the chapter provides a physical explanation of retention behaviors in terms of CF size for set state and mechanical stress for reset state.
Chapter 8 provides an insight into read noise and corresponding broadening of the current and resistance read distributions in single devices and 0.5 Mb arrays. Read noise is mainly composed by 1/f and RTN noise and we provide two simple analytical models to predict current level broadening and multilevel reliability. We also study the resistance broadening through a 3D finite-element model. After that, the chapter focuses on the resistance broadening in RRAM arrays, highlighting the presence of two main noise contributions, which are random walk (RW) and intermittent RTN. RW is characterized by a time decay, while intermittent RTN shows fluctuations only during limited random times. To account for this behavior, the chapter finally provides a Monte Carlo statistical model based on RW and intermittent RTN.
Finally, chapter 9 deals with the application of RRAM devices as synapses in a neuromorphic network. In the first part, we propose a 2-transistors 1-resistor synapse architecture, where the first transistor allows for communication between the neuron before the synapse (PRE) and the neuron after the synapse (POST), while the second transistor allows for plasticity of the resistive device, implementing the spike-timing-dependent-plasticity (STDP) protocol. After an electrical characterization of the 2T1R synapse, pattern learning simulations are provided, validating the synapse as a building block in a neuromorphic network. In the second part, the chapter presents an alternative approach, using a 1-transistor 1-resistor synapse and comparing deterministic and stochastic switching. Finally, we simulate complex networks with inhibitory synapses and color learning.

3D SAR IMAGING OF LAND ICE STRUCTURE: METHODS & PROCESSING

Francesco Banda - Supervisor: Prof. Stefano Tebaldini

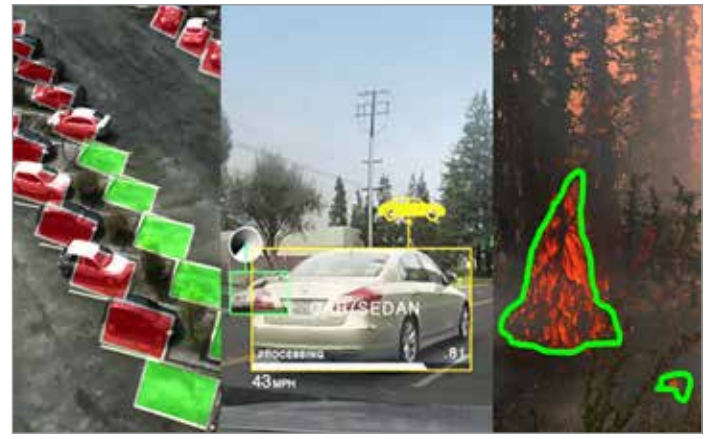
In this thesis, a novel research about the use of Synthetic Aperture Radar Tomography (TomoSAR) to characterize the 3D structure of ice sheets and glaciers is discussed.
TomoSAR is a 3D imaging tool which is now used in the radar remote sensing community, due to its potential of inferring information about complex natural scenarios on a large scale in 3D. The forthcoming European Space Agency (ESA) BIOMASS mission will have TomoSAR capabilities for the study of vegetated areas and land ice.
In this thesis, the new methodologies and the processing aimed at obtaining information about the 3D structure of land ice with TomoSAR are discussed. Results from the experimental data of the ESA IceSAR 2012 campaign are presented. It is proved that TomoSAR can effectively retrieve information about subsurface ice structures, as long as the scene is not affected by temporal decorrelation phenomena.

VISUAL ANALYSIS TOOLS FOR ENERGY AWARE HETEROGENEOUS NETWORKS

Luca Baroffio - Supervisor: Prof. Matteo Cesana

The potential of the Internet of Things is leading to a paradigm shift with an ambitious long-term vision, in which battery-operated sensing nodes are empowered with sight capabilities and are able to accomplish complex visual analysis tasks. According to such a vision, sensing devices such as smartphones, visual sensor nodes, smart cameras, will be able to perform complex visual analysis tasks in a cooperative fashion. Several heterogeneous applications might benefit from such paradigm shift. For instance, augmented reality systems would be more and more pervasive in our everyday lives, whereas future smart cities will exploit smart camera networks to improve citizens' quality of life, implementing advanced surveillance systems, monitoring the status of the environment and offering effective solutions for basics tasks such as finding available parking lots or improving the performance of public transport.
Unfortunately, this is out of reach with the current technology. In fact, computer vision algorithms are often computationally intensive and thus not suitable to be run on low-power, portable devices. Hence, such algorithms are often run on powerful centralized servers or resorting to cloud computing. According to the traditional paradigm, namely "Compress-Then-Analyze" (CTA), sensing devices acquire visual content in the form of images or video sequences, compress it resorting to traditional image or video coding algorithms such as JPEG or H.264 AVC, and transmit it to some central entities for further processing. That is, sensing nodes query the central processing nodes to perform a high-level visual analysis task by sending them pixel-level visual content.
Although being successfully implemented in a number of applications, the CTA paradigm has some limitations. The sink node relies on a lossy representation of the original signal, due to image or video compression, which contains coding artifacts that could possibly impair the results of the analysis. Furthermore, most visual analysis tasks require only a succinct representation of the acquired visual content in order to be performed. Hence, sending pixel-level representations of the content might not be the most rate- and energy-efficient solution, yielding a possibly large transmission overhead.
In recent years, with the advent of more and more powerful computing architectures and efficient computer vision algorithms, a novel approach is gaining popularity within both the scientific community and the industry. Such approach, denoted as "Analyze-Then-Compress" (ATC), moves part of the visual analysis directly on sensing nodes. According to ATC, sensing nodes acquire the content, extract semantic information from it in the form of visual features, that are subsequently exploited directly on the node or compressed and transmitted to a central unit in order to carry out a given high-level task. ATC and CTA represent concurrent paradigms that can be implemented to tackle distributed visual analysis. Differently from CTA, ATC is not affected by image compression artifacts that may impair the task and, sending just the data needed for the analysis, is more efficient with respect to transmission resources. Nonetheless, it is not usually possible to recover the actual pixel-level content from the compressed visual features sent according to ATC.

Table 1 compares the two paradigms and reports their main strengths and weaknesses. Unfortunately, ATC is currently out of reach with current technology. In fact, the "Analyze-Then-Compress" paradigm requires efficient solutions for feature extraction, coding and transmission.This thesis addresses a comprehensive set of new methodologies to empower sensor nodes with vision capabilities comparable to those achievable by power-eager visual analysis systems. Furthermore, the thesis provides several tools aimed at enabling the "Analyze-Then-Compress" paradigm for distributed visual analysis. Under the severe energy and bandwidth constraints posed by the ATC paradigm, it is imperative to optimize the extraction, the coding and the transmission of the visual features.
On the coding side, this thesis proposes several different architectures for feature coding. The statistical characteristics of the different types of visual features have been deeply investigated, and some ad-hoc coding modules have been developed, so as to achieve the best possible tradeoff between compression ratio and feature quality. The thesis puts a particular emphasis on the coding of features extracted from video sequences, developing methods and algorithms that efficiently exploit the temporal redundancy in the signal at hand. Furthermore, both local features, referring to interest regions within an image, and global features, capturing the properties of an entire image, have been considered. Such coding architectures play a fundamental role in the development of ATC, and such a thesis helps validating the effectiveness of such paradigm in the context of distributed visual analysis.
The extraction of visual features is subject to tight application-dependent requirements regarding computational and memory resources and bandwidth/delay guarantees. That is, visual features have to be extracted quickly, so as to not introduce a delay in the task processing, and without requiring too many computational and energetic resources. Still, the quality of the features has to be the highest possible, being crucial for the accuracy of the visual analysis tasks. Therefore, on the feature extraction side, this thesis addresses the design of energy-efficient tools for optimizing the operation of sensor nodes.



1.

"Analyze-Then-Compress" (ATC)	"Compress-Then-Analyze" (CTA)
High sensor computational burden	Low sensor computational burden
Only data needed for analysis is sent	Dispensable data is sent
Low/Medium bandwidth requirements	High bandwidth requirements
Low/Medium central computational burden	High central computational burden
Not affected by image coding artifacts	Analysis impaired by image coding artifacts
Pixel-level content unavailable	Pixel-level content available

Table 1 - Comparison of ATC and CTA paradigms for distributed visual analysis

A UNIFIED FRAMEWORK FOR ACOUSTIC SIGNAL ANALYSIS, SYNTHESIS, AND PROCESSING

Lucio Bianchi - Supervisor: Prof. Augusto Sarti

My research work proposes a unified framework that acts as a fundamental structure for a wide range of acoustic signal analysis, synthesis, and processing tasks. The framework emerges as a novel system of concepts and methods that encompasses the most recent advances in signal processing theory for a twofold purpose. I) It provides the foundation for the design and the evaluation of new signal processing techniques based on conventional acoustic field representations. II) It provides the required level of abstraction to introduce a novel representation for acoustic fields, based on the idea of decomposing them into a set of spatially confined wave objects.

Motivation

Spatial distributions (arrays) of microphones and loudspeakers provide the technological mean for capturing and reproducing acoustic signals (i.e. functions of both space and time). As they are conventionally employed, arrays are conceived as global sources of information, in the sense that their spatial extension is exploited to extract information about the acoustic field at a single point. Devices such as acoustic cameras and spherical microphone arrays emerged from this view: they do not attempt to extract punctual information at the microphone capsules. If the acoustic field is uniform in the region occupied by the array, or if the array can be considered compact, many effective techniques have been proposed to extract salient information from array data. However, there are situations in which the acoustic field evolves in space: this is the typical case of sound sources placed in the near field of the array. To tackle this scenario, in this work I propose to adopt the concept of the directional plenacoustic function, aimed at encoding the spatial evolution of the acoustic field as a function of both the position along the array and direction of propagation. The directional plenacoustic function allows me to rethink the use of arrays in a local fashion, i.e. by subdividing the array into smaller portions and by processing their data from a local perspective.

Goals

In all situations in which the spatial properties of the acoustic field can be considered uniform over the whole extension of the array, the global processing paradigm offers a valuable tool to design effective and efficient signal processing solutions. The first goal pursued during this work is to enhance those solutions. I pursue that goal by adopting signal processing methodologies based on Fourier theory, which allows me to devise techniques that exhibit high accuracy while being conceived to work with real-world array data. Next, we move to an higher degree of proposed innovation, shifting our mindset towards a local paradigm for array processing. In order to do that, we consider portions of a whole array: this step introduces resolution issues since we rely on less data and on a reduced spatial extension. However, we capture far more information on the acoustic field, being able to represent variations of the spatial structure of the acoustic field itself. Therefore, we set a trade-off between the achievable resolution and the capability of capturing the local evolution of the field. We provide tools, based on the mathematical theory of redundant representation of signals (frames), that allow us to find the optimal trade-off between resolution and locality.
We rethink array processing in these terms. During our work, we derived a new representation of acoustic fields, which emerges from a generalized version of Fourier theory that, based on the mathematical theory of frames, allows us to define a spatially local Fourier transform. We showed that this signal processing abstraction is supported by a strong physical intuition. As a matter of fact, we defined acoustic beams as the elementary wave objects at the roots of the spatially local Fourier transform.

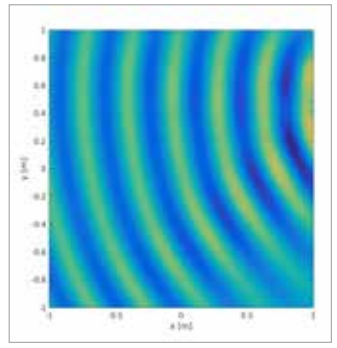
Methodology

In order to pursue our goals, we devised and followed a three-step methodology.
First, we identified a suitable set of physical models, ranging from conventional ones (plane waves, spherical waves, cylindrical waves) to more geometrically intuitive ones (acoustic rays). Beams emerge as wave objects that enable a physically accurate representation of acoustic phenomena maintaining, at the same time, the strong geometric intuition behind acoustic rays.
Second, on top of these physical models we derived specific representations of acoustic fields: this is accomplished by merging tools from mathematical physics and from the theory of signal representations. According to the specific wave object assumed as the basis for the representation, we end up with representations that are, in general, infinite-dimensional and continuous.
Three, by introducing some assumptions, we reduced infinite-dimensional and continuous representations to finite-dimensional and discrete ones. In this latter setting, we rely on the powerful tools provided by linear algebra in order to define signal processing operations (filtering, analysis, etc.). Furthermore, the properties of linear algebra allow us to discover and derive fast algorithms, suitable for the real-time implementation of processing operations.

Main contributions

i) Analytic derivation for a spatially-confined wave model (beam wave model).
ii) Derivation of a redundant representation for acoustic fields based on the beam wave model and on frame theory. On top of this representation I have defined a linear and invertible ray space transform that extracts directional plenacoustic information from array data.
iii) Design and validation of techniques for acoustic field analysis using microphone arrays. I have proposed statistical techniques to perform high-resolution analysis of array data. I have successfully exploited these techniques for the analysis of acoustic reflections inside auditoria as well as for the accurate localization of acoustic sources.
iv) Design and validation of techniques for acoustic field processing. I have proposed techniques for the generation of virtual microphones based on the plenacoustic representation, allowing me to extract individual sound signals from the mixture captured by a microphone array.
v) Design and validation of techniques for acoustic field synthesis with loudspeaker arrays. I have proposed techniques for the synthesis of acoustic fields with arbitrary geometries of the loudspeaker array, also taking into account non-idealities in the loudspeaker radiation pattern.

1. Measurement setup for the evaluation of sound field synthesis techniques.



2. Acoustic field measured with the setup in Fig. 1

ULTRA LOW-POWER ANALOG AND MIXED-SIGNAL SOCS =OR SMART SENSORS APPLICATIONS

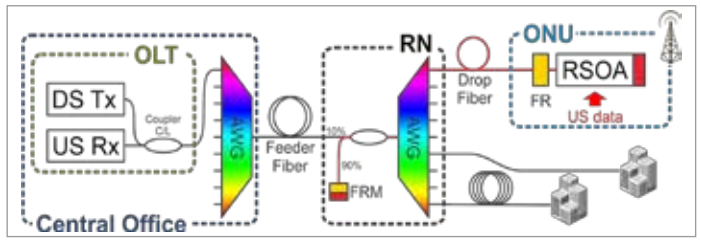
Stefano Brenna - Supervisor: Prof. Andrea L. Lacaita

The relentless miniaturization of microelectronic technologies is leading to drastically reduce power consumption thus making possible the design of sensor nodes for distributed sensing or the integration of multiple sensing systems in portable, consumer electronic devices. In these systems low- power low-noise front-ends have to acquire, digitize and transmit information. To improve energy efficiency, design solutions should be found to keep power consumption as low as possible and improving the efficiency of all the circuit blocks. A/D converters with energy efficiency better than 20 fJ/conversion-step must be investigated as well as high efficiency transceivers for short range radio links. To meet the recent trends of consumer and in general portable electronic applications, three different efficiency oriented designs were presented in this work. Two earth magnetic field sensing systems to support the development of indoor navigation systems and a multichannel wireless neural probing systems with state-of the-art efficiency, In particular: both a 3-axes Lorentz force based and a 3-axes AMR magnetic field sensing readout integrated circuits were designed in CMOS 350-nm to provide the signal amplification and digitalization. The Lorentz force based sensing system is the first presented in literature with an integrated readout electronics, works with a 3V supply, achieves a resolution of 28mGa and a programmable full-scale-range up to 24Ga with less than 1mW power consumption per channel. Thanks to better sensor characteristics, the AMR sensing system is designed to achieve 4mGa resolution, a, full scale range of drawing less than 200μW-per-axis from a 1.8V power supply. The neural probing system is fabricated in 130-nm CMOS process, features 64 channels, each comprising a low-noise amplifiers and a 10bit 6fJ/c-step efficiency ADC. The systems is provided with an UWB wireless link able to transmit a 20mbps bit stream to a 7m far receiver. The overall system power consumption is equal to 965/1μW from a, 0.5V supply, it is the lowest among multichannel (N>32) systems and it is achieved with the widest transmission range. Most of the applications described in the previous paragraph are typically characterized by the need of an analog front-end to read the design metrics (humidity, temperature, electric field, magnetic field, acceleration...). Indeed, the world we live is rich in information that come in the analog form and their measurement is possible thanks to a wide variety of sensors. In the most cases, the output of the sensors requires amplification and the conversion to a digital format which is easier to transmit with minimum losses. The combination of conditioning circuitry is called Analog Front-End (AFE) and can include amplifiers, mixers, filters and analog-to-digital converters (ADCs). A smart sensor is an entire System-on-Chip (SoC) that is devoted to the readout and the elaboration of a certain physical metric. Optimizing the design of these systems and in particular of the front-ends their adopt is a necessary task in all portable applications and the criteria that must be adopt to pursue the maximum efficiency can vary significantly depending on the class of sensor required by the application. This research work is focused on the analysis, design and optimization of three different SoCs, dealing with the three different class of sensors described above, every time aiming to define both general and sub-circuit oriented design guidelines to simplify the designer's work for the future development of similar applications.

OPTICAL TRANSMITTER BASED ON SELF-SEEDED ARCHITECTURES FOR PASSIVE WDM NETWORKS

Marco Brunero - Supervisor: Prof. Mario Martinelli

Future access networks capable of sustaining the increasing demand of capacity have recently been standardized by ITU-T under the Next Generation Passive Optical Network 2 (NG-PON2) recommendations; the proposed hybrid time and wavelength division multiplexing (TWDM) architecture is required to guarantee 10 Gb/s upstream rates per user over tens of kilometers between user and central office. WDM-based networks can greatly benefit from availability of colorless transmitters, reducing inventory costs and eliminating the needs of control layer for wavelength assignment. Due to the evolution of the radio technologies, a new fronthaul segment is appearing between remote radio head (RRH) and base band unit (BBU) to achieve a centralized radio access network (C-RAN) architecture. This optical network has to be optimized in order to make efficient usage of fibers and minimize deployment cost. Moreover, to deal with the high number of digital radio over fiber links per antenna site (one potential link per carrier, per radio sector, per mobile generation), this optical network requires the exploitation of a multiplexing technique. Based on common public radio interface (CPRI) radio, link data rates can range from 614 Mb/s to 9.8 Gb/s and the link length from each antenna and BBU hotel are comparable to the one required in access networks. A WDM network is, at present, one of the most interesting solutions, as is allow to setup a virtual point to point connection between each antenna and the corresponding BBU at the central office. As for access networks, the availability of colorless transmitters is a key factor for successful implementation of WDM networks applied to fronthaul.This PhD work has been focused on the study of a colorless solution for the upstream PON transmitter for access and fronthaul applications; it's based on the exploitation of a reflective semiconductor optical amplifier (RSOA) located at each optical network unit (ONU) or antenna and physically coincides with a portion of the network itself. The transmitter relies on a laser cavity, which comprises an RSOA active chip, placed at the ONU, whose reflective ends is one of the cavity mirrors, the drop fiber, which connects the ONU and the remote node, the remote node (RN) arrayed waveform guide (AWG) multiplexer and a mirror also placed at the RN. The AWG is the cavity wavelength-selective element. This WDM solution brings the advantage of self-selecting the wavelength by simply connecting the drop fiber to the multiplexing element at the remote node. No more management and control interface by the optical line terminal at the central office (CO) is required to set up the wavelength. The RSOA device in each ONU is the key element in the proposed transmitter as it plays the triple role of sustaining the cavity gain, of modulating the transmitted signal via its bias current and of bleaching the recirculating modulation inside the cavity.
The small E/O bandwidth of the RSOA, ~2 GHz in typical devices, is the main limit to the maximum upstream bit rate at which the proposed transmitter can reach good performance. In particular, with such small E/O bandwidth, direct modulation of the RSOAs with 10 Gb/s OOK is very challenging. RSOAs devices with E/O bandwidth up to 4 GHz, in C-band devices, are available but, due to the their design, they also present very high polarization dependent gain, higher than 20 dB. If the recirculating state of polarization is not aligned to the principle axis of the RSOA the total losses can be higher than the gain. Hence, we presented and studied a topology of the self seeded transmitter, based on a Faraday rotator mirror at the remote node and a Faraday rotator at the ONU; using this topology we proved that the input polarization is stable and always aligned to the principle axis of the RSOA. Hence we are able to use faster HPDG RSOAs for the self seeded transmitter and have a better performance when the upstream signal is at high bit rate, like 10 Gb/s.
With RSOA emitting in the C-band (i.e. 1550 nm) we were able to obtain good performance at 2.5 Gb/s and 5 Gb/s, with optical distribution network (ODN) links up to 50 km and 24 km respectively; at 10 Gb/s measurements with C-band RSOA have showed the great impact of the chromatic dispersion when SSMF are exploited for the optical distribution network, due to the frequency chirp of the output signal. In SSMF the dispersion coefficient is close to zero at 1300 nm; thus, exploiting O-band RSOA (i.e. emitting light in the 1300-nm band) we were able to directly modulate the RSOA at 10 Gb/s and obtain a BER below the $3{\cdot}10^{-3}$ FEC limit with commercial AWG, drop fiber link up to 1 km and feeder fiber link of 40 km, both with SSMF.
For legacy consideration with the previous PON generation, the ITU recommendation in NG-PON2 suggest the C-band for upstream signal from the ONU to CO, while the O-band is used by the already deployed GPON. Moreover, the vast majority of the fiber already deployed for access and metropolitan network is SSMF. Hence the self seeded transmitter in the C-band can represent an interesting solution for PON with bit rate limited at 2.5 Gb/s or 5 Gb/s while, for PON at 10 Gb/s more optimization is still required, primarily in the increase of E/O bandwidth of the devices, to allow longer link connections between ONUs and CO with BER below the FEC limit. In a green field environment, with the possibility to deploy dispersion shifted fiber, the C-band transmitter can be suitable even for a signal at 10 Gb/s, allowing the link length required by the recommendation. While permitting the reuse of the deployed SSMF, the O-band self seeded transmitter can be considered for access network applications only in a green field environment, where the O-band is not already reserved for signal of previous generation of PON solutions. On the other hand, the relatively new network segment for mobile fronthaul lacks of proper standardizations or recommendations, without indications on the dedicated transmission band for example; hence, the colorless transmitter for WDM networks studied in this thesis represents an efficient multiplexing technique to connect several RRHs with centralized BBUs. In particular, the O-band transmitter can support 10 Gb/s with 40-km long links over SSMF, fulfilling the basic requirements of future mobile networks.



1. Representation of the proposed self-seeded and colorless transmitter for WDM-PON networks.

CONTROL STRATEGIES FOR REDUNDANT AND MOBILE ROBOTIC MANIPULATORS SUBJECT TO MULTIPLE CONSTRAINTS

Giovanni Buizza Avanzini - Supervisor: Prof. Paolo Rocco

The collaboration between humans and robotic workers in productive scenarios is likely to become central in the industry of the next future. The combination of the best features of these two players, that is the dexterity and adaptability of humans compared to the repeatability and tirelessness of robots, could in fact be the key ingredient to maintain a high level of competitiveness in the manufacturing field of developed countries, despite their comparatively high labor cost.
New robots explicitly designed to work with humans are now available, opening new and largely unexplored possibilities for human-robot collaborative operations within the industry. However, a number of issues must still be solved before an industrial application of such collaborations is actually possible. The main problems are related to human safety and robot adaptability, and they require the development of new methods and technologies for robot control.
This thesis tackles the above issues by developing reactive and sensor-based control strategies that exploit the kinematic redundancy typical of collaborative robots in order to enhance human safety. The problem is first addressed considering traditional, fixed-base and not inherently redundant industrial manipulators. The goal is to develop a reactive control algorithm that is able to increase human safety in an industrial scenario, where a robot has to carry out a task in an environment that can be accessed by humans. This involves issues related to sensor perception of the robot workspace, danger assessment, redundancy exploitation and safe reactive control, which tries to maintain consistency with the robot task whenever it is safe to do so.
In particular, an overview on sensor-based control for industrial robots and on safety assessment method for human-robot interaction is first given. Redundancy-based and active control of robot exploiting exteroceptive sensors measurements is shown to be a viable approach to the issue of safety enhancement for humans. A newly conceived control strategy is then discussed, aimed at improving human safety by mean of assessing the level of danger induced by the robot in its surroundings. Methodologies to apply such strategy to non-structurally redundant robots are also considered. The proposed control strategy is then validated experimentally on an ABB IRB140 industrial robot that has to perform a pick and place task in an environment populated by humans.

1. ABB IRB 140 industrial robot performing a pick and place task in a dynamic environment populatedd by humans. The proposed control algorithm enhances human safety by exploiting sensors measurements

Having dealt with traditional robotic manipulators in the first part of the thesis, in the second part the focus is instead devoted to mobile manipulation, again aiming at allowing a closer interaction with humans in a dynamic scenario by leveraging sensor measurements and robot's redundancy. This second part builds upon the results of the first part in terms of danger assessment and sensor-based reactive control, while addressing specific issues related to mobile manipulation, such as autonomous navigation. The control paradigm is also shifted towards constrained and reactive optimal control, and specifically to Model Predictive Control (MPC).
Online constrained optimization is in fact emerging as a new paradigm for robot control, which allows to easily include constraints into the problem formulation (from the typical joint position and velocity constraints, to other kind of constraints as e.g. collision avoidance ones). At the same time the cost function to be optimized can include all the desired performance requests, which can be possibly fulfilled by exploiting redundancy.
At the beginning of this second part, a review of robot control through constrained optimization is given. Particular attention is devoted to different Model Predictive Control formulations for tracking problems. A methodology to apply such technique to the mobile manipulation context is then presented.
A specific redundancy resolution strategy for mobile manipulators is also proposed, so to allow for tracking both joint space and Cartesian space references. Specific constraints to be applied to the MPC problem are then formulated for the specific mobile manipulation case, in a way that maintains the linear-quadratic formulation of the problem.
Reactive navigation techniques for mobile robots are then reviewed, with the goal of augmenting the autonomous capabilities of the mobile manipulator. A specific procedure based on the well-known artificial vortex fields is then devised to exploit online only sensor measurements, and a method to include the autonomous navigation references in the MPC framework is proposed.
An experimental validation of the MPC approach to mobile manipulation detailed is finally presented. The validation is carried out on a KUKA youBot mobile manipulator.

2. a KUKA youBot interacting with a human operator in an unknown environment. The robot is controlled through the proposed Model Predictive Control strategy

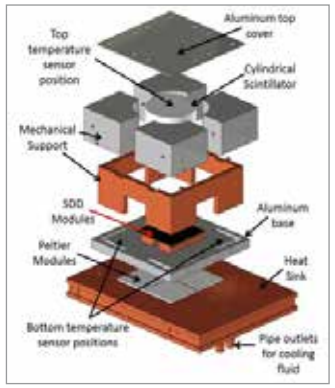
The MPC formulations are compared in different scenarios, involving interaction with the a-priori unknown environment and human operators, validating the efficacy of the proposed approach and of the autonomous navigation capabilities.
The main contributions of this thesis can finally be summarized as:

- - the development and the experimental validation of a reactive control algorithm for fixed-base industrial manipulators that enhances safety of humans in the context of shared workspace between humans and robots;
- - the development of a high level control framework for reactive motion planning of mobile manipulators, integrating environment perception, danger assessment and human collaboration in a constrained Model Predictive Control (MPC) approach;
- - the experimental validation of the proposed MPC approach, which is one of the first applications of Model Predictive Control to the high dimensional configuration space of mobile manipulators.

SILICON DRIFT DETECTOR ARRAYS FOR X- AND GAMMA-RAY DETECTION APPLICATIONS

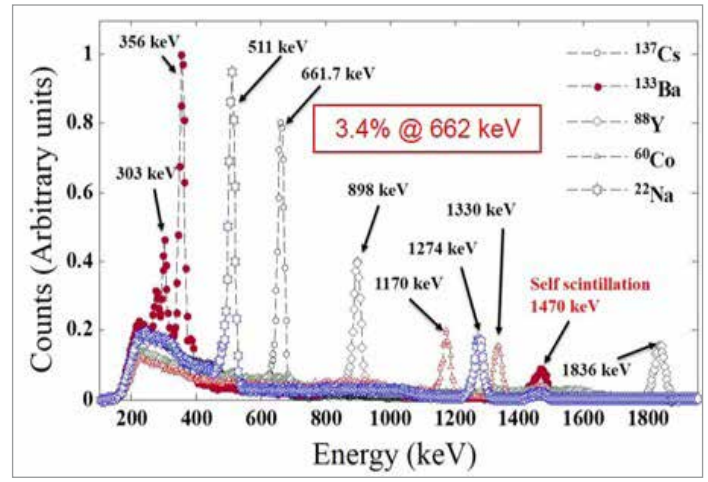
Arslan Dawood Butt - Supervisor: Prof. Carlo Fiorini

My Doctoral activity aims at the study and development of X- and gamma-ray detection systems based on Silicon Drift Detectors (SDDs) for diverse applications in the field of radiation detection instrumentation.
Silicon Drift Detectors, invented by E. Gatti and P. Rehak in 1983 are now widely used for applications demanding low-noise, high-rate X-ray detection solutions (i.e. EDX, XRF, etc..) in the typical range of 0.2-30 keV. However, these applications are always evolving with a continuous demand for better performance SDDs and readout electronics. Recent developments to meet these challenges have resulted in an ultra-low leakage SDD technology and an external CMOS based preamplifier solution with input trans-conductance optimized for very low noise readout of SDDs. In addition, many Application Specific Integrated Circuits (ASICs) have recently been developed to provide compact analog readout of preamplifiers to optimize energy resolution performances. The general trend in SDD readout is moving towards development of modular detection modules, each with a monolithic array of SDDs, preamplifiers and complementary support mechanics. This modularity provides the possibility of scaling of the overall detection system with the integration of multiple detection modules.
One such application demanding state of the art SDD performance is an upgrade of the SIDDHARTA experiment, which involves a large detector surface area composed of multiple SDD arrays to detect X-ray emissions of exotic atoms for study of strong nuclear interactions. This application involves operation of the SDD arrays at cryogenic temperatures to reduce SDD's charge collection time. Such reduced charge collection time is needed for implementation of a stricter timing logic to minimize collection of asynchronous background events. Another application within ARDESIA project involves development of compact SDD arrays for high count rate X-ray Absorption Fine structure Spectroscopy (XAFS). In this application, the SDD arrays are placed in fluorescence geometry to measure core-level binding energies of samples containing light atomic elements. Contrary to the common technique involving measurement of absorption spectra, Ardesia utilizes fluorescence spectra which can be very useful when dealing with diluted or supported (thick) samples.
In addition to X-rays, SDDs have also proved to be useful for

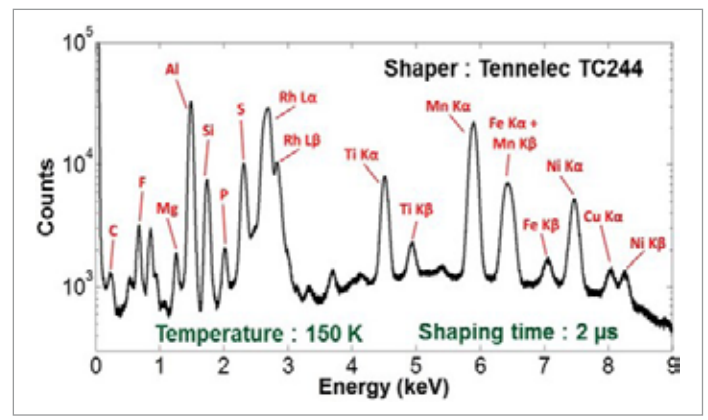


1. Exploded view of 3D model of mechanics developed for cooling down four SDD modules coupled with 2"×2" LaBr3:Ce crystal within the scope of ESA project.

scintillator readout to perform gamma-ray spectroscopy and imaging. Nominally for gamma-ray applications involving indirect conversion, Photo Multiplier Tube (PMT) represents the state of the art solution owing to the presence of signal multiplication. Although PMT has the obvious advantage of simplicity of use in lab environment and room temperature operation, SDD arrays provide the advantage of relatively compact (low mass/volume) detection module design with low voltage operation, compatibility with magnetic fields and a higher quantum efficiency. A new kind of detector which possesses advantages of both SDD and PMT, is the Silicon Photo-Multiplier (SiPM) device whose only disadvantage at the moment is a lower overall Photon Detection Efficiency (PDE) as compared to SDDs. One application of SDDs, supported by European Space Agency (ESA), involves a feasibility study to evaluate the use of SDD arrays to readout large LaBr3:Ce scintillators for planetary gamma-ray observations. In this scope, a wide nuclear transition region i.e. 150 keV to 15 MeV is considered to investigate chemical composition of planetary surfaces. This scintillator+SDD approach is also a good candidate for use as a Compton camera in gamma-ray observatories. In addition, large LaBr3:Ce scintillators readout by SDD arrays can in principle also be used for a possible reduction of Doppler Broadening effect. This Doppler Broadening effect occurs when radioactive sources moving at high/relativistic velocities with respect to the detector, emit gamma-rays which undergo an apparent shift in energy. This results in a broadening of the gamma-ray lines at the detector which can in principle be reduced by using a position sensitive gamma-ray spectrometer.
My Doctoral dissertation describes my scientific contributions in the study, design, development and characterization of prototype X- and gamma-ray detection systems for the above mentioned applications.



2. Gamma-ray spectra results obtained at -25 °C with 6 µs shaper peaking time using the gamma-camera developed for ESA project.



3. Low energy X-ray lines measured with an 8×8 mm2 detector. L and K lines of various elements including Fluorine and Aluminum are indicated. This measurement has been obtained at a temperature of 150 K in the scope of Upgrade of the Siddharta Experiment.

STRATEGIES FOR THE EFFICIENT MANAGEMENT OF THE CAPACITY IN FIXED-MOBILE CONVERGED NETWORKS

Anna Buttaboni - Supervisor: Prof. Massimo Tornatore

Nowadays customers can access telecom services via fixed-line networks or via mobile networks. Fixed-line broadband networks in Europe are currently dominated by ADSL. Other solutions like Fiber to the Curb (FTTC) with VDSL are widely used. The technologies which will deliver fiber access directly to the home, referred to as Fiber to the Home (FTTH) solutions, are the next step and first deployments have been started. For the mobile area, different network technologies are available and widely used. 2G (e.g., GSM, GPRS) and 3G (e.g., UMTS, HSPA) networks have been already installed and the deployment of the 4G LTE technology has started. So far fixed and mobile access networks have been optimized and evolved independently, with some contradicting trends: There is a tendency to centralize fixed networks and to decentralize mobile networks. A certain degree of convergence among the two network domains, typically referred as Fixed Mobile Convergence (FMC), has only been achieved at the service level with the introduction of all IP services (e.g., a practical case of FMC at service level can be found in smartphones and tablets which can access the same services through Wi-Fi and/or the cellular network). The development of a single convergent infrastructure for fixed and mobile networks will enable relevant savings in terms of Capex and Opex and will provide converged services to customers at reasonable costs in the years. The development of this FMC access network is driven by the requirement to combine optimal seamless quality of experience for end-users together with an optimized network infrastructure. Another motivation to merge together in a single and optimized structure both fixed and mobile traffic is related to the energy consumption of the current access network. In fact the access is the part of the network which is consuming the highest amount of energy. The integration of functionalities and equipment are expected to enable relevant energy saving, e.g, lowering the number of nodes of the access network. A promising network solution to develop such FMC access networks are Next-Generation Passive Optical Network (NG-PONs). In particular, Long-Reach PONs (LR WDM/TDM PON) which use both time-domain and wavelength-domain multiplexing can support a large number of different services (with different QoS requirements) originating from both fixed and mobile users. For these reasons, we consider LR WDM/TDM PON as the physical infrastructure for FMC backhauling.

The aim of this work is to propose, define and technically assess methods to efficiently manage the bandwidth in FMC networks. We first proposed dynamic bandwidth allocation methods for FMC networks: Dynamic Bandwidth and Wavelength Allocation (DBWA). These proposed methods have shown to have better performances, in terms of average packet delay, compared with existing DBWAs. Moreover, we evaluated the multiplexing gain introduced when different number of wavelengths are used in LR WDM/TDM PONs. Such results are compared with the case where a single wavelength is used to transmit (LR TDM PON). From these results we can provide some preliminary consideration regarding the design of LR WDM/TDM PON which can be used for the FMC backhauling.

Then, we studied DBWAs for network scenarios where transmission technologies installed at the ONUs are all tunable lasers (i.e., the transceivers have the ability of retuning to a different wavelength all with the same tuning time (TT)). We proposed TT-aware DBWAs in this uniform transmission technologies scenario (i.e., all the lasers are tunable). A DBWA is TT-aware if the scheduling is performed considering that the laser needs a certain amount of time to retune to another wavelength. Since the aim of our work is to study how to efficiently manage the bandiwidth in FMC where different types of traffic are present, we studied DBWAs in scenarios where different transmission technologies with different characteristics (i.e., different TTs) are used at the ONUs, considering that different services might be supported by different technologies. First, we assumed that the arrays of lasers can retune to every wavelength available in the network. However, to have a cheap device the number of wavelengths where an array of fixed tuned lasers can transmit (i.e., number of lasers of the array) has to be limited. For this reason, we evaluated the performance of the TT-aware DBWAs when the arrays of fixed tuned lasers can transmit over a limited number of wavelengths. Results show that using an array of lasers with a limited number of lasers provides an average packet delay which is only slightly higher (in the order of tenth of microseconds) than in the case where the arrays of tunable lasers have full tunability. Moreover, we proposed a new DBWA that takes into account the fact that real tunable lasers have different values of TT depending on the wavelength they have to retune to (i.e., higher distance between wavelengths requires higher TTs). To the best of our knowledge this is the first time that such a DBWA is proposed and analyzed.

A second part of work is static wavelength allocation in FMC networks. We evaluated if, in a FMC access-aggregation network, it is more convenient to isolate the different types of traffic (e.g., fixed traffic, 3G, LTE, ...) over different sets of wavelengths (i.e., VPON case) or to mix the traffic types over all the available wavelengths (i.e., converged TWDM case). To model such comparison we developed ILP formulations for both cases, based on the classical bin-packing problem. Through numerical evaluation we investigate the trade-off between the VPON case and the converged TWDM case, under different network scenarios. We can conclude that, in terms of total number of wavelengths assigned the converged TWDM case is always sligthly better (i.e., about 12%) than the VPON case. However, to use VPONs provides several benefits: i) At the OLT-side, there is a simplification of the traffic management method. ii) At the data-link layer, using VPONs to segregate the different types of traffic it is not necessary to have a convergent protocol in the access-aggregation segment of the network. iii) At the ONU-side, by using several VPONs in the same network it is possible to install cheaper transceivers (e.g., tunable lasers with smaller tuning range). Finally, we considered the scenario of Multi-Operator network-sharing for which we proposed a game theoretic (GT) collaboration strategy. The GT-based coopetition scheme proposed in our work allows operators to increase the traffic served during peak-hours. At the same time, our proposed approach aims to avoid that one operator fills all the unused bandwidth of the second operator causing a consistent traffic loss which in turns can lower the QoS experienced by the end-users of the operator while potentially causing a revenue loss. With this approach an operator can share its own bandwidth while preserving an amount of bandwidth for its own users. GT is a suitable tool to solve the problem of sharing the network capacity among different network operators since, in this manner, operators can decide whether collaborate or not according to traffic conditions (i.e., dynamically) and not following predefined agreements. Results show that our GT collaboration model is able to improve the amount of traffic served by the operators with respect to the case where operators not collaborate. At the same time, the GT collaboration have the lowest amount of traffic lost due to the collaboration with respect to the other collaboration methods evaluated in this work. The results of this research will be of impact for the future generations of high-speed broadband and mobile network infrastructure. An FMC access network will lead to significant network cost and energy-consumption reductions which will be key to address the profound transformations needed to face data traffic explosion in the medium to long term.

MODULAR SCHEMA BASED DATA WAREHOUSING OF EVOLVING, REDUNDANT AND INCOMPLETE DATA: APPLICATION TO BIOMOLECULAR KNOWLEDGE DATA INTEGRATION AND INFERENCE

Arif Canakoglu - Supervisor: Prof. Marco Masseroli

Heterogeneous data integration is an important persistent problem in different domains. It is highly challenging when the heterogeneous data are very numerous, fast evolving, from different and distributed sources, and need to be efficiently and comprehensively evaluated in order to answer complex queries in short time. Several approaches have been proposed to integrate data from multiple heterogeneous data sources, including information linkage, multi-databases, federated databases, mediator based solutions and data warehousing. The last one well supports applications with numerous data from various and dispersed sources. Off-line processing is used for mining comprehensively and efficiently the integrated data towards knowledge discovery. They leave open issues to solve, firstly, when the many heterogeneous data sources to be integrated are evolving in number, in type and also in their data schema (less rapidly). Another difficulty is that such sources may be complementary, but also overlapping, in the information that they provide. All these aspects require performing the integration by means of a well-defined but simple methodology, which is easily configurable and rapidly adjustable, in order to be able to cope with the source changes and to combine the partially redundant information from the different sources.

All these difficulties and requirements are typical in bioinformatics. Such complex scenario led us to choose the bioinformatics as the domain to demonstrate the efficacy and effectiveness of our research. Although bioinformatics is selected as test domain, by developing a domain independent abstracted and generalized data warehousing approach it can be straightforwardly customized and applied in all the other domains. This is paramount, in particular in the bioinformatics field where many questions can be addressed only by comprehensively analyzing different types of data, in order to collect evidence supporting and increasing the confidence of obtained results. As an example, the identification of biomolecular phenomena involved in a specific biological condition requires the evaluation of several different structural, functional and phenotypic characteristics of numerous genes resulted differentially expressed in a high-throughput gene expression experiment testing that biological condition.

In this Thesis, the above issues and challenges are addressed by focusing on the integration of controlled annotation data, expressed through different terminologies and ontologies, from different sources. Another difficulty is also having high quality and provenance evident data in the integration process from the multiple sources. In order to meet these requirements, in our work we performed the following steps: 1) abstraction and generalization of the main features and their associations; 2) design of a modular global data schema; 3) design of a multi-level data architecture and the metadata that describes it; 4) design and implementation of provenance recording and consistency checking of imported data; 5) development of a software framework for the automatic creation of a data warehouse. For the integrated data, we defined a multi-level modular global data schema, which is composed of multiple layers and interconnected modules. Each module represents a single feature or topic. In the case of biomolecular annotation data, which we focused on, a feature can be defined within two main group: biomolecular entity (i.e. DNA sequence, gene, transcript, protein), or biomedical feature (e.g. pathway, genetic disorder, etc.). The biomedical features of biomolecular entities are described as a multiple association of the latter ones with the former ones. These modules are defined in the multi-level abstracted data schema. Although our global data schema may seem rather complex, it is automatically created by our developed software framework according to the content of a well-defined XML configuration file, which describes the data features and sources to be imported and integrated.

In the developed software framework, we included general data loading and updating procedures that work guided by the content of the XML configuration file, so as to ease their adaptation to structural modifications of previously integrated data as well as to the integration of new data types and sources. Furthermore, the framework also supports full data provenance tracking, integrated data quality checking, data merge and redundancy removal procedures. By using the developed framework, we created a high quality data warehouse of numerous genomic and proteomic annotations, which we leveraged in several different projects and applications, including the detection of new annotations based on the integrated biomolecular entities and biomedical features; towards this goal, we applied the transitive relationship method on the integrated genes, their encoded proteins and the protein features in order to detect missing new gene annotations. Although this method may seem simple, it is very effective and used already in other domains, but it has not been previously applied to integrated annotation data in the biomolecular field. By combining data from multiple sources and their cross database identification information, we increase expressive power of our method and the quality of our detected annotations. This approach can correctly detect with good precision not only annotations that are already present in some databases on which the transitive relationship approach is not based, but also new valuable annotations not yet included in any database.

To ease access, query and extract the many valuable data integrated in the created data warehouse, we developed several different interfaces. The data warehouse is publicly accessible through a basic and a more advanced web interface at http://www.bioinformatics.deib.polimi.it/GPKB/; furthermore, we created a Web Service interface to the data warehouse (

BASEBAND UNIT HOTELLING ARCHITECTURES FOR FIXED-MOBILE CONVERGED NEXT-GENERATION ACCESS AND AGGREGATION NETWORKS

Nicola Carapellese - Supervisor: Prof. Massimo Tornatore

To enable and sustain the "Internet Society" in the next future, characterized by an exponential growth of bandwidth and Quality-of-Service requirements by users, communication networks must be continuously evolved, by resorting to novel technologies and architectural solutions to improve cost and energy efficiency.
In the realm of access and aggregation networks, such process causes severe issues to network owners.
In fact, from one side such networks are much more expensive to evolve and consume a relevant quota of the energy consumption, with respect to core/backbone; from the other side, they often constitute the bottleneck of the whole network performance.
Among the main trends that are expected to guide the evolution of such networks towards cost and energy efficiency, we focus on two promising principles.
The first one is the Fixed/Mobile Convergence (FMC), i.e., the concept of designing and optimizing networks as "a whole" resorting to infrastructure and equipment sharing among fixed and mobile networks.
The second one is the BBU Hotelling (also known as C-RAN, Centralized/Cloud Radio Access Network), i.e., the new mobile access paradigm in which base stations are splitted among BaseBand Units (BBU) and Remote Radio Heads (RRH) and BBUs are centralized into hotels, with the consequent introduction of the new "fronthaul" traffic.
In this PhD research work, we investigate some FMC architectures incorporating the concept of BBU hotelling, for next-generation access and aggregation networks.
After a survey of the relevant state-of-the-art BBU hotelling technological solutions, we make an energy-consumption comparison of some mobile network architectures that enable such schemes. Then, we identify optical WDM aggregation networks as the ideal substrate to perform hotelling, therefore we devise some alternative architectures. For these, a novel "BBU placement" optimization problem can be identified. An energy-minimization version of the model is formalized by Integer Linear Programming (ILP) and solved for multistage trees topologies. To deal with larger instances, a greedy-based heuristic algorithm is formulated, with a generic cost function. Finally, the model is further extended to consider the joint placement of hotels, the installation of aggregation electronic switches, and different options for fronthaul transport. The results show the cost and energy advantages of BBU hotelling architectures with respect to classical RANs, due to "consolidation" of BBUs into a few sites, and give insights on the interaction of BBU placement with other degrees of freedom, like electronic aggregation switching, wavelength routing and number of available wavelengths and fibers.
More in detail, the thesis is structured as follows.
In **Chapter 1**, we introduce the motivations behind this research, summarize the current panorama of related works and point out the main contributions of the thesis.
In **Chapter 2**, we focus on the role of BBU hotelling in converged network architectures. The main motivations behind this technique are described, and the critical drawbacks are detailed, mainly related to the transport of the new fronthaul traffic over the network infrastructure. A classification of the various architectural solutions for BBU hotelling is detailed, regarding BBU placement and implementation, and fronthaul transport.
In **Chapter 3**, we present an energy-consumption comparison of different optical network architectures for both mobile backhauling and fronthauling, performed in collaboration with *Orange Labs Networks, Lannion*. The main focus is about how much energy savings are allowed in a macro-cells based RAN, built on traditional mobile aggregation infrastructures with no fixed/mobile convergence, under the different combinations of choices for BBU placement and aggregation technologies that are described in the previous chapter. Finally, the energy results are combined with the analysis of the total length of fiber needed for each solution, to identify some basic tradeoffs.
In **Chapter 4**, we continue the investigation of the benefits that can be achieved by BBU hotelling solutions in terms of energy consumption, from an overall network perspective. With respect to the previous chapter, we shift our focus to an optimization approach that can be applied on a more generic fixed/mobile converged network, in which the location of hotels in not a priori given, but constitutes one of the outputs of the decision. To do so, we define some architectures for a FMC aggregation network based on WDM, in which BBU hotels can be potentially located into any of the intermediate nodes of the aggregation infrastructure. Therefore, a novel network optimization problem, jointly involving BBU placement and traffic routing, is defined and investigated.
In **Chapter 5**, the BBU placement problem introduced in the previous chapter for the optimization of energy consumption in a FMC WDM aggregation network is evolved towards two directions. First, a more generic optimization metric is considered, that jointly takes into account the dominating contributions of cost and energy consumption in such network. Second, a larger-scale aggregation network is considered, that comprises several stages of intermediate COs up to a single high-level Point of Presence (PoP) acting as edge node. As a consequence of the increased number of nodes, an ILP formulation would be too complex to be solved for realistic instances, therefore a greedy-based heuristic algorithm is proposed.
In **Chapter 6**, we introduce a novel optimization problem that takes into account more degrees of freedom of WDM aggregation network design. Differently from the previously defined problem, a multifiber infrastructure is considered and a new node architecture is adopted, that allows both the insertion of internal traffic (locally generated) and the transit of external traffic. The transit traffic can be either purely bypassed by an optical switch, or be aggregated/groomed by installing an additional electronic switch, independently from the presence of a hotel. Therefore, the decision of installing electronic switches in each node and the multifiber network dimensioning are optimized, together with BBU placement and traffic routing.
In this way, the model becomes more scalable to real-world larger networks and more interesting to investigate because of the non-trivial interaction among hotels/switches placement and traffic routing/grooming.
In **Chapter 7**, we draw the final conclusions and discuss the most relevant open issues, with a focus on both the key strengths and the drawbacks that FMC architectures and massive fronthaul deployment will put on future 5th-Generation (5G) mobile access technologies.

ON THE ROLE OF POLYHEDRAL ANALYSIS IN HIGH PERFORMANCE RECONFIGURABLE HARDWARE BASED COMPUTING SYSTEMS

Riccardo Cattaneo - Supervisor: Prof. Marco D. Santambrogio

Recent years have seen dramatic improvements in High Level Synthesis tools: from efficient translation of numeric algorithms to image processing algorithms, from stencil computations to neural networks, relevant application domains and industries are benefiting from research on compute and/or memory/communications network.

In order to systematically synthesize better circuits for specific programs and kernels, last decades' studies focused on the development of sound, formal approaches; one notable such framework is the Polyhedral Model and the associated code analysis technique, collectively called Polyhedral Model. Under this representation it is possible to compute dependencies, find loop bounds, and reorder instructions in a completely automated manner relying on the same set of sound and comprehensive assumptions of Polyhedral Model.

We reconsider this rich state of the art, and elaborate on different methodologies and implement the related toolchains to generate highly parallel and power efficient kernels running on reconfigurable hardware, with the aim of distributing the workload on multiple computational blocks while maximizing the overall power efficiency of the resulting heterogeneous system. We approach this problem in three different, interrelated ways.

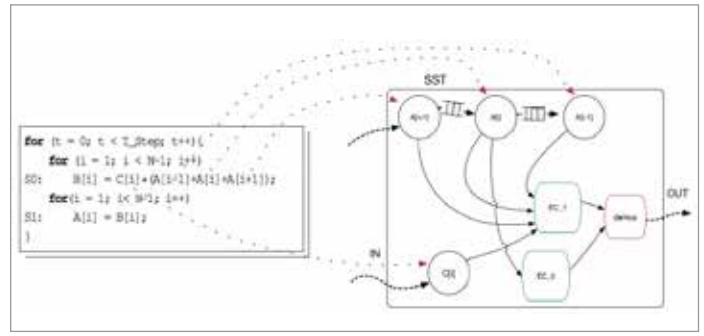
First of all, we dedicated our early efforts to the development of an accelerator-rich platform where the focus is on the coordination of multiple custom and software processors, via a novel Design Space Exploration phase. Specifically, the work elaborates on two relevant aspects: the effectiveness of Partial Reconfiguration to attain improved energy delay and throughput metrics, and the effectiveness of the heuristics chosen to realize the Design Space Exploration step, which feature both low complexity and good exploration times. We also extended the scope of this work by assuming that multiple computing elements are in place, and an adequate communication architecture is required to coordinate those accelerators.

Secondly, we focus on amply data-parallel codes, and develop a novel High Level Synthesis approach to using Polyhedral Model as a means to explicitly extract and isolate data and computation from affine codes in order to efficiently divide the workload among an arbitrary number of nodes, in the light of the current and foreseeable trend of adoption of reconfigurable hardware in the datacenter; towards energy proportional computing, we improve the current state of art in single core acceleration, as our methodology obtains near-linear speedup with the area at disposition to accelerate the given workload.

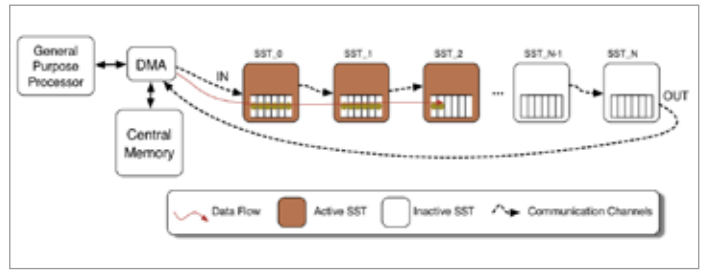
Lastly, we focus on a specific, and more restricted class of data parallel codes, namely Iterative Stencil Loop, as they play a crucial role in a variety of different fields of application. The computationally intensive nature of those algorithms created the need for solutions to efficiently implement them in order to save both execution time and energy. This, in combination with their regular structure, has justified their widespread study and the proposal of largely different approaches to their optimization. However, most of these works are focused on aggressive compile time optimization, cache locality optimization, and parallelism extraction for the multicore/multi processor domain, while fewer works are focused on the exploitation of custom architectures to further exploit the regular structure of Iterative Stencil Loops, specifically with the goal of improving power efficiency. This work introduces a methodology to systematically design power efficient hardware accelerators for the optimal execution of Iterative Stencil Loops algorithms on Field Programmable Gate Arrays.

As part of the methodology, we introduce the notion of Streaming Stencil Timestep, a streaming-based architecture capable of achieving both low resource usage and efficient data reuse thanks to an optimal data buffering strategy; and we introduce a technique, called Streaming Stencil Timesteps queuing, capable to deliver a quasi-linear execution time speedup with constant bandwidth. This methodology is adopted in a multi FPGA platform to demonstrate such scalability, a critical factor to take into account when designing large scale heterogeneous computing system.

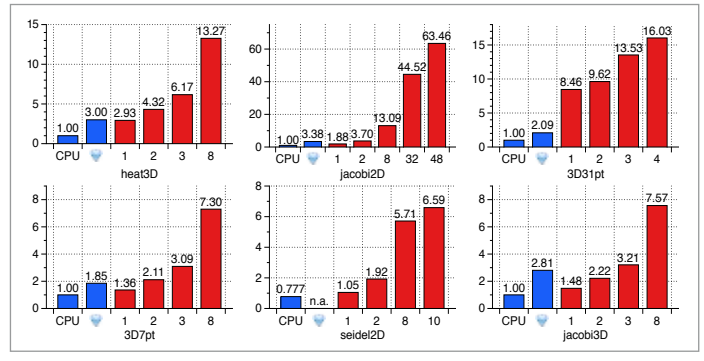
We validate all methodologies, approaches, and toolchains on significant benchmarks on either a Zynq-7000 or Virtex-7 FPGA using the Xilinx Vivado suite. Results, which are reported in each Chapter devoted to the respective subject, demonstrate how we are able to improve the efficiency of all the baselines we compare against; specifically, we improve the energy delay and throughput metric when using our accelerator-rich platform, and dramatically improve the on-chip memory resources usage using the Polyhedral Analysis-based methodologies, allowing us to treat problem sizes whose implementation would otherwise not be possible via direct synthesis of the original, unmanipulated High Level Synthesis code.



1. Fig. 1 – The SST approach, HLS-wise.



2. Fig. 2 – A high level overview of the scalability of the SST approach.



3. Fig 3. – Power efficiency results. Our approach allows to scale power efficiency with resource usage; plus, by efficiently exploiting both the heterogeneity of the platform and the scalability of the methodology, we always eventually outperform the power efficiency of a server class general purpose computing system.

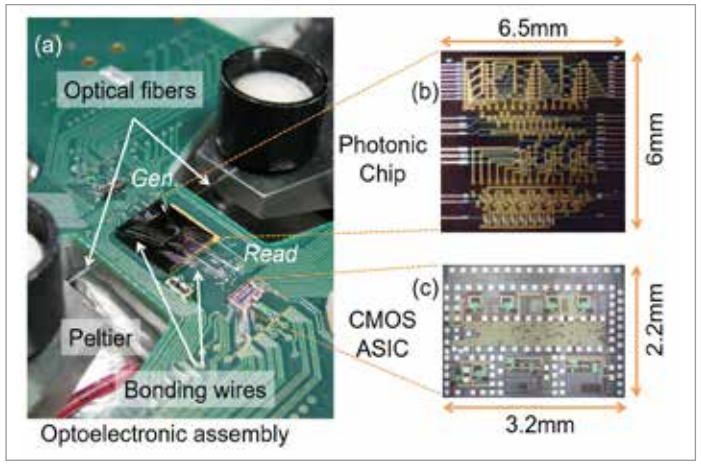
HIGH SENSITIVITY CMOS DEDICATED ELECTRONICS FOR INNOVATIVE IMPEDANCE-BASED SENSORS

Pietro Ciccarella - Supervisor: Prof. Giorgio Ferrari

Impedance measurement is a powerful tool exploited in many fields of research to inspect the properties of materials and detect a variety of phenomena. At the micro-scale, low-noise electronics is required for very small signals detection. CMOS integrated circuits offer advantages in terms of: (i) higher detection sensitivity, thanks to low parasitics and an optimal sizing of the front-end electronics, (ii) multichannel operation, (iii) higher bandwidth and (iv) a considerable miniaturization of the instrumentation. Aim of the research of this PhD work is the design and the implementation of CMOS electronics for innovative sensors.

Recently, the novel ContactLess Integrated Photonic Probe (CLIPP) allows a non-invasive light monitoring in Silicon Photonics. The photonic probe is an impedance-based photodetector and is able to measure the light induced change of the electric conductance of waveguides through a capacitive access without perturbing the characteristics of the optical field, exploiting the inherently photon interaction with intra-gap energy states localized at the waveguide interface.

A 32-channel CMOS ASIC has been designed to extend the CLIPP concept to multipoint optical chip monitoring and to achieve a higher detection sensitivity (thanks to lower parasitics) compared to bench-top instrumentation, better than 10 pS, i.e. -30dBm in optical intensity. The chip features four independent sensing chains based on a low-noise capacitive feedback transimpedance amplifier, a bias handling network based on sub-threshold transistors, double balanced square-wave demodulators and a low-parasitics multiplexer. The ASIC has been employed in an optoelectronic assembly (Figure 1) custom designed for photonic circuits monitoring and control, with the aim to implement feedback control to adjust in real-time the operating conditions of optical devices, correcting against temporal drifts thanks to thermo-optic actuators. An all-in-one electronic platform has been developed to manage the signal acquisition and the excitation sources for several CLIPPs and thermal actuators. Feedback control on SiP microring resonators, assisted by the CLIPP integrated inside the cavity, has been demonstrated: the resonant wavelength of microrings is automatically tuned and locked against wavelength drifts of the optical signal. Furthermore, the CLIPP is also able to monitor and discriminate suitably labeled signals, allowing one to tune and lock photonic devices to the wavelength of a channel, regardless of the presence of other signals, simultaneously coexisting in the same photonic device. On-chip real-time light-path tracking and reconfiguration of an 8x8 SiP switches matrix has been achieved by interrogating 24 CLIPPs, enabling sequential tuning of individual photonic elements and robust local feedback control. The non-invasive local feedback on microring resonators, light path tracking and optical circuits reconfiguration of a SiP switches matrix demonstrated in this work, pave the way for the large-scale integration of photonic microsystems.



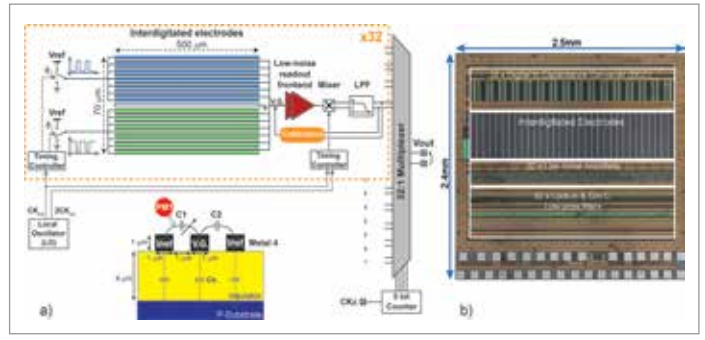
1. Photograph of the optoelectronic assembly (a) hosting the CMOS lock-in ASIC for non-invasive light detection in Silicon Photonics(c) wire-bonded to a photonic chip (b) onto the same printed circuit board. The photonic chip features complex structures such as ring-based optical filters and switches matrixes.

Microelectronics performing high sensitivity impedance measurements can be usefully exploited for airborne particulate matter (PM) detection. Standard detection methods lack portability, hampering low cost pervasive monitoring, whereas the recently demonstrated capacitive detection of PM10 directly in air is CMOS compatible and enables radical miniaturization of air quality monitors.

A monolithic CMOS PM detector (Figure 2) has been designed for the counting of the single airborne PMs by means of a capacitive measurement. The arrival of dust particles of small diameters down to PM1 is detected through integrated planar interdigitated electrodes, fabricated using the top metal layer of a standard CMOS technology without post-processing steps. Numerical simulations have been carried out to optimize the electrodes geometry estimating a capacitive change of $\Delta C = 0.7$ aF for a particle with 1 μm diameter and dielectric constant $\varepsilon_r = 2$. The integration of the electronics and the sensor electrodes on the same chip entails a strong reduction of parasitics with benefits in terms of capacitive resolution.

The required sensitivity has been achieved by using a lock-in architecture on 32 sensors operating in parallel, featuring a digital-to-capacitance converter (DCC) that reduces the equivalent capacitance of the differential sensor structure and a DC bias network that diminishes the low-frequency noise. An internally generated rail-to-rail square-wave by means of a ring oscillator, as excitation voltage, maximizes the amplitude of the capacitive signal and integrated Gm-C low pass filters, which avoid the settling time during a fast scan of the 32 sensors through the single analog output, set the signal bandwidth between 40 Hz and 750 Hz. The high gain of the low-noise capacitive-matched readout electronics is necessary to minimize the intrinsic kT/C noise added by the filters, limiting the input differential capacitance range at less than 1 fF. However, an auto-calibration network based on a Digital to Capacitance Converter manages the electrodes mismatches and automatically brings the readout front-end in dynamics. Moreover, process mismatch limits the common-mode rejection of the filters in cascade to single balanced mixers, affecting the ideal cancellation of the large low-frequency noise given by the high gain amplifier, but a DC current injected into the input node shapes the overall frequency response as a high-pass filter, resulting in a substantial reduction of the low-frequency noise.

The designed sensor features 32 channels with a resolution better than 65 zF rms allowing a real-time detection, counting and sizing of PM with an equivalent diameter down to 1 μm deposited on the active area of 1.15 mm^2. The miniaturization and low-power consumption of the resulting system go beyond the state of art of the actual PM detectors, paving the way to pervasive (smartphone-embeddable) high performance air quality monitoring and control strategies.



2. (a) circuit overview of the CMOS capacitive dust sensor with a quoted cross-section of the on-chip electrodes. (b) micro-photograph of the CMOS monolithic sensor for PM detection, implemented in AMS 0.35 μm standard technology.

ORGANIC COMPLEMENTARY LOGIC AND ANALOG CIRCUITS

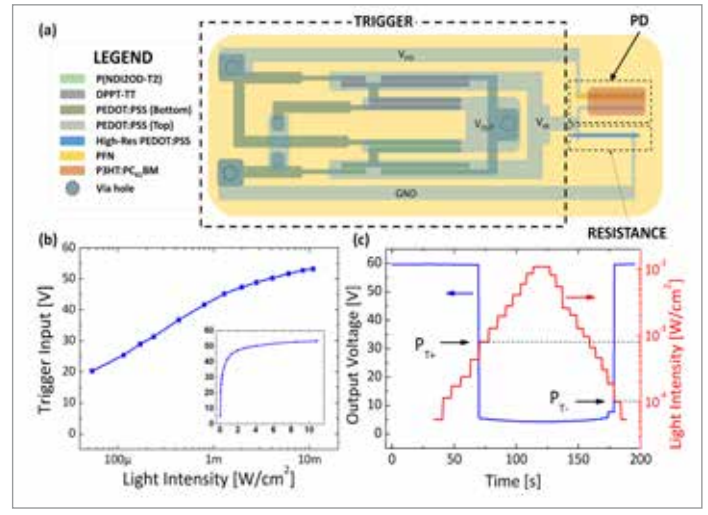
Giorgio Dell'Erba - Supervisor: Prof. Dario Natali

The tremendous development of organic electronics in the last 30 years can be attributed to the new perspectives envisioned through the study of organic materials. This development is reaching a consumer electronics stage, introducing very interesting properties like flexibility and transparency, along with solution-based manufacturing thanks to cost-effective mass production printing techniques. The main target applications are flexible displays, integrated systems for distributed and wearable sensing and interactive surfaces, with applications in automation, health-care, industrial diagnostic and security. To make organic electronics suitable for these applications, reliable and robust electronics must be implemented. This can be achieved by following the path paved by silicon electronics industry which found the best robustness in the development of complementary logic. Although the concept of complementary logic is simple, fabrication of such circuits with organic materials and scalable printing techniques is not straightforward. For this reason, most of the organic circuitry in literature is unipolar – bearing a direct consequence on the circuit performances mainly in terms of power dissipation and area occupation – or includes hybrid approaches for the fabrication - in which some of the process steps require, for example, high resolution photolithography. Printing by means of techniques inherited by the graphic arts (i.e. ink-jet printing, flexographic printing, bar coating, gravure, etc...) would enable industrial mass production with reduced capital investments and limited cost per unit area of the final product. Moreover, depending on the chosen deposition method, it would allow faster customization and time-to-market thanks to the use of digital pattern definition techniques. Anyway, the step towards all-printed circuits and systems is far from trivial: techniques in use still suffer resolution limitations, and the uniformity and yield requirements for industrial production can still be hardly guaranteed. Some recent works have tested different scalable printing methods for the fabrication of organic complementary circuits and have been able to successfully produce devices with good reproducibility and yield. On the other hand, these works still suffer some performance limitations in terms of charge carrier mobility or need to include some process steps which are hardly scalable .
The work, held at Center for Nano Science and Technology of the Italian Institute of Technology under the supervision of Mario Caironi, aims to demonstrate the feasibility of a simple fully-printed approach for the fabrication of real-life applications. Starting from the study of specific charge injection interlayers to allow ambipolar polymers to be proficiently used in complementary circuits, an approach for the fabrication of all-carbon-based, transparent, flexible complementary Organic Field Effect Transistors (OFETs) and circuits only through printing techniques will be presented. This fabrication process will be then ported to an ultra- thin substrate. Finally, a fully-printed integrated system comprising a printed photodetector and a printed analog circuit for twilight recognition is fabricated and measured.

1. A Fully-printed electronic circuit. Techniques employed for this sample fabrication are ink-jet printing and bar-coating

As an example, the latter, a photoactive switch system, comprising a fully printed photodetector, a fully printed Schmitt Trigger and a resistance, is reported in this thesis abstract. The fabricated photoactive switch detects the amount of environmental light power and triggers a digital ON/OFF signal at pre-defined threshold conditions. In particular, the OFF signal is provided when the light power exceeds a pre- set P_{T+} threshold and the ON signal is triggered when the light power drops under a second P_{T-} threshold, with $P_{T-} < P_{T+}$. This feature, intentionally designed, grants an intrinsic immunity to light fluctuations, feature that may be desirable for particular applications operating in a real-life context. Indeed, a commonly required behavior for devices responding to environmental stimuli is to trigger their state when a predefined threshold condition is met, and then reject any possible small signal fluctuations (i.e. the devices should hold their new state until the stimulus undergoes a large variation). The photo-active switch described in this work is provided with this advantageous capability, guaranteeing its suitability for integration in applications requiring to detect and react to environmental light conditions.

Here we individuate and investigate a possible example: a twilight switch, a device used in outdoor lighting to automatically activate illumination at night or in general when the environmental light is lacking. The immunity to fluctuations, which is a key feature in our implementation, effectively rejects small light intensity drops and peaks which are common in a real-life environment (i.e. due to the build-up of clouds or the passage of a body in front of the device).



2. (a) Full system layout with identification of the trigger, photodetector and load resistance areas. (b) Measurement of the trigger input voltage versus impinging light intensity (linear scale on abscissa in the inset). (c) Variation of the output voltage of the circuit (blue) upon change of the impinging light intensity (red).

A MODEL PREDICTIVE CONTROL APPROACH TO AIRCRAFT MOTION CONTROL

Luca Deori - Supervisor: Prof. Simone Garatti

Air transportation is nowadays a widespread service that plays an important role in modern world. It allows people and goods to reach far away locations across the whole planet with times and costs were not achievable before. Air transportation has, hence, become an essential means to boost economic and cultural exchanges among people and societies.
Air traffic is expected to increase rapidly over the next decades, and this growth is likely to be unsustainable for traditional rigidly structured Air Traffic Management (ATM) systems leading to the saturation of the routes, congestions, and delays.
In order to face the increasing air traffic, new concepts are needed to exploit more effectively the airspace, ensuring more efficiency and safety.
A possible solution that has been proposed in the literature rests on the concept of 4-D trajectories and Target Windows (TWs). TWs represent constraints on the 4-D trajectories requiring that the aircraft passes through a 2-D rectangle in the 3-D space within a given time interval. TWs are viewed as a key enabler of new ATM systems involving all different actors (airlines, airports, air navigation service providers) in the management process. TWs should allow for a more efficient use of the airspace, enhancing predictability of the aircraft trajectories, improving safety and airspace capacity.
In perspective, each aircraft will be assigned a sequence of TWs to meet, where the TWs altogether are designed with the twofold objective of better exploiting the airspace capacity and of avoiding conflicts. Coping with TWs is not straightforward though because TWs impose non trivial constraints on the aircraft motion in the time-space domain, which have to be satisfied while explicitly accounting for physical limitations on aircraft speed and accelerations and for other constraints related to passengers comfort. Furthermore, the aircraft motion is affected by uncertainty, in particular, the presence of wind, which leads to unpredictability of the aircraft trajectory.
The main goal of this work is to develop a motion control able to steer the aircraft along a reference trajectory, which is designed by means of an original approach so as to be compatible with both the aircraft motion capabilities and the TWs time space requirements. Tracking is performed robustly with respect to the wind disturbance and accounting for the aircraft physical limitations. Hence, by keeping the aircraft as close as possible to the reference trajectory, compatibly with the effect of the wind disturbance, TW specifications are satisfied also by the actual aircraft trajectory. The intermediate steps to achieve these goals are now discussed in details.
Given its capability of handling constraints on the input and on the state variables, a Model Predictive Control (MPC) approach to control design is adopted. MPC then involves minimizing at each sampled time instant a suitable finite horizon cost subject to suitable constraints so as to enforce trajectory tracking and the satisfaction of the aircraft physical limitations and passenger comfort requirements. The so obtained control action is applied at the current time instant only, and the process is repeated at every sampled time (receding horizon).
In formulating the finite horizon constrained optimization problem, a key issue for computational and solvability reasons is that of achieving convexity. Given that the model of the aircraft dynamics is non-linear, we apply feedback linearization so as to find a linear model with new input and state variables, whose dynamics exactly matches the original one. The so obtained linear model is then time discretized so as to be embedded in the MPC controller. The constraints on the aircraft physical limitations and passenger comfort rewritten with respect to the new state and input variables of the feedback-linearized model are non-convex. Hence, a major effort is put in reformulating these constraints as convex constraints, introducing some relaxation when needed. Such relaxations are appropriately designed so as to guarantee that the original constraints are satisfied at least for the first time instant of the current finite horizon. This way, the actually implemented MPC controller is guaranteed to satisfy the original constraints at all time steps, because of the receding horizon.
The idea of using feedback linearization followed by a convexification of the constraints is known in the literature. The contribution is to apply this methodology to address a control problem that was never tackled in this way before to the best of our knowledge. In this respect, the main results which will be obtained are the solution of a non-trivial global feedback linearization problem and the development of an ad-hoc method for the convexification of the constraints. This latter, in particular, is a problem that still lacks a general solution and its resolution will require specific analysis for each type of constraint.
TWs specifications usually involve large time and space scales, and for this reason they are hard to be handled in the finite horizon problem to be solved at each time step in MPC. Therefore, in order to address TWs specifications we resort to generate reference trajectories that comply with TWs requirements accounting at the same time for aircraft motion capabilities. To this purpose a method to suitably design such trajectories is developed. This method patches together simple motion primitives so as to eventually obtain a smooth trajectory that can be easily followed by the aircraft. Patching is obtained by solving a highly non linear system of equations by means of procedures specifically conceived to this purpose.
The obtained trajectory is then used as reference for the MPC controller, which has the critical task to keep the aircraft as close as possible to it, compatibly with the aircraft motion constraints and the disturbances. Note that the trajectory is composed by a path and a time law, and hence in order to track it the motion controller is required to steer the aircraft so as to make it reach the proper position at a proper timing.
To pursue the objective of tracking the reference trajectory, constraints on the aircraft position are included in the optimization problem. These constraints depend on the wind disturbance which makes the problem challenging because: (i) the wind disturbance depends non linearly on the aircraft position, and (ii) the wind disturbance has unbounded support, a fact that hampers the feasibility of the problem.
The former issue is addressed by replacing the original wind disturbance model with a local approximation, around the aircraft current position, which is accurate in the region of the airspace that the aircraft will be potentially flying into along the look-ahead time horizon of MPC. As for the latter issue, probabilistic - instead of robust - constraints are considered to avoid infeasibility, thus leading to a chance-constraint optimization program. More specifically, we require the controller to keep the aircraft position error small for all wind realizations except for a set of prescribed probability.
To handle the probabilistic constraints, we resort eventually to the scenario approach, which allows to find approximate but guaranteed solutions to chance-constraint optimization problems at low computational effort. In particular we rely on approaches, where the scenario approach has been tailored to the MPC framework. This way, we eventually obtain an optimization problem that can be efficiently solved at every time step and which returns in receding horizon a well performing, robust with respect to wind, control action as revealed by numerical simulations.

ON THE DESIGN OF AUTONOMIC TECHNIQUES FOR RUNTIME RESOURCE MANAGEMENT IN HETEROGENEOUS SYSTEMS

Gianluca Carlo Durelli - Supervisor: Prof. Marco D. Santambrogio

For many years the technological improvements of computing systems have been granted by the possibility to continuously increase the integration and frequency of the chips.
As everyone knows the Moore's states that the number of transistors inside a chips double roughly every 1.8 years; furthermore the Dennard's law postulated that the power density of the transistors stays constant as they shrink.
Building on these two postulates the increase in system performance from early '70s to mid '90s come as free food; unfortunately this period is over.

The continuous miniaturization of transistors lead to an unmanageable power density, with the failing of the Dennard's law; and later '90s processors were consuming more power per square inch that it was possible to dissipate causing thermal management issues and the subsequent stop in frequency increase.
This phenomenon, referred to as the reach of the power wall, caused a shift in processors design and marked the end of single processors systems in favor of multiprocessors designs.
From late '90s till nowadays system design has focused on multiprocessors to increase system performance; these solutions trade maximum frequency in favor of parallelism.
In fact the maximum frequency of the processors has been reduced compared to the single processor scenario; from the 4GHz of the latest single processor families, nowadays we have system that rarely exceed 3GHz frequency.
On the other side it is now common to have desktop, or even mobile solutions, featuring up to 8 cores, while high-end systems might have up to 64 or 128 cores on the same machine.

The shift towards parallel architectures lead to many performance issues since it was not possible anymore to get performance benefits for free.
In the single processors era, in fact, it was possible to observe performance improvements from one generation to the next one without any effort than recompiling the application code on the new machine.
In the multiprocessors era however, to achieve performance improvements it is necessary to restructure the application to take advantage of the underlying parallel architecture.
Furthermore communication and synchronization issues becomes fundamental and the best solution for one family of multiprocessors, might be completely wrong for new generations, causing the continuous need to maintain and tune the application code.

However, even if it partially mitigated the power issue, the shift to multiprocessors did not completely solved the problem; in fact nowadays we are facing similar situation, since the continuous shrinking of transistors size made the power wall problem reappear.
Today we refer to this problem as "dark silicon" which refers to the fact that although we are able to fit more and more circuitry on our devices we are seldom able to keep all of that circuitry active at the same time.
We see this phenomena in all the device and solutions as clock gating is adopted nowadays everywhere to shut down unused part of the system to avoid wasting power.

The always growing concern about power consumption caused a further shift in technology trends and nowadays solution are not using only multiprocessor solution, but are instead focusing on heterogeneous processors, with the possibility to exploit different computational unit each one with its own peculiarities.
This new trend in Heterogeneous System Architetures (HSA) recently developed due to the fact that nowadays computing system might be more efficient if the different processing units available can be exploited at their best.
The kind of accelerators is different from one system to the other and both HW accelerators and GPUs are used to attain a good performance efficiency.
The research for green computing is not only for environment related issues, but it is a strong requirement for improving overall performance of computing system.
It has been estimated that in order to reach the exascale computing performance we need a 60x improvement in performance efficiency, in order to avoid the need of a power plant dedicated for each computing facility.
In this context HSA architectures are playing a key role for reaching the exascale capability.

It is needless to say that this new increase in complexity in computing systems, further exacerbated the issues that system designer and application programmers have to take in account in order to improve both performance and power efficiency of their solutions.
Nowadays in fact the entire burden to tune an application for a given system and to decide whether a given task is suited to be offload to an accelerator, and to actually offloading it, is in charge of the application designer.
This situation is difficult to manage even when considering a simple application that has to run in a small system, and it is completely unbearable when multiple applications have to be considered and when they compete for the same resources.

The work done in my thesis tackles the problem of runtime resource management in heterogeneous systems, and in particular focuses on the definition of a framework for runtime resource management able to ease the development of new management techniques and to encompass a wide range of systems and applications.
Most of the state of the art works focus only on multiprocessors systems and provides a wide range of solutions to address specific problems; what we want to do in this work is instead looking at the problem from another perspective, focusing in particular on heterogeneous systems.
We tried to generalize the runtime management problem for our target architectures to identify a general framework that can be used to develop and evaluate new techniques to address a variety of problems ranging from resource partitioning to improve colocation, to power consumption management.

In particular this thesis propose a wide set of contribution aimed at helping system architects and programmers in realizing self-adaptive systems able to exploit heterogeneous resources.
In first place we developed a framework, which already provides all the backbone infrastructure to allow resource management.
In particular all the monitoring infrastructure is already in place and such framework can be easily configured for and ported to different systems. Furthermore all the communication between the different actors and the networking needed is already provided and does not have to be further customized when deploying this framework on other machines. The final user simply has extend and personalize it with more advanced polices, if needed.

The framework is then accompanied with an high level simulation environment that can be used to fast prototyping new runtime management policies.
Solutions designed using the simulator can be easily plugged in the framework, this allows the user to use the simulator to experiment with different architectures and applications without using a physical machine.

Finally the dissertation proposes multiple policies that have been developed thanks to the realized instruments and that improve the current state of the art regarding runtime resource management in heterogeneous systems. We propose a first heuristic to easily predict performance scalability of an application depending on the number of used threads and we exploited such method to manage resource partitioning across multiple applications. We then deployed the developed framework in the context of High Performance Computing exploiting FPGA clusters in order to provide QoS guarantees to the applications, improving the current solution available on the aforementioned clusters. Finally we designed a new runtime management policy that integrates multiple actuators (i.e. thread scaling, thread mapping, and DVFS control) to optimize power consumption under the presence QoS requirements in the case of asymmetric processors.

NARROW TRACK TILTING VEHICLES: OUT-OF-PLANE DYNAMICS ANALYSIS AND CONTROL

Simone Fiorenti - Supervisor: Prof. Sergio M. Savaresi

In recent decades, the electric vehicle market has seen considerable growth, set to increase in the coming decades, driven by several factors that led to address in this area an increasing attention.
One of most fundamental motivation is the necessity to find energy source alternative to the fossil fuels for two main reasons. First of all is the reduction of the pollutant emissions; some countries signed important agreements that impose the reduction of pollutant emissions. Second is that the fossil fuels represents a non renewable resource.
Despite some restricting issues, like performances, charging time and trip range, electric powertrain represents a good solution in terms of efficiency and emission reduction.
The new urban mobility model seems to be going towards light and small vehicles, shared by different people, with high energetic efficiency and space occupation benefits. Electrical vehicles are developing very fast in this new field, in particular so called *Narrow Track Tilting Vehicles* (NTTV).
Big automotive companies also confirm this new trend oriented to develop this kind of vehicles: a great example is the Concept E, realized by BMW (See Figure 1), that is collocated in the new electrical scooter market; another example is the Nissan Land Glider Concept (See Figure 2) presented in 2009 at Tokio Auto Show, a great example of Narrow Track Tilting Vehicle.

1. BMW Concept E

NTTV represents a solution between car and motorbike, trying to join safety and comfort of cars with the agility, maneuverability and small size of motorbike.
In order to tackle the car oversizing, this new vehicles have a narrowed track than cars, with a maximum two transportable passengers. This contributes to reduce the fuel consumption as well as redouble the maximum road capacity and improve the parking availability, particuliarly in the urban areas. Moreover electrical motors improve the efficiency compared to the thermal engines, helping to reduce the fuel consumption.

Nevertheless, narrow track vehicles have important stability problem during cornering; small size track can cause overturning. In order to guarantee a good safety level within wide range of speed, an effective engineering solution has to be designed.

2. Nissan Land Glider Concept

The best solution is allow the vehicle to tilt, like a motorbike, in order to perform the curve without overturning problems. Moreover, this characteristics provide to the NTTV as maneuverability as the motorbikes.
Compared to the two-wheeled vehicles, the possibility to tilt of four-wheeled NTTV is more critical and difficult in the different condition.
At high speed, tilting vehicle can reach safely high speed cornering and overturning can be avoided.
Low speed represent, instead, a more critical condition. Four-wheeled NTTV are typically heavier than two-wheeled vehicles; moreover they are less stable than traditional four-wheeled non-tilting vehicles. It is straight forward that at high lean angles we may have overturning problems and a lean angle dynamic control is necessary.

For these reasons, in this work we present two different lean angle control strategy oriented to two different condition: high speed and low speed.
In the first case, at high speed, a driver may have some difficult to maneuver a heavy tilting vehicle (for example during lane changing maneuver). In order to help the driver during cornering a roll angle dynamic control is developed.
The target is to speed up the roll dynamic in order to provide to the driver the feeling of a lighter vehicle.
In the second case, at low speed, we are interested in avoiding the overturning. A roll angle dynamic control is designed as well. Unlike to high speed, in this case the target is to avoid the vehicle overturning; the vehicle has to remain vertical (also with banking).
As introduced before, Narrow Track Tilting Vehicles represent a solution between car and motorbike: trying to merge safety and comfort of the car with the maneuverability of the motorbike.

However the diffusion of the motorbike in the overall circulating vehicle still growing up. Due to the technological improvements the power of the circulating motorbike engine is increasing and the safety problem is becoming increasingly important.
In order to improve the safety on the overall circulant vehicles, in the last 10 years electronic systems have been strongly developed.
A prominent example is the Bosh's Electronic Stability Control (ESC) effectiveness in reducing crashes is around 45 %. Another sophisticated electronic system available is an Anto-lock Brake System (ABS). For four wheeled vehicles, it allows, during braking, to avoid wheel lock and thus to maintain vehicle maneuverability.
In two wheeled vehicles the ABS prove beneficial for panic brake during a turn.
Safety can be improved also in acceleration conditions; as a result traction control system has been developed. Among the alternatives (e.g. engine torque control, wheel acceleration control) in the lasts years the slip based traction control has gained more attention, thanks to its several advantages with reference to the mentioned alternatives .
Following the path of the four-wheeled ones, the development of two-wheeled vehicle dynamic control systems is strongly headed on a model-based approach.
Indeed, the availability of a dynamic model of the system is preferred, since it gives a deeper understanding on system dynamics. Moreover, it can be easily used to test different control strategies, via closed loop simulations, as well as to perform sensitivity analysis w.r.t. changes in some relevant parameters.
This fact is confirmed by the numerous works which can be found in literature dealing e.g. with suspension ,, steer and global-chassis stability control for two-wheeled vehicles.
Following the mainstream, the present work focuses on the development of a model-based traction control system for powered two-wheeled vehicles.
Among the control possibilities, the traction control technique here referred to is based on the feedback of the longitudinal wheel slip.
Given the peculiar nature of the wheel slip dynamics, the employment of a closed loop slip controller, can be beneficial for vehicle performance and stability, since it would allow to keep the rear wheel slip in the "stable region", where traction force is maximum.
The possibility of a closed-loop slip controller to guarantee/improve both stability and performance of the vehicle is investigated in this part of the work.

OPTICAL INTERCONNECTS IN FUTURE GENERATION SWITCH ARCHITECTURES

Dario Giuseppe Garao - Supervisor: Prof. Guido Maier

Requirements on switching and interconnection fabrics inside switching equipment, high-performance computers and data-centers of the future are getting more and more demanding in terms of throughput and power saving. Optical technology provides many opportunities of improvement of both features compared to electronic counterparts.

The vast range of applications of optical interconnections demands for fabric architectures that are applicable to different interconnection scale, from the intra data-center infrastructure to the sub-millimetric System-on-Chip (SoC) and Network-on-Chip (NoC). Moreover, the architectures should be invariant to the implementation technology selected for the optical switching elements and the links between stages. For instance, the OMSN design should equally adapt to free-space optics, integrated waveguide photonics and fiber-optics implementations. Third, the architecture should be elastic in the specific topology selection, e.g. allowing the designer to add stages in a modular way and to configure the interstage-link patterns according to the blockingness degree (rearrangeable, strictly non-blocking, wide-sense non-blocking) needed by the application.

This work proposes a design approach for OMSNs that complies to all the three features: being suitable for different optical technologies, being modular and being applicable to a large family of switching architectures. It deals both with internal stage architecture and interconnections between stages, allowing a highly-modular implementation of the network. The design is the result of a systematic analysis of the properties of multistage switching architectures, on one side, and of optical switching systems on the other side. By exploiting the concept of module, in this work we show how such systematic approach leads to a flexible design of a vast class of multistage topologies. These include all the banyan networks and the extended generalized shuffle networks.

The modules are easily configurable. The configuration of each module is set according to the connectivity properties of the corresponding stage in a specific architecture. By joining properly configured modules, we can obtain the layout of the overall desired multistage architecture.

This layout can be used as blueprint for the physical implementation of the network. In order to simplify the discussion we refer to a planar implementation, in which optical signals propagate (in waveguides or free-space) over a plane parallel to the substrate of the network over which the switching elements are fabricated.

The planar layout is also common to non-optical technologies as for instance electronic IC technology. During this study we realized that our approach is in some aspects similar to a well known method adopted in Very Large Scale Integration (VLSI) and Ultra Large Scale Integration (ULSI) integrated circuit design, that is the Thompson grid model. Thus the proposed approach can be in principle applied also to the implementation of multistage switching networks in electronic Integrated Circuit (ICs). We briefly discuss the possibility of applying the proposed technique to the electronic case.

We define also a technique improvement to decrease the number of waveguide crossovers, while avoiding an excessive increase of waveguide bends. A quantitative analysis is conducted by an optical circuit simulator, to evaluate the overall performance of a basic multi-stage switching network, and also to estimate the improvement obtained by the modified technique. Two different integrated optics implementations are take into account, using Mach-Zender and micro-ring resonators 2 × 2 switching elements respectively.

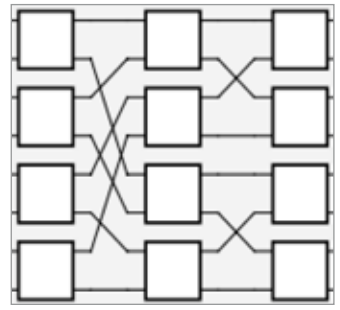
Furthermore, we propose a technique extension to increase the type of family of switching architectures can be designed by the introduced technique, using modified n x n switching elements.

In Figure 1 and Figure 2 are shown as example, respectively: a theoretical layout of a 8 x 8 Banyan Network and its optical layout with Mach-Zender switching elements generated by the proposed technique.

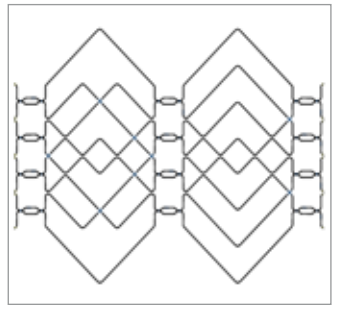
The structure of the thesis is as follows.

Chapter 1 introduces some definitions needed in the discussion. It summarizes some topics of the switching theory useful for better understand the following chapters.

Chapter 2 presents the proposed architectures. The property of such architecture are pointed out and are sustained by a set of theorems formalizing them.



1. 8x8 Banyan Network



2. Optical 8x8 Banyan Network

Chapter 3 gathers some extra arguments supplied with the previous chapter, as the physical parameter and the implementation issues.

Chapter 4 is dedicated to the scalability analysis and the technique improvement, to overcome some waveguide-optics implementation issues. Some improvements to the technique introduced in the previous chapters is proposed.

Chapter 5 reports the performance simulation results of two different integrated-optics implementations. Optical implementations by Mach-Zehnder Interferometer switches and Microring Resonator switches are considered.

The last chapter, Chapter 6, proposes a technique extension to increase the type of family of switching architectures can be designed by the introduced technique.

MINING UNIT TEST CASES TO SYNTHESIZE API USAGE EXAMPLES AUTOMATICALLY

Mohammad Ghafari – Supervisor: Prof. Carlo Ghezzi

Software reuse is the use of existing software, or software knowledge, to construct new software. Reusing existing software can decrease development effort while increases the quality of production, if mature previously tested assets are used. Developers cope with the complexity of modern software systems and speed up the development process by increasingly relying on the functionalities provided by off-the-shelf components and frameworks. However, understanding how to properly use APIs of large libraries is non-trivial and requires extensive developer learning effort. Moreover, real-world software libraries and frameworks may be underspecified or poorly documented. In addition, if documentation exists, it may include erroneous or out of date content.

With the rise of the open source movement since previous decade, an increasing quantity of source code has become available in public, and software developers have resorted to study and reuse existing source code to alleviate the aforementioned problems. However, manually exploring such source code, that is not necessarily designed in a reusable fashion, to find sufficient and adequate code examples for a given API is a difficult and time consuming activity.

In response to the increasing need for meaningful code examples, research has been focusing on finding or generating relevant examples to assist developers. Current approaches typically rely on external sources such as the source code of existing projects. Recently, good coverage of API usage examples in Q&A websites motivates researchers to use crowd sources as well. Nevertheless, there are still several cons which hamper the applicability of existing approaches in some practical scenarios. Among them, untrustworthiness, low quality, and inaccessibility of appropriate code examples are the most important ones. Besides, maintaining these resources is challenging partly because they are not explicitly linked to the APIs, and changes in the APIs are not reflected in the resources. Furthermore, existing approaches assume the availability of a reasonable sample of client code which use the API, whereas client code does not exist for private APIs or clients themselves barely exist for newly released APIs, or non-widely used ones.

According to my experience with a large number of unit test cases from a variety of open source software systems, unit test cases seem to be an additional source of significant API examples for the following reasons.

- **Tests are examples too**. Subtle dependency between APIs, incomplete API sequences, and wrong parameter values are common obstacles to use APIs successfully. Luckily, studying unit tests of an API conveys significant information of this kind and helps developers to learn the API. Specifically, unit test cases contain interesting information on: how to correctly instantiate classes, how to construct arguments for method calls, and the expected system state after the invocations of methods. Furthermore, unit test cases are not only helpful for API users but also help API developers. In software development, some private APIs intended to help the API developers during development process. Such APIs are internal and therefore there is no usage example for them in external resources. Also, they often lack of usage documentation due to busy schedule. Hence, unit test cases of private APIs could be a good resource of such API uses.
- **Tests are relevant**. Existing solutions adopt a plethora of ranking algorithms and filtration techniques to find a suitable example in external repositories. These techniques may easily produce irrelevant examples. This problem is mainly caused by the extremely large size and variety of the code they analyze. This aspect represents the dilemma of repository-based approaches. On the one hand, if they analyze small repositories, they may fail to find relevant examples. On the other hand, too large datasets may produce irrelevant examples. Conversely, the amount of code that exists in test suites is limited in size and, at the same time, it is highly relevant with respect to the tested API.
- **Tests are self-contained.** Developers would like to use self-contained and ready-to-run code examples that may be easily incorporated as they are in the current code being developed. Differently from examples obtained by mining external sources that may contain extraneous statements, unit tests have this feature being usually executable, simple, and concise snippets that are conceived to run in isolation.
- **Tests are trustworthy**. Developers are not only interested in runnable examples but often look for best practices. Nevertheless, software developers with differing knowledge often choose very different approaches to develop the same task. Thus relying on external repositories does not guarantee that examples representing best practices are actually picked. This is particularly true for untrustworthy repositories. Conversely, tests are written by the same organization or by the same developers that designed and implemented the tested classes. Thus, not only tests may they include the correct invocation protocols and patterns, but they may also represent valuable examples of best practices.
- **Tests are consistent**. Snippets of code mined from external code resources may be wrong or simply outdated. In this setting improper or not working suggestions may easily occur. Moreover, in team development, there is usually no usage documentation for newly program-specific APIs introduced by team members due to busy schedules. In contrast, APIs developed adopting modern software development processes (e.g., agile and test-driven methodologies) encourage the early generation of unit test cases in the development process. Given the early availability of unit test cases, they may provide useful information about these API uses.

In the light of these benefits, this research reuses the important information that is naturally present in test cases of an API to circumvent lack of usage examples for that API when other sources of client code are lacking. For this purpose, I inspected a large number of unit test cases manually. I observed that method calls are atoms to construct test cases from, but each method call may contribute to a different part of a test. I observed that dependencies between different parts of a test manifest themselves through object state changes verified in the oracle part of the test case. According to this observation, I could identify the most representative API uses (F-MUTs) within each test case. Then, I studied how several sub-scenarios are organized in one scenario to realize the type of relation among different F-MUTs within a test scenario and identifying relevant statements for each API use. I observed that methods accessing (read or write) the same object are usually tested together.

The investigation results inspired my API usage synthesis approach that leverages unit test cases produced by software developers during the development process to automatically generate API usage examples. I designed and implemented the proposed approach in a research prototype. It is fully automated and applies to object-oriented software and unit test cases in JUnit format. The approach requires API source code and its unit test cases, and uses classic static analysis techniques to extract the information required to construct the API usage examples. I applied this research prototype to real-life software systems.

I compared the output of the prototype to both human-written examples as well as a baseline approach. My studies show that the examples generated with this approach is relatively similar to examples that are constructed from test cases manually. That is, when there is no other source of client code except unit test cases, the proposed approach facilitates providing API usage examples which are helpful for both the API developers and the API clients.

TOWARDS ROBOT LEARNING FROM DEMONSTRATION

Amir Masoud Ghalamzan Esfahani - Supervisor: Prof. Luca Bascetta

Robots are used to perform many repetitive and precise tasks. However, the programming time and cost of a robot restricts the use of robots, especially in non-production-line uses. Although robot programming by demonstration, by which a robot can learn to perform a task from demonstrations, has been introduced to tackle this issue, still a major concern is how a robot can generalize task demonstrations across different conditions.

In this regard, many studies have been recently inspired by studies of psychologists and particularly by imitation learning in observational learning allowing a human to generalize an observation to a new goal point.

According to studies of observational learning, humans learn to perform tasks from demonstrations at three different levels: mimicking, imitation and emulation. Mimicking is the copying of a model's bodily movements. It is mentioned that mimicking must involve no conceptualization by the observer concerning the purpose of the action. Hence, it may not be possible to accomplish the task in a new environment through mimicking alone. Imitation learning is known as a goal oriented copying the form of an observed action. In imitation learning, an observer is assumed to recognize what the form of the model's movements is bringing about and use that to carry out the task; it is, therefore, analogous to a traditional DMP. Lastly, in emulation learning, the observer replicates the expected results of the model's action.

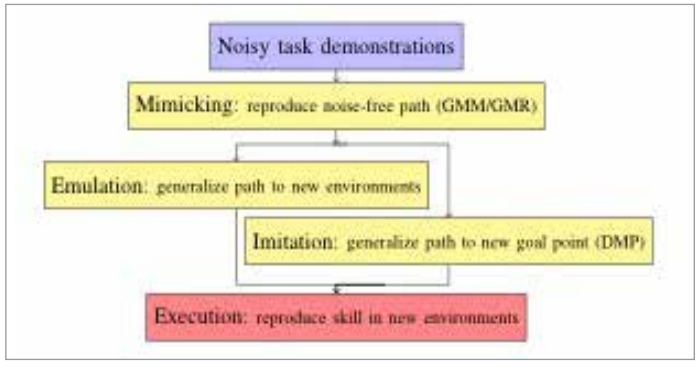
Based on the analogy between robot learning from demonstrations and human learning from observation, and according to different types of observational learning, including mimicking, imitation, and emulation, we propose a multilayered approach to robot learning from demonstration. This approach proposes a possible mathematical means of expressing different types of observational learning and implies how different types of learning from demonstration can be combined into an optimal policy generating necessary actions to perform the task.

Although the corresponding processes of learning at very early stage of human learning are very complex involving not only observational learning but also sensory-motor learning, this approach only focuses on the basic formalism of learning from demonstration. This approach enables a robot to learn a model to perform a task from noisy demonstrations and to generalize it to a new start and goal point as well as to different environments. The proposed approach has three levels of generalization, shown in Figure 1. At the first level, robot obtains a principal trajectory of a demonstrated task using noisy data of the task demonstrations; at the second level, a robot can replicate the task with different goal points; and finally it learns a control policy to avoid colliding static obstacles in the environment while performing the task. Using this approach, a non-expert user can teach a robot both a desired task model trajectory and control policy of responses to different static objects in an environment.

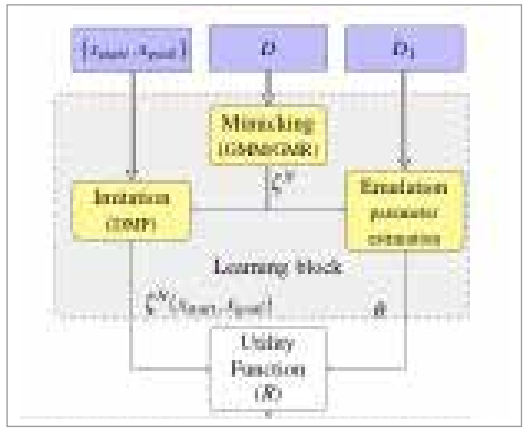
As shown in Fig. 2, for the mimicking level a regression method like Gaussian Mixture Model/Gaussian Mixture Regression is used. For the second level, dynamic movement primitives is used to generalize a demonstration to a new goal point. A generated trajectory with a new goal point and a model of teacher's response to an obstacle are combined into a utility function. The parameter of the utility function is computed using demonstrations. Finally, the utility function is used to generate a necessary trajectory to perform the task in a new environment.

This approach of robot learning from demonstration can be further extended to an environment with moving objects. Based on the studies of sensorimotor learning in cognitive psychology and neuroscience, many intelligent, context-specific responses of humans to different stimuli during a task execution are consistent with the theoretical framework of optimal control. Prediction provides the brain with expected sensory consequences. The corresponding motor commands are then used to obtain desired consequences. Accordingly, we may adopt the same perspective on the model obtained by multi-layered robot learning from demonstrations to generalize a set of demonstrations to a dynamic environment.

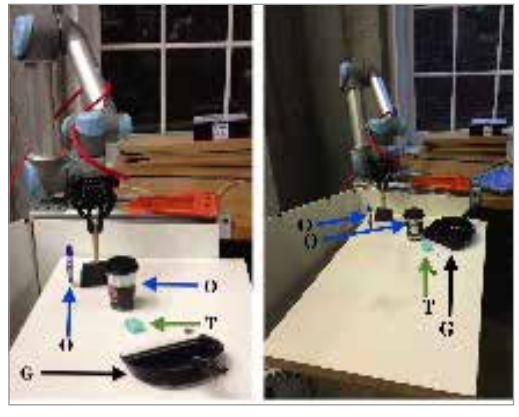
Two experiments with a da Vinci and a UR5 robot are presented to illustrate how the workflow proposed in this thesis allows a robot to automatically build a task model from a set of demonstrations. The example is shown in Fig. 3 is a 6-degree of freedom industrial robot learns from human demonstrations a model of sweeping a cube into a dustpan while avoiding some obstacles in the environment. The sweeping task is a typical example of a household task. It then replicates the task in environments with different conditions of environment, e.g. different location of the dustpan, cube and obstacles. The mean square error of the reproduced path for sweeping the cube into dustpan using the model obtained by this approach with the environment corresponding to the training set and the test set are compared which illustrate the effectiveness of the proposed approach with a good accuracy.



1. Overview of the proposed method inspired by human skill learning. In the mimicking step, we compute the average path from a set of demonstrations. We create a model that can be scaled to different start and goal positions (imitation) and learn a parameterized cost function modeling the appropriate response to different objects in the environment (emulation).



2. Incremental robot learning from demonstration scheme. D is a set of demonstrated trajectories corresponding with a task, D1 is a set of trajectories and their corresponding obstacle positions and xstart and xgoal are the start and goal point at every task execution.



3. Two different scenarios of sweeping task with unseen position of the dust- pan as well as unseen position of the marker and cup (obstacles): scenario 1 and 2 are shown left and right figure.

MEMS GYROSCOPES BASED ON PIEZORESISTIVE SENSE: A THEORETICAL AND EXPERIMENTAL ANALYSIS

Federico Giacci - Supervisors: Prof. Antonio Longoni, Prof. Giacomo Langfelder

In literature the alternative between AM mode-matched and AM mode-split MEMS gyroscopes is reported. Mode-matched gyroscopes, as they operate at resonance, feature a high gain but, at the same time, their bandwidth is limited by the -3dB width of the sense resonance peak. In addition to this, their gain is significantly sensitive to variations of pressure, temperature and process imperfections. On the other hand, mode-split capacitive gyroscopes feature a lower gain due to the off-resonance operation. The mismatch between drive and sense modes leads to an extended device bandwith, and to a high gain stability (in presence of variations of temperature, pressure,...). In this scenery, a radically different approach is represented by frequency modulated gyroscopes. While QFM gyroscopes suffer from poor temperature stability, LFM gyroscopes feature better insensitivity to temperature. In both the device topologies, the XY geometry is difficult to be designed as large out-of-plane motion is hard to be achieved with standard process thicknesses. This Work of Thesis has been developed within the fp7 European project NIRVANA, whose aim is the realization of inertial sensors based on a novel piezoresistive sensing and associate integrated electronics. The piezoresistive effect is given by silicon nanowires, also called nanogauges. In Chapter 1 the mail structural choices done for these devices are difghfgscussed. The drive motion is achieved through traditional comb-fingers actuation. The design of a differential structure with tuning-fork helps to reject common mode accelerations. Multiple frames are implemented so to reduce the quadrature error. AM mode-split gyroscopes have been designed, achieving all the advantages of this topology and compensating the low gain with an intrinsicly high sensitivity of these piezoresistive nanowire-based devices. The anti-phase Coriolis motion is transduced into differential stress on the nanogauges through a single central lever. The high sensitivity is both given by the change in cross-section between micro and nanometric parts, and by the lever, which de-amplifies the sense displacement and amplifies the force on the nanogauge. The operative characterization of these devices requires a flexible, low-noise discrete components characterization electronics. Chapter 2 reports the design guidelines for the implemented electronics. The primary drive oscillator actuates the MEMS gyroscope drive masses at resonance. It consists in a positive loop interfaced with the MEMS, that acts as frequency selecting element. The sensitivity of a gyroscope (both piezoresistive and traditional capacitive) is dependent on the drive displacement: controlling this quantity leads to a stable scale factor with respect to variations of temperature, package pressure and device-by-device. A control electronics has been therefore implemented. It continuously compares the effective displacement with a reference value, given externally, and reacts on the actuation force so to achieve the required drive motion. The differential resistance variation on the nanogauges, due to the anti-phase Coriolis motion, is readout through a Wheatstone bridge. The voltage difference between its branches is amplified and referred to ground by an Instrumentation Amplifier. After a conditioning stage, a laboratory Lock-In amplifier demodulates the signal. A drive and sense electronics for traditional capacitive gyroscopes has been also designed. The drive electronics has the same topology than the one used for piezoresistive gyroscopes. The sense electronics exploits charge-amplifiers to read-out the differential sense motion, and again an Instrumentation Amplifier, conditioning stage and laboratory Lock-In amplifier end the sense chain, so to get a voltage signal purely proportional to the angular rate. The noise of the device and the sense electronics of both piezoresistive and capacitive gyroscopes is theoretically analyzed. With discrete components electronics, the noise of capacitive gyroscopes is limited by feedback resistances and by the charge amplifiers voltage noise. In the case of piezoresistive gyroscopes the dominating contribution is given by the INA. An extended measurement campaign has been carried out so to characterize the first-run devices, analyzing advantages and drawbacks of this novel sensing technique and extracting some guideline for the next-run design. Chapter 3 reports all the experimental results. An experimental setup has been implemented for the operative characterization: a labview masters a rate table through a wafeform generator, directly interfaced to the motor of the rate table. The output signal of the sense electronics is sent to the laboratory Lock-In amplifier and its output acquired and post-processed. The piezoresistive gyroscopes feature a sensitivity linearity error of only some ppm due to their restricted operating range of stress. This constitutes a remarkable advantage with respect to capacitive gyroscopes, whose linearity error is of some % at full-scale. The measured cross-axis rejection is of about 55 dB. The temperature variation of both drive and sense modes is approximately by a factor -13 ppm/°C. The Allan variance graph, computed for both Z-axis and XY-axis devices, show an ARW till 2 mdps/ /, with no bias instability observed at 20s. The devices feature low noise levels (referring to consumer applications), to be mostly ascribed to the INA input voltage noise. The noise can be nevertheless reduced increasing the Wheatstone bridge bias and compensating the quadrature error. The increasing use of microelectromechanical system (MEMS) gyroscopes in several areas, such as automotive, consumer and military fields, leads to focus the attention on the repeatability of the performance of these devices in presence of shocks and/or vibrations that represent realistic environmental conditions. A piezoresistive gyroscope and an off-the shelf capacitive gyroscope have been compared in terms of noise floor worsening in presence and absence of vibrations. While the capacitive device shows a 65 times worsening of the ARW, the piezoresistive gyroscopes demonstrates a not noticeable worsening of its white noise. This can be ascribed to the high linearity of their response and to the characteristic modes of the structure within 50kHz, in correspondance of which the nanogauges are common-mode stressed or even they are not stressed. Finally, a piezoresistive and a capacitive devices, sharing the same structural choices, drive electronics, Instrumentation Amplifier, conditioning electronics, Lock-In and also overall system sensitivity, have been compared in terms of theoretical and experimental noise level. The results well match the predictions, and the capacitive gyroscope reveals a 10 times worse noise performance, due to its readout electronics. The use of charge pumps could decrease the capacitive gyroscopes noise: this however increases the power consumption. In Chapter 4 the possibility to integrate a circuit following the main guidelines adopted for the PCB board is discussed. The perspective is the implementation of a first version of integrated electronics for M&NEMS gyroscopes, so to highlight issues and challenges that can raise new considerations for the future fields of application of piezoresistive gyroscopes. A drive and sense electronics for capacitive gyroscopes has been also designed, to carry on the comparison between the two topologies. It is evident the the need of solutions both power saving and acceptably noisy. In conclusion, these devices are promising because they feature: intrinsicly high sensitivity, intrinsicly high response linearity, high vibration rejection, less noisy sense electronics (with respect to capacitive gyroscopes). Furthermore, their sensitivity is independent on the area occupation (whose reduction is however limited by the thermo-mechanical noise) and their input referred noise is independent on capacitive parasitics. Their performances, even with respect to capacitive gyroscopes, are still to be explored when interfaced with integrated electronics.

A METHODOLOGY AND A TOOL FOR QOS-ORIENTED DESIGN OF MULTI-CLOUD APPLICATIONS

Giovanni Paolo Gibilisco - Supervisor: Prof. Danilo Ardagna

One of the most pervasive changes happened in recent years in the ICT world is the appearance on the scene of cloud computing. The main feature of this new computing paradigm is its ability to offer IT resources and services in the same way fresh water or electric power is offered, as a utility. In a traditional environment, in order to make use of a software system to address some business needs, a company would need to acquire and manage the hardware infrastructure, different software stacks and, in many situations, develop their own software. Cloud computing changes this paradigm by offering all these elements as services that the user can acquire and release with high flexibility. Another key advantage brought by the adoption of the cloud paradigm is the shift of responsibility in the management of the portion of the software system that is acquired from the cloud provider. If, for example, a company decides to decommission its own physical infrastructure in favor of a new infrastructure offered by a cloud provider all the maintenance operations required by the hardware and some software systems, like OS acquisition and update, are delegated to the cloud provider. This allows the internal IT team of the company to focus on tasks that provide more value for the company.

Delegating management responsibility of part of the infrastructure to a third party, in this case a cloud provider, comes inevitably with a loss of control on the entire infrastructure. This change creates some new challenges for teams used to build entire software systems from the ground up. When faced with the selection of a cloud service the application developer has to take into consideration many new characteristics that he/she was probably not considering before. The wide variety of similar services, the lack of interoperability between APIs offered by different cloud providers and the lack of specific training for developers are just a few of the new challenges that the IT staff of a company has to face when considering a migration to a cloud infrastructure. My PhD work tries to address some of these issues by focusing on the support to the development of multi-cloud enabled applications with Quality of Service (QoS) guarantees. It embraces the model driven engineering principles and aims at providing development teams with methodologies and tools to assess the expected QoS of their applications early in the design stages.

To achieve this goal we adopt and enrich different component based and UML-like modeling technologies like the Palladio Component Model and MODACloudML extending them in order model deployments in a cloud environment, by introducing a new meta-model, with the goal to evaluate the application QoS and determine the optimal deployment in multi-cloud scenario. The integration of the new meta-model into state of the art modeling tools like Palladio Bench or Modelio allows software architects to use well known modeling approaches and specify a cloud specific deployment for their applications.

In order to ease the portability of both the model and the application, the meta-model uses three abstraction levels. The Cloud enabled Computation Independent Model (CCIM) allows to describe the application without any reference to specific cloud technologies or providers. The Cloud Provider Independent Model (CPIM) adds the specificity of some cloud technologies introducing concepts like Infrastructure and Platform as a Service (IaaS/PaaS) but still abstracts away the specificity of each particular provider. the Cloud Provider Specific Model (CPSM) adds all the details related to a particular cloud provider and the services offered allowing to automatize the deployment of the application and generate performance models that can be analyzed to assess the expected QoS of the application.

High level architectural models of the application are then transformed into a Layered Queuing Network performance model that is analyzed with state of the art solvers like LQNS or LINE in order to derive performance metrics and asses QoS. Software architects can use the result of the evaluation to refine their design choices. Furthermore, the approach automates the exploration of deployment configurations in order to minimize operational costs of the cloud infrastructure and guarantee application QoS, in terms of availability and response time. In the IaaS context, as an example, the deployment choices analyzed by the tool are the size of instances (e.g. Amazon EC2 m3.xlarge) used to host each application tier, the number of replicas for each hour of the day and the split of the incoming workload among all the available providers.

We have analyzed the problem of finding the optimal deployment configuration from a mathematical point of view and show that it is NP-hard. For this reason, we have proposed a heuristic approach to effectively explore the space of possible deployment configurations. The heuristic approach uses a relaxed formulation of the problem based on M/G/1 queuing models to derive a promising initial solution that is then refined by means of a two level hybrid heuristic algorithm. Solutions derived by this algorithm are validated against the LQN performance model to assess their QoS.

We have validated the proposed methodology by means of two industrial case study in the context of the MODAClouds project. These case studies used the tool developed in this work in order to evaluate different deployment configuration and reconsider architectural choices in order to better exploit characteristics of the cloud environment, like elasticity. Furthermore, we have performed a scalability and robustness analysis and showed that our heuristic approach allows reductions in the cost of the solution ranging from 39% to 78% with respect to current best practice policies implemented by cloud vendors. The scalability analysis shows that the approach is applicable also to complex scenarios since the heuristic algorithm was able to find the optimized solution of the most complex instance in approximately 36 minutes for a single cloud deployment and in 46 minutes for a multicloud scenario.

STOCHASTIC MODEL PREDICTIVE CONTROL WITH APPLICATION TO DISTRIBUTED CONTROL SYSTEMS

Luca Giulioni - Supervisor: Prof. Riccardo Scattolini

The goal of the Thesis was twofold. Firstly, it dealt with the analysis and the development of Stochastic Model Predictive Control algorithms (SMPC) for linear discrete-time systems with additive stochastic disturbances and probabilistic constraints on the states and the inputs. Secondly, it considered the development of distributed Model Predictive Control algorithms for uncertain linear discrete-time systems and extended the techniques described in the first part to the distributed framework.

Part I – Stochastic MPC

Stochastic Model Predictive Control (MPC) is nowadays a standard in many industrial contexts, due to its ability to cope with complex control problems and to the availability of theoretical results guaranteeing feasibility and stability properties. These reasons have motivated the many efforts devoted to develop MPC algorithms robust with respect to unknown, but bounded, disturbances or model uncertainties.
The problem of designing robust deterministic MPC schemes has many solutions, however, these algorithms can suffer from some issues. Firstly, feasibility, convergence, and stability properties are usually achieved by resorting, implicitly or explicitly, to a worst-case analysis, which may turn out to be very conservative or even impossible in the case of unbounded uncertainties and it may require the solution to difficult on-line min-max optimization problems, that are computationally very demanding. Secondly, the tuning of robust algorithms, for instance the off-line procedures required to compute robust positive invariant sets, can be computationally demanding or even impracticable for systems of medium-high order.
Thirdly, and perhaps most importantly, they do not consider the possible a-priori knowledge of the statistical properties of the disturbances, i.e. their distribution function, which can be assumed to be available in many problems. In this setup, if the uncertainties or the state and control disturbances are characterized as stochastic processes, constraints should be reformulated in a probabilistic framework.
Starting from some pioneering papers these reasons have motivated the development of MPC algorithms for systems affected by stochastic noise and subject to probabilistic state and/or input constraints.
Stochastic MPC (SMPC) has already been considered in several application fields, such as temperature and HVAC control in buildings,
power production, management, and dispatch in systems with renewable energy sources, cellular networks management, driver steering, scheduling, and energy management in vehicles, path planning and formation control, inventory control and supply chain management, portfolio optimization and finance.
In spite of this large number of applications of SMPC and the already available theoretical results, many tough challenges emerge in this setup, related to the development of methods with guaranteed stability and feasibility properties.
Indeed, while for bounded disturbances recursive feasibility and convergence can be established both in a deterministic or a stochastic setup, the more general case of unbounded noise poses more difficulties and some specific solutions and reformulations of these properties have been adopted.
Due to the large variety of available SMPC techniques, the literature on the topic is quite vast and sometimes not very consistent. Thus, the first part of this Thesis has been devoted to discuss the main key points and the many possible formulations of SMPC algorithms with the aim of clarifying the most important aspect about the topic. Moreover, an algorithm, denoted as probabilistic-SMPC or p-SMPC, has been proposed and its properties and applications thoroughly discussed.

Part II- Distributed MPC

In recent years, another very significant topic is represented by the development of distributed control algorithms, that allow to handle large-scale systems whose complexity is related to the presence of a high number of small or medium-scale subsystems interacting via inputs, states, outputs or constraints.
As discussed before, due to its flexibility, robustness, and the vast literature on the topic, among all the possible solutions, particularly interesting appear to be those based on Model Predictive Control.
The most important aspect, when controlling large-scale systems in a distributed framework, is that using local MPC controllers, the predicted trajectories of inputs, states and outputs are directly available and can be used as information to be transmitted to other local controllers to coordinate their actions. This data exchange can greatly simplify the design of a distributed control system and can allow one to obtain performances close to those of a centralized controller.
The goal of the second part of this Thesis has been to extend the SMPC technique proposed in the first part to the distributed framework. In particular, the first addressed problem has been the control of dynamically coupled linear systems with additive disturbances and subject to local probabilistic constraints. Then, the problem of designing a distributed controller for tracking reference signals has been discussed and a solution based on the well-established DPC algorithm, has been presented. The main idea of DPC is that, at every sampling time, each subsystem transmits to its neighbors the reference trajectories of its inputs and states over the prediction horizon and guarantees that the real values of their inputs and states lie in a specified invariant neighbor of the corresponding reference trajectories.
In this way, each subsystem has to solve an MPC problem where the reference trajectories received from the other controllers represent a disturbance known over all the prediction horizon, while the differences between the reference and the real values of its neighbors inputs and states can be treated as unknown bounded disturbances to be rejected. To this end, a robust tube-based MPC formulation has been adopted and implemented using the theory of polytopic invariant sets. Finally, the problem of tracking a reference signals in presence of probabilistic constraints has been discussed for dynamically decoupled subsystems and the presented algorithm applied to a mobile robot coordination problem.

Several extensions to the work presented in this Thesis are possible.
Concerning the p-SMPC algorithm, an approach for enforcing hard constraints on the input, while preserving the nice feasibility and convergence properties, should be studied and the effect of different control law parametrizations needs to be investigated.
Moreover, the proposed approach can be extended to encompass different uncertainty models, for example the case of systems affected by multiplicative disturbances, or different type of stochastic constraints, for example the so-called joint chance-constraints.
Concerning the topic of Distributed Model Predictive Control, while for the robust deterministic case results are well-established, in the stochastic setup a lot of work has still to be done. The stochastic distributed control algorithm for regulation should be revised, less conservative solutions need to be defined to handle the couplings and stronger theoretical results should to be derived. As for the Stochastic Distributed Control algorithm for tracking, the extension to the case of dynamically coupled systems could be the next step.

PRACTICAL ASPECTS OF MCSEM DATA PROCESSING AND IMAGING

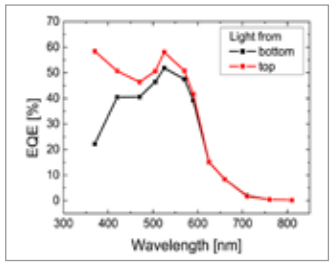
Andrea Gola - Supervisor: Prof. Giancarlo Bernasconi

The subject of my major research topic is in the field of signal processing for geophysical exploration, mainly for oil&gas discovery.
In recent years a new approach is gaining more popularity in the geophysical industry, that of *multi-physics* integration: the traditional and well studied imaging tools based on seismic data are increasingly complemented by results from other kinds of data - mainly gravity and EM – or even jointly processed before interpretation , with the aim to have a better understanding of different subsurface physical properties.
Along with EM methods for shallow terrestrial investigations (e.g., GPR), well known and studied, the technological development allow to extend EM sounding to deep investigations and in offshore scenarios. This dissertation is focused on mCSEM (marine Controlled Source Electromagnetic Method) data , used for large scale investigation with the aim of recovering , after proper processing steps, a subsurface electrical resistivity map to be used to complement other kinds of data, seismic images in first place. It has been proven that features of the estimated resistivity distribution contribute to identify either hydrocarbon accumulations or particular lithologies, greatly reducing exploration risk.
Being a relatively new method exploited by the industry, mCSEM has great potential but also present some peculiar characteristics that must be correctly taken in account, namely the inherent coarse spatial resolution, especially in depth, highly non-linear sensitivity with respect to model parameters and the underlying physics of the diffusion of very low frequency EM signals in conductive media.
I have explored different aspects of mCSEM data analysis, imaging and interpretation.
Firstly, a 2.5 dimensional, numerical sensitivity analysis conducted analyzing modeling results , exploiting an accurate frequency domain Finite Element Method (FEM) simulator has permitted to highlight the data sensitivity of the two measurable EM field components (at least in the considered, standard acquisition setup – Sea Bed Logging, i.e. inline electric field and crossline magnetic field) to high resistivity anomalies parameters, their depth and extension (horizontal and vertical). Main results are that the high sensitivity to lateral discontinuities is paired with a reduced one in vertical direction and a saturation-like effect with respect to resistivity and a rapid decaying with depth; the latter effects suggesting the method's potential in detecting even small resistivity variations but leading to the known problems that afflict mCSEM depth imaging.
Second, we have proposed a method for "quick&dirty" interpretation of underlying resistivity model from data and a reference response computed assuming knowledge of acquisition geometry and an hypothesis on the background resistivity trend (without the anomalies). The output of this processing step is called the *pseudo-image* and allow detection, inline localization and a first estimate of resistivity anomalies lateral extensions.
Analysis of pseudo-images variations with frequency made possible also to infer information on depth of anomalies, although in a very qualitative manner.
Furthermore we have compared pseudo-images with other scalar indicators known in literature, highlighting pros and cons, similarities and differences of each one; I have verified that the integrated analysis of these fast methods allow to strengthen or formulate further hypothesis on the unknown model, thus helping interpretation. An example is showed in figure 1 in which the application on a real dataset (provided by a major E&P company) detects and highlight the main resistivity model features such as abrupt discontinuities and spatial evolution of the mean trend along the acquisition line.
Third, a migration imaging technique inspired by known algorithms in reflection seismic, is investigated and applied in order to achieve satisfying localization in depth of the anomalies, although an accurate, direct resistivity estimation from this approach is not possible; instead the resulting images represent the spatial locations of apparent resistivity contrasts, i.e. model features expressed as measured EM field deviations from the background field computed on the same model assuming a known resistivity model.
As known in specialized literature, besides the very low frequencies involved, that greatly limit resolution, migration (wave focusing) in high attenuation media such as highly conductive, water-bearing sub-seafloor sediments is a difficult task; intrinsic attenuation due to absorption breaks the symmetry in time of the backward propagated field from the receiver array, leading to a resolution loss, image degradation (focusing pattern with lower contrast) and localization errors of anomalies in the image. The spatial resolution achievable by this migration imaging approach in typical acquisition setup and signal spectrum is experimentally evaluated on synthetic data and turns out to be of the order of some kilometers in horizontal direction; vertical resolution is heavily impaired by transmitted signal characteristic in frequency domain that yields a Point Spread Function of the imaging system (negative sidelobes) that tends to cancel response from anomalies that are near in depth.
I have applied the proposed migration method also on real data, obtaining encouraging results since main large scale-resistivity anomalies are correctly located . The output images provide information on resistivity model geometry, at least its top boundaries, and could serve as starting model for computationally-intensive inversion, helping counteract convergence rate and non-uniqueness issues.

INTEGRATION OF DIRECT-WRITTEN ORGANIC PHOTODETECTORS AND ORGANIC TRANSISTORS: TOWARDS PASSIVE PIXELS FOR PLASTIC LARGE-AREA IMAGERS

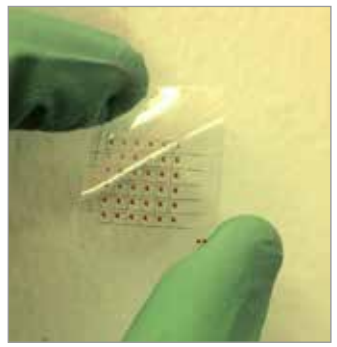
Andrea Grimoldi - Supervisor: Prof. Dario Andrea Nicola Natali

In the past few decades organic electronics has raised considerable interest because of solution processability, mechanical flexibility and tunable optical capabilities of organic semiconductors. In light-exploiting applications, organic semiconductors show outstanding characteristics with their high absorption coefficients and typical luminescence in the visible range of wavelengths. The development of solution based technology paves the way for large area, low cost and flexible electronics. Printing techniques mutated from graphical arts can also successfully cope with the need of patterning capability required to fabricate complex optoelectronic systems. Among them, inkjet printing allows fast prototyping and reduced waste of material.
The realization, characterization and optimization of organic inkjet printed vertical photodetectors (PD) aiming to their integration with switching elements to build an organic imager was the first topic of my research. Large area, non-fragile imagers are devices of great attractiveness for X-ray medical and industrial diagnostic, but a fully-direct-written matrix has not been demonstrated yet. Concerning these large area imagers, specific performances are requested for the light sensing element: external quantum efficiency (EQE) above 25%, reverse dark current density around 10 nA cm^{-2}, capacitance of 1 – 10 pF. I performed a careful optimization of inkjet printed photodetector prototypes. Thanks to the development of a properly functionalized bottom electrode, I was able to realize the first example of a fully-inkjet-printed, all-organic and semitransparent photodetector with EQE of 42 % (Figure 1) and dark current density of 3 nA cm^{-2} at 1 V reverse bias (best device), which results in specific detectivity reaches 6.2 × 10^{12} Jones.



1. External quantum efficiency of fully inkjet printed, semitransparent, organic photodetector.

These performances fulfill the specifics for a large area imager. I deeply investigated the non-trivial task related to the integration of the photosensitive element realized with an addressing, laser-ablated thin film transistor (TFT). Specifically, I identified a solvent for the spin-coated gate dielectric material compatible with the PD yet capable of yielding dielectric layers suitable to TFTs. I fabricated on the same substrate the PD with bottom electrode connected to the source terminal of a direct-written TFT. I verified the functionality of PD as photosensitive element and TFT as addressing one. I realized PDs with suitable dimensions of the active area in order to minimize the effect of disturb caused by the coupling of the gate signal through large parasitic capacitance. To demonstrate the versatility of this approach, I explored alternative pixel architectures stacking the photodetector on TFT dielectric or using an inkjet-printed diode as addressing element. The latter structure opens the way not only to the simplification of the process flow, but also to the fabrication of a crossbar array of pixels with high geometrical fill factor. Results were comparable to the one reported in the literature for which the range of detectable photons per frame goes from 10^6 to 10^8. Finally, I fabricated and characterized a direct-written organic imager taking care of the issues related to the realization of many pixels on the same substrate one next to the other, their electrical interconnection, the proper electrical insulation between conducting lines belonging to different planes and the driving signals characteristics for a proper readout (Figure 2).



2. Fully direct-written imager integrating organic photodetectors and organic transistors.

Side-aim of my work was to develop photodetectors with reduced dark current, extended spectral response or integrated in complex electronic circuits.
In organic photodetectors electrode/semiconductor interface plays an important role determining the amount of carriers injection and consequently the dark current. This is a key figure of merit that can enable or not the integration of a light sensor in system with demanding signal-to-noise ratio requirement. Moreover, interface traps can affect the response speed of the device. These considerations encouraged me to work on devices with a newly synthesized interlayer. A double aim was sought for the new layer: to reduce dark current and to improve the wettability of the top water-based electrode on a hydrophobic photoactive material. I fabricated and electro-optically characterized detectors with the aforementioned new interlayer.
Small molecules have some advantages with respect to polymers in particular they are not affected by polydispersity and are characterized by a simpler purification technique. I found out an approach that allows small molecule printing and I exploited it in order to extend the photoresponse to the near infrared range of wavelengths obtaining a devices sensible from 470 to 750 nm. EQE in excess of 10 % was achieved up to 710 nm both when light impinges from the top and the bottom side of the device.
In order to integrate a photodetector with an inkjet printed organic semiconductor-based Schmitt trigger typically operating with voltages of tens of volts, I selected a planar structure for the photosensitive element. I realized the photodetector with inkjet printing only and verified that characteristics fulfilled requirements. After having realized a circuit integrating the trigger and the photosensitive element (Figure 3), I validated the possibility to operate this system as dusk sensor by full-system measurements reproducing the exposure to sun that occurs during a day.

3. Optical micrograph of the fully inkjet printed dusk sensor comprising organic Schmitt trigger and planar organic photodetector.

OPTIMAL ENERGY MANAGEMENT OF SERIES HYBRID ELECTRIC VEHICLES

Jacopo Guanetti - Supervisor: Prof. Sergio M. Savaresi

Electrification in vehicles is a significant trend in recent years, motivated by economical reasons, improvement of fuel economy and reduction of CO_2 emissions. The range of technological solutions is widespread.
Micro and mild hybrid powertrains have no significant architectural difference from conventional vehicles; however, using Start/Stop systems, engine coasting systems and possibly some limited energy regeneration, they improve fuel economy with low impact on the vehicle purchase cost.
Full HEVs have a high degree of hybridization and a large battery pack, that allow both pure electric and pure thermal driving, as well as hybrid modes as ICE boosting and battery recharge. It is possible to achieve very significant improvements of fuel economy through an intelligent exploitation of the hybrid powertrain. However, projections are more favorable to plug-in HEVs and Extended Range Electrical Vehicles (EREVs), that compensate the increased technical complexity and costs with higher flexibility of use.
In plug-in HEVs, the battery can be recharged from the grid and the battery pack is typically larger than in HEVs, allowing to drive on electricity for significant distances. EREVs can be seen as a special kind of plug-in HEVs, conceived as EVs with an additional power unit that acts as a Range Extender. Pure EVs are currently appealing, due to absence of local emissions, low price of electric energy, good dynamic performance and low noise; they represent an excellent solution for urban mobility, both technically and economically.
A drawback is the battery technology, that is rapidly evolving but still offers limited energy and power density: as a consequence, the All Electric Range (AER) of EVs is pretty limited and the recharge time is far longer than the time to refuel in conventional vehicles. Range Extenders are therefore regarded as an enabling technology to push the spread of EVs, allowing occasional long trips. This motivates the interest for the series hybrid architecture, that EREVs implement. Despite the simplicity of this solution, in urban usage the Range Extender can represent a useless weight and encumbrance. Proposals and prototypes of plug-in Range Extenders have therefore been developed, in order to make it a component that can be attached to the car only when necessary. While Range Extenders are a solution to occasionally extend the use of EVs beyond the urban context, light EVs are a very convenient solution for mobility within the urban context. In addition to all the advantages of EVs, they are easy to use in congested areas and, due to the low weight, can use the energy stored on board in an efficient way. Electrical Power Assisted Bicycles (EPACs) can be seen as a special class of light HEV, that combine the electrical energy stored in the battery with human power. A natural goal is to actively control these systems in a way that enhances the experience of the user, for instance reducing the perception of fatigue.
This thesis is devoted to the series hybrid architecture, that is the basis of several promising solutions for future mobility, both in passenger vehicles and EPACs. As series hybrid vehicles, they can employ two power sources to meet the power demand from the vehicle. This represents a degree of freedom, that is common to all hybrid vehicles and that is the key to improve fuel economy and performance in general. To exploit this degree of freedom, a so-called energy management algorithm is necessary, to dispatch the power request between the available sources. While this can be done quite easily with simple rules, the challenge is to design this algorithm in a way that optimizes some performance measure, taking into account the peculiarity of each vehicle.

The energy management problem is formulated for a generic series HEV, with the goal of minimizing the monetary cost incurred by the user. To this aim, three major cost items are considered, fuel, grid energy and battery storage. Simulations show that this approach can yield nontrivial results with current energy prices, and thus can be a valuable alternative to formulations in terms of energy consumption. Moreover, it represents a convenient way to account not only for energy consumption, but also for battery aging, which can significantly affect the overall value of the vehicle.
The Least Costly formulation is even more convenient when further cost items are summed up. For instance, the Least Costly concept is applied to a peculiar hybrid architecture: an EV with off-board, on-demand Range Extender. Since in this scenario the Range Extender is rented only when necessary, the rental fee is also included in the cost function. Moreover, the switching between EV and EREV operation can happen only at the rental stations: therefore, the problem is formalized as a hybrid OCP, which is significant different from more classical energy management problems.
Based on Pontryagin's Minimum Principle, the Least Costly control policy is derived both in implicit and in explicit form. The explicit policy can be seen as a rule based policy that gives guarantees of optimality; it is appealing because, if compared to the implicit policy, it substantially reduces the amount of computations performed online, although it is based on some model approximations. For this reason, the implicit and explicit implementations are critically compared, also by means of simulations.
Moreover, both the implicit and the explicit Least Costly policies are based on some approximation. A sensitivity study shows that these approximations do not affect sensibly performance. These results are also discussed in view of the casual implementation of the Least Costly policy, identifying possible sources of performance deterioration and proposing real-time frameworks to handle such issues, both for the generic series HEV and for the plug-in Range Extender. Comparisons with the theoretical optimum show good performance in any reasonable scenario.
While the energy management of EPACs is a far less studied topic than in HEVs, the peculiar powertrain investigated here – a chainless bicycle – makes the problem even more original. Two major innovative aspects are the presence of the human in the role of a power source, and the exploitation of the electric transmission. The proposed approach is to minimize the exertion perceived by the cyclist, while guaranteeing a minimum state of charge of the battery at the end of the trip. To quantify the exertion, an index of fatigue based on heart rate is defined, also based on the biomechanics literature.
The energy management problem is formalized as an OCP, that is tackled with different numerical techniques. Analogously to what is commonly done in HEVs, an approximated technique is proposed, based on Pontryagin's Minimum Principle. Introducing some approximations on the human part of the model, the approach provides good accuracy, reduces the computational burden and concentrates the knowledge of future driving conditions into a single tuning parameter. This approach is also applied in real-time, introducing a scheme to adapt the tuning parameter to the driving conditions. The depletion of the battery, allowed by the plug-in concept, and the use of partial route information known in advance, are discussed; the case of harsh elevation profiles is also analyzed. Comparisons with the acausal optimum show good performance, both in nominal conditions and, with limited deterioration, in non ideal conditions, i.e. when the elevation information or the heart rate measurement are not available.

A VISUAL FRAMEWORK FOR THE EMPIRICAL ANALYSIS OF SOCIAL INFLUENCERS AND INFLUENCE

Ajaz Hussain – Supervisor : Prof. Chiara Francalanci

Social media have become pervasive and ubiquitous and represent a source of valuable information. The literature on social media makes a distinction between "influencers" and "influence". The former are social media users with a broad audience and the latter instead used to refer to the social impact of the content shared by social media users. Our claim is that while the information shared by influencers has a broader reach, the content of messages plays a critical role and can be a determinant of the social influence of the message irrespective of the centrality of the message's author. This thesis starts from the observation that social networks of influence follow a power-law distribution function, with a few hub nodes and a long tail of peripheral nodes. In this thesis we investigate the relation between 'content' and dynamics of social 'influence', by dealing with issues connected with influence and influencers on social media. The thesis considers information shared by "influencers" and seeks to show that the role played by the content of the message is higher than the centrality of authors for the spread and the reach of the message. The specific research questions which we concern about are stated as under:

- Is content shared by influencers is bound to have high influence?
- Is content a driver of social media influence?
- What are the characteristics of content that help to increase social media influence?

The research aims to understand how influential content spreads across the network. For this purpose, identifying and positioning hub nodes is not sufficient, while we need an approach that supports the exploration of peripheral nodes and of their mutual connections. Our claim is that Influential content can be generated by peripheral nodes and spreads along possibly multi-hop paths originated in peripheral network layers. We provide a conceptual visual framework and related software tool to the assessment of influence and identification of influencers.

We exploit a modified power-law based force-directed algorithm to highlight the local multi-layered neighborhood clusters around hub nodes. The algorithm is based on the idea that hub nodes should be prioritized in laying out the overall network topology, but their placement should depend on the topology of peripheral nodes around them. In our approach, the topology of periphery is defined by grouping peripheral nodes based on the strength of their link to hub nodes, as well as the strength of their mutual interconnections, which is metaphor of k-shell decomposition analysis. A part of our research focuses on the connections among Italian tourist destinations in order to understand complex dynamics of tourism and destination branding. We propose a visual approach to highlight the local multi-layered neighborhood clusters around hub nodes in aesthetic graph layout. The approach is tested on a large sample of tweets expressing opinions on a selection of Italian locations relevant to the tourism domain. We put forward few hypotheses that tie *specificity*, *frequency of tweets* and *frequency of retweets* and are tested on data samples of roughly one million tweets. Based upon our proposed approach, visualizing the complex tourism destinations networks, and understanding the dynamics of tourism, reveals many aspects in order to better understand tourism phenomena. The empirical and visual evidence raises theoretical challenges and encourages further research to understand the relationship between content and influence on social media. The main innovative aspect of our approach is that we use statistics (hypotheses) and visualization together. One can visually verify the proposed hypotheses on graphs. Overall, results highlight the effectiveness of our approach, providing interesting visual insights on how unveiling the structure of the periphery of the network can visually show the potential of peripheral nodes in determining influence and content relationship. Further practical implications to the study can be helpful in *co-branding*, *brand fidelity*, *brand promotion* strategies and provides a visual platform to explore the complex tourism network where people talk about different brands or tourism categories.

Another part of our research dedicated to provide a novel visual framework to analyze, explore and interact with Twitter '*Who Follows Who*' relationships, by browsing the friends' network to identify the key influencers based on the actual influence of the content they share. We have developed *NavigTweet*, a visual tool for the influence-based exploration of Twitter friends' network. It helps to identify the key players, and follow them directly through the *NavigTweet*. The user can explore its own Friend-of-a-Friend (FOAF) network in order to find interesting people to be followed. The top influencers are identified by both user-level (e.g. number of followers, number of tweets, etc.) and content-based (number of hashtags, number of URLs, etc.) parameters. To gather some insight into the user experience with the pilot release of *NavigTweet*, we have conducted a qualitative pilot user study. We report on the study and its results, with initial pilot release.

This thesis targets the research in two correlated perspectives, one perspective of the thesis focuses on the complex dynamics of tourism network and destination (brand) analysis, and the other focuses on behavioral perspective of content-based influence browsing and exploration through a software tool - *NavigTweet*.

We provide a conceptual visual framework and related software tool to the:

- Influence Assessment (using network analysis of Tourism dynamics).
- Influence Maximization (by focusing on spread and reach).
- Influencers Identification

The findings presented in this thesis are relevant not only theoretically but also 'practically'. Understanding what variables impact on the dynamics of information on social media platforms, like Twitter, is a prerequisite for marketers who seem interested in devising efficient social media strategies and optimizing the ways they engage with consumers on these platforms. In fact, the main innovative contribution of this work is on the different perspective on influence using a visualization approach, which is not focused only on the centrality of the author, also visually identify major key players in the network, like information spreaders and information sources.

The results of a pilot test and subsequent large-scale test of *NavigTweet*, where we present the results of feedback questionnaire collected through the survey. We found that pilot participants were positive about the functionalities and features of the tools along with novelty of the idea itself, and received favorable comments concerning *NavigTweet*. The preliminary feedback that we have obtained suggests that *NavigTweet* identified top-influencers more accurately by understanding influential content provided by Twitter users. On the basis of content-based influence, both user-level and tweet-level parameters play a critical role in order to identify the top-influencers among the network. *NavigTweet* can help general users in order to understand the influence dynamics by providing a visual exploration platform, by which users can browse through their own and FOAF networks at unlimited depth-level friends' network. Users can maximize their influence, through content-based ranking and scoring by comparing with other friends' score. As the provided ranking is relative, and it also provides the general guidance upon perimeters, to maximize the influence spread and reach.

TRACE CHECKING OF QUANTITATIVE PROPERTIES

Srđan Krstić - Supervisor: Prof. Carlo Ghezzi

Software engineering has dramatically changed over the past decade and many of the changes have challenged our most basic assumptions about the nature of the software products that we develop. The most important realization is that modern software has a very complex interaction with the environment in which it executes and it is often not safe to assume that the behavior of the environment is stable. Designing software that anticipates changes in the environment makes the software itself exhibit dynamic behavior that can only be observed at run time. This asks for verification techniques that complement design-time approaches and puts forward trace checking as a viable complementary choice for verifying modern software. Trace checking is an automatic procedure for evaluating a formal specification over a trace of recorded events produced by a system after execution. The output of the procedure states whether the system behaves according to its specification.

The goal of this thesis is to develop general and efficient trace checking procedures that support a broad class of quantitative properties. Quantitative properties can be seen as constraints on quantifiable values observed in an execution of a system. They typically express non-functional requirements, like constraints on resource utilization (e.g., number of computation resources, power consumption, costs), constraints on the runtime characteristics of the environment (e.g., arrival rates, response time), or constraints on the runtime behavior of the system (e.g., timing constraints, QoS, availability, fault tolerance). To formally express quantitative properties, we utilize SOLOIST – a specification language based on metric temporal logic (MTL) with additional aggregating modalities.

The first part of the thesis proposes two algorithms that implement the satisfiability procedure for SOLOIST. An efficient satisfiability procedure provides a general framework for building a SOLOIST verification suite that supports many verification use cases including trace checking. The implementation is a translation that reduces the problem of SOLOIST satisfiability to satisfiability of a particular logic supported by the SMT solver theories. The major difference between the two implemented procedures is the logic targeted by the translation and how it encodes the satisfying assignment of SOLOIST formulae. In order to use satisfiability procedure to perform trace checking, we make use of a technique called bounded satisfiability checking (BSC). In principle, BSC uses satisfiability procedure of a language to perform model checking of a system whose formalization is provided in a descriptive form (i.e., using a logical formula). Similarly, we use BSC to perform trace checking by representing the appropriate system with its execution traces.

The two procedures mentioned above are complementary: one is tailored to efficiently check dense traces, while the other can efficiently check sparse traces. Finally, we extensively evaluate the proposed complementary approaches and apply them to an extensive case study in the domain of cloud-based elastic systems.

The second part of the thesis focuses on the problem of distributed trace checking. We have witnessed in recent years that the amount of collected data is growing rapidly and the so-called BigData management problem is becoming relevant in many different fields of engineering. The volume of the execution traces gathered for modern systems increases continuously as systems become more and more complex. For example, an hourly page traffic statistics for Wikipedia articles collected over a period of seven months amounts to 320GB of data. This kind of traces can neither be stored nor processed on a single computing node. Therefore the second part the thesis aims at providing algorithms that rely on existing distributed computation frameworks (like MapReduce and Spark) to efficiently check SOLOIST specifications over very large traces using multiple interconnected computation nodes.

The thesis also contributes to the state of the art in MTL trace checking by proposing a novel decomposition technique for MTL formulae that can be applied as a preprocessing step for any trace checking algorithm. It is known that the lower bound for memory complexity of MTL trace checking algorithms is exponential in the size of the time intervals of MTL formula. This means that checking formulae with large time intervals would require a lot of memory resources. The decomposition we propose provides a scalable way of trace checking formulae with large time intervals. However, due to the known restrictions of the standard point-based MTL semantics, we facilitate the decomposition by proposing an alternative semantics for MTL, called lazy semantics. It possesses certain properties that allow us to decompose any MTL formula into an equivalent MTL formula with smaller time intervals. We show that lazy semantics does not hinder the expressive power of MTL: we prove that MTL interpreted over lazy semantics is strictly more expressive than MTL interpreted over point-based semantics. In other words, any MTL formula interpreted over point-based semantics can be rewritten using an MTL formula interpreted over lazy semantics. Moreover, there are MTL formulae interpreted over lazy semantics that do not have an equivalent formula that can be interpreted over point-based semantics. We have integrated lazy semantics and formula preprocessing into our distributed trace checking algorithm and its evaluation shows that the proposed approach can be used to check formulae with very large time intervals, on very large traces, while keeping a low memory footprint.

ANALYSIS OF THE WIENER PHASE NOISE ISSUES IN OPTICAL TRANSMISSION SYSTEM

Silvio Mandelli - Supervisor: Prof. Maurizio Magarini

In optical communications, local oscillators and propagation introduce multiplicative phase noise that must be taken in consideration, estimated and compensated at the receiver. In the literature such channels are dealt with by considering a symbol-spaced discrete-time model where the transmitted symbol is impaired by both AWGN and a multiplicative phase noise given by a first order Wiener process. The issues given by such channels are objects of several works in the literature. The aim of this thesis is to discuss some of them and try to extend those dissertations. First of all, in the literature this model is assumed by considering "small" phase noise but nobody has ever discussed how much "small" it must be. A statistical and mathematical analysis is derived by the author, and a threshold of validity of the so-called Discrete Model is worked out, proving that the assumption is correct in almost all practical scenario and it is conservative in term of performance simulation. The analysis of phase noise channels is then deepened by studying Bayesian tracking techniques to extract all the information about the transmitted symbols. An iterative demodulation and decoding scheme is proposed and compared to others in the literature. The major gain is given by the greater spectral efficiency obtained by not transmitting Pilot Symbols and still working better than other considered similar schemes. Bayesian tracking allows also to derive the information rate of the considered Discrete Model channel and to verify that the proposed algorithm can achieve it. The focus is then moved to analyze short reach access optical scenarios where, for spectral efficiency and receiver sensitivity, OFDM has been considered instead of single carrier systems. For cost, footprint and power consumptions requirements, Direct Detection has several advantages compared to Coherent schemes since the multiplicative phase noise introduced by the transmitting laser can be neglected. However, if dispersive compensating fibers are not used, Chromatic Dispersion impairs the signal and the phase noise cannot be cancelled. The author has proposed the literature analysis of this phenomenon, enlightening its weaknesses and comparing the new results of the thesis with experimental measurements given by other authors.

OPTIMAL CONTROL OF LARGE SCALE STOCHASTIC HYBRID SYSTEMS WITH A FINITE CONTROL SPACE

Giorgio Manganini - Supervisor: Prof. Maria Prandini

In recent years, the interest for the class of hybrid and possibly also stochastic systems has grown, motivated by the large variety of applications where continuous and discrete dynamics are strongly interacting and are affected by uncertainty.

Modeling, analysis, and control design of such systems raise severe methodological questions. This thesis addresses optimal control design for discrete time Stochastic Hybrid Systems (SHS) characterized by a finite control space, and relies on the theoretical frameworks of Markov Decision Processes, Dynamic Programming and Reinforcement Learning to this purpose.

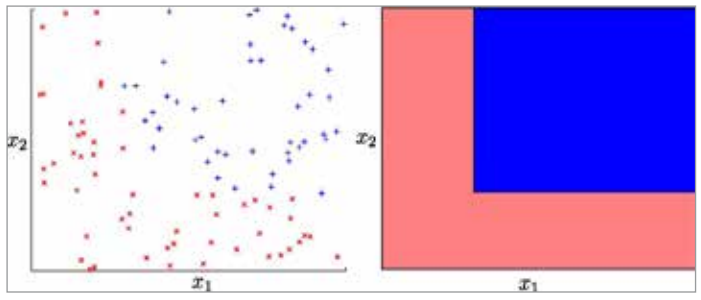
The main challenge herein is represented by the so--called curse of dimensionality, since, when the state dimension is large, classical approximate dynamic programming techniques may become computationally intractable, motivating the quest for efficient algorithms.

In this thesis we develop two control strategies for systems characterized by a finite control spaces and high dimensional continuous state spaces. Both the approaches share the same idea of effectively partitioning the state space into regions identified by the same optimal control action.

In the first part of the thesis we address the optimal control of discrete--time switched systems, described by a finite set of operating modes, each one associated with an affine dynamics, and with a random initial state. The objective is the design of the switching law so as to minimize an infinite--horizon expected cost, while penalizing frequent switchings.

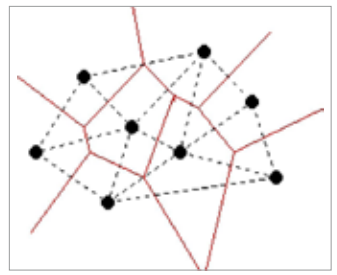
Starting from the observation that a control policy associates a mode to each state and, as such, can be viewed as a classifier, we propose a novel classification--based algorithm, called Two Stage Algorithm (2SA), that trains such classifier--like controller with a set of training data in the form of optimal state--mode pairs (see Fig. 1). In the considered setting, the generation of the training set involves solving a Mixed Integer Quadratic Program (MIQP) for each pair. The optimal switching law is computed off--line, which allows an efficient on--line operation of the system via a state feedback policy.

A key feature of the proposed approach is the use of a classification method that provides guarantees on the generalization properties of the classifier. Exploiting this peculiarity, we provide theoretical guarantees on the performance of the resulting policy: the sub--optimality of the approximate policy obtained by the 2SA can be made arbitrarily small, by suitably choosing the prediction horizon in the MIQP procedure and the number of sampled used in the training of the classifier. Interestingly, the number of samples required is independent of the state space dimension, which makes the algorithm amenable for large scale systems.



1. 2SA: pictorial representation

The algorithm is tested on a multi--room heating control benchmark problem to show the viability of the approach. Simulation experiments revealed that the 2SA can effectively approximate the optimal policy, and also results in an extremely fast on--line computational time.



2. . Voronoi and Delaunay diagrams associated with a set of particles

In the second part of the thesis we turn to the optimal control of more general discrete--time stochastic hybrid systems, and cope with the curse of dimensionality by searching for the optimal control policy in a restricted parameterized policy space. In particular we introduce a novel policy parametrization that adopts particles to describe the map from the state space to the action space, each particle representing a region of the state space that is mapped into a certain action through the associated Voronoi diagram (see Fig. 2). The locations and actions associated to the particles describing a policy can be tuned by means of a recently introduced policy gradient method with parameter-based exploration, called PGPE.

The main advantage of such parametrization is its ability to exploit a particle to represent an entire region of the state space. Note that, by changing the granularity of the particles over the state space, it is possible to achieve any desired accuracy in the policy approximation.

The task of selecting an appropriately sized set of particles is then solved through an iterative policy building scheme that adds new particles to improve the policy performance and is also capable of removing redundant particles. The iterative nature of the algorithm helps to alleviate its sensitivity with respect to the parameters initialization and to reduce the risk of getting stuck in local optima.

The computational complexity of the algorithm scales linearly with the problem size and the policy parameters. Moreover, a computational study on a classical benchmark example for SHS has shown that the number of samples required to estimate the gradient direction is not strictly related to the number of parameters (i.e., particles) and does not increase exponentially with the continuous state space dimension. Experiments on two benchmarks problems demonstrate the scalability and effectiveness of the proposed approach as the dimensionality of the state space grows.
Finally, we have investigated the use of second--order methods in place of gradient--based optimization, although not in the context of the proposed particle–based parametrization. We provided a method that estimates the Newton direction by sampling directly in the parameter space, together with a technique for variance reduction in the sample–based estimation of the Hessian matrix and a finite--sample analysis in the case of Normal distribution.
Future works will regard the application of this second--order approach to the proposed particle--based parametrization, and further developments may concern the investigation of Modified-- or Quasi-- Newton solution concepts.

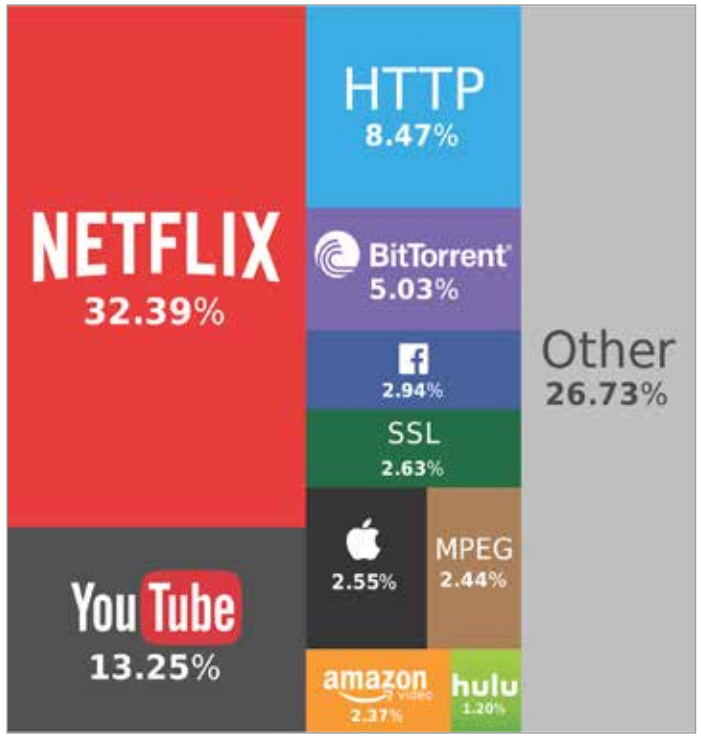
EFFICIENT IN-NETWORK CONTENT DISTRIBUTION: WIRELESS RESOURCE SHARING, NETWORK PLANNING, AND SECURITY

Michele Mangili - Supervisors: Prof. Fabio Martignon, Prof. Antonio Capone

It is undeniable that Internet, as we know it today, has undergone very deep changes through recent years: originally composed of only few nodes, its worldwide diffusion has been so remarkable that it quickly reached our households, and then became an indispensable means of communication available everywhere, at any time, in the palm of our hands. As a result of these momentous changes, there seem to exist a misalignment between the way people exploit the network today and the original design goals that drove the development of the Internet protocol stack.

Conceived in the late '60s, Internet was designed to solve the problem of resource sharing: in its dawn, the objective of the network was to let the scientific community remotely exploit the computational power which was very expensive at that time, and therefore the protocol stack was specifically engineered for a communication infrastructure. For this reason, IP packets contain the address of the remote location where the information can be found, although, on the other hand, nowadays Internet users care mostly about what content is carried by the packets themselves. In terms of bandwidth requirements, apps download, music streaming and photo sharing through social networks have all certainly had a remarkable impact, however, with no doubts the game changer has been video distribution, which nowadays accounts for more than 50% of the overall downlink traffic, in fixed access for North America.

To support the efficient and effective network distributions of digital contents, specialized technologies such as Content-Delivery Networks (CDNs) were specifically designed to attain this objective. Rather than making the origin server fulfill all the incoming requests, in a CDN copies of popular contents are replicated on surrogate servers deployed closer to the locations where users are actually demanding the data. Whenever a host requests one of these popular contents, the CDN infrastructure can transparently make its request be directly processed by a nearby CDN surrogate server, leading to two main advantages: (1) the origin server is offloaded and (2) the end-to-end latency is severely reduced, thus boosting the QoE especially for live streaming.



1. Global Internet Phenomena Report (2nd Half 2014). Per-application downlink traffic share, North America, Fixed Access. Source: Sandvine

CDN is a technology that was developed as an overlay on top of the TCP/IP protocol stack, in a way such that no dedicated support should be provided by the end-points, and yet, the lack of support from the network-layer protocol only confirms that these designs were an afterthought to make Internet content distribution scale.

At the network layer, from the early ages of Internet until nowadays, the IP protocol continues to survive, showing that the original Internet design goals were so general and universal, that not only there was not the need to evolve the protocol, but IP itself fostered the unprecedented development we discussed so far.

However, in recent years, the research community has started to question whether the misalignment between the protocol stack and users' needs should be resolved, and, to do so, it began to explore alternative research directions in which a better protocol stack is taken into account for the sake of boosting the efficiency of in-network content distribution.

This research effort is often motivated by the practical problem that the services provided by a content delivery network are usually very expensive, whereas remarkable savings would be obtained if we only could design a better network layer protocol. These type of solutions are known under the name of Information Centric Networks (ICNs), but we will often call them also Content-Centric Networks (CCNs), to emphasize the fact that they are specifically tailored to support the content distribution problem at its roots.

Despite that many different designs for ICNs have been formulated in the literature, they all share, as a common feature, the fact that they change the addressing space: the packets themselves contain the name of the content, rather than the host address of the machine providing the corresponding data. As a result of this choice, it becomes much easier to implement a distributed caching mechanism and, in some designs like NDN/CCN, potentially any network node (including intermediate routers) can behave as a cache, thus the content distribution capabilities of the network will be boosted.

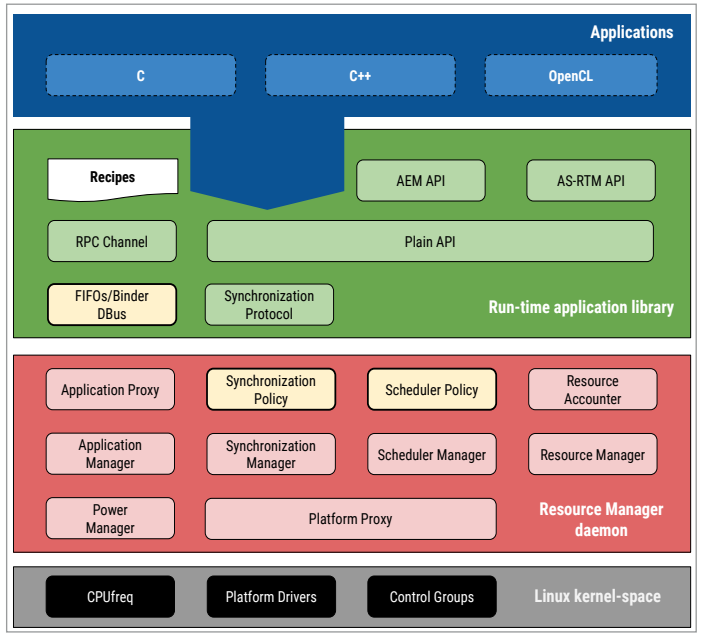
In this context, this Ph.D. thesis provides new contributions to the field, by tackling the general problem of supporting efficient digital content distribution. More in depth, we specifically focus on three complementary problems, namely: (1) wireless resource sharing, (2) network planning and (3) security aspects for content delivery. In particular, our contributions are the following:

- We extensively describe Named-Data Networking (NDN), the instance of ICN we take into account in this study. We introduce the NDN protocol stack, and we provide references to the state of the art concerning the security, routing and mobility properties of this paradigm.
- We take into account the wireless scenario, and we formulate a novel auction mechanism that is used to motivate wireless access point owners to lease their unused bandwidth and storage capacities, in exchange for economic incentives. This proposal can improve the content distribution capabilities of the network, while making the operator save significant amount of costs to run the servers of the distribution infrastructure.
- We study the centralized, off-line network planning problem, in the presence of distributed caches and we tackle (a) the optimal network design for the migration to an ICN, under a single-time slot; (b) we compare the performance bounds of a CDN with an ICN with evolving object popularity, and (c) we consider a virtualized CDN and study the stochastic network planning problem of one such infrastructure.
- Lastly, we consider the security implications on access control and content access trackability, and formulate ConfTrack-CCN, a cache-aware mechanism to foster the efficient enforcement of these security requirements, in an ICN.

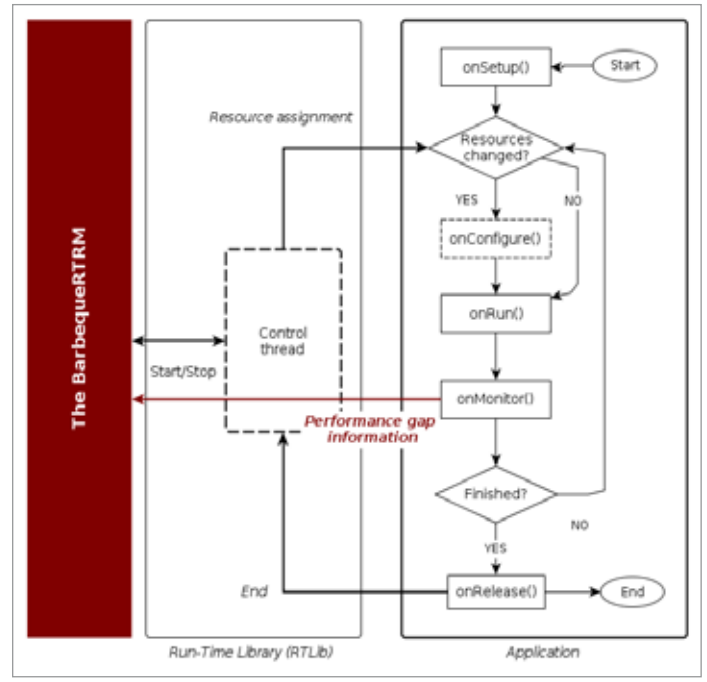
RUN-TIME RESOURCE MANAGEMENT OF MULTI/ MANY-CORE COMPUTING SYSTEMS

Giuseppe Massari - Supervisor: Prof. William Fornaciari

As VLSI chip manufacturers continues to deal with always higher density productive process, modern computing systems continues to grow in terms of delivered performance. However, the approaching of the physical limits of the silicon is leading us to face emerging problems in terms of power, thermal management and reliability. Before the CMOS integration process reached the 65nm (year 2005), the increase of performance delivered by the computing processors was also followed by a scaling of power consumption, proportional to the chip area reduction. In other words, the chip power density remained constant. This property, known as "Dennard's scaling law", does not hold for modern processors, featuring higher integration processes (e.g., 32, 22, 14 nm...). Accordingly, this increasing power density is also making always more challenging to find effective solutions to dissipating heat. As a consequence, modern processors' performance are bound by what is called the "power wall".
Although this trend, the processor industry is trying to prolong the life of the Moore's Law, and addressing the "power-wall" problem at the same time. The proposed solution had deeply affected the design of modern processors, increasing the complexity. first, by moving from single to multi-core processors, and then, by introducing many-core and heterogeneous architectures. What is worth noting is that whatever is the market segment, High-Performance Computing (HPC) or embedded/mobile systems, the computing platforms are now characterized by architectures featuring cluster-based topologies, processors with tens of processing units, and the availability of heterogeneous computing resources (e.g., CPUs, GPUs, FPGA). From the user side instead, the performance demand constantly grows, recently pushed by the explosion of the mobile devices market (smartphone and tablet). At the same time, this market field poses severe challenges in terms of energy efficiency improvement. Given this complex scenario, the problem of the computing resources management, and related task allocation, cannot be left uniquely in charge of a general-purpose operating systems (OS), any more. The outcome of this research has led to the development of the Barbeque Run-Time Resource Management (RTRM); a framework aiming at dynamically facing the aforementioned challenges, from a software perspective, integrating the control actions of the operating systems. More in detail, the framework aims at addressing the run-time problem of allocating system computing resources to the application workload, by taking into account multiple objectives at the same time, instead of managing them separately as the operating systems typically do. In Fig.1 it is shown layer view of the framework. Starting from the lowermost layer, it is possible to see how the BarbequeRTRM aims at being as much portable as possible, by building its resource management policies on top of already available resource management interfaces, provided by the Linux operating system. However, custom computing platform can be easily supported too, by providing suitable drivers. The core part of the framework consists of the "resource manager daemon", which is made by core components and plug-in modules. Such modules are for example resource management policies. This means that depending on the underlying target hardware platform, the resource manager can be configured to use a specific policy, hence addressing the specific criticality and weak points of a system. Finally, the upper layer highlights the presence of a run-time library, that the framework provides to expose the resource management services to the applications. Among these, the library aims at supporting the workload execution, by defining an execution model (Abstract Execution Model), though which the application can be notified about the amount of assigned resources, adapting itself accordingly.
The Abstract Execution Model is illustrated in Fig. 2. According to this model, the execution flow of the application can be rearranged into a set of "states". A control thread, completely transparent to the application, is in charge of managing the state transitions. In case of variations of resource assignment, the application is notified, such that it could reconfigure its parameter to match quality/performance objectives. On the other hand, a set of suitable interfaces (API) is provided, to allows the application to send requests to the resource manager, and collect information about the its real performance and requirements, at run-time. This constitutes a further helpful input information set, to drive the resource management choices, other than representing a remarkable difference with respect to the OS based resource management strategies.
The effectiveness of the framework, with its set of resource management policies, has been shown through several experiments performed on different real hardware platforms, targeting HPC and embedded mobile systems, running multi-threaded benchmarks and real applications. The experimental systems considered include single and multi-CPU homogeneous systems, a single CPU multi-GPU system, and an heterogeneous single CPU mobile-oriented platform. Currently, the framework is already available as an open-source project (BOSP: the



1. Layer view of the Barbeque Run-Time Resource Manager.



2. Application Abstract Execution Model.

BarbequeRTRM Open Source Project), including the resource manager, the application run-time library, several utility tools, benchmarks and sample applications.
In Fig. 3 it is shown is a use-case already exploiting the BarbequeRTRM. The system runs image processing applications and predictive models, to perform landslide detection. This critical task requires the hardware system to stay always on. The goal of the BarbequeRTRM is to manage the variability of the application performance requirements, under thermal and energy budget constraints, to guarantee such a requirement.

3. Landlisde detection system by HENESIS S.r.l. exploiting the BarbequeRTRM.

METHODOLOGIES FOR THE DEVELOPMENT OF CROWD AND SOCIAL-BASED APPLICATIONS

Andrea Mauri - Supervisor: Prof. Marco Brambilla

A new class of software applications, called crowd-based applications, is emerging. These applications use crowds, engaged through a variety of platforms, for performing tasks; the most typical application scenarios include fact checking, opinion mining, localized information gathering, marketing campaigns, expert response gathering, image recognition and commenting, multimedia decoding and tagging, and so on. Crowds have been used also for more creative tasks, including scientific discovery1 exploiting game-with-a-purpose (GWAP) approaches.
All these applications delegate to people the activities that are better performed by humans than by computers.
The common aspect of these applications is the interaction between the requestor (who poses a task), the system (which organizes the computation by mixing conventional and crowd-based modules), and a potentially wide set of performers (who are in charge of performing crowd tasks and are typically unknown to the requestor). The system may take multiple forms: in addition to crowdsourcing platforms (such as Amazon Turk or CrowdFlower) or question-answering systems (such as Quora or Yahoo! Answers), a recent trend is to use social networks (such as Facebook, Twitter or LinkedIn) as sources of human labor to be integrated in software applications, thanks to the availability of their programming interfaces.
Crowd-based computations undergo a new set of design principles and phases, as dealing with crowds introduces many peculiar aspects:

- Performers should be selected, possibly on the basis of their expertise on the task that they should perform.
- They must be reached, typically through an invitation system.
- They must be motivated through incentives, that include non-monetary ones, such as fun, self-esteem, altruism, visibility and reputation.
- Their work should be controlled, and in particular a task should be deemed as complete when it meets certain quality criteria, which take into account elapsed time, cost, and quality of result.
- Performers should be controlled, in particular by detecting malicious and/or incompetent performers, who should be banned from the computation.

Crowd-based computations can be part of a complex workflow and can be intertwined with conventional software computations; it is important to provide them with clear interfaces to the rest of the software, so that it becomes possible to discuss their requirements, design, testing, deployment, control, and maintenance, both from a local perspective and from a global, system-oriented perspective.
In this situation different issues arise: each type of scenario is characterized by a peculiar set of needs and requirements, that need to be mapped to the particular platform, that usually is not flexible, as it do not support a high-level, fine-tuned control upon posting and retracting tasks.
For instance if the requester wants to post and control a crowdsourcing task on Amazon Mechanical Turk he has to code the implementation with imperative and low-level programming language or using a specific framework.
If he wants to exploit the relations between people, he may prefer to use a social network as crowdsourcing platform. In this case the requester has to directly use the API provided by the social network.
The objective of this work is to develop methodologies for creating applications that leverage the knowledge of the crowd or social communities. The approach developed should be platform agnostic and allow the requester to create his application without having strong technical knowledge. This approach defines a design methodology, a specification paradigm and a reactive execution control environment for designing, deploying, and monitoring crowd-based modules. It is focused on data-centric applications, which apply structured human processing to large datasets; these applications represents the largest and currently most important class of crowdbased applications, as they include big data preparation, cleaning, and consistency checking.
The remainder of this thesis is structured as follow: Chapter 2 provides a detailed overview of the state of the art regarding crowdsourcing methodologies.
Chapter 3 describes the proposed model for building crowdsourcing application and the related design process.
Chapter 4 deals with the problem of controlling the work of the crowd, so that cost, quality and time constraints are satisfied.
Chapter 5 analyses the problem of building interoperable crowdsourcing applications, where the execution of the task is carried out upon different execution environments.
Chapter 6 presents a systematic approach to the design and deployment of crowd-based applications as arbitrarily complex workflows of elementary tasks
Chapter 7 addresses the problem of understanding how incentives influence the engagement of people performing a crowdsourcing task.
Chapter 8 describes the prototype (CrowdSearcher) that has been built in order to perform the experiments needed for the validation of this work.
Chapter 9 summarizes the results obtained and proposes possible future research directions.

ANALYSIS OF SECURITY ISSUES IN INFORMATION CENTRIC NETWORKING

Giulia Mauri - Supervisor: Prof. Giacomo Verticale

The Internet is rapidly changing. New problems are arising as a consequence of its architecture. When the Internet was designed, the main issue to solve was to connect two end points that were located far away in a fixed position. However, the Internet is growing in terms of its size and number of applications that run on it. Thus, this brings to the need of defining a new clean-state architectural approach for the new Internet. The research community is defining the properties and requirements of the Future Internet.

The current Internet is facing the increased traffic volume by using distribution technologies, such as P2P (Peer to Peer) and CDN (Content Delivery Network), that are based on a communication model of accessing data by name, regardless of the source location, and on employing caching and content replication. However, different content providers and P2P applications rely on their technologies without having a unified solution for content distribution over the Internet. Thus, it is not easy to optimize network efficiency and performance.

The Future Internet architecture will be probably defined by the Information Centric Networking (ICN) paradigm. Information is named at the network layer, thus making easier the content delivery to the users. Moreover, ICN solves some other issues of the Internet architecture such as mobility and security. Then, the in-network caching, that is one of the fundamental principles behind ICN, helps in improving the network efficiency and capabilities for information distribution. ICN is expected to evolve the Internet architecture by providing a network model more suitable for the current and future needs.

Firstly, the ICN paradigm uncouples information and location. Information can be located anywhere in the network but each information element is uniquely named. The receiver should know the name or the name's prefix of the information element it wants to retrieve. The network locates information in in-network caches. Then, the network is responsible for forwarding requests and responses on optimal paths.
The key concept of ICN is naming data objects: names are important for making forwarding decisions and for matching requests to responses. It becomes noteworthy the validation of the name-content binding: the content should carry the information requested with the corresponding name.

Secondly, the ubiquitous network storage allows every node to answer to the requests for the cached object without the need to verify the node authenticity. The in-network caching provides a lot of advantages: enabling sharing, making communication more robust, supporting retransmission, and fast reacting to disruption. Moreover, the in-network caching together with the communication model allows new options for transport services, new interconnection and business models.

In summary, the new ICN framework could provide a lot of advantages to today's Internet. Moreover, it seems the natural evolution of the latter.

The design of an efficient ICN architecture poses several technical challenges. A recent Internet draft presents an overview of the main open problems.
Figure 1 represents the open challenges in the Information Centric Networking scenario that will be presented in the following of this section and will be evaluated and solved along the chapters of this thesis.

1. The Information Centric Networking Open Challenges

Firstly, data object authentication is a fundamental ICN feature. Since data objects are replicated in network caches, they can be modified by malicious entities. Thus, ICN should provide a security mechanism to verify origin and integrity of contents. Then, it is also necessary to define a *trust management* infrastructure to distribute the publisher's public key to the consumers.
Moreover, data replication leads to a loss of control on content access and content dissemination. The content provider needs to know who accesses its contents, where and when its contents are used, and to revoke the access rights. The content provider also needs to update and synchronize its content.

Beyond data network security in terms of data confidentiality and integrity, the ICN domain introduces new *privacy* issues related to the protection of what data could reveal. The most relevant goal is the protection of data that could reveal information about an individual along with his or her physical, cultural, economic, social characteristics, or personal behaviors. Both the user requests and the cached contents have a unique name and can reveal a lot of information about the users. Meanwhile, these data are useful to improve the network performance, for example to define a caching policy based on user. Thus, it is necessary to find a trade-off between network performance and users' privacy.

The in-network *caching* brings along improved efficiency, better scalability, and increased network performance but also attracts new kinds of attacks such as cache pollution attacks by which a malicious content producer could control a massive amount of storage for spreading malware, junk, and other attacker controlled content at a low price.

Then, the communication model and data replication in the network caches should facilitate a seamless handover in a mobile scenario. A seamless transition in ICN ensures that the content retrieval does not suffer from intermittent connectivity. New challenges arise from the ICN mobile scenario. Especially, the *mobility* management should be coordinated between the network nodes and the users for optimizing caching policies and sizing.

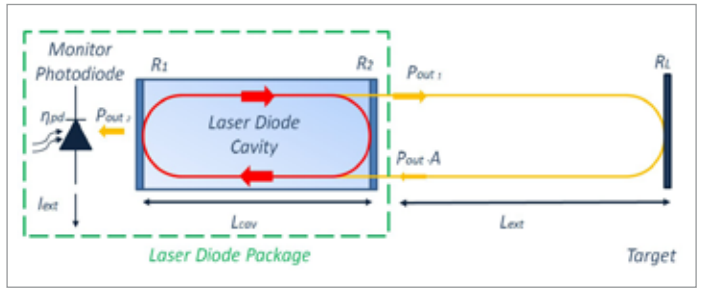
Highly related to mobility, there is the problem of *routing* based on names. ICN routing comprises name resolution, content discovery, and data delivery. There is not a common consensus on how to manage these steps and different solutions are provided in literature. However, specific challenges of ICN routing are still open.
The routing issues are strictly related to the *naming* convention. Indeed, two possible solutions have been proposed: hierarchical and flat namespaces. Each solution has its own advantages and drawbacks but also in this case there is not a definitive accepted proposal.

In this work, we cover some of the previous presented challenges in the Information Centric Networking scenario. All our proposals are original and, as far as we know, we are the first presenting such solutions.
Firstly, we consider the problem of trust management, deeply inspecting the issue of distributing and retrieving up-to-date keys. Then, we evaluate the content access control and staleness management.
Secondly, we describe the caching issues, in relation with users' privacy and cache pollution attacks.
Finally, we also provide an overview of some optimization models for a vehicular ICN scenario.

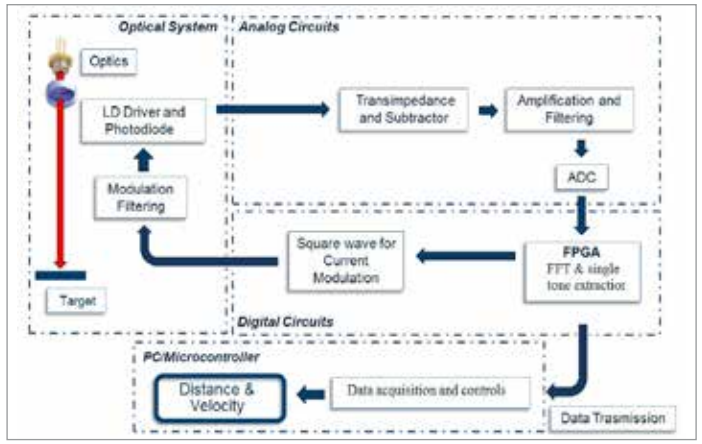
APPLICATIONS OF SELF-MIXING INTERFEROMETRY FOR SENSORS AND MEASURING INSTRUMENTS

Dario Melchionni - Supervisor: Prof. Michele Norgia

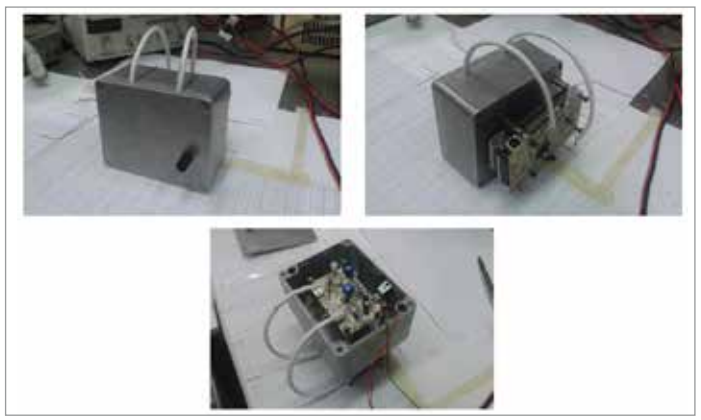
The purpose of this research project is to investigate the feasibility of contactless measurements using Self-mixing Interferometry. The reason for exploring this relatively new technique is the benefit of a simple setup. In addition, a good trade-off between production cost and obtainable accuracy, compactness and insensitivity to background noise creates many applications particularly for displacement measurements. This research work explores the potential of this optical effect and employs it to measure vibrations, absolute distance, liquid level, speed, holes depth, rotation, etc. Over the last few years Self-mixing Interferometry has taken hold as convenient alternative to other well-established optical measurement techniques. The novelty is represented by the combination of high accuracy, with compact, low-cost and user-friendly setup derived from the synthesis between physic phenomenon of Interferometry and recently discovered Back-injection configuration. In accordance with this, aim of the research is to develop new systems and instruments based on Self-mixing technique. There are two issues associated with this. The first issue is resolution: two methods are proposed to improving that. The first is a closed loop vibrometer which guarantees spatial accuracy down to few nanometers, high linearity, but operating range of about 100 µm. The other method exploits the frequency modulation component of interferometry signal and a fiber optic coupled laser to produce high sensitivity sensors. The FM Self-mixing signal detection is achieved using a *Mach-Zehnder* filter that maps the frequency variation into intensity one. The theoretical noise limit reachable is around tens of picometers. The second concerns about the development of novel measuring instruments: it includes the implementation of a device for velocity and absolute distance measurement. By means of a frequency-domain algorithm, it is able to retrieve both types of data simultaneously on a dynamic range of up to 2.5 m. The relative accuracy is 10^{-4} for distance and 10^{-3} for velocity. Another device studied and fabricated is an optical proximity sensor in the short range. Novel back-injection effect has been studied and used to reach continuous detection from 0 up to 80 mm. The prototype demonstrated good performances in terms of reliability and bandwidth. It is suitable for monitoring moving objects along a production chain by means of multiple detectors configuration. Finally another application is presented; liquid level measure is achieved by means of Self-mixing Interferometry. The contactless prototype designed by means of high speed electronic and custom optic is the first optical system able to monitor industrial filling process. Test campaign has validated the feasibility of this novel approaches. At present, the instruments and configurations described are well suited to carry out the measurements; however they represent just an initial stage: reliability will be improved in future works to meet industry-distributional constraints. Also, another problem is the laser source availability: few among many lasers tested show suitable behavior. In the future it would be interesting to fabricate custom lasers to solve this problem and enhance the performances.
In conclusion, this Thesis proves great qualities of Self-mixing Interferometry in many application fields. Even if there is a lot more that should be done, the author believes that it will be a suitable alternative to other optical measurement systems and, in some applications, even the only possible choice.



1.



2.



3.

DEALING WITH INCOMPLETENESS IN AUTOMATA BASED MODEL CHECKING

Claudio Menghi - Supervisor: Prof. Carlo Ghezzi

In the last few years, software systems overspread the human society. Software pervades every aspect of every day life: electronic banking, telephone and medical systems, are only few examples of highly computerized systems which are of common use. The massive development of software systems is continuously supported by the reduction of hardware costs, in particular memory costs, and the grow of the Internet which allows a constant communication between the software executed on the different devices connected to the network.
The software development industry is obviously affected by the dynamism of this technological environment. Developers are no longer talented individual people who write the whole software system in isolation. Conversely, nowadays software is developed by tens or thousands of programmers which interact, share code, ideas and components. Sometime developers do not even reside in the same physical location and use other software systems to communicate and integrate their work. In this setting, software development life-cycles evolve from being purely sequential and monolithic to iterative, incremental and agile. Instead of being obsessed by a complete elicitation of requirements, followed by a waterfall shaped development based on hierarchical teams of highly specialized engineers, in an agile approach requirements and solutions evolve through collaboration between self-organizing, cross-functional teams. These factors, in conjunction with the growing complexity of software systems and their increasing interaction, make software more vulnerable to errors and malfunctions. Errors may have devastating consequences. In 1994, the crash of a Royal Air Force (RAF) Chinook helicopter killed twenty-five passengers due to an error in the digital engine control system. In 1983, the warning system of the Soviet Union reported incoming US missiles from bases in United States. Luckily the officer recognized the false alarm and decided to do nothing. To prevent these errors, it is crucial to ensure that the software under development satisfies the properties of interest, i.e., its functional and non functional requirements.
Verification concerns a set of activities performed to guarantee that the software under development possesses the properties of interest. One of the most used verification techniques is testing. Testing requires to choose a representative set of input values that provide useful information about the behaviors exhibited by the running software and to check the correctness of the produced outputs. The set of test cases must provide enough evidence to give the developer confidence that the system is providing the desired behavior. This makes testing strongly dependent on the correct choice of the suite of test cases.
Another technique used to verify whether the software possesses the properties of interest is model checking. It is usually performed on a model which abstracts the behavior of the real system and usually reduces the risk of implementing a flawed design. Given a model of the system M and a formal property Φ, model checking exhaustively analyzes the state space of M to check whether all of the system behaviors satisfy Φ. The property Φ, usually expressed as a temporal logic formula, specifies the requirements the system must satisfy.
Model checking has matured to a stage where practical use is now often possible. It has already been used in practice in several application domains and has been adopted in many industrial settings. However, formal verification techniques are still not fully integrated with the current software development cycles, i.e., they do not support incomplete and evolving specifications.
The development process of any complex system can be viewed as a sequence of *design* decisions that make the system evolve from an initial, high-level model into a fully detailed and verified implementation. Typically, this process is performed by iteratively decomposing the model of the system into smaller functionalities. We would like the initial specification to allow a wide collection of possibly inequivalent implementations, which is constantly reduced during the design process, until one single implementation is determined. Thus, at each stage, the model may be deliberately *incomplete*, either because development of certain functionalities is postponed or because the implementation will be provided by a third party, as in the case of a component-based or a service-based system.
In the case of a postponed functionality, an implementation is usually provided at some later stage of the development process, possibly after exploring alternative solutions to evaluate their trade-offs. There are also cases in which the postponed functionality may become available at *run-time*, as in the case of dynamically adaptive systems. In this setting, the benefit of analysis performed using classical verification techniques, such as model checking, only appears at the end of a costly process of constructing a comprehensive behavior model, which contains a fully description of the system. Indeed, verification techniques usually do not support the verification of incomplete specifications.
My Phd work provides techniques that support the analysis of incomplete models designed during the software development. First, it proposes Incomplete Büchi Automata (IBAs) a novel modeling formalism that natively supports incompleteness. IBAs extend the well known Büchi automata (BAs) with unspecified states, called black box states, which encapsulate unspecified functionalities. Black box states can be (recursively) refined into other (Incomplete) Büchi automata. In order to analyze IBAs, we propose an *automata-based model checking technique* to verify if an IBA M satisfies a property Φ, written in Linear Temporal Logic. Due to the presence of black box states, the model checking procedure is modified to produce three different values: yes, if the model of the system satisfies Φ; no (plus a counterexample), if it does not; unknown, when the property is possibly satisfied, i.e., its satisfaction depends on the replacements, still to be designed, associated with the black box states. Whenever the property is possibly satisfied, a *constraint synthesis* procedure allows the computation of a constraint for the unspecified parts. A constraint concerns a set of sub-properties that must eventually be satisfied by the automata fragments (replacements) that will replace the black box states in the refinement process. The developer may use these sub-properties as guidelines in the replacement design. Finally, the work proposes a *replacement checking* procedure able to verify a replacement against the previously generated constraint. In this way, at each development step, only the new increment is considered in the verification activity.
The proposed approach has been implemented in the CHIA (CHecker for Incompete Automata) framework. CHIA is a prototype tool which supports the designer in the system development and its verification and it has been used to evaluate the approach over two practical examples.
The first is a classical computer science example and concerns the well known mutual exclusion system. The second concerns the evolution of a Pick and Place Unit (PPU). The PPU example is used to compare tools that analyze the evolution of automation system since it is a limited size example, but it provides a valuable trade-off between complexity and evaluation effort.
Finally, to analyze the scalability of the approach, in absence of a realistic benchmark suite, a set of random models with increasing size has been considered.
The evaluation compares the difference in terms of time and space between checking the replacement against the previously generated constraint (the corresponding sub-property) and the effort required to verify the refined model (the original model in which the new component is injected) against the original property Φ.

CHARACTERIZATION AND MODELING OF INNOVATIVE SOLID-STATE MEMORY TECHNOLOGIES

Halid Mulaosmanovic - Supervisor: Prof. Christian Monzio Compagnoni

The solid-state memory devices can be classically divided into two groups according to their ability to retain the stored data: volatile and non-volatile memories. While the former ones guarantee the retention as long as the device is powered and in the computer applications they allow for short-term data access, the latter ones typically meet a worst-case unpowered retention time of ten years and therefore allow for long-term data access. Furthermore, between the two types of devices exists an enormous performance gap, where volatile memories possess substantially shorter access times with respect to the non-volatile ones.
Over the past decade, both volatile and non-volatile solid-state memory devices have seen their physical size aggressively shrunk in order to meet the scaling process requirements. Although this led to a remarkable increase in device integration density, and consequently, to a reduction of the cost per bit, it pushed the respective technologies to their physical limits. On the one hand, Dynamic Random Access Memory (DRAM), as the most densely integrated volatile memory, is faced with ever increasing leakage currents, while the integration of the storage capacitor became the main issue for the ultra-scaled devices. On the other hand, Flash memory, today's leading non-volatile memory technology, is approaching the point where the discreteness of the matter and of the charge flows is severely impacting its performance.
The goal of this thesis is to investigate two innovative memory devices, one of volatile and the other of non-volatile nature, which appear to be promising candidates for the future replacement of DRAM and Flash technologies, respectively. Consequently, the dissertation is divided into two parts, with the first part dealing with the Thyristor-based RAM (T-RAM) cells as volatile and the second part with Hafnium-oxide based Ferroelectric Field Effect Transistors (FeFET) as non volatile elements.
T-RAM cell is a three-terminal device consisting of a nanoscale physical thyristor, featuring p^+–n–p–n^+ doped silicon regions and metal gate controlling by field effect the potential of its *p*-base. While its bistable static current-voltage characteristics is the starting point for the memory considerations, the real memory effect in the cell is achieved by dynamically modulating the hole concentration in the *p*-base. This concentration represents the physical parameter allowing information storage in the device: the presence of holes in the *p*-base leads the T-RAM cell to its high current (1) state during read, while a depleted *p*-base leads the cell to its low current (0) state.
On the other hand, FeFET is a semiconductor device featuring the same structure as the traditional transistor, with the exception of having a ferroelectric material embedded in the gate stack. It relies on the permanent polarization reversal of the ferroelectric to encode the two binary states. Recent discovery of ferroelectricity in HfO_2 thin films paved the way for a great scaling possibility of these devices, which, instead, is precluded for the classical ferroelectrics. This scaling led to the demonstration of the first functional 28 nm FeFET.

In Chapter 1, an overview of the main-stream volatile and non-volatile technologies, with a particular emphasis on DRAM and Flash technologies is provided. Then, novel memory concepts and their possible future applications are illustrated, discussing advantages and limits these devices are currently facing.

Chapter 2 introduces the gated-thyristor and its basic operating scheme. Here, its static bistable curve is analyzed, consisting of low-current and high-current regions. This leads to the identification of the main static parameters, such as switching voltage levels and the corresponding bistability window. Moreover, the central importance of its dynamic behavior is pointed out. In particular, the early-turn on of the cell, which is the most relevant phenomenon for the DRAM-like operation of the device, is studied by means of the dynamic sensing. It reveals that a premature device turn-on is possible for gate voltage rise times shorter than few microseconds and only if a high enough amount of holes is present in the *p*-base of the gated-thyristor. This feature makes the hole concentration in the *p*-base the physical element determining the memory effect.

A more detailed description of physical phenomena underlying the T-RAM cell operation, deduced from numerical simulations of the device, is provided in Chapter 3. Here, the electrostatics and charge flows in the device are carefully analyzed. It was possible to correctly reproduce both static and dynamic experimental behaviors of the cell.

Chapter 4 concludes the discussion on the volatile memory device providing an extensive experimental investigation of the T-RAM cell performance, including dynamic operation, reliability issues, data regeneration and disturb immunity. A trade-off between gate voltage biasing levels in terms of retention time of the two states was clearly identified and should be taken into consideration for future memory design. Proper biasing conditions could enable the stability of one state and the retention time of tens of seconds for the other one, which represents the net advantage over the conventional DRAM. Moreover, the cells were characterized by high endurance (cells were perfectly functional after 10^{10} write/read cycles).

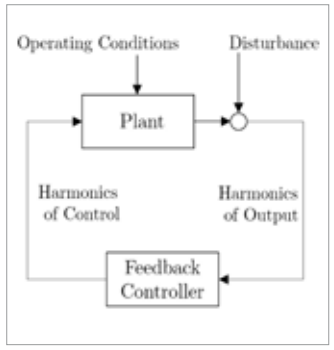
Chapters 5 and 6 deal with FeFET devices. In particular, Chapter 5 focuses on the introduction of ferroelectricity and of FeFETs as non-volatile memory devices. Moreover, a statistical analysis of the main device parameters in small NOR arrays is illustrated. Charge trapping phenomena and their impact on memory performance of FeFETs is discussed. In Chapter 6, a detailed analysis of switching phenomena leads to the evidence for single ferroelectric domain switching in ultra-scaled FeFETs. Using this finding, a possibility of multi-level storage is proposed and its initial feasibility is tested, highlighting the application possibilities of this innovative semiconductor technology in a future post-Flash scenario. Moreover, these results unveil a powerful and alternative way for studying ferroelectric switching physics.

In conclusion, the results in the thesis confirmed that although the two studied devices are in their initial research phase, they possess a considerable potential for replacing the conventional DRAM and Flash devices. Not only are they fully CMOS compatible, which would facilitate their path towards the semiconductor industry, but they also exhibit simple working principles and excellent performance features in view of their future memory applications.

ROBUST HARMONIC CONTROL FOR DISTURBANCE REJECTION: METHODS AND APPLICATIONS

Roberto Mura - Supervisor: Prof. Marco Lovera

Harmonic Control techniques aimed at reducing tonal disturbances have been extensively studied in the last few decades, with particular attention to a representation of the system as a linear model constructed in the frequency domain, the T-matrix model.
The precise knowledge of its elements is necessary for a proper functioning of the overall control system, and classical employed controllers resorting to the linear quadratic theory have not be framed to deal with model and parametric uncertainties or nonlinearities, possible causes



1. Feedback control for disturbance rejection

for degraded performance or instability of the closed-loop system. Adaptive control variants, coupled with a suitable offline or online identification method, have been employed in literature to handle this problem, but very little effort has been devoted to the analysis of the trade-off between robustness and adaptation in their deployment.

In this work, a discrete-time H∞ approach and a systematic methodology to the design of a robust Harmonic Control algorithm (Fig. 1) is proposed, for both SISO and MIMO system representations. The proposed control solution allows to account for model and parametric uncertainty in the control design problem, and provides a further benefit when dealing with the tuning problem, in particular when a multivariable plant is considered and different performance requirements are associated to the disturbance attenuation on each considered output. Indeed, specifications in terms of steady attenuation levels and desired transient performance can be directly incorporated in the robust problem statement. Moreover, the control design approach has been modified to deal with the explicit accounting for also predictable changes of the system, such as the ones induced by actuator characteristics and nonlinearities. Their role and effects on closed-loop performance and stability have been considered by resorting to analysis and synthesis frameworks, as the Describing Function and Linear Parameter Varying approaches represent. Both are introduced as modifications of the original T-matrix model.
While the first is used to analyze the cascade connection of a static nonlinear function with the plant matrix, the second allows to recast the problem both from a modeling and control design perspectives, combining the advantages of the robust control solution with the accounting of predictable plant changes handled within a gain-scheduling framework.
A validation of the proposed methodologies on a classical Harmonic Control application like the helicopter's rotor vibration problem has been included. Among the main problems affecting modern helicopters, vibrations generated by the main rotor are possibly one of the most important. Various derivations of Harmonic Control (HC) algorithms have been considered for many years as the most effective and valid approaches for the design and implementation of control laws aimed at reducing the rotor-induced vibrations and the improvement of rotor performance. In this context, the basic idea is to attenuate the vibratory components at the blade-passing frequency in the fuselage accelerations or in the rotor hub loads by adding suitably phased harmonic components to the rotor controls. Three case studies are proposed, including the general rotor-induced vibration problem, its variant based on semi-active lag dampers and the structural noise/vibration problem.

As a principal and most significant contribution to the state-of-the-art in the field of Harmonic Control algorithms for disturbance attenuation, this study presents a first comprehensive and detailed perspective of the robustness properties of the closed-loop system, including in particular:

A detailed systematic and automatic approach to design robust harmonic control laws for a SISO representation of the system, described as a linear (discrete) quasi-static model constructed in the frequency domain. Contextually, the control synthesis is obtained by defining a suitable optimization algorithm produced to automatically find control matrices coherent with classical control architectures;

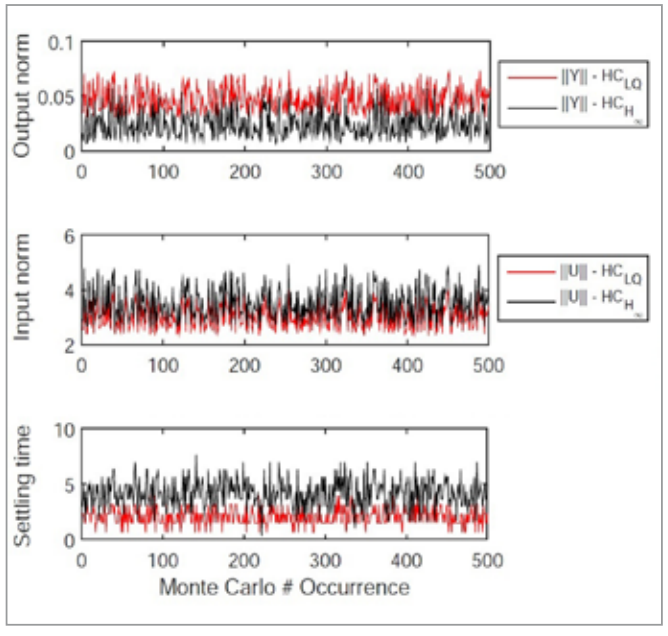
An extension of the provided robust approach to the multivariable case, by resorting to decentralized and decoupling strategies with the aim of eliminate cross-interactions which could degrade the closed-loop performance. A two-step decoupling strategy is proposed in order to obtain a compensator able to diagonalize the T-matrix plant and to use, similarly to the robust single variable case, the same instruments and tools;

An investigation on the application of a useful tool as the Describing Function in the T-matrix framework, in order to understand the role of loop nonlinearities on the closed-loop performance and stability. This in view of the fact that the Describing Function is very used as a standard mathematical tool for analyzing limit cycles in closed-loop controllers, and it naturally well fits in the T-matrix system representation;

A Linear Parameter Varying approach to the modeling and control synthesis of the system affected by both predictable and unpredictable uncertainties. In particular it is shown how the framework can provide a suitable means to account for variations in the plant such as the ones induced by, e.g, changing conditions or actuator nonlinearities;

A set of validation experiments of the proposed control strategies in the rotor vibration problem on helicopters. The Harmonic Control system is evaluated in three different case studies with the aim to confirm the results obtained with the numerical tests used when presenting the proposed control methods (Fig.2 and Fig.3).

2. . Agusta A109 MKII fuselage - experimental facility



3. A numerical example - stability and performance

METHODS AND TOOLS FOR EFFECTIVE SMART BUILDINGS DEPLOYMENT

Alessandro A. Nacci - Supervisors: Prof. Donatella Sciuto, Ing. Massimo Valla

Smart buildings are **living environments** that have the capability to **self-organize themselves** given some policies. Having a clear statement of what a smart building is a fundamental step since the smart-building term does not have a shared meaning for all the different interested stakeholders. In fact, the potential final users call smart a building that can be remotely accessed to turn devices on and off, even though there is in fact no actual automation involved. From the researchers and IT experts side, a building is smart when it is responsive to its inhabitants and it is able to adapt autonomously in sophisticated ways, e.g., using intelligent machine learning algorithms to predict user occupancy and control the heating system. In industry, smart is generally used simply as a marketing term to describe programmable technologies. Differently, in this thesis, a smart building is intended as **a distributed control system**, where dozens of distributed computation, sensing and actuation modules are exploited to increase the **comfort** and **safety** of occupants, while managing the building **energy efficiency** at the same time. From the architectural point of view, a smart building is a layered system composed of a **hardware level**, **a software abstraction layer** and a **user interaction layer**. More precisely, at the hardware level, a network of sensors is generally used to collect useful data (such as rooms temperature and humidity, users location, etc.): the sensor network is indeed usually composed of occupancy, temperature, luminosity, humidity, weather sensors and appliances monitors. An actuators network is then used to modify the building behavior. Typical actuators are the HVAC systems, lights, windows and smart appliances (which are defined as the appliances that can be controlled remotely). All such hardware is coordinated by a central Building Management System (BMS); some sort of human-understandable interface is then exposed to the buildings occupants troughs LCD panels, web browsers or mobile devices.

The smart building topic has been widely explored in the last 25 years, both form the academia and the insustry. State of the art works have discussed whether or how the so called **Ubiquitous Computing** have become available in everyday and how the advancements is this discipline would translate the smart-home from a vision to reality. In the last years, several challenges have been already addresses, especially from a technical point of view (low power sensors/actuators design, communication protocols, software interfaces design, etc.). These new technologies have also introduced new challenges: within this thesis work, some of the principal open challenges will be faced, proposing new theoretical frameworks, technologies and tools to move forward the actual implementation of smart buildings. In particular, after three decades of works on this topic, there are three interesting topics that has still to be faced to have smart buildings parts of our daily life.

The first problem is called the **Occupancy Detection Problem** (or Indoor Location Problem): knowing the position of people with a building is a crucial information, since it enables smart behaviors like lighting on the room when someone is inside or shutting the lights off when nobody is. Nowadays, a good occupancy detection technology, cheap but accurate at the same time, is still missing, making this one of the most relevant hot topics in all the ubiquitous computing conferences during the last years. With respect to this topic, this thesis experiments the usage of the iBeacon technology and the Bluetooth Low Energy protocol as a baseline tool for a novel occupancy detection technique.

The second problem is the **Building Programming Problem**: as said, the main contribution given by academia and industry have been basically focused on providing the enabling hardware technologies for the development of the smart building concept. In fact, in order to achieve the aforementioned high level goals (comfort, safety and energy efficiency), basic sensing infrastructure or means to actuate home appliances have been developed, solving also issues of interoperability, reliability of domestic technologies, ambiguity in sensing and other low-level problems. Thanks to this work, nowadays we have a plethora of different hardware solutions that enables the creation of an actual smart building from a theoretical point of view. Unfortunately, this is not actually true, since in order to have an actual smart building, it is necessary to coordinate all these components. The coordination of all these components is not trivial and needs some pieces of software that express the so called management policies of the building complex distributed system. Up to now such policies were written by technical people, the building-managers, that, thanks to their knowledge were able to accomplish this task. Anyway, in order to make that building technology widespread and democratic, it is necessary to simplify the programming interface in order to let also people with a non-technical background to use this technology. In order to do that, recent works underlined how it is nowadays fundamental to focus on the interaction of humans with the smart-things network, i.e., the sensors, the actuators and the smart appliances. Even if this could seem matter of interaction designers, many technological aspects has still to be faced: for instance, it is still necessary to find the right abstraction level to model a smart building, since users are not able to directly interact with sensors and actuators. Finding the right abstraction level means to find the right model, that is sufficiently expressive but simple to be understood at the same time. This thesis introduces a layered infrastructure, composed of different layers and different components, with the main goal of easily handling complex smart buildings behavior. On top of this building model, this thesis proposes experiments the usage of a Trigger-Action paradigm as programming paradigm for smart building.

The third problem is the **Building Behavior-Description Problem**: since a smart building is actually autonomous in the decision about its behaviors, and, consequently, in the decision influencing the life of the occupants, people are generally scared about the fact that their life is strongly influenced by something they cannot understand and directly control. For this reason, finding ways to describe to the occupants how the building is going to behave, is a crucial aspect that has still to be solved; within this context, we propose a lightweight simulator that, in our vision, is an essential tile to predict and show the future behavior of the smart building. As it will be, EASIM simulator, differently from the existing simulation solutions, we focused on the creation of a fast simulator able to perform many simulations in a very short time, in order to be a useful technology to explore different situations a building can face.

Given the aforementioned concepts, ideas and theories, this thesis will present a variety of methods and tools for an effective smart building deployment, i.e. technologies able to address the three main problems exposed above. In particular, this thesis will introduce a possible solution to the occupancy detection problem (BlueSentinel, an open source iBeacon based system for the indoor location), a easy to use building programming interface (BuildingRules, a trigger-action conflicts free framework aiming at simplify the expression of building management policies) and a a novel fast and efficient simulation platform (EASIM, a SystemC simulation framework for smart building to easily evaluate different policies to control the energy consumption of a smart space). These three technologies will be finally proved to be effective through two different case of studies showing an easy integration with smart grids and an easy user safety management system.

DEVELOPMENT OF A GAMMA-RAY DETECTION MODULE FOR MULTIMODAL SPECT/MR IMAGING

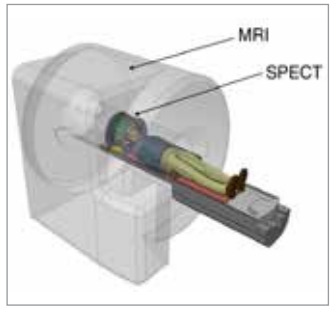
Michele Occhipinti - Supervisor: Prof. Carlo Fiorini

INSERT is the name of the European project (7-th framework program of UE commission) dedicated to the realization of an integrated SPECT/MRI for enhanced stratification of brain tumors (more specifically gliomas). The objective of the multimodal medical device is to combine the anatomical and versatile information from MRI with the high sensitivity of SPECT systems to track radiolabeled biological processes, with the final purpose to provide higher quantity and quality information for the diagnosis, patientspecific radio/chemo-therapy and early assessment of treatment efficacy. The range of possible applications extends from in vivo studies for physiological and pathological studies and drug delivery to cardiovascular functionality and diagnosis, oncology, neuro-imaging and neuro-oncology.

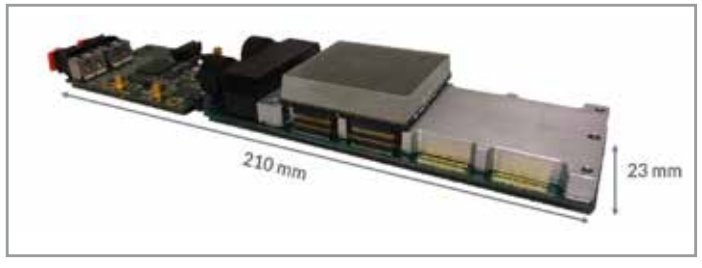
The synchronous combination of SPECT and MR is rather unexplored in literature mainly due to the major technical constrains related to mutual compatibility of the scanners. The SPECT insert is required to be compact, do not interfere with the MR acquisition, presents no significant susceptibility to magnetic field and RF signal from MR coils and also to detect multiple radiotracers within the same imaging session.
The SPECT insert, composed by a population of gamma cameras organized in a ring shape, is going to be realized in two versions: a preclinical system for biological research on small animals and a clinical one for the human brain/ neck region imaging.

The overall work is shared by a community of European partners and Politecnico di Milano represents the project leader, being responsible for the design and development of the basic gamma camera composing the SPECT ring. The doctoral thesis has concerned the design, development, qualification and optimization of the detection module, with particular focus on the conversion block of the device and on the signal elaboration for planar image reconstruction. I collaborated to the electronic and mechanic design and gave a contribute in the exploitation of a cooling strategy necessary to aim towards the required system performance.
The detection module has been build following an Anger architecture (a monolithic inorganic scintillator optically coupled to a photodetectors matrix), balancing good imaging performance with a limited number of electronic readout channels which represents a fundamental aspect for the realization of a large clinical SPECT with a plurality of gamma cameras.
The presence of high magnetic fields in the commercial available MR scanners has prevented the project to adopt conventional photomultiplier tubes (PMTs) in an integrated SPECT architecture, due to their high sensibility to the EM fields and to their bulky dimensions. This issue required the study and selection of an adequate photodetector among the family of the solid-state detectors, which are insensitive to magnetic fields and typically present a compact structure. In a preliminary phase of the project, a campaign of experimental tests with Silicon Drift Detectors (SDD) and Silicon Photomultipliers (SiPM) has allowed to elect the second as the photodetectors for the INSERT module.
Another important requirement for the SPECT insert is to permit multiple radiotracers detection for a better stratification of the biological processes in only one diagnostic acquisition. This implies that adequate energy resolution capability is needed together with the standard SPECT imaging performance. The design of the gamma camera was supported by analytical and statistical models and implemented with Montecarlo algorithms to forecast the possible system outcomes and to monitor the implications of the project decisions on the final spectroscopic and imaging performance. In order to improve image quality and to better address the initial requirements, a sophisticate and practicable statistical reconstruction method has been developed and implemented. The main issues in scintigraphy image reconstruction are related to the algorithm employed: the center of gravity method, widely diffused in general practice, presents intrinsic limitations that prevent the recovery of the entire field of view of the imaging camera and produce uniformity and linearity distortions, mainly at the edges of the reconstructed image.
The achievement of the project requirement has stimulated the study and development of a statistical reconstruction algorithm based on maximum likelihood, with better output performance. The most original characteristic of the method implemented is that the light response function (LRF), required from the statistical method for reconstruction, is retrieved from a calibration dataset obtained directly from the acquisition of an experimental uniform irradiation of the gamma camera.
The state of the art picture for the INSERT project depicts the realization of a first gamma detection module prototype, characterized in its imaging and energy capabilities.

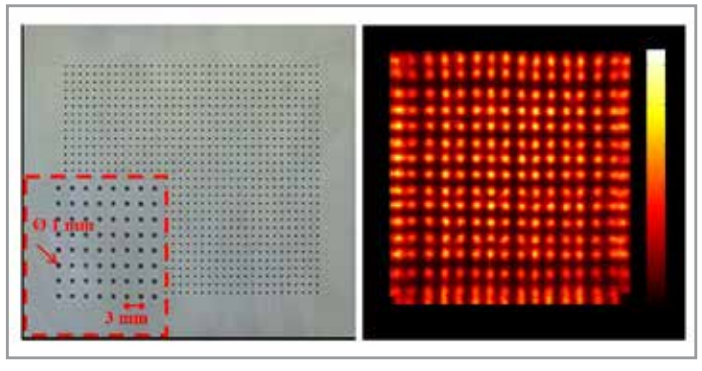
Tests with the preliminary gamma-camera prototype have produced the first experimental results: the intrinsic spatial resolution in the center of a 50 mm × 50 mm field of view has been measured around 1.4 mm FWHM, with an energy resolution of 16.6 % at 140 keV (Tc-99m). The simulated outcomes present good correspondence with the experimental ones, thus providing a reliable tool for the prediction of the final system performance: approximately 1mm FWHM of intrinsic spatial resolution and an energy resolution slightly higher than 12 % at 140 keV.



1. Conceptual draw of the MR-compatible, clinical SPECT insert, dedicated to human neck/head sector multimodal tomography.



2. INSERT gamma-detection module (preclinical configuration). The CsI(Tl) crystal is coupled to 4 SiPM tiles, covering a surface of slightly more than 50 mm × 50 mm. The 8 mm connectors stack separates the tiles from the ASIC board (two ASICs are mounted below, not depicted in figure) and provides internal volume for the cooling block. The DAQ board (on the left) is connected to the ASIC board through a 80 pins, edge mounted connector.



3. Figure 3: (left) Top view of the 4 mm thick, lead grid of holes employed for the estimation of the INSERT prototype intrinsic resolution.(right) Maximum Likelihood (ML) reconstructed images of the experimental grid irradiation (V_bias = 32 V, τ_shaping = 7.5 μs).

OPERATOR PRECEDENCE LANGUAGES: THEORY AND APPLICATIONS

Federica Panella - Supervisor: Prof. Matteo Pradella

Operator Precedence Languages (OPLs) were introduced in the 1960s by Robert Floyd to support deterministic and efficient parsing of context-free languages. Recently, interest in this class of languages has been renewed thanks to a few distinguishing properties that make them attractive for exploiting various modern technologies in two main contexts: automatic software verification techniques, as model checking, and parallel and incremental parsing of programming and data-description languages.

This thesis provides a complete theory of OPLs and investigates the properties that allow for their application in these different fields.

Along a first line of research, we complement the results on this class of languages that have been proved in the last half a century, which characterized them in terms of equivalent classes of grammars, recognizing automata and a Monadic second-order logic; the study of their algebraic properties, furthermore, has qualified them as the largest class of deterministic context-free languages enjoying closure under all main language operations (Boolean ones, concatenation, Kleene * and others), strictly including renowned families of formalisms as parentheses languages and Visibly Pushdown Languages (VPLs). In this dissertation we extend research on OPLs to the field of omega-languages, i.e., languages consisting of strings of infinite length, which can model the behavior of systems with never-ending computations (such as operating systems, control systems, web services). We introduce an automata and Monadic second-order logic-based characterization for this class of languages and we prove their closure properties and the decidability of the emptiness problem, showing that they admit a decidable model checking problem. Furthermore, we study logic formalisms simpler than Monadic second-order logic to define suitable subclasses of OPLs.

On a second line of investigation, this dissertation deals with a further property enjoyed by OPLs that is not exhibited by other families of deterministic context-free languages such as LR and LL, namely their local parsability. Local parsability means that parsing of any substring of a string according to a grammar depends only on information that can be obtained from a local analysis of the portion of the substring under processing and hence is not influenced by parsing of other substrings. The lack of this property implies that parsing algorithms for, e.g., LR and LL languages are inherently sequential and cannot exploit the speedup achievable by a parallel execution on modern multi-core computing platforms: in fact, if an input string is split into several parts, analyzed in parallel by different processing nodes, the parsing actions may require communication among the different processors, with considerable additional overhead.

This thesis studies and exploits the local parsability property of OPLs to enable efficient parallel parsing of data description languages (such as, e.g., the JSON standard data format) and programming languages (as, e.g., Lua and JavaScript) and presents a schema for parallelizing also the lexical analysis phase.

The algorithms for parallel parsing and lexing have been implemented in a prototype tool (PAPAGENO), which we validated with an extensive experimental campaign, showing that they achieve significant, near-linear speedups on modern multicore architectures, overcoming state of the art sequential parsers and lexers generated by, e.g., Bison and Flex. We exploit the local parsability property enjoyed by OPLs also for efficient parallel querying of large structured and semi-structured documents: we examine, as a case study, an extension of the parallel OP parsing algorithm allowing parallel XPath querying of XML documents on multicore machines.

CELLULAR NETWORKS PUSHED-TO-THE-LIMIT

Ali Parichehreh Teroujeni - Supervisor: Prof. Umberto Spagnolini

The surge of mobile data traffic is nowadays evident and companies have already forecasted future growths. 1.9 billion smart phone subscriptions were registered in 2013 and a 10x growth of mobile data traffic is expected before 2019. The dramatic increase of mobile broadband access is not only a technological matter but it rather reflects a deep societal change. In the upcoming years, it is expected that dramatic users mass and different traffic types will be transported over the cellular network, including video, voice, web surfing, social media, customized services and massive amount of device to device (D2D) communications. These massive mobile users with data-intensive applications at high Quality of Service (QoS) strictly push the current radio access technologies, like LTE and its advanced version, to their extreme limits.

In this context, massive on-board passengers of the high speed public vehicles such as trams, buses, high-speed trains and aircrafts can better highlight the mentioned phenomenon. In the mentioned scenarios massive on-board UEs with intensive traffic (generated by data-intensive services e.g., video streams, online games, etc.) at the high speeds cause a crucial condition for the wireless access network operators.

The most tangible approach of improving the network capacity is cell densification (i.e., shrinking cell sizes and increasing their numbers in congested areas). However, cell densification implies increased interference management and coordination complexity and handover (HO) signaling overhead besides capital/operational expenditures for the radio access network operators.

Another approach that avoids additional expenditures with reduced HO overhead and failure rate, is offloading cellular traffic among mobile users whenever the involved users are in the same vicinity. This, of course, can be achieved by tailoring D2D communications underlying the cellular network, which allow two or more nearby devices communicate with each other (without a base station involved or with limited involvement). This is demonstrably a dramatic step from the conventional cellular network architecture toward the fifth generation (5G) wireless networks.

The commenced quest for 5G multi-tier network accommodates design and development of novel algorithms and communication schemes especially in the context of D2D communications in extreme conditions (e.g., massive UEs with massive traffic in high mobility scenarios). In the following we briefly look at the multi-tier architecture used for design and development of novel D2D communication schemes and algorithms designed in this thesis.

Multi-tier Wireless Network Architecture

As mentioned above multi-tier network architecture (here two-tier is considered by convention) is a promising solution accommodating D2D communication as a means of alleviating the cumbersome overhead via traffic offloading from the core network to the mobile devices. In this thesis, a two-tier cellular network architecture is envisioned with so-called macrocell tier and device tier. The macrocell tier involves eNB-to-device communications as in a conventional cellular system. The device tier involves D2D communications among UEs inside the macro cells whenever necessary or based on the UE proximity (e.g., on-board the vehicles). Generally, the following types of communication at device tier can be considered.

Device relaying: Device relaying communication is an ad-hoc short range low power relaying within the device tier that provides a seamless connectivity especially at cell edge in confrontation with frequent discontinuities (e.g., HOs) and poor coverages. This two-hop communication link between mobile users at and eNB via intermediate relay node can enhance the robustness of communication link at the cell edge.

Direct D2D communication: Direct D2D communication is designed for traffic offloading especially for large content distribution among massive UEs. Each two devices within a device tier can have direct communication in one-hop or multi-hop via intermediate relay nodes.

Dissertation Contributions

The main contributions of this dissertation are the design and analysis of protocols for traffic offloading into the two-tier wireless network. In particular, the use of D2D communication underlying the cellular network is investigated for providing a fault tolerant seamless communications via traffic offloading in extreme condition. An experimental study is investigated for high speed train (HST) as a realistic example of wireless connectivity in extreme conditions to model the faults and discontinuities occurred in different states especially at the cell edges where cell changes happen. The statistical models of the faults are exploited and D2D communication is used to design pragmatic solutions for providing seamless wireless connectivity. Efficient traffic offloading mechanism for delivering mobile content in the device tire (via direct D2D communication) is investigated, and finally a simple but effective resource allocation mechanism is designed for device tier.

LTE in Extreme Conditions: An Experimental Study

An experimental study on the performance of the LTE cellular networks in extreme conditions is firstly investigated where the goal is mining the statistical properties of LTE performance in high mobility scenarios i.e., massive UEs are moving at high speeds and downloading contents from FTP server. In particular, an ad-hoc in-lab experimental set up is conducted and statistical properties of the frequent HOs, and LTE service time for a fixed size packets are mined as a function of cell load (i.e., number of UEs on-board the HST). To the best of our knowledge, this is the first experimental study on exploring the statistical properties of the HO-induced failure particularly in the 4G cellular networks. Beside that the two-tier network architecture is empowered via multi-cell access mechanism provided through deploying fixed beam directional antennas at the relay nodes. This new architecture alleviates the HO signaling overhead from the current serving cell while exploiting higher cellular network physical resources. The proposed scheme provides multi-cell access through a cost effective static beamforming.

Seamless Wireless Connectivity in Extreme Condition: A Case Study on High Speed Trains

A novel architecture is proposed by tailoring D2D communication to provide a seamless wireless connectivity on-board the high speed vehicles (in particular, high speed train). The challenging task investigated here is the service discontinuity caused by frequent HOs and manage this high-traffic sources that are spatially concentrated in less than 200m (i.e., the train length) and move at a speed up to 500km/h. The statistical properties of the HO-induced discontinuity is elaborated to design an ad-hoc offloading mechanism among the carriages of the HST. Device relaying scheme is the tailored concept on-board the HST to provide a seamless wireless connectivity, especially at cell edges where HO takes place.

Traffic Offloading in Device to Device Communication

It is widely acknowledged that the HO-induced latency is lower when the contents of the downlink traffic are closer to the mobile users, (namely communication without radio bearer allocation). In massive scenarios, some mobile devices can be elected to form a distributing cache that traffic can be locally cached depending on the traffic popularity statistics. End users request their required content through this ad-hoc distributed cache instead of the macrocell tier and via core network. This accordingly mitigates the HO-induced latency in downlink direction. However, advanced transmission techniques tailored in the PHY layer of the massive devices can impose variant range of performance from totally orthogonal channels to the conventional interference scenario for simultaneous end-to-end transmission. This accordingly causes congestion at some CA nodes considering different popularity statistics of the content in elected nodes. Therefore a novel ad-hoc offloading mechanism is proposed to balance the traffic among mobile nodes and maximize the cloud relay utility.

FEMTOSCOPE ARRAY FOR CORRELATION AND SPECTROSCOPY (FARCOS): SILICON DETECTOR LAYERS RESPONSE MAPPING AND HIGH-DYNAMIC RANGE VLSI FRONT-END DESIGN

Tommaso Parsani - Supervisor: Prof. Chiara Guazzoni

FARCOS (Femtoscope ARay for COrrelation and Spectroscopy) is a novel compact, modular and versatile telescope array based on two Double Sided Silicon Strip Detectors (DSSSD), 300 µm and 1500 µm thick with a high degree of segmentation (32 strip per side, 2 mm pitch) and CsI(Tl) scintillators acting as calorimeters, 60 mm thick. FARCOS features high angular and energy resolution and it is able to reconstruct the particles momentum at high precision for different physical cases in heavy-ion collision nuclear physics experiments at intermediate energies (10-100MeV/u). Full particle identification in terms of energy, charge and mass is performed by exploiting, together with the standard identification techniques (such as ΔE-E and ToF), Pulse Shape Analysis (PSA) techniques both on the CsI(Tl) signals and the DSSSD. The most challenging feature, not implemented in similar systems such as MUST2 and HiRA, is the implementation of PSA to silicon detectors signals to perform charge identification of particles stopping in the first detection layer. The research activity consists in the development and the characterization of the frontend electronics to be coupled to the DSSSD detectors.

In order to identify the frontend design specifications we performed a deep characterization of the silicon detectors. Firstly we investigated the fundamental physical parameters such as dark currents and capacitances (bulk and interstrip), studied as a function of the applied bias voltage and of the signal frequency. The next step was a detailed amplitude and position detectors response mapping and to this aim we exploited the monochromatic pulsed ion beams available at the DEFEL line of the 3 MV Tandetron accelerator of the INFN-LaBeC facility (*Laboratorio di Tecniche Nucleari per i Beni Culturali* – Sesto Fiorentino, FI, Italy). That line offers a surgical test probe with a profile down to better than 100 µm × 100 µm with a variable and finely controllable number of particles in each pulse, down to an average value much below one ion per trigger. We performed different measures accelerating protons at 1, 3 and 5 MeV and Li^{3+} at 7.5 MeV impinging on both the junction- and the ohmic-side of the detectors. A special attention is given to the impact of inter-strip beam incidence on the shape of the induced signals, among the main causes that can greatly spoil the overall identification capabilities. A deeper study has been performed by using an infrared pulsed laser at 705 nm and 904 nm wavelengths, with a spot size of less than 10 µm FWHM and performing a scan in the interstrip region. The measured output signal shapes have been fully explained by means of the Ramo theorem and mutual Coulomb interaction of the mobile carriers.

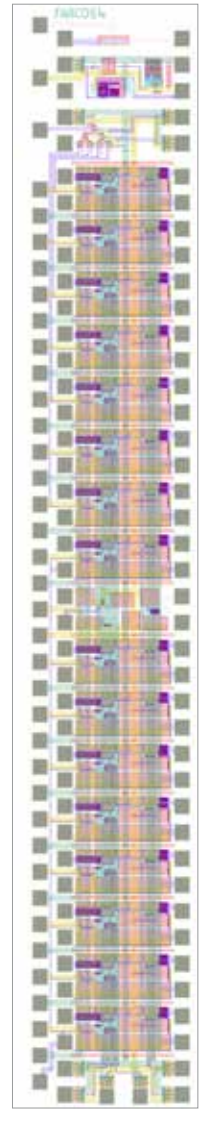
The detector characterization results have been used as design specification for the development of a multichannel CMOS VLSI frontend. The choice of a VLSI solution is needed for compactness and low power consumption for a number of readout channels operating in vacuum exceeding 2000 (for the foreseen 16 telescopes of the FARCOS system). The circuit, based on Charge PreAmplifier (CPA) topology, is able to amplify both the signal polarities for both the silicon detector thicknesses without amplitude and shape distortion, as required for PSA capability. The forward gain stage is a telescopic cascode with a source follower output stage.

To cope different experimental scenarios we implemented also a selectable gain feature – with corresponding energy ranges of 90 MeV to 500 MeV – by means of a selectable feedback capacitance. In order to guarantee a proper phase margin in all the gain configuration and for the different coupled detector we implemented also a selectable compensation network. The frontend will be AC-coupled to the detector by an external capacitance and the DC feedback network – that also fixes the DC operating point – will be either an integrated MOS biased in the triode region or an external resistance. The full chip also allocates the bias voltage generation for the CPA channels, the gain and compensation selection circuitry, a monostable pulser and a diode-based temperature sensor. The circuit has been produced in different 8-channel and 16-channel prototype versions using the AMS C35B4C3 technology, a low-cost low-flicker noise single-well process with minimum channel length of 0.35µm. The 16-channel layout view is presented in Fig. 1. One CPA channel is 980 µm × 310 µm and the full chip is 980 µm × 6500 µm.

The qualification of the circuit, both stand-alone and coupled to the detectors, has shown an amplitude integral-non-linearity below 0.5% for the different designed energy ranges, rise times below 20 ns (20%-80%) for the different gain and detector configurations and energy resolution below 10 keV FWHM. The results are well described by the post-layout simulations and they are fully compatible with the experimental requirements.

We have also produced a multichannel PCBs allocating 32-channels of the frontend circuits. Each PCB is able to carry out the signals of one side of the silicon detectors. The CPA analog output signals are amplified by dedicated fully-differential line drivers (one per channel) in order to drive the signal along the meters-long connections to the digitizers located outside the vacuum chamber. The 32-channel PCB houses also the frontend supply generation from an external supply and all the slow control circuitry for gain and compensation selection.

The characterization of the VLSI frontend coupled to the detector with an ion beam is under completion in order to probe the Pulse Shape Analysis capability of the circuit.



1. Layout view of the 16-channel version of the VLSI frontend. One CPA channel is 980 µm × 310 µm and the full chip is 980 µm × 6500 µm.

INFORMATION RATE OF CHANNELS WITH MEMORY BY BAYESIAN TRACKING

Simone Pecorino - Supervisor: Prof. Arnaldo Spalvieri

Increasingly, in many engineering fields, a more appropriate characterization of the dynamical systems leads to a model where an accurate description of the underlying dynamics is obtained by including elements of non-linearity and non-Gaussianity. In the context of digital transmission systems the computation of posterior probabilities of the hidden Markov state of such systems is often required. In these cases the exact Bayesian tracking of a posteriori probabilities of the state often becomes not feasible and, therefore, resorting to approximated methods becomes necessary. Among the techniques that realize approximated tracking, the extended Kalman filter, the state-space quantization, and the particle filtering method can be taken into account. Specifically, this last technique is a sequential Monte Carlo approach based on point mass representations of probability densities. Particle filtering, that can be seen as a generalization of the traditional Kalman filtering, can be applied to any state-space model, starting from the definition of the state transition and observation models and without restrictive assumptions. For these reasons, it has found widespread application in many research fields.

The thesis focuses on computing the information rate of Markov channels, where the a posteriori probabilities cannot be exactly tracked, such as Wiener's phase noise channel, autoregressive moving-average phase noise channel, and Gauss-Markov fading channel. Starting from the definition of mutual information, that is very difficult to compute for these channels, approximated Bayesian tracking can be used to calculate upper and lower bounds to the information rate. The goal of the work is to obtain upper and lower bounds as close as possible to each other, in order to compute an information rate close to actual one. Some successfull examples, that point out the goodness of the state tracking methods, are shown.

The Shannon information rate of a Markov channel can be achieved also by ideal demodulation, which cannot be often adopted in practical system. The iterative demodulation and decoding schemes try to reach the performance of ideal demodulation. An other goal of the work is the study of the Bayesian tracking methods to extract all the information about the transmitted symbols in a new iterative demodulation and decoding scheme. A reported analysis of some complexity reduction techniques makes deployable the proposed scheme in a real receiver.

REINFORCEMENT LEARNING: FROM THEORY TO ALGORITHMS

Matteo Pirotta - Supervisors: Prof. Luca Bascetta, Prof. Marcello Restelli

This thesis is mainly based on the idea that the design of an algorithm must be supported by theoretical results. Instead of starting from an algorithm and analyzing its properties and guarantees, we will derive theoretical results and we will investigate their applicability to design practical algorithms. All the analysis will be performed in the reinforcement learning framework, i.e., a mathematical framework for learning by interaction. This thesis summarizes recent results in different reinforcement learning topics and builds on these results to provide novel algorithms with an attention to performance guarantees. The thesis is organized in three parts, one for each studied topic.

We start investigating the recent advances in the framework of safe reinforcement learning, i.e., the design of algorithms with strong performance guarantees. Historically, reinforcement learning approaches have focused on the recovering of a function representing the value (goodness) of an action in a situation. Such description contains all the information required to compute the optimal (or more promising) action in each state. Despite the successful applications, these methods are sometimes difficult to deploy on real systems because the learning process is not guaranteed to improve overtime or do not offer per-iteration guarantees. The motivations for the need of guaranteed algorithms are twofold. The design of such algorithms is a theoretical challenge that has called the attention of a small but active community of researchers. Critical domains, e.g., car driving, industrial robotics, nuclear plants, dam control and so on, and domains that involve the human presence or interaction are reluctant to exploit learning techniques because they are perceived like something uncontrollable and somewhat "magical". They are used to classical control theory where the focus is put on the stability, robustness and sensitivity. We think that providing algorithms with theoretical guarantees about the performance improvement and, in some form, ways to limit or control the changes overtime, can be a way to introduce reinforcement learning in industrial and daily-life applications. In the first part of the thesis we will investigate and derive how theoretical results can be incorporated in algorithmic structures in order to derive methods with guaranteed performance improvements. We will suggest algorithms that, by exploiting a lower bound to the expected performance gain, are able to guarantee a monotonic performance improvement overtime.

However, to face many real applications reinforcement learning must be able to handle problems with multiple objectives. Think to economic systems (production versus budgeting), water resource problems (electric energy production versus irrigation supply), robotic systems (energy consumption versus rapidity), drug dosage (benefits versus collateral effects), just to mention a few. Such problems can be modeled as multi-objective reinforcement learning problems where the classical concept of optimality (i.e., to maximize the cumulative reward) is replaced by the one of Pareto optimality, i.e., a set of behaviors providing a compromise among the different objectives. Despite the successful developments in reinforcement learning theory and a high demand for multi-objective control applications, multi-objective reinforcement learning is still a relatively young and unexplored research topic. In particular, although gradient algorithms have become a standard approach both in single-objective reinforcement learning and multi-objective optimization, there are a few works that exploit gradient techniques in conjunction with multi-objective reinforcement learning. We will present two new approaches, called radial and Pareto following, that perform gradient-based policy search procedures aimed at finding a set of non-dominated policies. Moreover, we will derive an algorithm that performs a single gradient ascent run that at each step generates an improved continuous approximation of the entire Pareto frontier.

One of the core components of reinforcement learning is the reward signal that tells to the agent how well she is behaving. The reward is an extrinsic signal that incorporates only and all the information about a specific goal. It is in general hand-tuned by humans that are experts of either the domain or the learning algorithm. The design of the reward function is critical, not always trivial and sometimes frustrating.

Since early 2000s, reinforcement learning literature has focused on the problem of recovering the goal of an agent by observing her behavior in the environment. This subfield of reinforcement learning is named inverse reinforcement learning. In the third part of this thesis we will focus on this topic by deriving algorithms that do not require a complete information about the reward. On the contrary, they are able to recover the reward signal by observing a limited set of agent-environment interactions. The main focus will be posed over the theoretical properties and motivations, that has guided the design of the algorithms. We will show that the study of inverse reinforcement learning is a natural consequence of the work done in multi-objective optimization by deriving a clear connection between the two areas. We will exploit this connection to derive algorithms (CPIRL and PGIRL) with strong error guarantees that differ for the amount of information required. We will present an algorithm (GIRL) that removes the assumption of linear parametrization, leading to better generality. Moreover, we will show that PGIRL is an efficient implementation of the more general nonlinear-reward algorithm (GIRL) in the particular case of linear reward parametrizations.

SCALABLE FORMAL VERIFICATION OF UML MODELS

Mohammad Mehdi Pourhashem Kallehbasti - Supervisor: Prof. Luciano Baresi

Today's software systems are huge and complex, therefore abstractions and models are used to cope with their complexity and pave the way for efficient software development. Model-Driven Engineering (MDE) is a software development methodology that has gained significant interest over past years. MDE simplifies the design process, boosts productivity, and promotes communication between individuals and teams by standardization of terminology and models. The Unified Modeling Language (UML) is a general-purpose, platform-independent modeling language, that has been managed by the Object Management Group (OMG). During Model-Driven Development (MDD), UML often plays a key role and provides a significant set of diagrams to describe static (structural) and dynamic (behavioral) aspects of the system. UML models, as artifacts of the design phase, materialize the prospective software system at an abstract level. Defects in models are a significant concern as model transformations and code generation may amplify and propagate errors and make them become harder to detect and trace. Rigorous specifications allow the designer to reason on the models and detect defects and inconsistencies early in the specification process without propagating and amplifying them further. This can be done using model checking, a formal verification technique, where the system is formally specified and the expected behavior (property) is usually expressed in temporal logic. There have been many different attempts to ascribe UML with a (more) formal semantics, but the wideness of the language has often led the authors to only concentrate on some diagram types, while neglecting the key characteristics of UML, that is, the rich set of diagram types and the freedom with which the designer can model a system. These formalizations, which aim at a specific type of diagram, fail to capture the behavior of systems from different, orthogonal points of view. As a semi-formal language, UML has a formal syntax and an informally defined semantics. In fact, its semantics is intentionally underspecified to provide leeway for domain-specific refinements. On the other hand, a too general semantics hampers formal verification owing to the lack of required precision. These ambiguities mainly rise from sequence diagrams, which are often used to capture the most significant scenarios that describe how the components of a complex system interact. Focusing on sequence diagrams, many researchers propose their interpretations, among which only few present formalizations that are amenable for formal verification. Moreover, their formalizations are based on fixed semantics with limited interoperability. The second step towards formal verification is to use a verification technique to verify the formally specified model, while state explosion has always been a crucial problem. If we reduce the formal verification of UML models to satisfiability checking of LTL formulae, as the UML model becomes more complex, the resulting LTL formula grows in size accordingly and makes the whole verification process slow, and at a given point impossible. Complete and Bounded Satisfiability Checking (BSC) are the two main techniques towards checking whether or not an LTL formula is satisfiable. In BSC, LTL formulae are suitably translated into formulae of another decidable logic, such as (Propositional) Boolean Logic, which precisely capture ultimately periodic models of the original formulae of length up to a bound *K*; produced formulae are then fed to a solver (e.g., a SAT solver) for verification. BSC is the fastest approach to reveal system errors, but to make it complete and verify properties, a big enough *K* must be determined, and this is expensive. However, complete approaches suffer much more from the state explosion problem and bounded approaches are preferred especially when complete ones give up in returning verification results for huge formal models. Still, scalability is a problem even in bounded approaches and it hampers the verification of big formal models. The two parallel contributions of this thesis aim at the formal specification and scalable verification of eFT-UML (extensible Formal, Timed UML) models. eFT-UML integrates a significant number of diagram types in a single coherent framework including Class Diagram, Object Diagram, Interaction Overview Diagram, State Machine Diagram, and Sequence Diagram (SD). The goal of eFT-UML is to overcome the limitations of separated diagram types by means of a precise set of shared events. In addition, it borrows from MARTE (the UML Profile for Modeling and Analysis of Real-Time and Embedded Systems) the notion of time. This means that eFT-UML is particularly suited for the specification of timed systems. The formal semantics of eFT-UML is based on the TRIO metric temporal logic, which gives us the flexibility and composability required to specify the semantics of a complex notation. The first contribution is thus a flexible modular formalization for eFT-UML elements with many meanings, and it mainly concentrates on Sequence Diagrams. We studied the most significant semantic proposals, organized them into a single coherent framework, and proposed a solution to interpret SD in a compositional and modular way, in order to fulfill OMG's ambition of keeping UML useful in many domains. Users can decide the interpretations of the key aspects of interest and the result is a complete and coherent semantics; then, provided some simple constraints are respected to avoid making inconsistent decisions, our framework accommodates all other aspects. The proposed theoretical approach is implemented on top of our verification toolset, **Corretto**[1], where the produced temporal logic formula is fed to Zot[2] our bounded model/satisfiability checker and allows the user to easily play with the different semantics. The user can simulate the behavior of the diverse semantics and verify the satisfiability of the properties of interest. S/he can thus understand how the system behaves or work on the satisfiability of the properties and then obtain the guarantees the system must offer. The same theoretical approach can be used for SM. While working on a solution for making the verification of UML models scale properly, we have identified an interesting solution that suites a wider class of verification problems that can be reduced to TRIO, therefore Constraint LTL (CLTL)/LTL formulae. The second contribution of this thesis is thus a bit-vectorbased encoding of these CLTL/LTL formulae, which has allowed us to move a significant step forward in tackling the scalability of their formal verification. The proposed encoding for LTL and CLTL are implemented as plugins called bvzot and ae2bvzot in Zot our satisfiability/model checker. We compared bvzot with the state of the art model checker NuSMV, that contains implementation of several LTL encodings, and also compared ae2bvzot with the only existing tool for bounded satisfiability checking of CLTL, that is another plugin of Zot, called ae2zot. The experimental results witness a significant increase in the size of analyzable models, not only for our formalization of UML models, but also for numerous other models that can be reduced to bounded satisfiability checking of LTL/CLTL formulae.

[1] https://github.com/deib-polimi/
[2] https://github.com/fm-polimi/zot

TOWARDS A SAFE INTERACTION BETWEEN HUMANS AND INDUSTRIAL ROBOTS THROUGH PERCEPTION ALGORITHMS AND CONTROL STRATEGIES

Matteo Ragaglia - Supervisor: Prof. Paolo Rocco

Summary: In the past few years the need for more flexibility in industrial production has implied, in the field of industrial robotics, a growing attention towards the possibility of making humans work directly in touch with robots. As a matter of fact, it is today a common opinion that Human-Robot Interaction (HRI) represents the key factor that will facilitate industrial robots to spread in SMEs.

Nevertheless, HRI introduces a series of safety issues which are uncommon in industrial settings where physical separation between robot and human workspaces is typically enforced. In order to achieve safe and efficient HRI, this thesis was developed around two main goals:

- enhance the perception capabilities of a typical control system of an industrial robot by integrating information coming from different exteroceptive sensors, like for instance RGB and depth cameras;
- develop reactive control strategies and trajectory generation algorithms that not only rely on the information acquired by the aforementioned sensors, but that also guarantee human workers' safety by satisfying safety standards and regulations.

From the perception perspective, two main problems have been approached. At first we have developed sensor fusion strategies able to merge information coming from several RGB cameras, or from multiple depth cameras, or from both kind of sensors. Then we have chosen a simple yet effective human kinematic model and we have developed algorithms able to detect, track and predict human motion on the basis of the information acquired via sensor fusion.

Switching from the perception domain to the control perspective, we initially approached the problem of formalizing safety requirements and regulations in a mathematical way. To this purpose "safety constraints" have been formalized in order to express collision avoidance requirements with respect to both a-priori known obstacles and obstacles perceived at runtime. On the basis of these safety constraints, several safety-oriented control strategies and trajectory generation algorithms have been developed in order to exploit the information acquired from the perception system.

Finally, also the problem of safety in physical HRI has been investigated, with a peculiar focus on lead-through programming (LTP).

GEOSTARE: GEOSYNCHRONOUS SAR FOR TERRAIN AND ATMOSPHERE OBSERVATION WITH HIGH REVISIT

Andrea Recchia - Supervisor: Prof. Andrea Monti Guarnieri

The single location where we can learn the most about our planet is found nowhere on Earth but high up above it. The ability to fly satellites into space has changed all our lives in many ways, but the single greatest innovation has been the availability of new ways of seeing the world that satellites leave behind.
Early pictures of the Earth seen from space became icons of the Space Age, and encouraged an increased awareness of the precious nature of our common home. Today, images of our planet from orbit are acquired continuously; they have become powerful scientific tools to enable better understanding and improved management of the Earth and its environment.
Earth Observation images show the world through a wide-enough frame so that complete large-scale phenomena can be observed to an accuracy and entirety it would take an army of ground-level observers to match.
A single satellite image has the potential to show the spread of air pollution across a continent, the precise damage done in a region struck by an earthquake or forest fires, or the entire span of a 500-km hurricane from the calmness of its eye to its outermost storm fronts.
Earth Observation provides objective coverage across both space and time. The same space-based sensor gathers data from sites across the world, including places too remote or otherwise inaccessible for ground-based data acquisition. And because Earth Observation satellites remain in place for long periods of time, they can highlight environmental changes occurring gradually.
Looking back through archived satellite data shows us the steady clearing of the world's rainforests, an apparent annual rise in sea level approaching 2 mm a year and the depletion of the ozone layer by atmospheric pollution.
In the long term, this monitoring of the Earth's environment will enable a reliable assessment of the global impact of human activity and the likely future extent of climate change.
Synthetic Aperture RADAR (SAR) imaging has been widely recognized in the last thirty years as the most powerful technology for the observation of the Earth's surface from space.
SAR Interferometry (InSAR) has provided the scientific community with large scale, high accuracy measurements concerning phenomena such as land subsidence, earthquakes, landslides, volcanic activity, flooding and water vapor distribution in the atmosphere.
The main limitation of current space-born SAR missions is the revisit time, which cannot be made lower than few days even for modern satellite constellations. This limitations prevents the possibility to monitor and study fast events, with time scales of hours or few days. Besides the intrinsic difficulties in monitoring fast varying events, a further limiting factor for LEO-SARs may arise in non-urban areas due to temporal decorrelation. The electromagnetic response of natural targets is progressively modified over time under the action of phenomena such as wind-induced motions, vegetation growth, or local humidity changes. The resulting correlation loss between signals collected at different times prevents the retrieval of interferometric information, hindering the monitoring of any kind of events in areas where no stable target is present.
It follows after these considerations that the availability of a very frequent (one day or less) repeat-pass space-borne interferometric SAR system would make it possible to use well known InSAR methods for the investigation of short temporal scale phenomena such many hydro-geological, geophysical as well as meteorological events.
It is then clear that technologies able to shorten the revisit time while granting wide geographical coverage and fine spatial resolution would play a crucial role in the enhancement of SAR system capabilities to monitor Earth from space.
The exploitation of space-borne SAR systems operating in a Geosynchronous Orbit (GEOSAR) would make possible revisit times of 12-24 hours, or even much less, by employing satellite constellations. GEOSAR has significant potential advantages over conventional Low-Earth Orbit systems (LEOSAR), but also challenges to overcome. The main problem for a GEOSAR system is the extremely weak signal power available at the receiver due to the propagation losses. The issue can be overcome either increasing transmitted power and antenna size, making the system realization technologically challenging, or increasing the integration time, introducing the signal decorrelation problem.
This thesis is aimed at studying the feasibility of a long integration time (from several minutes to hours) GEOSAR system, focusing on both the system performances assessment and the development of accurate and efficient techniques for the processing of the acquired data. The first decorrelation source considered is the varying atmospheric delay which induces a random Atmospheric Phase Screen (APS) in the received signal. If not APS compensated, the azimuth compressed GEOSAR data will appear out of focus and blurred. Since even very stable targets will be affected by APS decorrelation, an algorithm for the estimation and correction of APS has been developed and tested over simulated data.
The second decorrelation source is the imaged scene itself. Targets such as canopies change their back-scattered phase on timescales of seconds due to their motion. On longer time scales, changes in dielectric properties of targets, contribute to generate phase fluctuations in the collected signals. Unlike APS, such fluctuations cannot be compensated and hence performance models to evaluate their impact on focused GEOSAR images have been derived.
The obtained performance models have been finally exploited to generate the first demonstrative GEOSAR products. The demo products provide to the potential data users a first glimpse of the appealing GEOSAR applications including the near real time monitoring of earth deformation phenomena like volcanoes, glaciers, landslides and building deformation in urban areas. The demo products are based on the preliminary design of an GEOSAR system currently under ESA study, named Geosynchronous SAR for Terrain & Atmosphere with short Revisit (GeoSTARe). The considered system is a potential candidate for a future ESA Earth Explorer mission and justifies further study of the concept.

SINGLE-PHOTON COUNTING INSTRUMENTATION FOR THE NEAR INFRARED WAVELENGTH RANGE

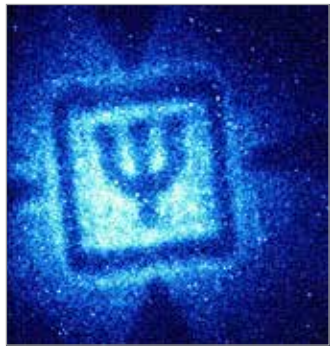
Alessandro Ruggeri - Supervisor: Prof. Alberto Tosi

The measurement of very faint (down to the single-photon level) and ultra-fast (up to 100 GHz bandwidth) optical signals in the near-infrared wavelength range up to 1700 nm is fundamental for many applications, such as: characterization of new materials and single-photon sources, 3D laser ranging, secure transmission of secret keys in telecommunications, singlet-oxygen dosimetry in photodynamic cancer therapy, lifetime measurement in time-resolved spectroscopy, non-invasive testing of VLSI circuits, testing of optical fibers.

A few single-photon detectors are available, but microelectronic ones have the advantages of high reliability, robustness and compactness. Single-Photon Avalanche Diodes (SPADs) are becoming increasingly widespread thanks to their good Photon Detection Efficiency (PDE), low timing jitter and low noise, with only moderate cooling requirements. In particular, InGaAs/InP SPADs are suitable to detect photons from 900 nm to 1700 nm, but they need dedicated circuits in order to be properly operated, mainly to limit the intrinsic dark counts and another detrimental noise, called afterpulsing. The widespread use of these devices is limited by the size and the high power dissipation of currently-available circuit solutions. While this is still acceptable when they are employed in a development laboratory, it is intolerable where portability and size become key requirements for their widespread diffusion.

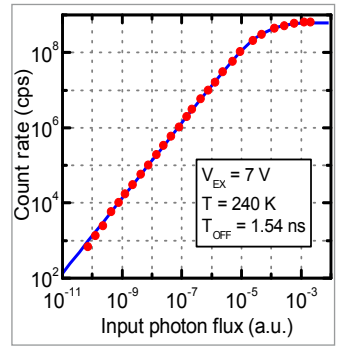
To this aim, my Ph.D. research topic was focused on the design of novel and compact single-photon detection systems, primarily based on InGaAs/InP SPADs, capable to perform photon counting and timing measurements.

Starting from a detection module previously developed at Politecnico di Milano, a new FPGA-based control unit with improved functionalities has been developed. Particular attention was given to the delay circuits implemented inside the FPGA, employed to generate the programmable width of SPAD enabling windows. This newly system has been employed to extensively characterize state-of-the-art InGaAs/InP SPADs, designed at Politecnico di Milano, obtaining excellent results. These tests were also very useful to optimize the operating conditions for the subsequent experimental exploitation of this module in scientific experiments in collaboration with different international research groups: we performed diffuse optical spectroscopy of highly absorbing materials, in collaboration with Dipartimento di Fisica at Politecnico di Milano (Italy); we collected infrared images of silicon chip samples using entangled photons at different wavelength (see fig. 1), in collaboration with Glasgow University (UK) and Heriot-Watt University (UK); we reconstructed centimeter-resolution depth images of objects at stand-off distances of up to 1 km, with Heriot-Watt University (UK).



1. Ψ letter engraved in gold over a silicon background, imaged through dual-wavelength ghost-imaging technique (in collaboration with Glasgow University and Heriot-Watt University).

2. Compact InGaAs/InP SPAD System-in- Package, which integrates detector, cooler and front-end electronics in a TO-8 can.



3. Count rate vs. photon flux for the gigahertz gating setup. Count rate is linear over more than five orders of magnitude, and saturates at 650 Mcps.

A novel integrated active quenching circuit (AQC) able to gate InGaAs/InP SPADs in well-defined time intervals with fast edges (down to 200 ps) has been extensively characterized and exploited. The differential read-out circuit and the fast integrated comparator guarantee very low timing jitter (about 30 ps) and the ability to detect photons even on the gate rise/fall edges. This integrated circuit is the first step towards compact multi-channel fast-gated SPAD systems. Two compact solutions exploiting this integrated AQC have been developed and fully characterized. The first one is a complete USB-powered Time-Correlated Single-Photon Counting (TCSPC) system, based on a 50 µm silicon SPAD, and housed in a very small case. The second one exploits an InGaAs/InP SPAD, mounted inside a tiny TO-8 package with the integrated circuit and a thermo-electric cooler (see Fig. 2). These systems guarantee flat sensitivity in the gate window for reconstructing the optical signals with low distortion and wide bandwidth. In the second system, based on an InGaAs/InP SPAD, the afterpulsing probability is below 1% even at high count rate (about 1 Mcps), thanks to the very short (< 1 ns) quenching time. This is independent form the gate width, which is programmable from DC (i.e. free-running operation) down to several hundreds of picoseconds.

To solve the typical trade-off between maximum count rate and afterpulsing probability in InGaAs/InP SPAD, a new gigahertz sinusoidal gating setup has been conceived, designed and tested. With the novel balanced configuration, the capacitive couplings of the high frequency gate signal through the SPAD are compensated. The extensive characterization of the developed system running at 1.3 GHz demonstrated excellent temporal response (down to 65 ps), very high count rate (up to 650 Mcps, see Fig. 3) and low afterpulsing probability (< 1.5 %). Starting from this system configuration, a novel compact, stable and highly-programmable system was conceived and developed. Its broad operating bandwidth (from 900 MHz to 1.4 GHz) allows the user to easily synchronize with different laser systems, depending on the specific application. Additionally, it guarantees long-term stability of the performances, thanks to a feedback stabilized system. Finally, the small size of the module will enable its exploitation in practical high-speed applications, such as quantum key distribution systems for secure telecommunications.

USER INTERFACE AND INTERACTION IN AN AMBIENT ASSISTED LIVING SYSTEM: GESTURAL AND VOCAL INTERACTION FOR THE INHABITANT, DATA VISUALIZATION FOR THE CAREGIVER

Hassan Saidinejad - Supervisor: Prof. Fabio Salice

It is widely accepted that ICT can play an important role in improving the quality of life and well-being of the elderly people especially for an independent living. Assistive Technology and Ambient Assisted Living are two examples of ICT technologies seeking to help elderly and disabled people. Ambient assisted living can be best understood as the marriage of two fields: assistive technology and ambient intelligence. Ambient intelligence is based on ubiquitous computing paradigm applied in an environment such as a home. This work is done in the context of an ambient assisted living system.

The Assistive Technology Group (ATG) is a multi-disciplinary group in Politecnico di Milano. The main goal of ATG is to provide innovative ICT-based and sustainable solutions addressing the problem of fragility, especially for elderly and disabled people. BRIDGe is an ambient assisted living system developed by ATG with the main goal of providing mutual reassurance for its users: the person inside the home, the family and caregivers outside the home. It is devised to be modular, interoperable, low-cost, and personalizable among its other characteristics. These features prepare the ground for a need-based, social-technological coordinated service design for the users of BRIDGe: the social counterpart, thanks to its domain expertise, interviews the users in order to elicit their genuine needs; the technological counterpart proposes an appropriate solution for the identified need.

User interface and interaction is one of the main aspects in an ambient assisted living system. The person inside the house, the family and caregivers outside the house are considered users of the BRIDGe system. Interface and interaction design for these users has been the main interest of this work. This work addresses three identified user interface and interaction related needs. The first two needs concern the person, i.e. the user inside the house with the main interest of Home Control. Gestural interaction for an elderly-disabled user with residual hand motor capability, and vocal interaction for a similar user profile with residual voice capability are the two corresponding solutions. The third need concerns the caregiver, i.e. the user outside the house with the main interest of understanding data collected from inside the house. An interactive visualization tool is the response to this need.

There have been two main reasons that motivated the research on gestural interaction in the context of Bridge AAL system. The first reason was to investigate the gestural interaction as an alternative and more natural interaction modality for older adults, in particular, for home control purposes. The second reason, that motivated also the focus on static hand poses, concerned a subject with motor problems with a residual ability in the fingers of the right hand. This subject was interested in performing home control with her residual motor abilities.

In brief, a mainly hand pose-based gestural interaction system was developed using Microsoft Kinect. Two user studies were done: one to investigate gestural interaction preferences in general, and the other to understand elderly preferences regarding gestural interaction. Moreover, a gesture-based home control demo application was developed in which the user is able to perform standard domotic control using static and dynamic hand movements.

The naturalness of the gestural interaction in reality strongly depends on the gesture recogntion system and the restrictions it imposes on the interaction of the user with the system. Such restrictions (position of the user, speed of hand movement, complexity of static hand poses, etc.) could adversely affect learnability, ease of use, and user interaction errors. Suffering from such restrictions, the gesture recognition system presented in this work was not appreciated as an alternative natural interaction modality for elderly users.

The main motivation for the research on vocal interaction has been the identification of a number of subjects with severe motor issues and compromised manual capability. With the residual speech capability they would like to be able to perform some standard home control. As most of the subjects are older adults, a more "natural" vocal interaction has been requested. Furthermore, some subjects have degraded speech that could be troublesome for the speech recognizer. Thus, robustness is also to be added to the requirements. Finally, a low cost and affordable solution is needed. So, in brief, the problem is to have a low cost speech recognition solution that is robust to degraded speech and that provides a "natural" flow of interaction with the user.

An android-based application was developed that exploits the freely available google speech recognition. An inexactness toleration mechanism was added to the google speech recognition to make it more robust against degraded speech. In order to make the interaction more "natural", flexible command creation is allowed in which the user freely issues a command. Based on a simple command model the required information is extracted from the sentence and the command will be executed. Moreover, the command goes through a disambiguation pipeline that responds to user commands in a more intelligent manner (e.g. by using context information like the location of the person, if available). Finally, in order to have a more natural interaction, a ubiquitous distant speech acquisition system based on distributed microphone nodes in the environment was studied.

In particular, the inexactness toleration mechanism proved to be effective in compensating the dysarthric characteristics of elderly-disabled speech and increasing the accuracy of the speech recognizer. The unstructured and flexible command creation mechanism was appreciated as it was making the interaction more natural. However, a dialog manager seems to be necessary to cope with user irritation and frustration. Distant speech acquisition system is an opportunity for a ubiquitous automatic speech recognizer. The primary results based on simulations show that the final obtained speech signal can be recognized with an acceptable accuracy.

Bridge AAL system offers the opportunity to have a huge amount of data collected by various sensors in the indoor environment. The challenge is how to extract meaningful information from these data and how to make sense of it. It can be too complicated to build a comprehensive model of the person's indoor life comprising all aspects of life and well-being of the person. The care professional needs a tool to gain insight into the collected data in order to identify trends, detect anomalies and possible deviations from daily "norms" of the person.

BridgeViz is an effort towards an interactive data visualization tool for the Bridge project. A web-based interactive visualization tool has been developed with overview and detailed visualizations. A variant of the sunburst diagram was developed as the overview visualization and a variant of the node-link diagram for the detailed visualization. Furthermore, a demo version was developed and a user study was conducted to evaluate the developed tool. The participants were asked to perform a series of pre-defined tasks and to fill a questionnaire.

Interactive data visualizations could be beneficial for the care professionl in particular and other non-technical stake holders in the well-being of the person in general to extract meaningful information from the data of an AAL system. The interactive capabilities of the tool are appreciated as useful mechanisms to explore data.

TOWARDS IMPROVING PROGRAMMABILITY OF HETEROGENEOUS PARALLEL ARCHITECTURES

Michele Scandale - Supervisor: Prof. Giovanni Agosta

Parallel Computing has been considered an effective approach to combine performance and power efficiency for a long time. Starting from High Performance Computing (HPC) to modern embedded systems the employment of heterogeneous parallel architectures is becoming the common case, since it is a good tradeoff in terms of power efficiency.
The exascale objective for the next generation of HPC systems is constrained to a target power envelope ranging from 20MW to 30MW. The existing "Green"' HPC systems are not yet able to reach the such power efficiency although they already employ modern heterogeneous parallel architectures.
Ultra-low-power hardware platforms are gaining an increasing traction, as they may represent the key component to allow future HPC systems to match the required power efficiency.
The programmability of such systems is a critical aspect that has an huge impact on the reachable power efficiency and the effort required to reach such target. Programming parallel architectures is a complex task, since many hardware features are directly exposed to the programmers.
Programming frameworks that try to hide such complexity exist, however they either provide only sub-optimal performance with respect to hand tuned implementations, or they are limited to specific application domains.

This dissertation tackles challenges related to the programmability of heterogeneous parallel architectures, acting on both existing and future programming models and hardware architectures.
We present OpenCRun, an OpenCL runtime implementation supporting a range of platforms with very different architectures characteristics, such as X86 multicores and embedded parallel accelerators.
In the context of ultra-low-power architectures we report the joint effort between hardware and software developers towards the PULP platform, showing the benefits of selected ISA extensions and their compiler support to maximize the power efficiency. At cluster level the energy savings ranges between 39% and 66%, with an area increment is only 2.3% and power consumption increased by 18%. On average the cluster is 47.8% more energy efficient than the initial architecture.
Moreover, to improve functional and performance portability of OpenCL code between GPGPUs and embedded many-core accelerators with explicitly managed memory such as PULP and STHorm, we propose a code transformation technique, work-item coalescing, that bypasses the limitations of the embedded platforms, allowing code developed for GPGPU to be ported seamlessly, as well as a memory transfer optimization technique to tune the resulting code to improve performance. Our results on two case studies show the effectiveness of the proposed technique, which allowed the code, developed for NVIDIA devices and not designed with portability to smaller-scale ones in mind, to effectively run on a STHorm target with a speedup up to 88x for non computational intensive kernels and a speedup up to 4x for computation al intensive kernels.
Finally, to increase the abstraction level in a more radical way, leveraging Shared Virtual Memory that is expected to be available in future architectures, we present a method to transparently implement shared function pointers in heterogeneous platforms with two or more ISAs, a building block for enabling full C++ support across heterogeneous ISAs. Our proposal, based on the trampoline technique for indirect calls, allows to preserve compatibility with the host ABI, while introducing only minimal overheads in terms of performance (2.7% - 5.5%) and code size (≈11%). We provide then a fallback solution to implement device call to function not available on the accelerator itself (e.g. operating system services) based on RPC mechanism where the compiler is responsible to generate stubs and skeletons, and the heterogeneous linker decides at link-time whether stubs shall be used or not depending on the presence of the target function for a given device. Enabling the full C++ features allows smoother migration of applications towards the heterogeneous architectures since no additional constraints are applied to the code that will be offloaded to the accelerators, reducing the initial porting effort through an iterative and incremental process of re-engineering of the code.

DEVELOPMENT AND CHARACTERIZATION OF A LOW NOISE MULTICHANNEL READOUT ASIC FOR X- AND GAMMA-RAY SPECTROSCOPY APPLICATIONS

Filippo Schembari - Supervisor: Prof. Carlo Fiorini

When dealing with X and γ-ray spectroscopy, nuclear physics experiments and medical imaging applications, the main focus is undoubtedly optimizing the energy resolution of the overall detection system. Thanks to the technological advances the microelectronics industry has encountered in the last fifteen years, readout front-ends are often implemented as multichannel application specific integrated circuits (ASICs) so as to cope with the increasingly high number of units in the detection modules. Under an architectural and operative perspective, monolithic crystal scintillator-based γ-ray readout electronics significantly differs from X-ray front-ends. A first difference might be represented by the data multiplexing strategy that, in the first case, needs to take into account that all the detectors of the matrix are hit by photons once a γ-ray is absorbed in the scintillator. This requires that all the readout channels analog-outputs are sequentially multiplexed after the first event is detected by a channel of one of the ASICs the detection module comprises. X-ray readout front-ends could require instead sparsification of events or, in high throughput applications, a high frequency asynchronous sequential multiplexing. Other important differences may also rely on the filter shaping times and gains, because different energy ranges have to be accommodated and because of the crystal scintillator, which pushes the electronics to exploit quite long processing times to minimize the ballistic deficit effect. In addition, most readout front-ends designed so far do not provide any analog-to-digital (A/D) conversion of output signals, making the analog transmission to the data acquisition system (DAQ) more susceptible to external disturbances, such as for instance the high electromagnetic fields in synchrotron facilities or in multimodality PET/MRI and SPECT/MRI medical imaging systems. All these constraints are often met by designing specific ASICs which may hardly find a versatile use in other applications. Aware of all these impairments, one of the main questions that have spurred my doctoral activity was: why not merge all the aforementioned peculiarities, thus paving the way towards a more general-purpose approach, by designing a fully programmable integrated front-end in principle suitable for both X-ray and γ-ray detection modules and, moreover, able to perform on-chip data digitization? The reference device outlined for the circuit design and characterization that followed is the silicon drift detector (SDD), because of its excellent noise properties in the typical range 0.2 to 30 keV, for applications such as the *energy dispersive X-ray spectroscopy* (EDX) and *X-ray fluorescence* (XRF). Thanks to its high quantum efficiency, SDDs are suitable and competitive alternatives to PMTs, PIN diodes and SiPMs also for the readout of scintillators. The ASIC designed, SFERA (SDD Front-End Readout ASIC), is intended to process the signals coming from solid state detectors coupled to CUBE preamplifiers. The technology adopted is the AMS 0.35 µm 3.3 V CMOS C35B4 because of its low flicker noise and cost, crucial whereas a high number of channels needs to be integrated. SFERA analog section is a stack of 16 readout channels, each of which comprises a 9th order complex-conjugate poles semi-Gaussian filter, that exploits a sub-optimum pulse shaping with a narrow time duration to maximize the throughput, followed by a peak detector (PKS) and a dedicated digital logic to synchronize PKS phases and reject piled-up events. SAs gains and peaking times can be defined according to the target application. The former, in case the chip is coupled to a 25 fF feedback capacitance CMOS preamplifier, can accommodate input energy ranges of 10, 16, 36 keV for X-ray detection and a 13800 and 20000 equivalent e^- in case of scintillation light readout. As for the peaking time, it can be set among 0.5, 1, 2, 3, 4 and 6 µs. A fast SA with 200 ns peaking time is also integrated in each channel to accomplish pile-up rejection purposes. Another programmable feature of SFERA is the data multiplexing protocol. Namely, a *polling-X*, *polling-y* and *sparse* readout are foreseen. The developed data acquisition system is made of commercial National Instruments™ hardware. A FlexRIO FPGA module manages all digital signals to/from SFERA while data digitization is performed by an NI-6115 ADC board. Custom aluminium boxes hosting the SDDs, CUBE preamplifiers and Peltier-cells allowed to cool the system down to -35 °C. State-of-the-art spectroscopic performance have been obtained using a 10 mm^2 single SDD, showing an energy resolution of 122.1 eV FWHM on the Mn-Kα emission line of an ^{55}Fe X-ray source (equivalent to an ENC of 3.2 e^-_{rms}) at 4 µs peaking time, which is among the best results measured so far with a silicon drift detector. An excellent energy resolution of 130.1 eV (ENC of 6.2 e^-_{rms}) has also been observed at the shortest 0.5 µs peaking time, already suggesting a possible clue about the adequacy of SFERA in high-rate X-ray applications. 3×3 SDD matrices with 8×8 mm^2 single units have also been employed, showing an average energy resolution of 134.3, 133.5 and 133.9 eV FWHM on the Mn-Kα line and a minimum resolution (one of the nine channels) of 130.3, 130.1 and 130.8 eV at 3, 4 and 6 µs peaking time respectively. Even better results have been measured on a single unit of another 3×3 matrix with the same format and technology, namely 127.2 eV FWHM, which is the lowest resolution achieved to date with these matrices, thus confirming the excellent noise performance achievable with SFERA coupled to SDD-arrays. The high-rate characterization of the chip aimed to validate the PUR logic efficiency and provide indications about the energy resolution degradation with the input count-rate. The measurements have been carried out pulsing a single channel of SFERA with a digital detector emulator, so as to arbitrarily tune the input count rate of the emulated signals from an SDD coupled to the CUBE preamplifier. Operated at the shortest 0.5 µs peaking time, the chip exhibits a PUR efficiency of 42 % at 1 Mcps input count-rate (420 kcps output count-rate), that overcomes the state-of-the-art performance achieved with the high-throughput HTRS ASIC, instead providing a maximum efficiency of 40 % at 800 kcps input-count rate at 0.6 µs peaking time. In the end, the dissertation focuses on the 12-bit SAR ADC embedded in SFERA, which represents a valid alternative to external converters in low-power and compact detection modules and furthermore ensures better immunity against external interference in environments with strong electromagnetic fields, such as synchrotron facilities. The electrical characterization of such ADC has shown an effective resolution of 10.75-bit, thus satisfying the original 10-bit ENOB specification. No differences in the output spectra have indeed been observed between what obtained with the external NI-6115 12-bit ADC and the embedded SAR, reason for which it represents a good candidate as the digitizer to be integrated in multichannel readout ASICs implemented in AMS 0.35 µm CMOS technology, for applications requiring less than 11-bit resolution up to 5 MS/s conversion rate. In the end, we might reasonably conclude that SFERA is a high-performance multichannel integrated front-end for the readout of signals coming from CMOS preamplifiers coupled to solid-state detectors. By exploiting a high degree of parameters programmability as well as high-order pulse shaping, different data multiplexing strategies and on-chip data digitization, it has a great potential for versatility together with also competitive performance with the current state-of-the-art readout electronic solutions, which makes it suitable for both X and γ-ray spectroscopy and imaging applications.

ENHANCEMENTS IN SPECTRUM MANAGEMENT TECHNIQUES FOR HETEROGENEOUS 5G FUTURE NETWORKS

Vincenzo Sciancalepore - Supervisors: Prof. Antonio Capone, Prof. Albert Banchs

The steady increase of traffic demand in current cellular networks requires new technologies and network architectures able to handle such demands. The growth foreseen cannot be only sustained by increasing the spectrum assigned to mobile radio networks. In fact, spectrum availability is already scarce in the ranges of practical interest, and spectral efficiency achieved by today's technologies, such as Long Term Evolution (LTE and LTE-A), is already close to Shannon's capacity limit. Therefore, the only viable approach to capacity increase is network densification with shorter distances between base stations and a large number of small cells. Unfortunately, network densification may seriously degrade the network performance due to interference increase thus decreasing efficiency in spectrum utilization. Therefore, interference control in dense network is a key enabler for future 5G networks.

In this context, it has been shown that the use of a scheme that does not rely on frequency reuse (frequency-reuse-1) is fundamental to have enough flexibility in resource usage and it can potentially improve efficiency. This implies that neighboring base stations (BSs) should be allowed to transmit on all available time-frequency resource blocks simultaneously, thus causing strong interference to each other's users. Interference must be then managed at each base station or at a coordination level between base stations. To this aim, several techniques have been proposed to independently cope with interference or low spectral efficiency in RANs, such as beamforming, MIMO or many others. None of them assume that these issues are highly intertwined because reducing the impact of either one strongly affects the others, thus seriously impairing the cellular network performance. This contrasts with interference mitigation and/or cancellation techniques that have been used for many years in the past, which basically exploit orthogonality of frequency and/or spatial resources. More recently, advanced solutions have been designed which actively reduce or cancel interference when orthogonality cannot be guaranteed.

In the framework of new generation high density networks, the complex management of resources requires a fine grain control architecture able to take decisions based on a global system view. Moreover, the programmability of control functions is a key feature for defining resource management algorithms tailored to the specific network scenario and the policies defined by the operator, as envisioned by the SDN paradigm. All the network elements belong to the same administrative domain and we neglect the security measures which must be implemented in practice to prevent malicious access of the control functions and to avoid unauthorized disclosure of sensitive information from customers.

Our aim is to present novel optimization mechanisms as SDN applications, aimed at enhanced wireless MAC operations in a district, which perfectly comply with the wireless software defined future network architecture requirements.

A high-speed backhauling enables the SDN controller to correctly gather all cellular user information, such as channel status (CSI), in real-time and to properly take scheduling decisions, which are promptly translated into ABSF patterns which each base station must apply for the next transmission period. As extensively proved by the literature, such centralized solutions achieve near-optimal solutions in terms of system spectral efficiency and thus, energy efficiency. However, this condition holds only when simplistic assumptions are taken into consideration. The huge explosion of mobile applications we are witnessing requires that network operators must introduce traffic guarantees for their customer contracts, burdening the network capacity. The compound effect of including traffic constraints with user and base station network densifications could be even further dramatic when considering complexity and delays. That is the reason why we are forced to consider a distributed approach, which must be enabled when network conditions are no longer under the SDN control. A distributed eICIC proposal with local decisions would not only be aligned with the well accepted self-organizing network concepts, but also allows to make ABSF and user scheduling decisions jointly---rather than assuming worst case conditions for the user scheduling process---which allows for further improving performance.

Note that one critical aspect in the design of the distributed scheme is to limit the amount of information exchanged between base stations as well. Therefore, we design a semi-distributed mechanism, which reduces the computational burden from a centralized SDN controller while drastically abating the signaling overhead. In particular, the ABSF coordination of local schedulers (base stations) is aided by the local controller, which supervises the ABSF decisions and drives the system to the best possible performance without requiring overall statistics and imposing centralized decisions. This makes our approach a practical and effective solution of ABSF that can be implemented in real networks.

On the other hand, the huge volume of traffic for mobile devices seems to be strictly required by the appearance of a great number of web and smartphone applications. This popularization of smartphones and the ensuing explosion of mobile data traffic lead to conventional cellular networks which are currently overloaded, and even worse in the near future. A large portion of that traffic consists in the distribution of content updates such as social network updates, road traffic updates, map updates, and news feeds (e.g., waze, an app for a social network for navigation, includes all the above mentioned features). Along with the appearance of such applications, some schemes have been recently proposed to offload the traffic generated by them in the cellular network. In particular, the device-to-device (D2D) paradigm has been proposed to assist the base station in the content distribution: with D2D communications enabled, the base station delegates a few interested mobile users (content injection) to carry and opportunistically spread the content updates to others interested upon meeting them (content dissemination).

Indeed, opportunistic communication exploits the daily mobility of users, which enables intermittent contacts whenever two mobile devices are in each other's proximity. These contacts are used to transport data through the opportunistic network, which may introduce substantial delays. However, the type of content concerned by cellular offloading may not always be entirely delay-tolerant. In many applications, it is indeed critical that the content reach all users before a given deadline, lest it lose its relevance or its usability. Therefore, the design of opportunistic-based cellular offloading techniques faces serious challenges from the intermittent availability of transmission opportunities and the high dynamics of the mobile contacts. In order to find the best trade-off between the load of the cellular network and the delay until the content reaches the interested users, any opportunistic-based offloading design must answer crucial questions such as, how many copies of the content to inject, to which users and when.

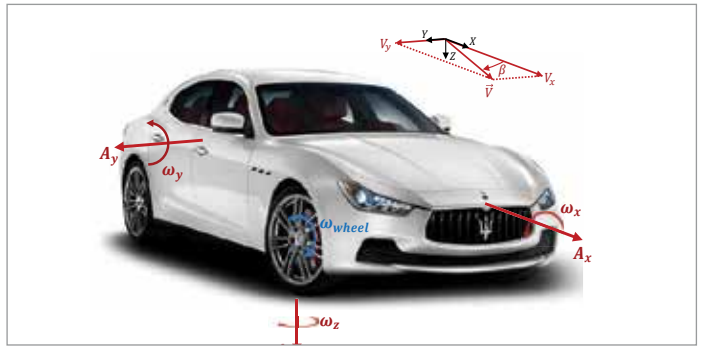
While most of the currently available offloading proposals focus on the characterization of content dissemination and the design of content injection strategies, they largely neglect the optimization of radio resources in the injection phase, i.e., the process of injecting a content in a subset of the mobile user population, which produces bursty and periodic traffic. Some existing work partially considered the impact of opportunistic resource utilization in the content injection strategies but their analysis is restricted to a single cell and does not consider the

interference caused by other cells, which is a key limiting factor for the deployment of dense and heterogeneous networks that are expected to appear in 5G cellular systems. In line with the 5G networks view, we leverage the heterogeneity of technologies in the network to implement a novel D2D-based offloading mechanism, which also tackles the cellular traffic offloading issue from a different and unexplored perspective: the inter-cell interference coordination problem. The rationale behind our approach is twofold: (i) interference is a key factor in future networks, where the single cell study case is not representative of a real network; (ii) content injection operations are impacted by network speed, which, in turn, strongly depends on intercell interference.

INERTIAL MEASUREMENT BASED WHEELED VEHICLE STATE ESTIMATION

Donald Selmanaj - Supervisor: Prof. Sergio Matteo Savaresi

Road vehicles are complex systems, composed of several subsystems, which interact among each other. Powertrain, transmission, brakes, steering system, suspensions, tires and other elements coexist in a vehicle and increase the overall complexity of the system. Each module is responsible for a specific function; nevertheless, all of the subsystems influence the vehicle dynamic behavior.
A principal goal of the control systems, and perhaps the most important, is the quality of the driving experience; the latter is perceived by the passengers in terms of performance and safety. Electronics is having a huge impact on the vehicle subsystems, in particular on the three major mechanical subsystems, the chassis system, the propulsion system and the interior system. They are all evolving from mechanical systems towards mechatronics systems that would not function without electronic control. The goal of the transformation is to increase comfort, safety as well as improving driving performance, fuel consumption, emissions and production processes. An essential part of the mechatronics evolution is the integration of the components of a system.
The chassis subsystem has seen the most evident development in automotive and has taken enormous advantages of the mechatronic miniaturization; it involves the longitudinal behavior of a vehicle as well as the lateral (turning) behavior and the vertical behavior (Figure 1). Components and control systems acting on the wheel steer, wheel speeds, wheel brakes, wheel torques, suspensions, camber angle and aerodynamic surfaces are all part of the chassis subsystem; they all aim at increasing the driving performances, such as reducing braking distance (longitudinal dynamics) and increasing vehicle speed on curve (lateral dynamics), increasing the passengers comfort (vertical dynamics) and increasing the vehicle safety.
This dissertation is concerned with state and parameter estimation problems for wheeled vehicles. The methods are meant to provide essential quantities for the vehicle chassis control. All the presented algorithms propose reliable solutions that aim at being independent from the vehicle at hand. In particular, the dissertation presents methods for estimating the planar dynamic state of a four-wheeled vehicle and for identifying dangerous conditions based on accelerometer measurements.
In four-wheeled vehicles, the knowledge of the longitudinal speed and the sideslip angle (i.e. angle between the velocity vector and the vehicle longitudinal axes) is crucial for the correct functioning of the chassis control systems. A method that uses the wheel speed measurements and the longitudinal acceleration measurement solves the problem of estimating the vehicle speed; the method makes use only of kinematic considerations and is independent from vehicle and road grip parameters. The sideslip angle estimation method is based on the kinematic approach which has the advantage of being independent from vehicle parameters, but is sensitive to measurement errors and in particular measurement offsets. A heuristic term is joined with the observer and solves the issue of divergence on straight drivings suffered by solutions based on the kinematic approach; it ensures the sideslip zero-convergence on straight drivings. Although the heuristic makes the observer robust to measurement errors additional components that estimate and compensate for the measurement offsets, the vehicle roll angle and the inertial platform displacement are included in the overall algorithm. The overall algorithm including longitudinal speed, sideslip and additional components is tested on experimental data. Results show the effectiveness and the reliability of the proposed solutions in all the driving and road conditions.
One parameter influencing the vehicle planar dynamics is the road grip: it influences the dynamic response of the vehicle and its limits. Chassis control systems are scheduled to take into account for the limitations of the road grip. A method for classifying the road grip and distinguishing between high and low grip conditions is introduced. The method is based on the RLS (Recursive Least Squares) algorithm, which is modified to reduce the effect of measurement errors on the estimate and to add a bias to the estimate. The major advantages are the fast classification of the grip change and the low sensitivity to measurement errors. The method uses inertial measurements and the estimates of the longitudinal speed and sideslip angle. Experimental data show the effectiveness of the method.
Recognition of dangerous states is important for activating safety systems. A method for distinguishing between dangerous and normal driving conditions is presented. The method is a two-step method and each step is based on self-organizing neural networks. The main advantages of the method are: (i) it automatically manages a large number of sensors and large tuning set of data; it does not require dangerous condition data to be tuned. The method, which is independent from the vehicle and the sensors type, is tested on a motorcycle in a simulation environment.



1. Vehicle chassis control. X, Y, Z are respectively the longitudinal, the lateral and the vertical axes. A_x , A_y , A_z are the vehicle accelerations; ω_x , ω_y , ω_z are the vehicle angular rates; V_x, V_y are the are the vehicle velocities; β is the sideslip angle and the ω_{wheel} are the wheel angular speeds.

A COMPREHENSIVE FRAMEWORK FOR THE DEVELOPMENT OF DYNAMIC SMART SPACES

Adnan Shazada - Supervisor: Prof. Luciano Baresi

We are living in an era where computers are becoming more and more ubiquitous in our everyday lives. There is an enormous technological growth in terms of increase in computing power, memory sizes and miniaturization of devices over the past few years. A plethora of inter-connectable devices such as tablets, smart watches, wearable gadgets, and sensing equipment are now available in a ready to use packaging. With all this advancement, Mark Weiser's vision of ubiquitous and pervasive computing world, where technologies interweave themselves into the fabric of everyday life in an invisible manner, seems to be realized in coming years.

Many efforts have been put to manifest pervasive computing in bound (physical or conceptual) environments to convert them into **smart spaces**. Smart spaces can be regarded as technology augmented intelligent environments with the ability to understand user/human needs within a space and react accordingly in order to provide contextualized services to the inhabitants. A smart space can be personal such as home, office, and assisted health care system, or a more open public place such as a shopping mall, park, train terminal and airport. Typically, a smart space comprises multiple autonomous entities (with different capabilities) that are heterogeneous in terms of functionality, communication protocols, and execution models. These heterogeneous entities co-exist and collaborate with each other to help inhabitants accomplish their tasks. Another important characteristic of these spaces is openness as different entities move while being connected to the space (other entities) and can join or leave the system unannounced.

Despite all these advancements, the pervasive computing vision is still reality only in parts, as of today. There have been some limited deployments such as smart homes and automation solutions for indoor spaces. But, they are yet to be widely exploited in physical spaces in the same manner as the ubiquitous Web and smartphones have been used to effectively revolutionize/change our daily lives. Given the state of all these sophisticated smart objects and technologies, one must ask why do we not already live in smart pervasive spaces (environments) where all these available devices are seamlessly providing us the services to enhance our efficiency and improve our lifestyles? One of the reasons for this is that all these hardware equipment and smart objects are one of the many prerequisites for the realization of such environments. The growing complexity of the devices also requires the *software development process* for such systems to be re-modeled. As the smart spaces are inherently complex, dynamic and open in their nature, the required software solution needs to be able to *adapt* itself to the always changing conditions of the spaces without any (or limited) external human intervention.

We conducted a comprehensive survey of the state of the art to investigate where do existing solutions and approaches lack and what is the right approach for building diversified smart spaces. Numerous solutions have been proposed by the researchers to solve different aspects related to the realization of these smart spaces, but there is still a gap between what is available today and the need for an effective end-to-end development framework. Many smart devices, off-the-shelf autonomous objects, and architectural solutions are available, but these systems often work in isolation and do not put an effort to build on top of/in complement to existing solutions. Most of the existing systems do not provide the support for the continuous validation and testing of alternative solutions in order to move seamlessly from design to actual deployment. Moreover, existing solutions also lack the balance between the control and the autonomy of smart space components. The majority of these solutions are developed through centralized architectural control loops (which are prone to have scalability issues) or completely decentralized and autonomous adaptation solutions (such as bio-inspired systems, which lack the control over the adaptation). Although, all these approaches are interesting and useful for certain kind of scenarios, there is a strong need for a decentralized yet manageable general solution that can be used to develop various kinds of smart spaces. We have established that in order to conceive all kinds of spaces, a framework needs to have the following properties: ability to abstract over heterogeneity, support to integrate different components and protocols at runtime, facilitation of incremental development of the smart spaces, cater scalability and dynamism of the components and the flexibility to use or replace alternative solutions or subsystems. Therefore, the goal and major objective of the thesis can thus be described as:

"Provide a self-adaptive framework that enables the developers to design, implement and validate dynamic smart spaces in an incremental fashion by offering programming abstractions that are suitable for the whole development life-cycle."

The thesis employs an **incremental development process** to create smart spaces in order to fill in the gap caused by the inability of existing smart spaces to evolve over the time and to deal with changing requirements. The group-based self-organizing framework offers the same abstractions throughout all the development phases while providing means for both the seamless integration of various components and the utilization of existing systems. The framework provides a fixed software backbone that allows the developer to move seamlessly from a fully virtual/simulated solution to a completely deployed system in an incremental manner. It provides interfaces both to surrogate system components through external simulators, and to ease the deployment of physical elements. In this way, the framework enables early evaluation of "incomplete" systems, assessment of alternative solutions and deployment of the system in a continuous and incremental manner. The proposed solution also integrates the conventional component-based control (autonomic computing) and the inherent self-adaptive capabilities of the bio-inspired (firefly-based) ecosystem. The framework eliminates the individual shortcomings of these approaches, that is, lack of inherent situated awareness and mobility, for autonomic approaches, and lack of control over the self-organization in bio-inspired systems. The integrated approach provides the required autonomy to the individual components/groups and at the same time ensures the desired level of distributed control at various granularity levels. Moreover, the proposed firefly-based adaptation ensures the uniform distribution of workload for each component in order to ensure efficiency of the system and avoid congestion.

Four different case-studies (with varying needs and characteristics – from static and closed spaces to more open and dynamic spaces) are selected to assess the feasibility of proposed concepts, fulfilment of system requirements, and various quantitate and qualitative measures (metrics). An evaluation plan is devised, which describes all the metrics (corresponding to the identified requirements in result of literature review) that are required to be measured for each of the four case-studies (smart office, modern greenhouse, energy efficient public building, and public park). The results show that the framework incurs a small (messaging and bootstrapping) delay, which is not significant for the systems of this scale and hence it provides a practical solution for the issues identified in the survey of the state of the art. The results also demonstrate that the presented solution complements the existing growth of the smart objects and plethora of software solutions by providing a framework to integrate/utilize the available solutions as a step forward towards the efficient end-to end development of smart spaces.

MIXED-INTEGER PROGRAMMING MODELS AND METHODS FOR BILEVEL FAIR NETWORK OPTIMIZATION AND ENERGY COGENERATION PLANNING

Leonardo Taccari - Supervisor: Prof. Edoardo Amaldi

This thesis addresses two relevant optimization problems, with application in telecommunications and energy systems, with mixed-integer programming (MIP) methods.

Part I focuses on a bilevel multi-commodity flow problem subject to max-min fair flow allocation, which arises in telecommunication networks with elastic demands. The problem is motivated by routing in Internet Protocol (IP) networks, where there are no prescribed demands to be satisfied, since the network provides a best-effort service. The network operator aims at maximizing a utility function (total throughput), by selecting the routing paths, while the bandwidth is allocated *fairly* by the transport protocol. Accordingly, we define the maximum-throughput Unsplittable Flow Problem subject to Max-Min Fair flow allocation (UFP-MMF) as the bilevel problem where, at the upper level, the routing paths maximizing the total throughput are sought, while, at the lower level, the flow is allocated to each origin-destination pair maximizing a fairness measure. In this work, we analyze the problem from a theoretical standpoint, and investigate various MIP-based solution approaches.

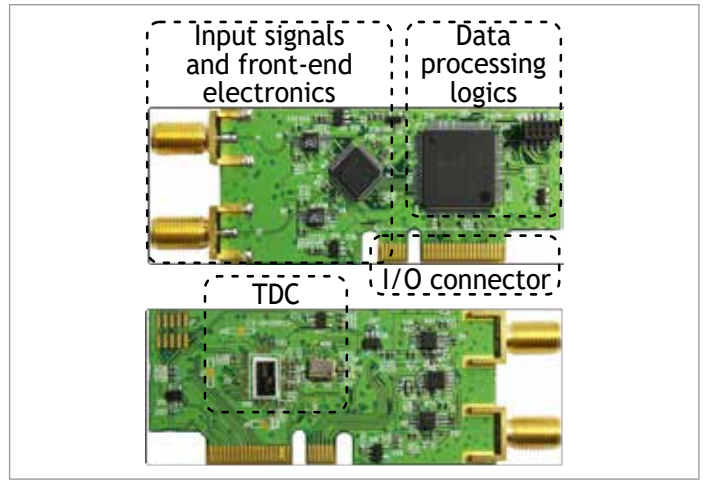
The bilevel problem can be cast as a single-level problem exploiting the concept of *bottleneck arcs*. Two different approaches are then developed: a branch-and-cut algorithm based on an arc formulation, where subtour elimination constraints are separated, and a branch-and-price algorithm based on a path formulation. We also propose a heuristic based on local search that provides good feasible solutions in a short computing time. Computational results are reported on different topologies from the SND library. The best exact method is the branch-and-cut algorithm with separation of generalized cutset inequalities, although optimality remains out of reach on a good fraction of the instances. The heuristic approach is effective in obtaining close-to-optimal solutions for the hardest instances in short computing times.

Part II addresses an operational planning problem in energy cogeneration system with thermal storage, where one has to determine the operations of a network of (co)generation units (i.e., on/off status and production level), over a given time horizon, in order to satisfy users demands and minimize the operating costs. We propose mixed-integer programming approaches, focusing in particular on some basic variants of the problem. For the variant with constant production upper and lower bounds, a dynamic programming-based polynomial algorithm is described. In presence of uncertainty in the data, we also introduce a revised Γ-robust formulation, inspired by the one in the lot-sizing literature. Finally, we discuss real-world cases, where the problem is solved as a MINLP or as a MILP, exploiting a piecewise-linear approximation. The problem can typically be solved with exact methods for time horizons between one day and a few weeks. When annual economic incentives have to be taken into account, we also propose a MILP-based rolling-horizon heuristic that is applied to larger-scale problems provided by an Italian energy company.

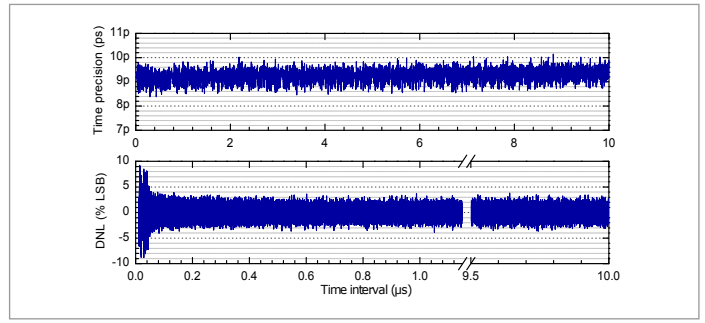
TIME MEASUREMENT INSTRUMENTATION FOR SINGLE-PHOTON COUNTING APPLICATIONS

Davide Tamborini - Supervisor: Prof. Franco Zappa

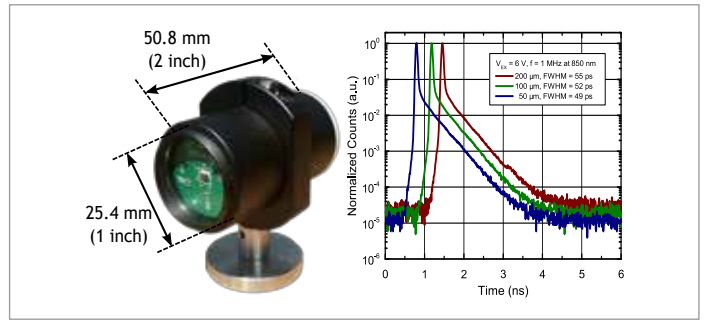
Many applications in medicine, biology, chemistry, engineering, and industry rely on precise time-interval measurements, in order to reconstruct very-low intensity and ultra-fast (at picosecond level) optical waveforms. In these applications, optical signals consist of just few photons per event and the discrete nature of the signal itself prohibits any analog acquisition. In many cases signals are also very fast, therefore a photodetector with very wide bandwidth would be required, if analog sampling were employed.
To overcome these limitations, the reconstruction of ultra-fast, time-resolved optical waveforms is carried out by means of the Time-Correlated Single Photon Counting (TCSPC) technique, which is based on the detection and "counting" of single photons composing the optical signal of interest, together with the "timing" measurement of their arrival time.
The TCSPC systems' requirements are very demanding: the single photon detector has to provide high detection efficiency (better than 50%), low noise (less than 100 counts per second), large active areas (at least 50 µm diameter) and low timing jitter (tens of picoseconds at the most), while the time-measurement resolution has to be very good (tens of picoseconds or better) with very low differential nonlinearity (DNL of about 1% LSB or better). Some TCSPC modules are commercially available, but most of them are bulky and power consuming, thus limiting the chance to develop multichannel systems to just one or very few channels, at the most. On the other hand, dense arrays of single photon detectors and time-measurement circuits have been developed to address more and more applications demanding many parallel channels. However those arrays reach neither the resolution nor the linearity required by most high-end TCSPC applications.
My PhD research activity aimed at designing and developing new state-of-the-art, multi-channel TCSPC instrumentation, with reduced power consumption and dimensions, for overcoming current TCSPC limitations.
The first goal was to develop a single-channel time measurement systems suitable for exploiting the TCSPC technique at best, starting from the time measurement core itself, namely a time-to-digital converter (TDC) microelectronic chip developed at Politecnico di Milano, able to provide top-level performance (10 ps resolution, 17 ps rms precision, and 1% LSB DNL) and with an architecture suitable for parallelization. My first activity consisted in validation and test of the employed TDC in real TCSPC applications, (e.g. optical time-domain reflectometry and fluorescence lifetime measurements).
The following step was the development of multi-channel, low-power TDC systems, from conceiving a novel the TDC card board (see Fig. 1), able to exploit the TDC chip performance with low power consumption and form factor suitable for easy integration into multi-channel instruments. The TDC card was employed as building block for dual- and eight-channels time measurement systems. These multi-channel instruments reach state-of-the art performance with a significant reduction of power consumption, and provide also an auxiliary input common synchronization channel able to implement different operating modes, to either increase the overall conversion rate or to improve the timing resolution (Fig. 2 shows the eight-channel module performance, reaching 1.25 ps resolution).
The final goal was to demonstrate and develop high-performance, extremely compact, single- and multi-channel TCSPC systems able to comprise both the time measurement core and the single photon detector, within a compact "detection head". To this purpose, I developed a first module based a 16x1 linear array chip of CMOS SPADs and TDCs. Finally, I conceived and designed a novel, hand-held system able to host different single-photon detectors and the TDC chip (see Fig. 3), proving state-of-the art (50 ps FWHM) performance.
These modules open the way to many new applications, especially in biomedical and industrial fields, which require high performance but also a compact, portable, low power instrumentation.



1. Picture of both sides of the TDC card, showing the input connectors with the front-end electronics to provide to the TDC the proper signals. Then, a data processing logics read TDC data, perform its processing and provide measurement result on the I/O connector.



2. Timing precision (top) and DNL (bottom) of the eight-channel time-measurement module, operating in averaging mode. It is able to measure a time interval with a precision of 9 ps rms and a DNL better than 1.5% LSB rms (i.e. 18 fs rms), over the whole 10 µs measurement range.



3. Picture of the hand-held, single-pixel, complete TCSPC system (left) and its temporal response (right) to a sharp laser pulse at 850 nm at different detector active area, resulting in an overall response of just 50 ps FWHM.

DEVELOPMENT OF AN ASIC FOR SIPM READOUT IN SPECT APPLICATIONS

Paolo Trigilio - Supervisor: Prof. Carlo Fiorini

An important role in the field of medical imaging is being played by multimodality systems, capable of correlating information obtained by means of different traditional imaging techniques that were once independently performed on patients.
Particular interest is assumed by the simultaneous combination of *functional* and *morphological* analyses, finalized to allow the physicians to track biological processes or detect eventual pathologies while being at the same time aware of the anatomical structures of the object under investigation. The aim of this approach is to achieve a cutting-edge image quality, opening the way for new medical scenarios.
The INSERT project, funded by European Community under the FP7-HEALTH programme (Grant agreement 305311), aims to the development of a novel Single Photon Emission Computed Tomography (SPECT) system compatible with a pre-existing magnetic resonance (MR) apparatus, with the goal to provide an accurate brain tumour (glioma) clinical diagnoses which could lead to the choice of proper personalized treatments.
This SPECT imager, realized both in a preclinical and clinical version, will be composed of a ring of several independent gamma camera modules, and will be hosted in the center of the MRI bore.
In order for this system to operate in the strong static magnetic field (up to 9 T) and under the harsh electromagnetic transients of an MRI scanner, the adoption of solid state detectors instead of photomultiplier tubes, traditionally used for SPECT, is mandatory. The INSERT SPECT module therefore features Silicon Photomultiplier (SiPM) photodetectors, devices currently experiencing a dramatic increase in popularity thanks to their constant performance improvements over the last decade.
The present doctoral thesis focuses on the design, realization and test of SiPM-readout application specific integrated circuits (ASICs) for SPECT systems. State-of-the-art SiPM front-end circuits mainly address the exploitation of the excellent timing properties of this device, allowing to achieve the sub-nanosecond temporal resolutions necessary for Positron-Emission-Tomography (PET) and for Time-of-Flight PET. The ASICs described in this work target the achievement of high spectroscopy performances at the relatively low gamma energies (in the 100 to 300 keV range) used in SPECT.
The single detection module relies on the Anger architecture, i.e on the use of a continuous scintillator crystal read by a photodetector matrix. The rays coming from the object under test pass through a collimator block, with the purpose to allow only photons with specific directions to reach the crystal. When a photon impinges on the scintillator, a visible light flash is generated and spread all over the various elements of the photodetector array in a non-uniform unique manner, depending on the coordinates of the hit point. The amount of light measured by each array pixel contains the information about the position of interaction point of the photon in the crystal.
Despite the use of reconstruction algorithms on the raw data collected by the photodetector matrix allows to distinguish much smaller details than the pixel size, requirements on spatial resolution still demand a large number of electronic readout channels to cover a field of view suitable for clinical applications. An ASIC represents a compact solution to deal with the large pixel number; its purpose is to provide an analog signal conditioning before the conversion to digital format for successive elaboration.
The INSERT ASICs should extract the information related to how the initial photon energy, converted into visible light by the continuous scintillator, distributes over the SiPM matrix. The main request is thus to maximize the signal to noise ratio, in order to guarantee the needed energy and spatial resolution performances. Angus, a 36-channel ASIC in CMOS 0.35 μm technology has been designed to this purpose. The complete clinical module makes use of two chips to read its 72 SiPM pixels, whereas the preclinical version of the instrument features a 36-pixel detection module that can be read by a single ASIC.
A low input impedance front-end stage, capable of coping with the large capacitances of the 8 mm × 8 mm SiPM pixels of the INSERT module with a limited current consumption, has been developed. The successive filter stage is a programmable RC circuit, in turn followed by a peak stretcher to provide an external analog to digital converter with a stable voltage value. The chip was presented in an oral presentation at the IEEE NSS-MIC 2014, in Seattle.
The development of the preclinical module based on the realized ASIC has been completed, and first tests prove its correct operation; bi-dimensional scintigraphies on mice have been successfully performed. The achieved performances are close to the fulfilment of the clinical and preclinical requirements; the use of better SiPMs detectors, currently under production, is expected to fill the gap. Minor issues present on the first Angus version led to a redesign of the chip. This new release is expected to further improve the results.
In addition, tests have been performed to assess the MRI compatibility of the operating detection module, with a very encouraging outcome. Though the most severe MRI imaging sequences significantly corrupt the acquisition of input pulses which emulates events, the involved mechanisms causing this signal degrading have been understood with a good degree of clarity.
The author, and the team he belongs to, are confident that the use of the second Angus version along with minor modifications in the surrounding electronic system will lead to the complete solution of the issues. Conversely, no disturbance induced by the electronics has been noticed in the MRI operation.
In parallel, a novel single channel readout ASIC has been developed. This prototype chip, named Angus GI and again implemented in CMOS 0.35 μm technology, features a self-triggered gated integrator architecture, realizing a quasi-optimal filtering of the input signal with a shorter processing time than the one needed by the Angus RC filter.
Tests on this chip are currently going on; preliminary obtained results have been divulged in an oral presentation at the IEEE NSS-MIC in November 2015.
The thesis is organized according to the following structure:

- Chapter 1 provides a short introduction to multimodal imaging systems, presenting the INSERT project and the motivations to the design of ASICs for this application.
- Chapter 2 deals with Silicon Photomultipliers, illustrating their operating principle and listing their main parameters. An overview of the main SiPM readout strategies reported in literature is afterwards presented.
- Chapter 3 focuses on gamma spectroscopy with SiPMs, with direct application to the INSERT project. The main contributions to energy resolution degradation are spotted, and the criteria followed in the filter choice are shown.
- Chapter 4 describes the design of Angus, presenting each circuital block and its operation in the complete electronic system of the detection module.
- Chapter 5 reports about the main measurement results performed with Angus, concerning energy resolution, spatial resolution and MRI compatibility of the overall system.
- Chapter 6 illustrates the redesigned version of Angus, aimed to solve technical issues noticed during the testing phase of the first chip, and presents the development of the self-triggered gated integrator prototype, Angus GI.

EVALUATING FORENSIC EXAMINATIONS IN A COURT OF LAW THE DIKE MODEL

Alessandro Trivilini - Supervisor: Prof. Licia Sbattella

Abstract

The DIKE (Description of Interrogations by Knowledge Extraction) project –composed of a model, a corpus, and a prototype– aims at analyzing examinations in a court of law. DIKE is based on a multi-dimensional, multi-level conceptual model, which represents several aspects of examinations according to psychological, juridical, and linguistic theories. A multi-dimensional, multi-level, HMM-based classifier implements such a model and permits to annotate new examinations; for training such classifier we created a new audio/textual annotated corpus, based on recordings and transcriptions of real examinations collected from Italian trials, and annotated with the sentence- and utterance- level labels defined into the DIKE model. Finally, the DIKE exploration tool analyzes annotated examinations and generates their profiles as well as profiles of those partaking in the examination dialogue.

Motivation and main contribution

The main motivation behind this project comes from the intention to define a conceptual model to tackle the issue of analyzing examinations in forensic contexts. Our goal is to support experts in improving their examination strategies and teachers in improving the effectiveness of their training activity. The DIKE model is based on a general interdisciplinary model adopted and defined by our research group to analyze the interactions between two people within a critical and stressful context.

The general model lays on interdisciplinary competences (linguistics, psychology, and forensics) and has been thought to fit various contexts (i.e. didactic, clinical, and juridical). Our approach puts the emphasis on a multidimensional analysis of the linguistic and prosodic traits of speech and dialogue, by means of a stochastic model.

We decided to focus DIKE on forensic examinations mainly because of the growing interest in forensic sciences for dialogue analysis. Moreover, we managed to get real audio recordings and transcription, which permitted us to build an ad-hoc corpus. Finally, previous collaborations of one of the authors with judicial authorities provided a privileged and strategic access to the knowledge we needed.

This work aims at the design of an original, conceptual model conceived through the exploitation of linguistic, juridical, and psychological theories. Thanks to this model, it is possible to retrieve useful information that permit to understanding how a forensic interrogation develops.

The first contribution of DIKE is teaching how to improve interrogation methods, given the current lack of tools and models that allow evaluating interrogations in an objective way. As of today, professionals recur exclusively to a few, basic rules drawn from juridical literature to conduct interrogations, and to a personal –that is subjective– set of skills for their subsequent evaluation. By means of the DIKE exploration tool, a more accurate and detailed subsequent analysis of interrogations would allow professionals to improve their existing examination techniques and to define some anew.

The second contribution of DIKE is making it possible to analyze interrogations in details. In particular, DIKE permits to define an interrogation profile and to identify a profile for the interrogation style in presence of and depending on those partaking in the dialogue. Moreover, DIKE makes it possible to identify and study profiles of examiners and examinee.

The third contribution of DIKE is a corpus made up of real forensic examinations, annotated according to the DIKE conceptual model, which can be used to conduct new analyses and conceive new interpretation models.

The model representing forensic examinations

We divide the participants to the examination into Actors and Roles. In particular, the Actors taking part to examinations are: A - person undergoing examinations, with the following Roles: Victim, Defendant, Witness, Investigator; B - persons conducting the examinations, with the following Roles: Judge, Prosecutor, Counselor.

The Prosecutor and the Defendant's Counselor are the competing parties. Each of them provides a list of examinees (the Victim is by definition a Prosecutor's examinee). The Investigator's Role includes the investigation police department, police agents, and various experts.

shows the multi-layered structure of the DIKE model. Each layer leverages output from all the preceding layers. The model input is composed of features generates by the Feature calculation phase, which works on aligned audio-textual sentences.

Level 0 takes into account the process of audio and textual features calculation. It is the entry level of the conceptual model, and it is used to extrapolate all needed features by starting from real forensics examinations. Some of such audio features are intensity variations, pitch variation, silences, speech rate, etc.

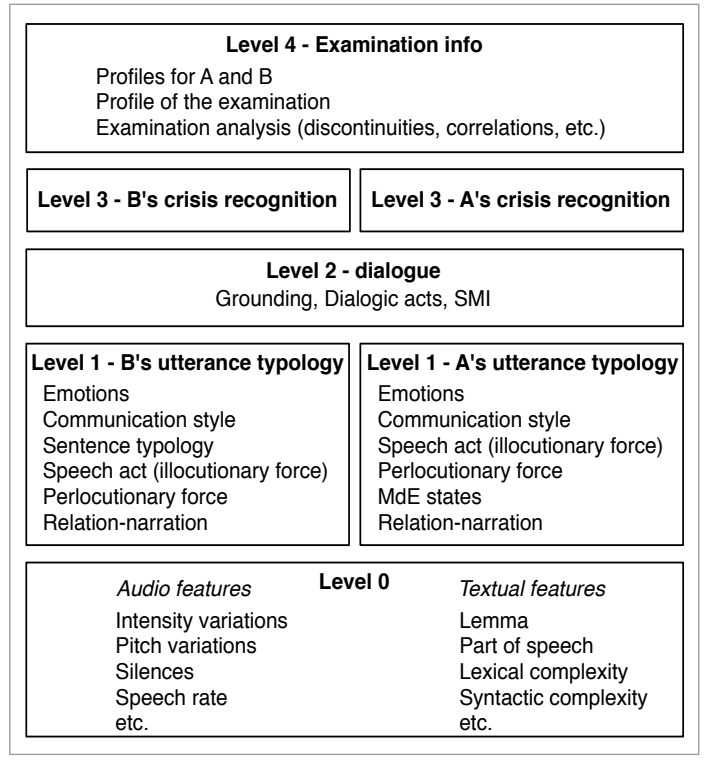
Some of selected textual features are lemma, part of speech, lexical complexity, syntactic complexity, etc.

Level 1 leverages audio/textual features, and generates six different utterance characteristics for B and A. Such characteristics represent information extracted analyzing a single Actor's utterance (i.e., without considering interaction with the other Actor), and are, for A: emotions, Communication styles, Illocutionary speech acts, Perlocutionary speech acts, MdE, and Relation-narration. For B the same list holds, but Sentence typology substitutes the MdE.

Level 2 leverages both the outputs of Level 1, and the SMIs associated to utterances' verbs for A and B. Level 2 considers the whole sequence of utterances (i.e., the whole examination) and associates to each of them a Grounding state, a Dialogic act state, and a SMI state.

Level 3 considers some dialogue exchanges in the past, starting from the current utterance, and checks whether A (or B or both) can be declared as experiencing a state of Crisis. Notice that crises of A and B are searched by means of two specialized models.

Level 4 leverages some "boundary examination information" such as the A's Role profile (for the whole examination), the B's Role profile (for each of the B's utterances, by remembering that B represents a group of persons), and the interrogation profile itself (by identifying and analyzing discontinuities, correlations, etc.) It is in charge of generating the final



1. The DIKE model

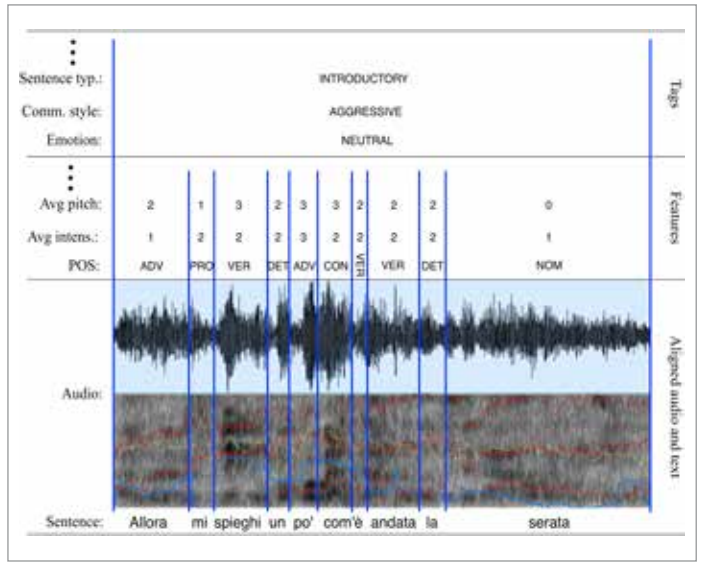
output by providing Actors' profiles and the whole examination profile, by means of statistics.

The corpus

To our knowledge, a corpus of real forensic examinations –built by considering forensic, psychological and linguistic perspectives– does not exist. Thus, as a part of this project, we collected real examinations from Italian trials and defined an audio/textual corpus, tagged with sentence- and utterance-level labels. The dataset of real examinations we collected is composed of about 30 Gbyte of recorded audio files and 20 Mbyte of textual transcriptions (see). Human experts used the well-known Praat application to add 11 annotation tiers –in accordance with the conceptual model– to sentences and utterances. So far, 36 examinations have been annotated, containing 1398 utterances and 2136 sentences.

The classifiers

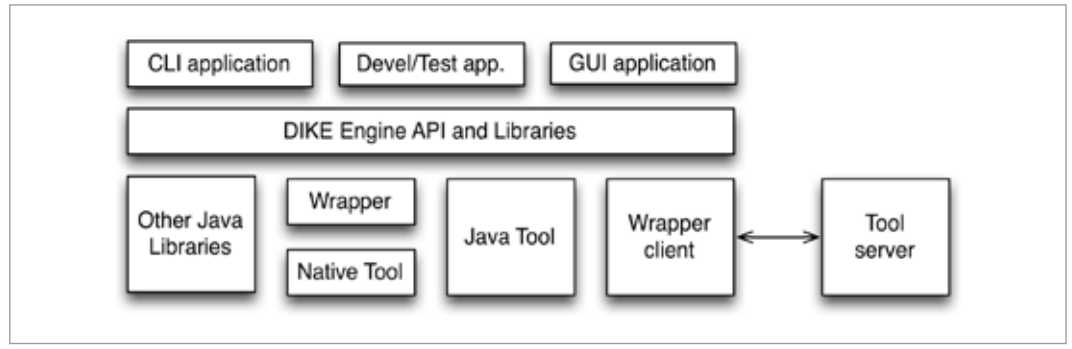
The Level 0 generates a huge set of audio and textual features, associated to each word of the aligned audio/textual input data. Two feature categories are considered: categorical textual features, such as recognition of misspellings, recognition of broken sentences, POS tagging, etc.; and continuous audio features, such as pitch, intensity, prosodic emphasis, speech rate, etc.
Level 1 leverages the audio/textual features generated by Level 0, and generates six different utterance characteristics for B and A, by means of two sets of HMMs.
Level 2 is based on output of Level 1 and, by means of three HMMs, generated three characteristics to each utterance of the whole examination.
Level 3 is based on output of Level 2 and, by means of two sets of HMMs, is devoted to the recognition of the "crisis" state for A or B.
Finally, Level 4 generates the final output, consisting of profiles associated to A, B, and the whole examination.
The classifiers were trained by means of our corpus, and tested adopting a k-fold cross-validation approach. Results were promising but highlighted that the corpus is still too small for some characteristics.



2. Graphical representation of an excerpt of the corpus

The prototype

shows the main architecture of DIKE, the prototype we are developing. DIKE is composed of three layers: an application layer, a core layer, and an NLP layer.
The application layer defines three different way to access the DIKE functionalities: a command-line application will permit to use DIKE as a part of scripts and other NLP pipelines; a development/test application, mainly used for testing DIKE during the development phase; and an application with graphical user interface, mainly for demoing the DIKE functionalities.
The core layer contains all the models, algorithms, and functionalities provided by DIKE. A well-defined API (Application Program Interface) will permit to easily add new applications to the system.
The NLP layer contains the NLP tools we leverage, and the glue code for using them in the DIKE core. Many NLP tools are available as Java libraries, thus we decided to write DIKE using such a language; this means that native-code tools need wrapper code for Java interoperability; finally, wrapper code is needed to interoperate with NLP tools that are implemented as separated servers.



3. The DIKE prototype

Conclusions

The DIKE project aimed to propose a methodology for the computational analysis of forensic examinations. To reach that goal, we defined an interdisciplinary conceptual model, created an annotated corpus, and demonstrated that a classifier can be trained and that the results, in terms of accuracy, are promising. We plan to release the corpus, along with the annotation manual we developed, to the scientific community. The prototype, coded using MATLAB, Java, and Python, has been developed and tested under Windows, Linux, and OS X, and will be released as open source.
We hope the DIKE project will offer new research opportunities to develop novel methodologies and tools in the field of dialogue analysis in forensic settings. DIKE is still an ongoing project. We plan to modify and improve several areas of the project. First of all, the corpus needs to be expanded. We own many recordings that, for budget reasons, were not cleaned and annotated; moreover, new recordings –exhibiting more emotive dialogues– are needed; finally, the overall quality of the annotation could be improved, leveraging what we learned creating the current version of the corpus. Secondly, the classifier could be improved. More sophisticated features could be explored and tested, new classifiers could be experimented, and a more optimized version of the code could reduce the computation time.

MODELING AND QUERYING GENOMIC DATA

Francesco Venco – Supervisor: Prof. Stefano Ceri

DNA is the molecule that encodes the instructions necessary for the development and functioning of all living organisms. DNA is copied and passed from parents to ospring; sometimes errors or mutations happen and are passed to the future generations, sometimes giving to them advantages.

We define as DNA sequencing the process of determining the precise order of nucleotides within a DNA molecule; in the last years new techniques emerged, making the sequencing process faster and cheaper. Sequencing is permitted by computers, and the result are files, in the majority of cases containing elements with a specific position on the genome.
Such information is further analyzed by biologists using various algorithms and heterogeneous techniques.

Genomic is a young science and for sure suffers a certain degree of fuzziness: it is evolving fast, new techniques are created and old ones become more precise or are abandoned. Our intuition however is that, no matter how it was produced, the final output of a NGS experiment is information on the DNA. Our very initial goal was to try to build a data model that could be flexible enough to adapt to changes and new data types, but at the same time structured enough to permit interoperability. We thus propose the the Genometric Data Model as a general abstract representation of genomic regions. The position of a gene, a mutation,... all these data types, no matter how heterogeneous, have one thing in common: they contain information related to specific regions of a DNA molecule.

We define a **genomic region** a well-determined portion of the genome, qualified by a quadruple of values, called region coordinates:

- **chromosome**.
- **left end**, encoding the index of fist base on the chromosome
- **right end**, encoding the index of last base on the chromosome
- **strand**

Genomic experiments always produce collections of homogeneous regions, i.e. containing the same type of information. In GDM, we define a structure that mimes this situation and we call it **sample**.
All the regions belonging to the same sample must have some common characteristics;
they all share the same structure, i.e. region schema, and they will have some data in common. For example, suppose that we are considering regions extracted from a specific file, obtained by sequencing transgenic mouse brain cells treated for ChIP-Seq analysis. At the end of the analysis, we will end up with a BED file or similar, showing the peaks for the desired protein binding positions.

Each peak can be obviously be modeled as a region, with probably some score attribute; all regions will have different coordinates along the genome and different values for the score, but the schema is necessarily the same. But what about all the other information? A good investigator will record for sure the basic characteristics of the original biological sample, the treatment used, the algorithms parameters, etc. All this data, however, do not belong to specific regions in the bed file, but to all the regions as a whole. We call this kind of information **meta-data**.

Sometimes it is better to group samples together: suppose that we made the same ChIP-Seq experiment on 10 different mice and we obtained 10 samples.
As samples are collections of regions, a data-set is a collection of samples, all sharing the same region schema rs but presenting different ids.
Ideally, each data-set corresponds to a collection of related experiments or information. It could be as simple and restricted as a of group bed files containing gene positions coming from different sources, or broad like all the files coming from a public project. Each data-set is only described by a unique name.
The global view of the GDM is illustrated by figure [francesco_venco_fig1.jpg].

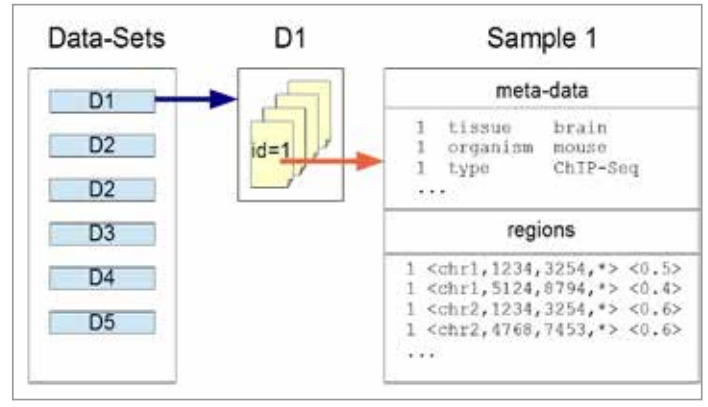
To fully exploit our model, we designed and implemented the GenoMetric Query Language, or GMQL in short. What we would like to be able is to pose questions like "Given a set of genes, select all the ones that are bound by the protein P in the condition C, then count the number of mutations found in each gene for a pool of patients"

GMQL operations form a closed algebra: results are expressed as new datasets derived from their operands; thus, operations typically have a region-based part and a metadata part, the former consists of regions, the latter one traces the provenance of each resulting sample. Most GMQL operations, although defined upon two connected data structures, are intuitive for readers who are knowledgeable of relational algebra.

Some operations are unary, like SELECT, which keeps in the result all the dataset samples which existentially satisfy the metadata predicate, and then selects those regions of selected samples which satisfy a predicate on the region's schema.
Other operations requires two operands, like the JOIN: a meta join predicate builds new samples from pairs of samples, one of the right and one of the left dataset. Then, a genomic join predicate, dealing with distal properties of regions, selects the regions to include in these new samples.

We implemented our system on a cloud computing framework; the Hadoop distributed file system is used to store the files and implement the GDM. In the first implementation Apache Pig is used to implement GMQL. In particular, we make heavy use of user defined functions in Java to efficiently tackle region operations based on distance. Recently, the group started working on a Flink and a Spark based implementation. A more refined GMQL compiler is now able to optimize the queries exploiting a DAG representation of the language and its double nature, executing first the cheap meta-data operations.

Out work is the first step to meet the need of managing in an easy, intuitive and scalable ways growing repositories of genomic data, and the joined biological projects with scientists of IEO and IIT have conrmed the strong aspects of our approach.



1. GDM representation: a collection of data-sets, each one being a collection of samples, containing regions and meta-data attribute-value pairs. More in details, each data-set (e.g. the data-set called "D1") contains a set of samples, each one with a different numerical id. Each sample is a collection of meta-data and regions. Meta-data are in the form of triplets composed by the id and the attribute-value pairs, while regions are composed by the id, the region coordinates and a list of values determined by the data-set schema.

COLLECTION, PROCESSING AND ANALYSIS OF ENVIRONMENTAL DOMOTIC SENSORS DATA FOR BEHAVIOR DRIFT DETECTION

Fabio Veronese - Supervisor: Prof. Fabio Salice

Demographics are showing an unprecedented growth in the number of the elderly people. The number of older persons aged 60 or over is projected to be 2 billion in 2050, three times the number in 2000. World Health Organization (WHO) considers health, safety, independence, mobility, and participation as five higher level needs of the older adults. It is widely accepted that ICT can play an important role in improving the quality of life of the elderly people. Assistive Technology (AT) and Ambient Assisted Living (AAL) are examples of ICT helping elderly and disabled people. AT is defined as "any device or system that allows an individual to perform a task that they would otherwise be unable to do, or increases the ease and safety with which the task can be performed". AAL can be best understood as the marriage of AT and ambient intelligence, which comprises artificial intelligence applied to home environments.
The Assistive Technology Group (ATG) is a multi-disciplinary group in Politecnico di Milano. The main goal of ATG is to provide innovative ICT based and sustainable solutions addressing the problem of fragility, especially for elderly and disabled people. BRIDGe is an AAL system developed by ATG with the main goal of providing mutual reassurance for its users: the person inside the home, the family and caregivers outside the home. It is devised to be modular, interoperable, low-cost, and customizable. Inside the BRIDGe project part of the research focus is devoted to data collection, processing and analysis for behavior drift detection. In this doctoral thesis four main declination of such focus will be addressed: the proficient collection of indoor human localization data, the HA (Home Automation) data processing to ensure dependability, the generation of synthetic HA datasets, and the analysis of data to create a correlation between HA status and Activities of Daily Living (ADLs).
Localizing the inhabitants of a Smart Home, most of times, is not a naive task. First of all, the sensors commonly used for HA do not provide enough information for a precise localization, only presence or motion in specific areas. Furthermore it is not possible to identify the person, not even to differentiate the inhabitants. These main reasons motivate the introduction of an Indoor Localization System designed specifically for Smart Home applications. Part of the Localization System was based on LAURA – LocAlization and Ubiquitous monitoRing of pAtients. Such system leverages a solid RF technology over the 802.15.4 IEEE standard.
A set of unobtrusive devices, battery powered, deployed in the environment dynamically and autonomously create and maintain an RF network. Among those devices some are given to patients, while the rest is statically placed in known locations. Such a structure enables to exploit a zero-configuration adaptive distance estimation method, which leverages the received power of the exchanged signals, to compute the distances of a person with respect to the known-position devices. Through a multilateration procedure, distances can be used to infer the person's coordinates in space. The presented system reports several limitations in precision and performances: some of these questions have been addressed successfully, proposing a new discrete multilateration method, and adopting a different frequency. Indeed a lower frequency permitted to avoid undesired effects and to obtain more precise localization.
Where localization is meant to be used for mutual reassurance the focus is also on dependability features. The main idea is to exploit already-in-place devices, as a source of redundant knowledge: crossing this information with the one from the Localization system enables to identify faults. In particular within this work two different approaches are detailed: a sharp model based and a probabilistic fault localization method. While most of the design effort for dependability is usually devoted to manage faults generated by components of the system (e.g., hardware or software, technically named natural faults), in this work the focus is also on those generated by the users, called human-made faults. Moreover, the system is provided with a fault localization method. This, based on the system status in case of inconsistency – and thus of fault – is able to point out the device not operating in a correct manner, and the building area where the fault is having place. A validation of both methods on real-world and simulated data proved their effectiveness. Furthermore, the need of leveraging HA data to recognize motion and presence triggered the attention for Pyroelectric InfraRed (PIR) sensors. The interest for those devices is also motivated by their unobtrusive functioning, low costs and easy installation. Nonetheless their behavior in terms of sensitivity to movement is often not sufficiently characterized: a piece of this thesis is devoted to the proposal of a model for PIR sensors characterization. Such method, relying on empirical data, considers not only the geometry of the device sensitivity, but also how the user's motion affects the output. This model is supported by experimental results in indoor environments.
Smart technologies development is nowadays oriented toward intelligent services for the dweller. Designing the Artificial Intelligence which plays behind the scenes of such smart devices requires large datasets for several reasons: training of machine learning algorithms, customization, testing, validation, etc. Usually such tasks are carried out on real world data, requiring devices purchase and installation, the identification of test users, protocol definitions, and long acquisition times. The BRIDGe project cannot be immune to those questions. To speed up the development and reduce costs, a behavior simulator can digitally reproduce environments and behaviors of the dwellers, in controlled conditions and in a short time. The novelty of the technique proposed is to leverage the balance of two competitive factors, the Time Dependency and the Personal Needs, to create an ADL scheduling. The resulting software was called SHARON (Simulator of Human Activities, ROutines and Needs), and was validated to reproduce human routines contained in real-world dataset. The obtained results showed a good match with the original distributions especially when looking at the more regular ADLs.
Aiming to the detection of behavior drifts, one of the core questions is to determine what such deviations are. Such task is challenging from many different perspectives: how to represent the behavioral routine information, to identify semantic domains, to define the detail level needed, etc. A possible approach is the definition of an atomic quantity, an elementary entity, which composes the behavior of a person, his or her routines. Such entity can be identified in the set of the Activities of Daily Living (ADLs). Indeed the identification of an ADL requires to be temporally defined, described by a specific set of sensors readings, in a certain area of the house. The final part of this thesis describes a methodology developed to discover ADLs instances along the daily life of a Smart Home inhabitant, represent them, and associate them to specific semantic contexts. The novelty of such contribution is a method that relies only on the data characteristics, by representing the HA sensors status in entities called snapshots, without needing any previous training. The distance designed specifically for those representation is enough for identifying clusters of homogeneous activities instances. Experimental results showed satisfying precision and recall. Moreover the representation of activities through BoW provides an interesting way to attach semantics to the identified sets of snapshots. A preliminary work concerning the identification of a correspondence between different instances of the same activity contained in different clusters, returned limited results.

AUTOMATIC VERIFICATION AND INPUT DESIGN FOR DYNAMICAL SYSTEMS: AN OPTIMIZATION-BASED APPROACH TO THE DETECTION OF NON-INFLUENTIAL INPUTS

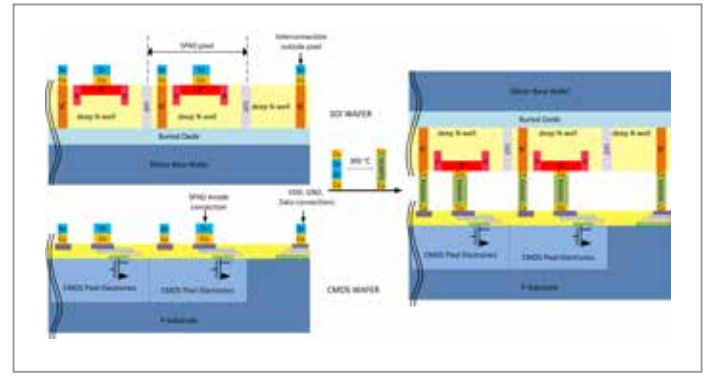
Riccardo Maria Vignali - Supervisor: Prof. Maria Prandini

The aim of this thesis is to develop novel methods for verifying the correct functioning of a system, which possibly involves interacting continuous and discrete dynamical components.
In general terms, the goal of system verification is to evaluate if a system behaves as desired. Commonly, the desired behavior is referred to as a *specification* and is associated to the evolution in time of some variables of the system (state or output).
In particular, *reachability* and *safety* specifications have been extensively studied in the literature and deal with checking if the state reaches a desired *target* set and if it remains within some given *safe* set, respectively.
In this work, we address verification problems where the inputs have to be designed so as to make the system satisfy either a reachability or a safety specification. Among the possible solutions to the problem, we look for the one that minimizes the time to satisfy the specification and that corresponds to the maximal number of *non-influential inputs*, i.e., inputs that can take an arbitrary value in their range without compromising the satisfaction of the specification. Because of the latter requirement, the resulting verification problem is non standard and, hence, novel verification techniques are needed to tackle it.
Detecting non-influential inputs can be particularly useful in testing complex control systems. In this context, it is common to model an undesired behavior of the control system in terms of a specification and then to detect the inputs sequence that originates that behavior. Once such a sequence has been identified, appropriate countermeasures can be taken to improve the control system design. For complex systems with many inputs, it may be the case that the identified sequence is not easy to interpret, so that it is difficult to understand the cause of the undesired behavior and to improve the control system design.
To ease the diagnosis of the system misbehavior, it can then be useful to identify the inputs that are actually influential and the input sequence with minimum length, since shorter sequences are usually simpler to interpret.
The problem of designing the inputs of a dynamical system so as to make its evolution satisfy a specification has been extensively treated in literature, and effective solutions have been developed for a variety of classes of systems. Many of these techniques are simple derivation of the ones developed in the *model checking* context, where the aim is typically to check if a system satisfies a certain specification, without input design.
For discrete systems, it has been shown that checking if a discrete system S satisfies a specification can be done by means of a reachability test on the enlarged system obtained by making S interact with the *test automaton* that translates the specification. The resulting reachability test can then be tackled by means of efficient model checkers, like SPIN, to name a few.
For continuous state systems, these model checking techniques are not directly applicable, since they would require to analyze an infinite number of possible evolutions of the system. For this reason, alternative methods have been proposed in literature. Some of them rely on the computation (either exact or approximated) of the reachable set of the system under test, so as to consider in a compact way all its infinite possible. Other common approaches are based on abstracting the continuous system to a discrete one, and then analyzing the latter via model checking techniques.
Many of these methods can be extended to the class of *hybrid* systems, i.e., systems with interacting discrete and continuous dynamics.
Our resolution approach for the detection of non-influential inputs does not specifically rely on any of the methods that can be found in literature but has been strongly influenced by other works on related topics like. In particular, the techniques that we develop are based on optimization, that appears to be the most suitable approach to address the requirement of maximizing the number of non-influential inputs.

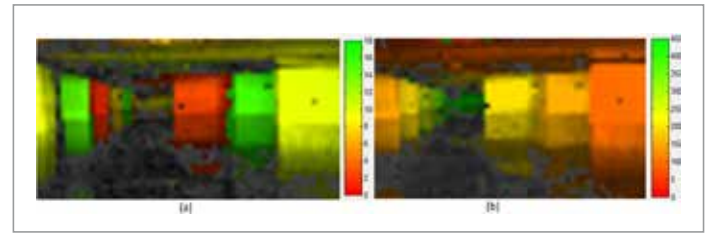
SPAD IMAGERS AND SYSTEMS FOR 2D IMAGING AND 3D RANGING

Yu Zou - Supervisor: Prof. Franco Zappa

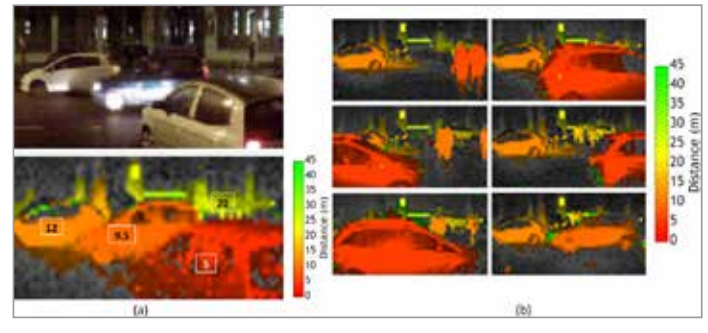
Aim of my PhD work was to develop CMOS image sensor and 3D vision system based on Single-Photon Avalanche Diode (SPAD) detectors and indirect time-of-flight techniques, targeting safety and security applications. The SPAD is the viable detector that provides single-photon sensitivity and short integration time, which are specific constrains of many applications in the safety and security scenario. Apart from the detector itself, it is important to conceive and develop the full CMOS SPAD imager and the overall electronics for both the 3D camera and the illuminator.
In my PhD research activity I actively worked in all these topics, designing and validating the individual components and then the overall 3D indirect Time-of-Flight (iTOF) system.
At first, the thesis presents the design and development of an innovative–BackSPAD sensor chip, based on a vertical wafer-bonded assembly of the chip with backside-illuminated SPADs together with the chip containing the electronics, as shown in Figure 1. A silicon-on-insulator (SOI) wafer containing custom SPADs, to be back-illuminated, flipped upside down and wafer-bonded on a standard 0.35 µm CMOS wafer integrating the analog front-end circuit, in-pixel digital processing and readout electronics. Such a BackSPAD approach provides reduced pixel's pitch (because the in-pixel electronics is no longer side-by-side with the detector, but under it), increased pixel-density (i.e. the overall number of pixels), and improved collection efficiency (thanks to the higher fill-factor and reduced blind area). Beyond the "detection" advantages (i.e. due to the SOI SPAD chip), the proposed BackSPAD approach provides also "processing" advantages since it allows us to employ a scaled CMOS technology for the readout electronics, thus positively further shrinking the pixel pitch, and speeding up speed and in-pixel processing capability.
Furthermore, the chance to have two separated wafers allow to customize the SPAD detectors in the SOI wafer (e.g. enhanced spectral sensitivity in the near-infrared, up to 1 µm wavelength), thanks to thicker active volume and to the backside illumination of the active area.
Then, the thesis presents the design and development of a complete 3D ranging system tailored to automotive applications, able to provide centimeter depth resolution over a 40° × 20° field-of-view up to 45 m with just 1.5 W of average active illumination at 808 nm. The imager is based on a CMOS SPAD array chip of 64×32 pixels, able to perform lock-in time-of-flight computation of individual single photons emitted by the illuminator and detected back by the SPAD pixels of the camera. The distance is measured by the continuous-wave (CW) indirect time-of-flight (TOF) technique.
In order to avoid the distance ambiguity (aliasing), I propose a double frequency (5 MHz and 8.33 MHz) modulation technique. For example, with a modulation period of 120 ns, the maximal non-ambiguous range would be limited to 18 m; hence an object at 20 m would be mistaken as an object at 2 m (see Figure 2, left). Instead, through such double-frequency technique, the depth acquisition is correct and the both the maximum detectable distance and the precision are improved (see Fig. 2, right).
The thesis analyzes and quantifies the error sources in the SPAD CW-iTOF system and their impacts on distance measurements accuracy and precision. For instance, this 3D SPAD system can achieve 60 cm precision at 40 m distance. Furthermore the thesis reports on the complete validation and characterization of the camera in many different environment conditions, both indoor and outdoor, with both static and dynamics scenes. The on-road tests show that the system can detect different objects (pedestrians, bicycles, cars, and light trains), providing their 3D distance map and 2D information to help identify those objects, as shown in Figure 3. A discussion on present state-of-art 3D iTOF cameras available on the market and a comparison on the developed CMOS SPAD 3D system eventually conclude the thesis.



1. Cross-section of SOI wafer and CMOS wafer before and after bonding process.



2. (a) Distance measurement performed with single frequency (max. ambiguous range 18m) shows aliasing problem. (b) distance measurement performed with double-frequency method achieves a maximum distance of 45 m.



3. Outdoor measurements with the developed iTOF system.